

Міністерство освіти і науки України
Харківський національний університет радіоелектроніки

Факультет Інформаційно-аналітичних технологій та менеджменту
(повна назва)
Кафедра Інформатики
(повна назва)

КВАЛІФІКАЦІЙНА РОБОТА
Пояснювальна записка

рівень вищої освіти другий (магістерський)

МЕТОД ПЕРЕДБАЧЕННЯ ЗМІН У КІЛЬКОСТІ ПРОСЛУХОВУВАНЬ
АРТИСТІВ НА ОСНОВІ ДОСЛІДЖЕННЯ ЇХ ОСТАННІХ ДІЙ
(тема)

Виконав:
здобувач 2 року навчання,
групи ІНФМ-24-1
Суровикін Ю. В.
(прізвище, ініціали)

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

Тип програми освітньо-професійна

Освітня програма Інформатика
(повна назва освітньої програми)

Науковий керівник проф. Машталір С.В.
(посада, прізвище, ініціали)

Допускається до захисту

Завідувач кафедри інформатики _____
(підпис)

Кобилін О. А.
(прізвище, ініціали)

2025 р.

Харківський національний університет радіоелектроніки

Факультет Інформаційно-аналітичних технологій та менеджментуКафедра ІнформатикиРівень вищої освіти другий (магістерський)Спеціальність 122 Комп'ютерні науки
(код і повна назва)Тип програми освітньо-професійнаОсвітня програма Інформатика
(повна назва)

ЗАТВЕРДЖУЮ:

Зав. кафедри _____
(підпис)

« _____ » _____ 2025 р.

ЗАВДАННЯ
НА КВАЛІФІКАЦІЙНУ РОБОТУздобувачеві Суровикіну Юрію Віталійовичу
(прізвище, ім'я, по батькові)1. Тема роботи Метод передбачення змін у кількості прослуховувань артистів на основі дослідження їх останніх дій

затверджена наказом університету від 14 листопада 2025 року № 1045Ст

2. Термін подання здобувачем роботи до екзаменаційної комісії 19 листопада 2025 р.

3. Вихідні дані до роботи часові ряди прослуховувань виконавців, жанрові мітки релізів, літературні джерела щодо методів аналізу часових рядів та жанрової класифікації, програмні засоби для реалізації моделі прогнозування у веб-середовищі, а також вибірки даних для тестування стабільності жанрових переходів та точності прогнозу.

4. Перелік питань, що потрібно опрацювати в роботі _____

1. Аналіз сучасних методів прогнозування часових рядів у музичній аналітиці.

2. Аналіз літературних джерел щодо апробації моделей визначення жанрової динаміки та поведінки слухачів.

3. Формування покрокового алгоритму розрахунку тренду, нормалізації даних, визначення стабільності жанру та прогнозування прослуховувань.

4. Візуалізація сформованих покрокових алгоритмів та процесів обробки даних.

5. Розробка програмного застосунку, що надасть змогу виконувати жанровий аналіз, формувати місячні ряди та прогнозувати майбутню кількість прослуховувань виконавця на основі отриманої моделі.

5. Перелік графічного матеріалу із зазначенням креслеників, схем, плакатів, комп'ютерних ілюстрацій (п.5 включається до завдання за рішенням випускової кафедри) актуальність проблеми прогнозування прослуховувань у цифровій музичній аналітиці, об'єкт та мету дослідження, постановку задачі, блок-схеми алгоритмів нормалізації, побудови тренду, жанрових переходів та формування прогнозу, приклад структури вхідних даних, ілюстрацію головного екрану розробленого веб-застосунку із вибором виконавця, ілюстрацію глобального прогнозу та окремих жанрових графіків, приклад виявлених жанрових переходів та їх впливу на тренд, висновки, перспективи та апробацію роботи.

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

Найменування розділу	Консультант (посада, прізвище, ім'я, по батькові)	Позначка консультанта про виконання розділу	
		підпис	дата

КАЛЕНДАРНИЙ ПЛАН

№ з/п	Назва етапів роботи	Строк / терміни виконання етапів роботи	Примітка
1	Отримання завдання на кваліфікаційну роботу	29.09.2025	
2	Аналіз завдання, підбір літератури	30.09.25-07.10.25	
3	Аналіз літератури з досліджуваної проблеми	08.10.25-14.10.25	
4	Дослідження особливості методів	15.10.25-20.10.25	
5	Розробка алгоритмів	21.10.25-27.10.25	
6	Програмна реалізація	28.10.25-05.11.25	
7	Обґрунтування отриманих результатів	06.11.25-11.11.25	
8	Оформлення пояснювальної записки	12.11.25-14.11.25	
9	Перевірка на нормоконтроль	19.11.25-10.12.25	
10	Перевірка на плагіат	20.11.25-10.12.25	
11	Рецензування	21.11.25-10.12.25	
12	Підготовка презентації та доповіді	21.11.25-22.12.25	
13	Занесення роботи в електронний архів	21.11.25-22.12.25	
14	Попередній захист кваліфікаційної роботи	01.12.25-22.12.25	

Дата видачі завдання 29 вересня 2025 р.

Здобувач _____
(підпис)

Керівник роботи _____
(підпис)

проф. Машталір С.В.
(посада, прізвище, ініціали)

РЕФЕРАТ

Пояснювальна записка до кваліфікаційної роботи: 78 с., 5 табл., 8 рис., 50 джерел.

АНАЛІТИКА ДАНИХ, ПРОГНОЗУВАННЯ, МУЗИЧНІ СЕРВІСИ, ЖАНРОВІ ПЕРЕХОДИ, ПРОСЛУХОВУВАННЯ, РЕГРЕСІЙНИЙ АНАЛІЗ, ЛІНІЙНА РЕГРЕСІЯ, КОВЗНЕ СЕРЕДНЄ, СТАБІЛЬНІСТЬ ЖАНРУ.

Об'єктом дослідження є динаміка змін кількості прослуховувань музичних виконавців на онлайн-платформах.

Предметом дослідження є методи прогнозування змін популярності артистів на основі аналізу їхніх музичних дій та жанрових переходів.

Метою дослідження є створення методу передбачення змін у кількості прослуховувань артистів із використанням історичних даних про їх релізи та поведінку слухачів.

Використано методи лінійної регресії, ковзного середнього та побудову матриці схожості жанрів. Алгоритм аналізує історичну активність виконавця, стабільність його жанрового напрямку та зміни у творчості.

Наукова новизна роботи полягає у поєднанні методів прогнозування часових рядів із жанровим аналізом, що дозволяє підвищити точність передбачення змін у популярності виконавців.

Взаємозв'язок з іншими роботами полягає у використанні принципів регресійного аналізу та статистичних методів обробки часових даних, адаптованих до задач музичної аналітики.

Рекомендації щодо використання результатів роботи стосуються впровадження методу в аналітичні системи музичних платформ для прогнозування популярності артистів.

У результаті дослідження розроблено сервіс прогнозування змін кількості прослуховувань артистів на основі аналізу їхніх останніх дій та змін.

ABSTRACT

Explanatory note to the qualification work: 78 pages, 5 table, 8 figures, 50 sources.

DATA ANALYTICS, FORECASTING, MUSIC SERVICES, GENRE TRANSITIONS, LISTEN COUNTS, REGRESSION ANALYSIS, LINEAR REGRESSION, MOVING AVERAGE, GENRE STABILITY.

The object of the research is the dynamics of changes in the number of listens of music artists on online platforms.

Subject of research is the methods of predicting changes in artist popularity based on the analysis of their musical activity and genre transitions.

The aim of the research is to develop a method for forecasting changes in the number of artist listeners using historical data on their releases and listener behavior.

The methods used are linear regression, moving average, and the construction of a genre similarity matrix. The developed algorithm analyzes the historical activity of the artist, the stability of their genre direction, and creative changes over time.

Scientific novelty of the work lies in combining time series forecasting methods with genre analysis, which increases the accuracy of predicting changes in artist popularity.

Interconnection with other works lies in the application of regression analysis principles and statistical processing of time series data, adapted to the field of music analytics.

Recommendations for using the results of the work concern the implementation of the developed method in analytical systems of music platforms for predicting artist popularity and improving recommendation accuracy.

As a result of the research, a forecasting service was developed to predict changes in the number of artist listeners based on the analysis of their latest actions and creative trends.

ЗМІСТ

Перелік умовних позначень, символів, одиниць, скорочень і термінів	8
Вступ.....	9
1 Огляд основних методів Вашого теми.....	11
1.1 Метод лінійної регресії	11
1.2 Метод ковзного середнього	13
1.3 Метод аналізу жанрової стабільності.....	15
1.3.1 Розрахунок жанрової стабільності артиста.....	15
1.3.2 Використання стабільності у матриці порівняння жанрів ...	16
1.4 Порівняння обраних методів із сучасними підходами прогнозування популярності музики	18
1.4.1 Моделі часових рядів	18
1.4.2 Нейронні мережі та глибокі моделі	19
1.4.3 Гібридні контентно-поведінкові моделі.....	19
1.4.4 Поведінкові та соціальні моделі	20
1.4.5 Обґрунтування вибору методу.....	21
1.5 Постановка задачі дослідження.....	21
2 Математичні моделі прогнозування змін у кількості прослуховувань	23
2.1 Загальна структура математичної моделі	23
2.1.1 Математичне формулювання задачі прогнозування	23
2.1.2 Застосування методу ковзного середнього	25
2.1.3 Моделювання лінійної регресійної моделі тренду	27
2.1.4 Моделювання стабільності жанру виконавця	28
2.2 Формування матриці жанрових зв'язків	30
2.2.1 Моделювання стабільності жанру виконавця	30
2.2.2 Моделювання стабільності жанру виконавця	31
2.2.3 Використання матриці у прогнозній моделі	32
2.3 Нормалізація та фільтрація аномальних значень.....	33
2.4 Інтеграція параметрів моделі в єдину систему прогнозування	34

2.5	Блок-схема побудови прогнозної моделі	36
2.6	Інтеграція компонентів прогнозної моделі	37
3	Комп'ютерна модель прогнозування змін у кількості прослуховувань.....	40
3.1	Обґрунтування вибору середовища програмної реалізації.....	40
3.2	Програмна реалізація	44
3.3	Інструкція користувача	59
3.4	Дослідження та тестування розробленої моделі.....	61
	Висновки.....	71
	Перелік джерел посилання	73

ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ, СИМВОЛІВ, ОДИНИЦЬ, СКОРОЧЕНЬ І ТЕРМІНІВ

БД – база даних

Стрімінг – модель розповсюдження контенту з передаванням у реальному часі

API – Application Programming Interface (програмний інтерфейс прикладного програмування)

DOM – Document Object Model (об’єктна модель документа)

Express.js – серверний фреймворк JavaScript для створення вебAPI

JS – JavaScript (мова програмування для веброзробки)

JSON – JavaScript Object Notation (формат обміну структурованими даними)

Node.js – серверне середовище виконання JavaScript

REST – Representational State Transfer (архітектурний стиль вебінтерфейсів)

TS – TypeScript (мова програмування на базі JavaScript з підтримкою статичної типізації)

UI – User Interface (інтерфейс користувача)

ВСТУП

Сучасна музична індустрія стрімко змінюється під впливом цифрових технологій та глобальних онлайн-платформ. Завдяки розвитку потокових сервісів, таких як Spotify, Apple Music, YouTube Music чи SoundCloud, музика стала не лише засобом самовираження, а й джерелом аналітичних даних про поведінку мільйонів слухачів у реальному часі. У цій новій реальності дані про прослуховування, релізи, жанрові переваги та активність артистів утворюють складну динамічну систему, поведінку якої можна моделювати та прогнозувати.

Важливою задачею для лейблів, продюсерів і самих виконавців є розуміння факторів, що впливають на зміни популярності артистів. Часто прослуховування змінюються не лише через маркетингову активність або сезонні коливання, а й через творчі рішення: зміну жанру, частоту випуску пісень, колаборації чи стилістичні експерименти. Визначення закономірностей між цими діями та реакцією слухачів дозволяє побудувати прогноз майбутньої динаміки прослуховувань і дає змогу артистам планувати свою діяльність більш стратегічно.

Актуальність роботи полягає у зростаючій потребі музичної індустрії в аналітичних інструментах, здатних передбачати зміни популярності виконавців. Попри значний розвиток систем рекомендацій, більшість із них спрямовані на індивідуальні вподобання користувачів, тоді як проблема прогнозування популярності самого артиста залишається менш дослідженою. Використання методів аналітики даних для виявлення впливу творчих рішень на кількість прослуховувань є актуальним напрямом дослідження, що поєднує наукові підходи до аналізу часових рядів і прикладні завдання індустрії розваг.

Сучасні підходи до прогнозування популярності у музичних системах базуються на статистичних і машинних моделях. Використовуються методи регресійного аналізу, згладження ковзним середнім, нейронні мережі, а також алгоритми прогнозування часових рядів. Проте більшість таких моделей розглядає загальні закономірності змін у потоках прослуховувань без врахування

контексту – тобто жанрових характеристик і творчих дій самих виконавців. Тому для досягнення більшої точності необхідно поєднати класичні статистичні методи із доменною специфікою музичних даних.

У даній роботі пропонується метод, що враховує як історичні показники активності артистів, так і якісні фактори – стабільність жанру, частоту релізів, а також переходи між жанрами. В основі підходу лежить поєднання лінійної регресії з ковзним середнім, що дозволяє згладити коливання даних та виділити довгострокову тенденцію. Для уточнення прогнозів застосовується матриця схожості жанрів, яка дає змогу оцінити, наскільки зміна жанрового напрямку може вплинути на слухацьку базу. Такий підхід дозволяє виявити як позитивні, так і негативні наслідки творчих змін артиста, що має практичну цінність для управління його кар'єрою.

Наукова задача дослідження полягає у створенні алгоритму, здатного прогнозувати зміни у кількості прослуховувань артистів на основі аналізу часових рядів, статистичних характеристик та контекстних факторів їхньої діяльності. Запропонований метод дозволяє інтегрувати дані про творчість, жанрові тенденції та поведінку слухачів у єдину модель, що забезпечує підвищену точність передбачень у порівнянні з традиційними методами лінійного прогнозування.

Отже, у роботі виконується аналіз сучасних підходів до прогнозування трендів у музичній індустрії, розробляється власний метод аналітики та створюється прототип сервісу, який на основі дій артиста визначає ймовірний напрям зміни його популярності. Отримані результати можуть бути використані у системах музичної аналітики, платформах потокового прослуховування та у маркетингових стратегіях артистів і лейблів.

1 ОГЛЯД ОСНОВНИХ МЕТОДІВ ПРОГНОЗУВАННЯ ЗМІН У КІЛЬКОСТІ ПРОСЛУХОВУВАНЬ

1.1 Метод лінійної регресії

Метод лінійної регресії є одним із базових статистичних підходів до аналізу та прогнозування числових залежностей. Його суть полягає у встановленні лінійного зв'язку між незалежною змінною та залежною величиною. Лінійна регресія широко використовується у прогнозуванні часових рядів, економічному моделюванні, аналітиці даних і машинному навчанні завдяки простоті реалізації та інтерпретованості результатів [1, 2]. У загальному вигляді рівняння простої лінійної регресії має наступний вигляд.

$$y = a + bx, \quad (1.1)$$

де y – прогнозоване значення залежної змінної;

x – незалежна змінна (наприклад, номер місяця або час);

a – зміщення;

b – коефіцієнт нахилу, відображає силу та напрямок зв'язку між змінними.

Для визначення параметрів a та b використовується метод найменших квадратів та обчислюються оптимальні значення параметрів [3].

$$S = \sum_{i=1}^n (y_i - a - bx_i)^2, \quad (1.2)$$

$$b = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2}, \quad (1.3)$$

$$a = \bar{y} - b\bar{x}, \quad (1.4)$$

де $\bar{x}$ та $\bar{y}$ – середні значення змінних x та y відповідно.

У контексті прогнозування змін у кількості прослуховувань артистів метод лінійної регресії використовується для побудови трендової моделі залежності кількості прослуховувань від часу. Кожна точка у часовому ряді відповідає загальній кількості прослуховувань певного жанру або виконавця у конкретному місяці. Отримана регресійна пряма дозволяє оцінити напрямок тенденції і зробити прогноз на наступні періоди [4].

Перевагою методу є його простота, мала обчислювальна складність і можливість швидкого отримання аналітичного результату. Крім того, модель лінійної регресії легко інтерпретується, що важливо для аналітичних систем, де потрібно пояснити причини зміни популярності. Недоліками методу є обмежена точність у випадках, коли дані містять різкі коливання або нелінійні залежності, а також чутливість до викидів – аномальних точок, що можуть суттєво вплинути на результати.

Для підвищення точності прогнозів у роботі метод лінійної регресії поєднується з попереднім згладжуванням даних методом ковзного середнього, що дозволяє зменшити вплив короткострокових коливань та виділити основний тренд. Такий підхід забезпечує стабільніші результати при прогнозуванні динаміки прослуховувань артистів у різних музичних жанрах.

Вибір цього методу обґрунтовується тим, що у задачі прогнозування прослуховувань артистів важливо зберегти баланс між точністю, інтерпретованістю та стійкістю до зміни даних.

На відміну від складних алгоритмів прогнозування, таких як ARIMA, SARIMA чи нейронні мережі типу RNN та LSTM, лінійна регресія не потребує великої кількості навчальних прикладів або глибокої параметризації. Це робить її особливо зручною для випадків, коли дані мають обмежену кількість спостережень, наприклад, коли аналізується діяльність одного артиста протягом кількох років.

У системах музичної аналітики, де прогноз має бути не лише точним, а й пояснюваним, надмірно складні нейромережеві моделі часто створюють «чорну скриньку» – результат можна отримати, але складно зрозуміти, які саме фактори

спричинили зміну тренду. Лінійна регресія, навпаки, дозволяє чітко простежити зв'язок між часовими показниками та зростанням або спадом популярності, що важливо для прийняття управлінських рішень продюсерами або маркетинговими аналітиками.

Крім того, у музичних потокових сервісах характер прослуховувань часто має лінійні тенденції в межах середньострокових періодів – наприклад, стабільне зростання після виходу нового альбому чи поступовий спад після завершення промокампанії. Тому використання лінійної регресії як базової моделі дозволяє побудувати аналітичну основу для подальшого доповнення більш складними елементами, такими як вагові коефіцієнти жанрової стабільності чи поправки на сезонні впливи.

У поєднанні з попереднім згладжуванням даних методом ковзного середнього цей підхід дає змогу усунути короткострокові коливання та аномалії, залишаючи лише загальну тенденцію. Таким чином, лінійна регресія виявляється оптимальним компромісом між простотою реалізації, стійкістю до шумів і достатньою точністю прогнозів у динамічних середовищах потокових платформ.

1.2 Метод ковзного середнього

Метод ковзного середнього є одним із найпоширеніших інструментів статистичного аналізу часових рядів. Його основне призначення полягає у згладжуванні короткострокових коливань даних і виявленні основної тенденції зміни показників у часі. Цей метод широко застосовується для аналізу фінансових, економічних, соціальних та поведінкових даних, а останнім часом – і для прогнозування активності користувачів у цифрових медіа, зокрема на музичних платформах [5, 6].

Суть методу полягає у тому, що для кожного моменту часу t обчислюється середнє значення показників за певне «вікно» попередніх періодів

$$\bar{y}_t = \frac{1}{n} \sum_{i=t-n+1}^t y_i, \quad (1.5)$$

де $\bar{y}$ – згладжене значення у момент часу t ;

$\bar{y}$ – фактичні значення показника;

n – розмір ковзного вікна.

Вибір довжини вікна n визначає ступінь згладжування: при малих значеннях модель реагує швидше на зміни, а при великих – краще відображає загальну тенденцію. У задачах прогнозування музичних прослуховувань оптимальним виявляється невелике вікно, що дозволяє збалансувати чутливість і стабільність прогнозу [7].

У межах розробленого методу прогнозування кількості прослуховувань артистів ковзне середнє використовується як попередній етап перед застосуванням лінійної регресії. Це дозволяє усунути шумові коливання, спричинені окремими вірусними треками або випадковими піками активності, та виявити загальну тенденцію розвитку артиста. У результаті прогноз стає більш плавним, стійким і краще відображає реальну динаміку популярності.

Вибір методу ковзного середнього зумовлений його ефективністю у виявленні глобальних тенденцій і усуненні короточасних шумових впливів без застосування складних статистичних моделей.

На практиці у музичній аналітиці дані про прослуховування часто містять різкі коливання, пов'язані з виходом окремих хітів, медійною активністю або колабораціями артистів. Такі сплески не відображають реального довгострокового тренду, тому їх необхідно згладити для коректного аналізу загальної динаміки популярності [6, 7].

Альтернативними підходами до згладжування є експоненційне згладжування, сезонні моделі типу Holt-Winters або фільтрація Калмана. Проте всі ці методи потребують ретельного налаштування параметрів і попереднього статистичного аналізу, що не завжди виправдано в умовах швидкої обробки великих потоків даних. Метод ковзного середнього не потребує навчання або

складних гіперпараметрів, забезпечує високу стабільність результатів і легкість реалізації навіть у режимі реального часу [8].

У поєднанні з лінійною регресією ковзне середнє відіграє роль фільтра, що готує дані до подальшого прогнозування. Такий підхід дозволяє уникнути переобчислення складних моделей при появі нових точок у часовому ряді, що особливо важливо для музичних сервісів, де дані оновлюються щодня.

Крім того, ковзне середнє забезпечує гнучкість – довжину вікна можна змінювати залежно від жанрових особливостей або частоти релізів артиста. Наприклад, для активних виконавців, що випускають музику щомісяця, достатньо короткого вікна, тоді як для виконавців із рідшими релізами доцільно застосовувати більш тривале вікно, щоб відобразити реальний тренд розвитку слухацької бази [9].

Таким чином, метод ковзного середнього є універсальним інструментом попередньої обробки часових даних, який забезпечує стабільність, масштабованість та інтерпретованість результатів, що робить його невід’ємною частиною побудови системи прогнозування змін кількості прослуховувань артистів.

1.3 Метод аналізу жанрової стабільності

1.3.1 Розрахунок жанрової стабільності артиста

У прогнозуванні популярності виконавців ключову роль відіграє стабільність їхнього творчого напрямку. Артисти, що дотримуються певного жанру протягом тривалого часу, зазвичай формують стійке ядро слухачів, тоді як різкі зміни стилю часто призводять до втрати частини аудиторії.

Для кількісного аналізу цього аспекту використовується показник жанрової стабільності, що визначає послідовність у виборі напрямку [10].

Метод полягає у порівнянні жанрової належності кожного нового релізу артиста з попереднім. Для кожного жанру формується послідовність появ у

часовому ряді. Потім обчислюється кількість змін між жанрами, що дозволяє оцінити ступінь стабільності.

$$S = 1 - \frac{C}{N-1}, \quad (1.6)$$

де C – кількість переходів між жанрами у релізах артиста;

N – загальна кількість пісень.

Отриманий коефіцієнт S приймає значення від 0 до 1, де 1 це абсолютна сталість жанру, а 0 – кожен новий реліз відбувався у новому жанрі.

Розрахунок цього показника дає змогу зробити висновки про стабільність творчої лінії виконавця. У нашій системі цей параметр виступає одним із вагових коефіцієнтів під час побудови прогнозу майбутніх прослуховувань. Якщо артист демонструє високу стабільність, модель прогнозу підсилює очікуване значення популярності; якщо стабільність низька – прогноз коригується вниз із врахуванням ризику втрати слухачів [11].

Метод є обчислювально простим, не потребує навчання моделі та може бути застосований до будь-якого виконавця незалежно від кількості його релізів. Це робить його гнучким компонентом у структурі загальної системи прогнозування, що поєднує статистичні й поведінкові фактори.

1.3.2 Використання стабільності у матриці порівняння жанрів

Окрім індивідуального розрахунку, коефіцієнт стабільності відіграє важливу роль у побудові матриці жанрових зв'язків.

Ця матриця є основою для оцінки того, наскільки схожими є різні музичні напрями за слухацькою реакцією. Її елементи відображають імовірність успішного переходу між жанрами – наприклад, перехід з року до гранджу часто не призводить до втрати слухачів, тоді як зміна року на електроніку може бути ризикованою [12].

Для побудови матриці використовується історія жанрових переходів великої кількості виконавців. Кожен перехід оцінюється не лише за частотою зміни жанру, а й за успішністю, тобто зміною кількості прослуховувань після переходу. Узагальнена формула елемента матриці має наступний вигляд.

$$M_{ij} = \frac{\omega_{ij}}{\max(\omega)}, \quad (1.7)$$

де ω_{ij} – середня успішність переходів із жанру i у жанр j ;

$\max(\omega)$ – максимальне значення серед усіх переходів.

Таким чином, значення M_{ij} близьке до 1 означає високу жанрову сумісність, тоді як низьке значення вказує на невдалий або рідкісний перехід.

Отримана матриця використовується при прогнозуванні для корекції коефіцієнта стабільності: якщо артист переходить у жанр, схожий на попередній, його стабільність частково зберігається. Якщо ж перехід відбувається у далекий за стилем напрям – прогнозована кількість прослуховувань зменшується [13].

Подібні підходи до моделювання схожості зустрічаються у системах рекомендацій музики та аналітичних інструментах потокових сервісів. Проте у нашій роботі така матриця не є окремою системою рекомендацій – вона виконує допоміжну функцію для підвищення точності прогнозу шляхом урахування контексту жанрових змін [14].

На відміну від класичних моделей кластеризації або нейронних автоенкодерів подібності, запропонований підхід не вимагає великих вибірок і може динамічно оновлюватися на основі нових релізів.

У підсумку, матриця жанрових зв'язків забезпечує адаптивність моделі прогнозу – вона враховує не лише історію конкретного артиста, а й загальні тенденції у взаємодії жанрів у музичному просторі. Це дозволяє більш точно оцінювати, як зміна стилю може вплинути на майбутню популярність виконавця.

1.4 Порівняння обраних методів із сучасними підходами прогнозування популярності музики

1.4.1 Моделі часових рядів

Класичні статистичні методи прогнозування, побудовані на основі аналізу часових рядів, тривалий час залишаються базовими у різних галузях [15]. Вони передбачають використання автокореляційних залежностей між минулими та майбутніми значеннями, що дозволяє виявляти загальні тренди й сезонні коливання. Такі моделі мають добру точність у випадках, коли дані демонструють регулярну повторюваність, але їх застосування у музичній сфері часто обмежене.

Дані про прослуховування музики не є стаціонарними, оскільки поведінка слухачів залежить від позасистемних чинників: виходу нових релізів, рекламних кампаній, змін у музичній моді, появи вірусних відео [16]. У таких умовах традиційні моделі часових рядів часто не встигають адаптуватися до нових тенденцій, тому результати прогнозу швидко втрачають актуальність. Крім того, ці підходи здебільшого оперують лише числовими значеннями і не враховують якісні характеристики, такі як зміни жанру чи особливості творчої поведінки артиста [17].

Обраний у цій кваліфікаційній роботі метод лінійної регресії з попереднім згладжуванням ковзним середнім забезпечує стабільність при невеликих обсягах даних і дозволяє інтегрувати додаткові показники, наприклад стабільність жанрової поведінки. Він не потребує перетворення рядів у стаціонарну форму, має нижчу чутливість до шумів і краще пристосований до коротких часових послідовностей, що характерні для молодих виконавців. Таким чином, у нашій задачі лінійна регресія виступає універсальним і водночас надійним інструментом.

1.4.2 Нейронні мережі та глибокі моделі

Сучасні дослідження у сфері прогнозування популярності активно застосовують методи глибинного навчання [18]. Ці підходи здатні виявляти приховані закономірності у великих обсягах даних, зокрема у взаємозв'язках між аудіофічами, текстовими характеристиками пісень та поведінковими даними користувачів [19]. Проте такі системи мають високу складність побудови, вимагають великих обсягів даних для навчання і значних обчислювальних ресурсів.

У випадку музичних артистів часто доводиться працювати з обмеженою кількістю релізів, а дані про прослуховування оновлюються нерівномірно. Для таких умов глибокі моделі виявляються надлишковими. Крім того, результати, отримані нейронними мережами, важко інтерпретувати [20]. Користувачеві або аналітику складно пояснити, чому алгоритм зробив той чи інший прогноз.

Запропонована у цій кваліфікаційній роботі модель є прозорою, її можна перевірити та проаналізувати на кожному етапі розрахунку. Вона базується на відкритих статистичних операціях і враховує поведінкові характеристики артиста через показник жанрової стабільності. У такий спосіб забезпечується зрозумілий зв'язок між діями виконавця і змінами його популярності. На відміну від нейронних мереж, що створюють складні структури, наша модель є відкритою та аналітично обґрунтованою [20].

1.4.3 Гібридні контентно-поведінкові моделі

Іншим напрямом розвитку прогнозних систем стали гібридні моделі, які поєднують статистичні закономірності та змістовний аналіз музичних творів [21]. У таких підходах використовується інформація про мелодійну структуру, темп, ритм, тексти пісень і навіть емоційні характеристики.

Поєднання цих параметрів дає змогу більш точно оцінити короткострокову реакцію слухачів на новий контент.

Попри це, гібридні моделі мають низку практичних обмежень. Вони передбачають наявність повних даних про треки, тоді як у реальних потокових системах така інформація не завжди доступна. Крім того, складність багатомодальних архітектур значно ускладнює їх використання для швидких оновлень прогнозу [22].

Наш метод об'єднує статистичну частину, що базується на часовому ряді прослуховувань, із поведінковою складовою, представленою жанровими змінами артиста. Це дозволяє зберегти баланс між точністю та швидкістю розрахунку. Таке рішення поєднує не текстові ознаки, а кількісну та змістову поведінку виконавця, що краще відображає реальну динаміку змін популярності [22].

1.4.4 Поведінкові та соціальні моделі

Останніми роками все більшої популярності набувають моделі, які аналізують соціальні взаємодії користувачів і реакцію аудиторії на музику [23]. Вони базуються на аналізі даних із соціальних мереж, взаємних вподобань і користувацьких плейлистів. Такі системи здатні передбачати сплески популярності на основі вірусного ефекту або соціального резонансу.

Хоча цей підхід ефективно пояснює раптові зміни у кількості прослуховувань, його застосування у загальних аналітичних моделях має низку труднощів. Дані з соціальних платформ є фрагментованими, неповними, часто недоступними через політику конфіденційності. Крім того, соціальна динаміка не завжди корелює з довгостроковою популярністю артиста [24].

Запропонований метод не потребує зовнішніх джерел. Усі обчислення здійснюються на основі даних самої музичної платформи: кількості прослуховувань, дати релізу, належності до жанру. Це забезпечує автономність і

сталість роботи системи. При цьому, навіть без урахування соціальних факторів, алгоритм враховує їх опосередкований вплив через зміну жанрової стабільності та поведінкові характеристики слухачів [24].

1.4.5 Обґрунтування вибору методу

Порівняльний аналіз показує, що поєднання ковзного середнього, лінійної регресії та коефіцієнта жанрової стабільності є найбільш раціональним підходом для задачі прогнозування змін у популярності виконавців. Ця комбінація забезпечує рівновагу між простотою реалізації, інтерпретованістю результатів і достатньою точністю.

Лінійна регресія дозволяє оцінювати загальний напрям розвитку артиста, ковзне середнє згладжує короткочасні коливання, а аналіз жанрової стабільності враховує творчі рішення, що безпосередньо впливають на поведінку слухачів. Такий підхід дає змогу не лише робити прогноз, а й пояснювати його, що є критично важливим для аналітичних систем і управлінських рішень у сфері музичного продюсування [25, 26].

Таким чином, запропонований метод поєднує властивості статистичного моделювання та поведінкового аналізу, є стійким до обмежених вибірок, масштабованим і придатним до впровадження у реальні аналітичні сервіси музичних платформ [27, 28].

1.5 Постановка задачі дослідження

Таким чином, передбачення змін у кількості прослуховувань музичних виконавців на основі аналізу їхньої творчої активності, жанрових переходів і поведінкових закономірностей є актуальним завданням сучасної музичної аналітики.

Прийнято рішення щодо розроблення програмного сервісу прогнозування, який дозволяє визначати ймовірні зміни популярності артистів на підставі аналізу історичних даних про їхні релізи, жанрову динаміку та реакцію слухачів.

Об'єктом дослідження є динаміка змін кількості прослуховувань музичних виконавців на онлайн-платформах.

Метою дослідження є створення методу передбачення змін у кількості прослуховувань артистів із використанням історичних даних про їх релізи та поведінку слухачів.

Для досягнення мети необхідно вирішити такі завдання:

- провести аналіз літературних джерел щодо сучасних підходів до прогнозування популярності музичних виконавців;
- дослідити та порівняти основні методи аналізу часових рядів, регресійних моделей і поведінкових факторів слухачів;
- розробити алгоритм, що поєднує методи ковзного середнього, лінійної регресії та коефіцієнта жанрової стабільності для прогнозу змін у кількості прослуховувань;
- визначити та сформувати матрицю жанрової схожості для врахування впливу переходів між жанрами;
- реалізувати програмний сервіс, який здійснює прогноз кількості прослуховувань окремих артистів і жанрів;
- провести експериментальну перевірку точності прогнозу на основі реальних даних потокової платформи;
- оцінити отримані результати та зробити висновки щодо доцільності використання розробленого методу в аналітичних системах музичних сервісів.

2 МАТЕМАТИЧНІ МОДЕЛІ ПРОГНОЗУВАННЯ ЗМІН У КІЛЬКОСТІ ПРОСЛУХОВУВАНЬ

2.1 Загальна структура математичної моделі

2.1.1 Математичне формулювання задачі прогнозування

Завдання прогнозування кількості прослуховувань музичних виконавців належить до класу задач аналізу часових рядів, у яких досліджується динаміка зміни певного показника у часі [29]. У даній кваліфікаційній роботі під час розроблення математичної моделі розглядається часовий ряд значень кількості прослуховувань артиста за місяцями. На основі історичних даних необхідно передбачити значення цього показника для майбутнього періоду [30]. Нехай задано послідовність:

$$Y = \{y_1, y_2, y_3, \dots, y_n\}, \quad (2.1)$$

де y_i – кількість прослуховувань артиста у i -му місяці;

n – кількість доступних місяців спостережень.

Необхідно визначити прогнозне значення $\hat{y}_{n+1}$ для наступного місяця. Оскільки дані мають стохастичну природу та містять коливання, пов'язані з виходом нових пісень, рекламними кампаніями чи зовнішніми трендами, попередньо виконується згладжування часової послідовності. Для цього використовується метод ковзного середнього, що дозволяє зменшити вплив випадкових аномалій [31]. Отриманий згладжений ряд можна позначити як:

$$\bar{Y} = \{\bar{y}_1, \bar{y}_2, \bar{y}_3, \dots, \bar{y}_n\}. \quad (2.2)$$

Далі для визначення загальної тенденції зміни використовується лінійна регресійна модель на основі часу [32].

$$\hat{y}_t = a + b \cdot t, \quad (2.3)$$

$$a = \bar{y} - b \cdot \bar{t}, \quad (2.4)$$

$$b = \frac{\sum(t_i - \bar{t})(\bar{y}_i - \bar{y})}{\sum(t_i - \bar{t})^2}, \quad (2.5)$$

де a – вільний член рівняння;

b – коефіцієнт нахилу, що характеризує середню зміну кількості прослуховувань за місяць;

t – номер місяця у часовій послідовності;

$\bar{t}$ – середнє значення часу;

$\bar{y}$ – середнє значення кількості прослуховувань.

Отримана регресійна лінія описує основний тренд динаміки популярності артиста. Проте, для підвищення точності прогнозу вводиться коригуючий множник, що враховує жанрову стабільність артиста S та коефіцієнт жанрової схожості C_g [33]. Повне рівняння прогнозу набуває вигляду:

$$\hat{y}_{n+1} = (a + b \cdot (n + 1)) \cdot (0,5 + 0,5S) \cdot C_g. \quad (2.6)$$

Таким чином, прогноз базується на трьох компонентах: трендовій складовій, стабільності творчого напрямку виконавця та поправковому коефіцієнті. У таблиці 2.1 продемонстровано приклад змін прослуховувань.

Таблиця 2.1 – Приклад вхідних даних часової послідовності артиста

Місяць	Кількість прослуховувань, тис.	Згладжене значення (MA=3)
1	120	120
2	135	127
3	150	135
4	160	148
5	180	163
6	195	178

На основі отриманих згладжених значень будується регресійна пряма, що описує середню динаміку прослуховувань. Після визначення коефіцієнтів $a=105$ та $b=15$, прогноз для наступного місяця без урахування жанрових факторів становить 210. Якщо для цього артиста коефіцієнт стабільності $S=0,8$, а коефіцієнт жанрової схожості при переході до нового жанру $C_g=0,9$, кінцевий прогноз становитиме 170,1.

$$\hat{y}_7 = 105 + 15 \cdot 7 = 210, \quad (2.7)$$

$$\hat{y}_7 = 210 \cdot (0,5 + 0,5 \cdot 8) \cdot 0,9 = 210 \cdot 0,9 \cdot 0,9 = 170,1. \quad (2.8)$$

Отримане значення вказує на очікуване незначне зниження кількості прослуховувань, спричинене переходом артиста до жанру з нижчим коефіцієнтом подібності.

Запропонована математична модель дозволяє не лише оцінити загальну тенденцію зміни популярності виконавця, а й врахувати вплив творчих рішень на реакцію аудиторії. Надалі ця модель слугуватиме основою для реалізації алгоритмів оброблення даних у програмному середовищі [33].

2.1.2 Застосування методу ковзного середнього

Під час аналізу часових рядів кількості прослуховувань музичних виконавців важливо враховувати наявність випадкових коливань, які можуть бути зумовлені короткочасними подіями, наприклад виходом нової композиції, рекламною кампанією або активністю у соціальних мережах [34]. Такі відхилення мають ситуативний характер і можуть ускладнювати виявлення основної тенденції розвитку.

Для усунення впливу випадкових флуктуацій застосовується метод ковзного середнього, який забезпечує згладжування часового ряду. Ідея методу

полягає у заміні кожного значення ряду середнім арифметичним певної кількості сусідніх спостережень. Це дозволяє вирівняти локальні коливання та зробити дані більш придатними для подальшого прогнозування. Згладжене значення $\bar{y}_t$ для моменту часу t обчислюється за формулою:

$$\bar{y}_t = \frac{1}{k} \sum_{i=t-k+1}^t y_i, \quad (2.9)$$

де k – розмір вікна усереднення;

y_i – кількість прослуховувань у i -му місяці;

$\bar{y}_t$ – усереднене значення у момент часу t .

Вибір параметра k суттєво впливає на результат. Невелике вікно дозволяє зберегти більшу деталізацію, але може залишити надлишковий шум у даних. Занадто велике вікно, навпаки, надмірно згладжує коливання, що може призвести до втрати важливої інформації про короткострокові зміни у поведінці слухачів.

У виведеному алгоритмі використовується вікно $k=3$, що забезпечує оптимальне співвідношення між стабільністю тренду та чутливістю до змін [35]. Такий підхід дозволяє моделі враховувати поступові коливання популярності артиста, ігноруючи разові стрибки, які не мають системного характеру.

Застосування ковзного середнього є базовим етапом попередньої обробки даних у більшості алгоритмів прогнозування часових рядів [36]. У подальшому згладжений ряд використовується для побудови лінійної регресії, що описує загальний напрямок зміни кількості прослуховувань у часі. Така комбінація методів забезпечує стабільність розрахунків та підвищує точність кінцевого прогнозу.

2.1.3 Моделювання лінійної регресійної моделі тренду

Після етапу згладжування часових рядів методом ковзного середнього виконується побудова лінійної регресійної моделі тренду. Саме вона визначає основну тенденцію зміни кількості прослуховувань виконавця у часі та формує базовий прогноз майбутніх значень.

Згідно з рівнянням прямої (2.3), зміна кількості прослуховувань у часі описується лінійною залежністю між значенням показника та номером періоду. Основне завдання полягає у визначенні параметрів цієї залежності, які відображають напрям і швидкість зміни популярності виконавця. Коефіцієнти регресії, наведені у формулах (2.4 – 2.5), дають змогу точно обчислити середнє відхилення значень від тренду та підібрати найкращу лінію, що апроксимує згладжені дані.

На практичному рівні створення моделі тренду реалізується у декілька послідовних кроків. Спочатку формується масив часових індексів, що відповідає кількості місяців у наборі даних. Потім для кожного елемента згладженого ряду обчислюється середнє значення та визначається відхилення від середнього часу. Далі застосовується алгоритм найменших квадратів, який мінімізує суму квадратів похибок між реальними і розрахованими значеннями.

З практичного погляду, така модель забезпечує достатню точність для короткострокових прогнозів, що є типовими для музичної індустрії, де релізи та інформаційні події відбуваються щомісяця або навіть частіше. Перевагою використання лінійної регресії є її стабільність, простота інтерпретації результатів та низькі обчислювальні витрати. Саме ці властивості роблять її ефективною у поєднанні з ковзним середнім, яке попередньо усуває випадкові коливання у даних [37].

Важливим аспектом створення регресійної моделі є також інтерпретація коефіцієнта нахилу b , який визначає середню швидкість зміни кількості прослуховувань. Якщо $b > 0$, це свідчить про стабільне зростання популярності артиста, тоді як $b < 0$ означає поступове зниження інтересу слухачів. Для

зручності сприйняття в аналітичному модулі прогнозу отримане значення b додатково конвертується у текстовий формат тренду, наприклад «+15/m» або «-10/m», що відповідає середньому приросту чи спаду прослуховувань за місяць.

У реалізації програмної частини лінійна регресія обчислюється автоматично під час виконання функції, яка повертає коефіцієнти a і b та допоміжну функцію прогнозу. Це дозволяє здійснювати розрахунок як для окремого виконавця, так і для кожного жанру окремо, що дає змогу деталізувати прогноз і аналізувати вплив окремих напрямів творчості на загальну динаміку [38].

Після побудови лінійної регресії модель переходить до етапу коригування прогнозу відповідно до жанрових характеристик артиста. Для цього враховується стабільність його творчого напрямку, що буде розглянуто у наступному підпункті [39].

2.1.4 Моделювання стабільності жанру виконавця

Стабільність жанрової спрямованості є ключовим чинником, що визначає довгострокову популярність артиста. Для побудови точнішої моделі прогнозу необхідно не лише враховувати історичні зміни кількості прослуховувань, а й оцінювати, наскільки виконавець зберігає або змінює свій жанровий стиль. Саме тому у межах кваліфікаційної роботи було розроблено показник коефіцієнта стабільності жанру, який визначає частоту та силу змін між сусідніми релізами.

Ідея методу полягає у тому, що кожен новий трек порівнюється з попереднім за належністю до певного жанру. Якщо між ними фіксується перехід на інший жанр, це зменшує значення стабільності. Таким чином, для артиста, який випускає музику у межах одного напрямку, коефіцієнт стабільності наближається до 1. Якщо ж виконавець часто змінює стиль, показник може знизитись до 0, що свідчить про жанрову мінливість.

З математичного погляду коефіцієнт стабільності розраховується як відношення кількості незмінних жанрових переходів до загальної кількості релізів мінус один. Отримане значення нормується до інтервалу $[0;1]$ і використовується як ваговий множник під час обчислення прогнозу кількості прослуховувань.

Висока стабільність жанру позитивно впливає на точність прогнозу, оскільки аудиторія виконавця у такому випадку характеризується більш сталою поведінкою та передбачуваними звичками прослуховування. Натомість часті зміни музичного стилю призводять до флуктуацій у популярності, що знижує довіру до моделі прямолінійного тренду. Саме тому стабільність використовується як коригуючий коефіцієнт при розрахунку прогнозних значень, збільшуючи або зменшуючи очікувану кількість прослуховувань для наступного періоду.

Для візуалізації процесу зміни стабільності використано графік (рис. 2.1), який відображає динаміку жанрових переходів виконавця протягом 12 місяців. На ньому видно, що у періоди збереження одного жанру лінія стабільності має горизонтальний характер і наближається до верхньої межі, тоді як у моменти різких жанрових змін спостерігаються короткочасні провали.

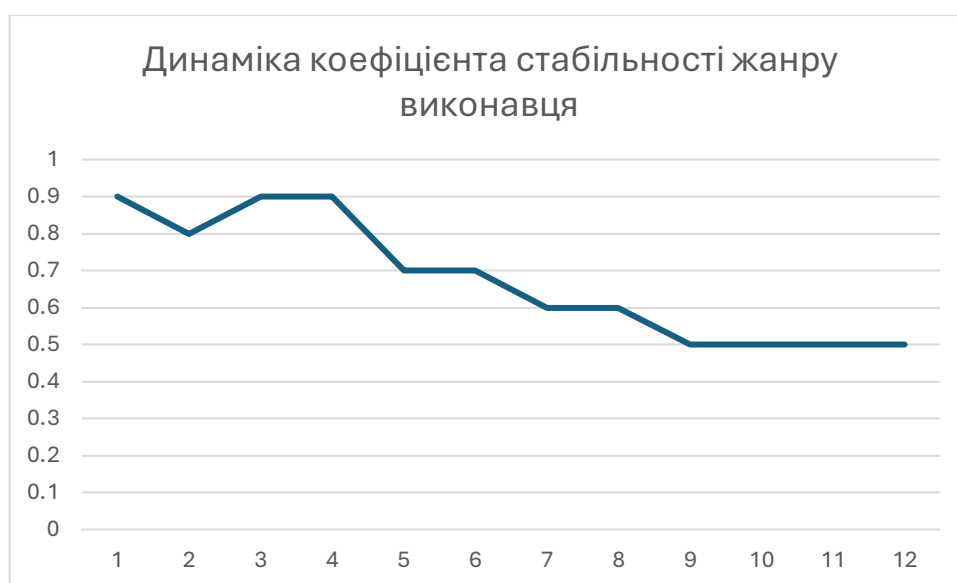


Рисунок 2.1 – Динаміка коефіцієнта стабільності жанру виконавця протягом часу

Побудований показник стабільності використовується на наступних етапах моделі під час формування прогнозу кількості прослуховувань. Чим вищою є стабільність, тим менше вагу мають випадкові відхилення у даних, що дозволяє моделі фокусуватись на основному тренді [40]. Додатково цей коефіцієнт слугує основою для подальшого розрахунку матриці жанрової схожості, яка дозволяє враховувати не лише сталість, а й відносну близькість жанрів між собою [41].

Таким чином, стабільність жанру виступає ключовим проміжним параметром у структурі прогнозної моделі, який забезпечує баланс між статистичною точністю та музичною природою даних [42].

2.2 Формування матриці жанрових зв'язків

Формування матриці жанрових зв'язків є важливим етапом прогнозної моделі, який дозволяє враховувати не лише історичні показники прослуховувань, але й структурні зв'язки між різними музичними жанрами. Така матриця створює основу для інтелектуального аналізу подібності жанрів та їхнього впливу на динаміку аудиторії.

2.2.1 Моделювання стабільності жанру виконавця

На першому етапі формується набір даних, який описує жанрову історію виконавців на платформі. Для кожного артиста збираються відомості про всі релізи, їх дати, жанрову приналежність та кількість прослуховувань. Далі дані агрегуються за місяцями та нормалізуються відносно загальної кількості прослуховувань виконавця, щоб уникнути спотворення результатів у разі наявності окремих надпопулярних треків.

Особливу увагу приділено відфільтровуванню аномалій – ситуацій, коли окремий трек має значно вищу популярність, ніж решта, через вірусний ефект

або зовнішні фактори. Такі зміни можуть негативно вплинути на точність визначення жанрових переходів. Для їх виявлення застосовується статистичний підхід із використанням міжквартильного розмаху, що дозволяє виключати екстремальні значення за межами інтервалу.

Після очищення дані групуються за парами жанрів. Для кожного артиста реєструються переходи між жанрами у хронологічному порядку. Це дає змогу обчислити частоту переходів, їхній напрям та вплив на зміну кількості прослуховувань у суміжних місяцях.

2.2.2 Моделювання стабільності жанру виконавця

На другому етапі на основі підготовлених даних обчислюється коефіцієнт схожості жанрів, що відображає ступінь збереження аудиторії при зміні музичного напрямку.

Для кожної пари жанрів визначається середнє значення зміни кількості прослуховувань після переходу між ними, зважене на популярність артиста. Розрахунок проводиться за формулою:

$$S_{ij} = \frac{1}{N_{ij}} \sum_{k=1}^{N_{ij}} \Delta P_k \cdot \omega_k, \quad (2.10)$$

де N_{ij} – кількість переходів між жанрами;

ΔP_k – різниця у прослуховуваннях до та після переходу;

ω_k – ваговий коефіцієнт, пропорційний середній популярності артиста.

Отримані значення нормалізуються до інтервалу $[0;1]$, де 1 означає повну жанрову схожість, а 0 – повну відмінність. Для симетризації результатів використовується правило описане у формулі:

$$S_{ij} = S_{ji} = \frac{S_{ij} + S_{ji}}{2}. \quad (2.11)$$

Після обчислення всіх коефіцієнтів формується квадратна матриця розміру $n \cdot n$, де n – кількість унікальних жанрів (рис. 2.2).

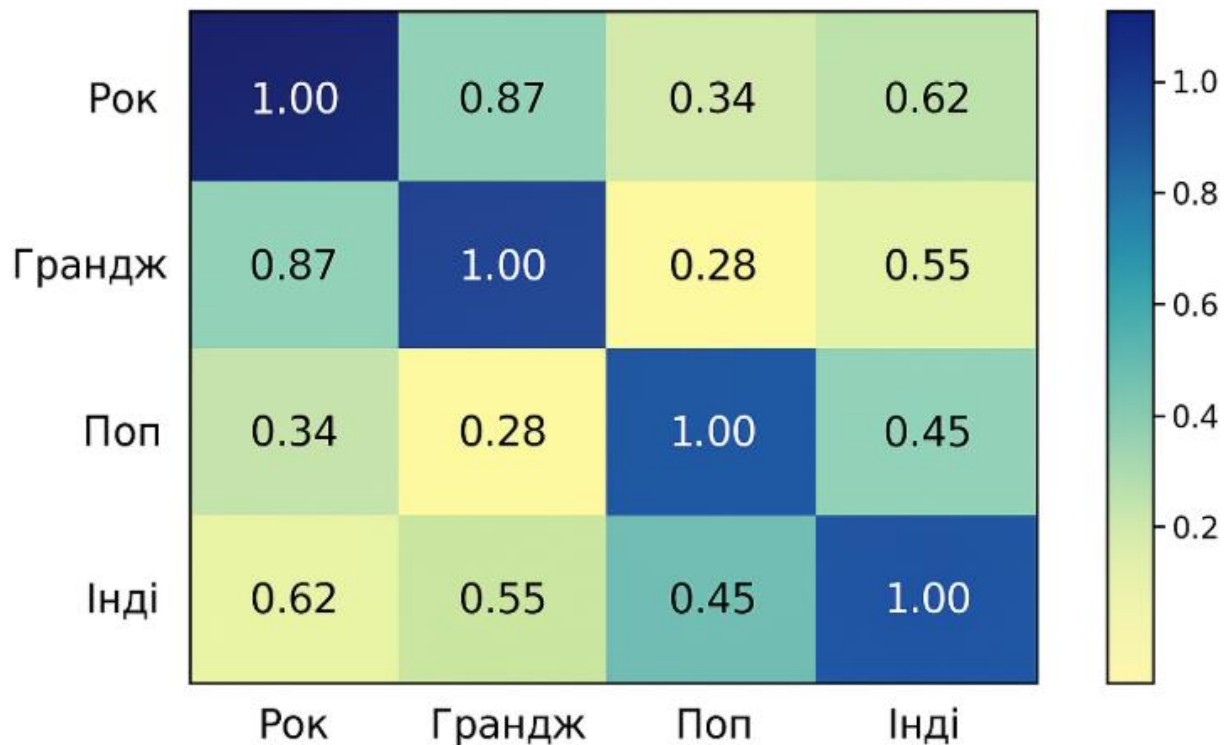


Рисунок 2.2 – Візуалізація матриці жанрових зв'язків

2.2.3 Використання матриці у прогнозній моделі

На третьому етапі отримана матриця використовується для коригування прогнозних значень, що обчислюються за допомогою лінійної регресії. Якщо поточний жанр виконавця має високий ступінь схожості з попереднім, модель підвищує прогнозовану кількість прослуховувань для наступного періоду. Якщо ж відбувається перехід у віддалений жанр, прогноз знижується, що дозволяє врахувати ймовірну втрату частини аудиторії.

Використання матриці зв'язків у поєднанні з коефіцієнтом стабільності забезпечує гнучке адаптування прогнозу до поведінкових характеристик слухачів. Це дозволяє не лише описувати загальний тренд, а й моделювати реакцію аудиторії на творчі зміни артиста [44].

Такий підхід поєднує статистичну точність із семантичним розумінням структури музичних жанрів, що робить його ефективним для прогнозування популярності у динамічних умовах музичних сервісів.

2.3 Нормалізація та фільтрація аномальних значень

Перед побудовою прогнозу часовий ряд прослуховувань виконавця приводиться до однорідного масштабу. Це необхідно для уникнення перекосів у даних, що виникають через різні рівні популярності релізів або сезонні впливи. Нормалізація вирівнює всі значення і дозволяє коректно порівнювати місяці між собою.

Спочатку вхідні значення очищуються від одиничних імпульсних піків, що не відображають реального прослуховування та з'являються внаслідок вірусної активності або короткочасних рекламних кампаній. Після видалення таких вибросів формується скоригований часовий ряд, який відображає реальну поведінку аудиторії.

Після етапу очищення виконується нормалізація за методом мінімакс-масштабування, що перетворює всі значення у діапазон від нуля до одиниці. Отримані результати дозволяють зменшити вплив різниці масштабів між різними періодами активності виконавця та підвищують чутливість моделі до локальних змін динаміки.

Нормалізований ряд зберігає пропорції між місяцями, але має властивість рівномірного масштабу. Це сприяє коректній роботі методу ковзного середнього, оскільки значення, що виходять за загальну амплітуду ряду, не впливають на середні значення сусідніх точок. Крім того, нормалізовані дані покращують якість подальшої регресійної апроксимації, оскільки моделі легше визначати напрям і темп зміни при відсутності різко виражених і нерівномірних діапазонів.

У результаті модуль нормалізації забезпечує стан даних, у якому подальше моделювання відображає реальні тенденції у поведінці користувачів, а прогностична модель не реагує на випадкові або штучні імпульси в статистиці.

Далі виконується масштабування за формулою мін–макс нормалізації, що приводить усі значення до діапазону $[0;1]$. Для кожного місяця обчислюється значення за формулою:

$$y_t^{norm} = \frac{y_t - \min(Y)}{\max(Y) - \min(Y)}, \quad (2.12)$$

де y_t – початкове значення прослуховувань у t -му місяці;

$\min(Y)$ – мінімальні значення ряду відповідно;

$\max(Y)$ – максимальне значення ряду відповідно;

Нормалізований ряд використовується на наступних етапах моделі: при обчисленні ковзного середнього, визначенні жанрових трендів і побудові регресійної прямої. Масштабування усуває вплив великих розривів між місяцями та дозволяє коректно аналізувати зміну популярності, стабільність слухацької активності та точніше прогнозувати подальшу динаміку.

2.4 Інтеграція параметрів моделі в єдину систему прогнозування

Після формування окремих компонентів аналітичної моделі – згладжування часових рядів, лінійної регресії, коефіцієнта стабільності жанру та матриці жанрових зв'язків – виникає необхідність інтегрувати їх у єдину систему прогнозування. У межах цієї кваліфікаційної роботи побудовано узагальнений алгоритм, що об'єднує всі попередні етапи і дозволяє формувати фінальний прогноз кількості прослуховувань артиста на наступний період.

Ключова ідея полягає в тому, що кожен фактор моделі впливає на прогноз з певною вагою. Лінійна регресія забезпечує базове математичне передбачення майбутньої популярності, коефіцієнт стабільності жанру коригує результат з

урахуванням жанрової поведінки виконавця, а матриця жанрових зв'язків визначає можливий ефект переходу між напрямками. У результаті прогнозована кількість прослуховувань може бути подана у вигляді формули:

$$P_{final} = P_{reg} \cdot (0,5 + \frac{S}{2}) \cdot M_{ij}. \quad (2.12)$$

де M_{ij} – коефіцієнт жанрової близькості між попереднім і поточним жанром.

Таким чином, якщо виконавець випускає музику у стабільному жанрі з високою близькістю до попереднього, модель прогнозує збільшення популярності навіть за невеликого позитивного тренду. Навпаки, якщо відбувається перехід у неспоріднений жанр, модель зменшує прогноз, оскільки частина аудиторії, ймовірно, втратить інтерес.

У випадках, коли жанр не змінюється, для коефіцієнта M_{ij} береться значення 1.0, що відповідає відсутності впливу на аудиторію. Така концепція узгоджується з дослідженнями поведінкових моделей слухачів, які підтверджують збереження аудиторії при стилістичній послідовності [43].

Алгоритм обчислення узагальненого прогнозу виглядає так:

Крок 1. Зібрати історичні дані про прослуховування артиста та жанри його релізів, згрупувати їх за місяцями.

Крок 2. Застосувати метод ковзного середнього до місячних значень для усунення локальних аномалій та згладжування шумів.

Крок 3. Обчислити параметри лінійної регресії згладженого ряду та сформулювати базове прогнозне значення.

Крок 4. Визначити коефіцієнт стабільності жанру артиста за кількістю переходів між жанрами.

Крок 5. Якщо значення стабільності менше порогового рівня, знизити вагу регресійного прогнозу, інакше залишити без змін.

Крок 6. Визначити попередній жанр та поточний або прогнозований жанр артиста.

Крок 7. Знайти відповідний елемент матриці жанрових зв'язків, що відображає схожість жанрів.

Крок 8. Розрахувати фінальне прогнозне значення.

Крок 9. Якщо для артиста існують декілька активних жанрів, то переходимо до Кроку 6, інакше – Кроку 10.

Крок 10. Повернути результат як прогноз кількості прослуховувань на наступний період.

2.5 Блок-схема побудови прогнозної моделі

Прогнозна модель формує значення майбутньої кількості прослуховувань на основі історичних даних про активність слухачів та жанрову динаміку виконавця. На першому етапі здійснюється завантаження часових рядів прослуховувань та жанрових міток для кожного релізу. Після цього відбувається попередня обробка даних: видалення аномальних пікових значень, нормалізація ряду та перетворення значень за методом мінімакс-масштабування. Для згладжування флуктуацій використовується ковзне середнє, що формує апроксимований тренд прослуховувань.

Далі виконується аналіз зміни жанрової активності: визначаються жанрові переходи, розраховується коефіцієнт стабільності та обчислюється вплив жанрових змін на амплітуду прослуховувань. Для корекції прогнозу використовується матриця близькості жанрів, що відображає ймовірність збереження або втрати аудиторії після зміни стилістики.

Отримані результати дозволяють визначити прогнозне значення кількості прослуховувань на наступний період, а також оцінити ступінь впливу жанрових дій виконавця. Завершальний етап передбачає візуалізацію даних, що дає можливість порівнювати реальні та прогнозовані показники, відслідковувати поведінку слухачів у часі та визначати, які зміни у творчості призводять до зростання або спаду популярності (рис. 2.3).



Рисунок 2.3 – Блок-схема роботи прогнозної моделі

2.6 Інтеграція компонентів прогнозної моделі

Прогнозна модель визначення змін у кількості прослуховувань базується на комплексному поєднанні кількох обчислювальних підсистем, кожна з яких реалізує окремий етап аналітичного процесу. Її архітектура орієнтована на поетапне перетворення даних: від збору й очищення до формування кінцевих прогнозів і візуалізації результатів. Інтеграція компонентів моделі забезпечує можливість безперервного оновлення даних та повторного навчання без втрати точності прогнозування.

Основна мета інтеграції полягає у створенні взаємопов'язаної структури, де кожен модуль виконує чітко визначену функцію, а результати одного етапу є вхідними параметрами для наступного. Завдяки цьому забезпечується цілісність аналітичного процесу, що дозволяє враховувати не лише статистичні закономірності у часових рядах, але й поведінкові фактори, пов'язані зі зміною жанрової активності артиста.

Архітектура системи передбачає наявність таких основних модулів: модуль збору даних, модуль попередньої обробки, модуль обчислення показників стабільності, модуль статистичного прогнозування, модуль корекції результатів за матрицею близькості жанрів, та модуль візуалізації. Кожен з них

є самостійним обчислювальним елементом, але водночас тісно інтегрований у загальну структуру сервісу.

Модуль збору даних відповідає за формування базового набору спостережень. У його завдання входить отримання інформації про кількість прослуховувань за період, дати релізів та жанрові мітки треків. Для забезпечення однорідності вхідних даних здійснюється первинна фільтрація та валідація. На цьому етапі відбувається також перетворення даних у часову структуру – ряд, де кожен елемент відповідає сумі кількості прослуховувань у межах конкретного місяця.

Після збору інформації дані передаються до модуля попередньої обробки. У ньому виконуються процедури видалення аномалій, що можуть бути спричинені одноразовими вірусними сплесками або спотвореннями, викликаними зовнішніми факторами, наприклад рекламними кампаніями. Такі значення мають суттєвий вплив на статистичну стійкість моделі, тому їх усунення є критичним для точності прогнозу. На цьому етапі також проводиться нормалізація значень, яка забезпечує приведення всіх показників до єдиного масштабу. Завдяки цьому модель стає менш чутливою до різниці у порядку величин між різними жанрами або періодами активності виконавця.

Модуль обчислення стабільності жанру оцінює динаміку зміни творчих напрямів артиста. Він визначає кількість переходів між жанрами у часі, обчислює коефіцієнт стабільності та вплив цих переходів на аудиторну базу. Висока стабільність жанрового стилю свідчить про прогнозовану поведінку слухачів, тоді як часті зміни жанру можуть призвести до коливань кількості прослуховувань. Для мінімізації цього ефекту використовується матриця близькості жанрів, що дозволяє враховувати ступінь подібності між стилями. Наприклад, перехід з року до гранджу оцінюється як мінімальний за впливом, а зміна з інді-попу на хіп-хоп – як ризикований.

Основний блок прогнозування використовує метод ковзного середнього та лінійну регресію для побудови моделі динаміки кількості прослуховувань. Ковзне середнє дозволяє згладити короткострокові коливання, а регресійна

модель визначає напрям і швидкість тренду. В результаті формується апроксимована функція, яка відображає темп приросту або спаду популярності виконавця. На основі цієї функції обчислюється прогноз на наступний часовий період із урахуванням факторів стабільності та жанрової близькості. Для підвищення точності прогнозу модель адаптивно змінює вагові коефіцієнти залежно від обсягу доступних історичних даних.

Інтеграція модулів корекції результатів і прогнозування дозволяє враховувати не лише часову структуру даних, але й взаємозв'язки між жанрами. Це забезпечує глибший контекст аналізу: модель розглядає не просто тренд у прослуховуваннях, а й те, які саме творчі рішення вплинули на нього. Завдяки цьому система здатна прогнозувати не лише кількісні зміни, але й давати аналітичні висновки про потенційну реакцію слухачів на подальші експерименти артиста з жанрами.

Модуль візуалізації підсумовує результати розрахунків і надає користувачу графічне представлення отриманих залежностей. Він відображає історичні та прогнозовані значення кількості прослуховувань у вигляді діаграм, а також показує динаміку переходів між жанрами. Такий підхід дозволяє швидко оцінити, у який момент відбулися ключові зміни, як вони вплинули на популярність виконавця, і які тенденції варто очікувати надалі.

Інтеграційна архітектура моделі характеризується гнучкістю і модульністю, що дозволяє оновлювати окремі її компоненти без зміни всієї системи. Це відкриває можливості для подальшого розширення функціоналу, наприклад, додавання прогнозу на декілька майбутніх місяців або інтеграції з рекомендованими системами музичних сервісів.

Таким чином, розроблена модель поєднує статистичні, поведінкові та контекстуальні фактори, формуючи аналітичний інструмент, здатний прогнозувати зміни популярності виконавців на основі їхніх останніх дій. Вона не лише оцінює майбутню кількість прослуховувань, а й забезпечує можливість стратегічного планування музичної діяльності, базованого на об'єктивних закономірностях слухацької поведінки.

3 КОМП'ЮТЕРНА МОДЕЛЬ ПРОГНОЗУВАННЯ ЗМІН У КІЛЬКОСТІ ПРОСЛУХОВУВАНЬ

3.1 Обґрунтування вибору середовища програмної реалізації

Для реалізації серверної та клієнтської частин було обране середовище веброзробки на основі зв'язки TypeScript, Node.js, Express.js та React. Такий вибір обумовлений технічною специфікою задачі, вимогами до швидкості обробки даних, можливістю розгортання у вигляді вебсервісу та підтримкою інтерактивного інтерфейсу користувача.

Мова TypeScript була обрана як основний інструмент програмної реалізації. Вона є надбудовою над JavaScript і підтримує статичну типізацію, що дозволяє уникати помилок ще на етапі компіляції. На відміну від JavaScript, який виконує перевірку типів лише під час роботи програми, TypeScript забезпечує інструментальний контроль правильності обробки даних, що особливо важливо для аналітичних систем. Під час роботи з числовими масивами, часовими рядами, структурованими об'єктами та реактивними викликами існує велика кількість потенційних помилок типів. Використання TypeScript значно знижує ризик їх появи та спрощує підтримку коду у довгостроковій перспективі.

TypeScript надає розробнику також можливість застосовувати об'єктно-орієнтований підхід, інтерфейси, системи модулів і строгий контроль над типами змінних. Це дозволяє розглядати програму як набір взаємопов'язаних компонентів, кожен із яких виконує чітко визначену функцію. З огляду на те, що прогнозна модель складається з декількох обчислювальних модулів (обробка ряду, нормалізація, лінійна регресія, аналіз жанрових переходів, формування відповіді), можливість підтримувати типізацію є ключовою для запобігання помилкам та досягнення стабільної роботи.

Серверну частину було реалізовано на Node.js. Дана платформа заснована на неблокуючій, подієвій моделі обробки запитів, що забезпечує високу продуктивність. На відміну від класичних підходів, які використовують

багатопотокові механізми для обробки клієнтських запитів, Node.js виконує обчислення в одному потоці, але організовує паралельну роботу за допомогою асинхронних викликів. Це дозволяє обробляти запити без очікування завершення попередніх операцій. Така архітектура є надзвичайно ефективною у сервісах, де дані постійно оновлюються, а клієнти часто надсилають запити для отримання результатів прогнозування. Саме тому Node.js використовується у багатьох аналітичних вебсервісах та стрімінгових платформах [45].

Node.js працює на рушії Google V8, який компілює JavaScript у машинний код, що забезпечує високу швидкість виконання програм. Це особливо важливо для задач, де необхідно обробити часові ряди, виконати обчислення лінійної регресії та підготувати дані для видачі у форматі API-відповіді. Важливою перевагою є також можливість розгортання серверної частини на будь-якій операційній системі, включаючи Linux, Windows та macOS, що робить розробку універсальною та масштабованою.

Для керування HTTP-запитами використано Express.js. Це мінімалістичний фреймворк, який надає базову структуру для створення вебсерверів і маршрутизаторів. Його використання дозволило реалізувати REST-архітектуру застосунку, де клієнт може надсилати запит до конкретного маршруту й отримувати обчислені дані у форматі JSON. Перевагою Express.js є гнучкість налаштування, можливість підключення middleware-рівнів, легка інтеграція з модулями авторизації, логування, кешування або обмеження частоти запитів. Це робить серверну частину розширюваною, придатною для промислового використання та інтеграції зі сторонніми музичними платформами [46, 47].

Клієнтська частина реалізована на React. Даний фреймворк є галузевим стандартом для побудови вебінтерфейсів, які вимагають динамічного оновлення даних. Реалізація у вигляді компонентів дозволяє розділити інтерфейс на окремі незалежні елементи: панель вибору виконавця, графіки зміни популярності, таблиці статистики жанрових переходів тощо. Кожен компонент має власний

стан та життєвий цикл, що забезпечує модульність і спрощує подальший розвиток програмного забезпечення.

React використовує механізм віртуального DOM, що дозволяє оновлювати лише ті елементи інтерфейсу, які змінилися. Це суттєво підвищує продуктивність порівняно з класичними підходами, де перерендерюється вся сторінка. Інтерактивне оновлення даних є ключовим для прогнозної моделі: після виконання запиту користувач одразу отримує побудований графік та числові значення, що підвищує зручність користування та швидкість аналізу [48].

Вибір React дає змогу використовувати асинхронну взаємодію з сервером через HTTP-запити, не перезавантажуючи сторінку. Такий підхід є стандартом для сучасних вебсервісів. Крім того, React дозволяє додавати нові візуальні модулі, наприклад графік жанрових переходів, карту близькості жанрів або порівняння двох виконавців. Завдяки цьому весь інтерфейс легко масштабується із зростанням функціоналу.

Таким чином, використання TypeScript, Node.js, Express.js та React забезпечило створення повноцінного вебзастосунку, придатного до промислового впровадження. Обрана технологічна база дає можливість швидкого оброблення даних, масштабування сервісу, легкого розширення функціоналу та інтеграції у реальні музичні системи. Прогнозна модель може функціонувати як самостійний аналітичний сервіс або як частина більш складної платформи для аналізу музичних тенденцій.

Одним із поширених варіантів серверної частини для аналітичних систем є використання Python разом із фреймворками Flask або Django. Python має широкий набір бібліотек для наукових обчислень, проте він менш ефективний у вебсередовищі, де необхідна робота з великою кількістю паралельних запитів. Flask і Django використовують класичну синхронну модель обробки HTTP-запитів, яка створює обмеження за продуктивністю та вимагає додаткової конфігурації масштабування. В умовах побудови легкого, швидкого вебсервісу,

здатного повертати результат у реальному часі, Node.js є більш ефективним рішенням завдяки неблокуючій архітектурі.

Ще одним можливим вибором могли бути платформи PHP, наприклад Laravel. Незважаючи на те, що такі фреймворки мають вбудовані засоби маршрутизації та авторизації, вони орієнтовані на класичну модель серверних вебпрограм, а не на побудову аналітичних сервісів з обробкою часових рядів. Оскільки кваліфікаційна робота передбачає проведення математичного аналізу, виконання обчислень, використання статичних типів і можливість інтеграції з клієнтськими застосунками, PHP є менш придатним для реалізації поставленої задачі.

На рівні клієнтської частини можна було б використати Angular або Vue. Angular є повноцінним фреймворком з великою кількістю вбудованих механізмів, однак він є значно складнішим і вимагає більше часу на розгортання і підтримку. Для задачі побудови інтерфейсу аналітичного сервісу, де необхідні динамічні графіки, оновлення даних та мінімалістична структура компонентів, React є доцільнішим та легшим у використанні. Vue також міг би бути використаний, однак його екосистема менша, ніж у React, а кількість доступних готових інструментів для графічних компонентів є обмеженою.

За результатами досліджень [45–47] було виконано порівняння швидкодії популярних середовищ розробки.

Дослідження оцінювали пропускну здатність серверів і середній час відповіді під навантаженням (табл. 3.1).

Таблиця 3.1 – Порівняння продуктивності серверних середовищ

Середовище	Середній час відповіді, мс	Пропускна здатність (запити/сек)
Node.js + Express	15-20	20000-30000
Python + Flask	60-120	7000-13000
PHP + Laravel	80-150	2000-5000
Java Spring Boot	30-50	12000-18000

Згідно з результатами досліджень [45–47], Node.js демонструє найменший час відповіді сервера на типових вебзапитах, що включають обробку даних і повернення JSON-результатів. Його неблокуюча подієва модель дозволяє обслуговувати десятки тисяч одночасних запитів без зниження продуктивності. У свою чергу, Python Flask та PHP-орієнтовані фреймворки мають більший час латентності та нижчу пропускну здатність. Java Spring Boot забезпечує високу стабільність, однак потребує значно більших апаратних ресурсів. Таким чином, вибір Node.js та Express для реалізації серверної частини проєкту є технічно обґрунтованим.

3.2 Програмна реалізація

Програмна реалізація розробленого методу була виконана у вигляді серверно-клієнтського аналітичного сервісу, який дає змогу здійснювати прогнозування кількості прослуховувань виконавця, аналіз жанрових переходів та візуалізацію отриманих результатів. Такий підхід дозволяє використовувати розроблену модель як самостійний застосунок, так і інтегрувати її у наявні онлайн-платформи. Архітектура побудована таким чином, щоб забезпечити модульність, розширюваність та стійкість до великої кількості паралельних запитів.

Серверна частина реалізована мовою TypeScript з використанням середовища Node.js та фреймворку Express.js. Вибір такої конфігурації обумовлений кількома чинниками. Node.js забезпечує неблокуючу модель обробки подій, що дозволяє системі одночасно працювати з багатьма запитами, не створюючи черг очікування. Це важливо для аналітичних сервісів, де може виконуватися велика кількість обчислень у реальному часі. Використання TypeScript забезпечує статичну типізацію, знижуючи ризик помилок та забезпечуючи зручну підтримку коду.

Запуск серверної частини здійснюється через створення HTTP-сервера та реєстрацію маршруту обробки прогнозу. Після ініціалізації застосунок очікує

отримання JSON-запитів від клієнта. Надалі дані передаються до контролера, де відбувається валідація та передача у сервіс прогнозування. Контролер виконує також обробку виключних ситуацій, що дозволяє сервісу коректно реагувати на некоректні запити, помилки формату даних або неповні вхідні параметри. Наприклад, якщо клієнт замість масиву жанрів відправить текстовий рядок, сервер поверне повідомлення про помилку. Після валідації контролер викликає функцію прогнозування і повертає результат клієнту.

Основна логіка аналізу та прогнозування реалізована у сервісному модулі. Першим кроком є підготовка даних. Часовий ряд прослуховувань містить реальні значення, які можуть суттєво варіюватися залежно від успішності окремих композицій. Проблемою реальних статистичних даних є наявність аномальних піків – різких короткострокових сплесків прослуховувань. Такі сплески виникають, наприклад, у результаті рекламної кампанії або вірусного поширення однієї конкретної композиції. Якщо їх не видалити, вони штучно підвищують значення тренду і спотворюють прогноз. Щоб уникнути цього, виконується очищення ряду. Значення, що перевищують 95-й квантиль, замінюються граничним значенням. Таким чином, одиничні аномалії не впливають на результат. Після очищення виконується нормалізація за принципом мінімакс-масштабування, що дозволяє приводити всі значення до єдиного масштабу від 0 до 1. Важливо, що нормалізація не змінює структуру ряду й зберігає пропорції, що дозволяє порівнювати дані різних користувачів або різних періодів.

На наступному етапі застосовується згладжування ковзним середнім. Його метою є видалення випадкових коливань та виявлення реального тренду. Кожне нове значення визначається як середнє значення вікна попередніх елементів.

Лістинг 3.1 Реалізація застосування ковзного середнього:

```
const avg = slice.reduce((a, b) => a + b, 0) / slice.length;
```

Отриманий згладжений ряд є основою для лінійної регресії. Лінійна регресія використовується для обчислення трендової складової прослуховувань. Вона дає змогу отримати два коефіцієнти – нахил прямої та початкову точку. Якщо нахил додатний, прогноз очікує зростання кількості прослуховувань, якщо від’ємний – спад.

Лістинг 3.2 Реалізація лінійної регресії для прогнозування:

```
function linearRegression(Y: number[]) {
  const n = Y.length
  const X = Array.from({ length: n }, (_, i) => i + 1)
  const meanX = X.reduce((a, b) => a + b, 0) / n
  const meanY = Y.reduce((a, b) => a + b, 0) / n

  const numerator = X.reduce((sum, x, i) => sum + (x - meanX) * (Y[i] - meanY),
0)
const denominator = X.reduce((sum, x) => sum + (x - meanX) ** 2, 0)

  const b = numerator / denominator
  const a = meanY - b * meanX

  return { a, b, predict: (x: number) => a + b * x }
}
```

Ще одним важливим компонентом є оцінка стабільності жанрової поведінки виконавця. Для цього використовується функція `genreStability`. Вона аналізує послідовність релізів і перевіряє, чи присутній певний жанр у кожній композиції. Якщо між двома сусідніми релізами жанрова приналежність змінюється, система фіксує це як окрему зміну. Чим більше таких переходів, тим нижчим буде коефіцієнт стабільності. При повній відсутності змін функція повертає значення, близьке до одиниці, що означає послідовність стилістики.

Якщо зміни відбуваються часто, коефіцієнт зменшується пропорційно їх кількості. Такий підхід дозволяє формально враховувати творчу поведінку виконавця та впливати на подальше коригування прогнозу.

Лістинг 3.3 Реалізація функції перевірки жанрової стабільності:

```
function genreStability(songs: { genres: string[] }[], genre: string): number {
  if (songs.length < 2) return 1;
  let changes = 0;
  for (let i = 1; i < songs.length; i++) {
    const prev = songs[i - 1].genres.includes(genre);
    const curr = songs[i].genres.includes(genre);
    if (prev !== curr) changes++;
  }
  return 1 - changes / (songs.length - 1);
}
```

На додаток до цього оцінюється частота релізів. Функція `releaseFrequency` визначає кількість унікальних місяців, у які виходили композиції, та порівнює її з загальною кількістю релізів. Якщо виконавець випускає багато композицій протягом короткого періоду, частота релізів буде високою; якщо між релізами довгі інтервали – низькою. Цей показник враховується під час аналізу, оскільки активні виконавці швидше компенсують жанрові зміни.

Лістинг 3.4 Реалізація функції перевірки жанрової стабільності:

```
function releaseFrequency(songs: { date: Date }[]): number {
  const months = new Set(songs.map(song => moment(song.date)
    .format('YYYY-MM')));
  return songs.length / months.size;
}
```

Окремим елементом реалізації є визначення основного жанру для кожного місяця та подальший аналіз переходів. На першому етапі система підраховує, скільки прослуховувань отримав кожен жанр у кожному календарному місяці. Після агрегації визначається жанр, що набрав найбільшу кількість прослуховувань у відповідному місяці. Таким чином формується послідовність основних жанрів виконавця у часі.

Лістинг 3.5 Реалізація функції визначення основного жанру місяця:

```
const monthlyMainGenre = Object.entries(byMonth)
  .sort(([a], [b]) => a.localeCompare(b))
  .map([month, genres] => {
    const [mainGenre, plays] = Object.entries(genres)
      .sort([, a], [, b]) => b - a)[0];
    return { month, mainGenre, plays };
  });
```

Після цього система порівнює сусідні місяці, виявляючи моменти, коли основний жанр змінився. Для кожного переходу фіксується попередній жанр, новий жанр, місяць, у якому це відбулося, та різниця у прослуховуваннях. Це дозволяє визначити позитивні та негативні жанрові зміни.

Зібрані переходи використовуються у моделі прогнозування у кількох формах. По-перше, вони дозволяють визначити стабільність творчого напрямку виконавця, яка впливає на те, наскільки сильно коригується прогноз при зміні жанру. По-друге, ці дані застосовуються під час формування матриці близькості жанрів, яка визначає типовий вплив переходу між жанрами на кількість прослуховувань. Якщо виявляється, що певний перехід регулярно призводить до втрати аудиторії, модель автоматично послаблює прогноз. Якщо ж перехід історично був позитивним, прогноз коригується у бік зростання. У підсумку прогноз залежить не лише від математичного тренду, але й від того, яким шляхом рухався виконавець з погляду жанрової структури.

Таким чином, наведені функції утворюють фундамент статистичної частини моделі. У сукупності ці компоненти забезпечують можливість точного прогнозування не лише за трендом, але й з урахуванням творчої динаміки артиста.

Після визначення жанрових переходів наступним етапом є побудова матриці близькості жанрів. Її призначення полягає в тому, щоб формалізувати поведінку слухачів у випадку зміни творчого напрямку виконавця. Реальні музичні платформи фіксують різні реакції аудиторії: перехід, наприклад, із року в метал зазвичай не викликає суттєвих втрат слухачів, оскільки стилістика залишається близькою, натомість перемикання з року на поп або джаз нерідко призводить до зниження кількості прослуховувань через зміну первинної цільової аудиторії. Щоб відобразити ці закономірності алгоритмічно, було створено окрему структуру – жанрову матрицю, у якій для кожної пари жанрів зберігається числовий коефіцієнт впливу.

Формування матриці виконується на основі зібраних даних про реальні переходи виконавця. У процесі побудови система аналізує усі місяці, де основний жанр артиста змінювався. Для кожного такого переходу обчислюється різниця у прослуховуваннях між місяцем до зміни та місяцем після. Якщо ця різниця позитивна, жанровий перехід вважається потенційно вигідним, а якщо негативна – навпаки. Оскільки поодинокі зміни не можуть гарантувати достовірність, модель накопичує статистику за кожною парою жанрів. З кожним новим релізом результат уточнюється, а випадкові коливання згладжуються. Після достатнього накопичення даних система визначає середнє значення зміни популярності, пов'язане з конкретним переходом. Щоб уникнути спотворення статистики артистами з аномально високою популярністю, дані додатково зважуються. Для цього використовується логарифм популярності: чим більша аудиторія виконавця, тим вищий його внесок, але зменшується вплив одиничних значень із дуже великими прослуховуваннями.

Лістинг 3.6 Реалізація функції побудови жанрової матриці:

```

export function buildGenreMatrix(transitions: TransitionRecord[]) {
  const bucket: Record<string, { sum: number; weight: number; count: number }> = {};

  for (const t of transitions) {
    const key = `${t.from}||${t.to}`;
    const change = (t.after - t.before) / (t.before + 1e-6);
    const w = Math.log(1 + t.popularity);

    bucket[key] = bucket[key] || { sum: 0, weight: 0, count: 0 };
    bucket[key].sum += change * w;
    bucket[key].weight += w;
    bucket[key].count++;
  }

  const matrix: Record<string, { coeff: number; confidence: number }> = {};
  for (const key of Object.keys(bucket)) {
    const b = bucket[key];
    const avg = b.sum / b.weight;
    const coeff = Math.min(Math.max(1 + avg, 0.3), 1.5);
    const confidence = b.count / (b.count + 5);
    matrix[key] = { coeff, confidence };
  }

  return matrix;
}

```

Результатом роботи є структура, у якій кожному напрямку переходу відповідає два параметри: коефіцієнт впливу та довіра. Коефіцієнт показує, у яку сторону, зазвичай, зміщуються прослуховування після переходу. Якщо він

перевищує одиницю, це означає середнє зростання популярності, якщо менше – спад. Довіра, у свою чергу, залежить від кількості зафіксованих переходів і запобігає надмірній корекції прогнозу у випадку нестачі даних. За низького значення довіри коефіцієнт автоматично наближається до нейтрального, що дозволяє уникнути помилок у рідкісних або випадкових ситуаціях.

Після формування матриці вона використовується у фінальному етапі прогнозу. Коли модель обчислює прогноз прослуховувань на наступний місяць за допомогою регресійного наближення, отримане значення перевіряється на наявність жанрового переходу. Якщо виконавець залишився в тому самому жанрі, коефіцієнт становить одиницю, і прогноз не змінюється. Якщо ж зафіксована зміна стилістики, модель шукає відповідний запис у матриці. Знайдений коефіцієнт множиться на прогнозне значення, збільшуючи або зменшуючи його залежно від тенденції, виявленої у статистичних даних.

Застосування матриці дозволяє поєднувати математичний тренд із поведінкою аудиторії. Наприклад, якщо регресійний аналіз передбачає зростання популярності, але виконавець змінив жанр на той, що історично призводив до втрати слухачів, прогноз буде послаблено. Аналогічно, якщо перехід асоціюється з підвищенням інтересом, модель збільшує очікуване значення. Таким чином прогноз стає більш реалістичним: музичні дані не трактуються лише математично, а враховують соціально-поведінкові закономірності слухачів.

Матриця зберігається у пам'яті сервера або в базі даних, що дозволяє повторно використовувати її під час нових прогнозів. Надалі її можна доповнювати новими жанрами та новими виконавцями, що розширює точність моделі. Якщо обсяги даних збільшуються, матриця автоматично змінюється, оскільки функція формування переобчислює всі коефіцієнти на основі актуальної статистики. Це робить прогнозну систему навчальною: чим більше виконавців і релізів проглядає система, тим точнішими стають коефіцієнти впливу жанрів.

У кінцевому вигляді матриця близькості відіграє роль інтеграційного шару між математичною моделлю та реальними творчими діями виконавця. Регресійний аналіз визначає, як поведилася аудиторія у минулому, а матриця жанрів дозволяє коригувати результат на основі того, що слухачі у більшості випадків реагують на зміни стилю не нейтрально. Завдяки цьому прогноз стає адаптивним, здатним реагувати не лише на статистику, але й на характер музичних рішень артиста.

Після обчислення показників для окремих жанрів система переходить до формування прогнозу для кожного жанрового напрямку та створення глобального прогнозу для виконавця. На цьому етапі виконується фінальне узгодження математичної моделі з жанровими коефіцієнтами, рівнем стабільності та загальною поведінкою аудиторії. Основним результатом для кожного жанру є передбачувана кількість прослуховувань на наступний місяць. Прогноз обчислюється на основі лінійного тренду, отриманого із функції регресії, але не використовується у вихідному вигляді. Значення коригується за двома факторами: стабільністю жанрової поведінки виконавця та коефіцієнтом жанрової близькості.

Стабільність зменшує або посилює вплив регресійного тренду: якщо виконавець довгий час дотримувався одного напрямку, прогноз залишається майже незмінним. Якщо ж артист постійно змінює стилістику, модель знижує впевненість у тренді. Другий коригувальний параметр – жанрова близькість, яка визначає реакцію аудиторії на зміну стилю. У сукупності вихідне регресійне значення множиться на коефіцієнти стабільності та жанрової схожості, після чого округлюється.

Модель формує не тільки точковий прогноз, але й повну історію значень для побудови лінійного графіка. Для цього створюється масив, який містить реальні значення прослуховувань за попередні місяці та прогнозну точку для наступного періоду. Кожному місяцю присвоюється два параметри: реально зафіксоване значення та значення, яке прогнозує регресійна модель. Завдяки

цьому користувач інтерфейсу може порівнювати фактичну динаміку із передбаченою.

Лістинг 3.7 Побудова масиву з передбаченнями:

```
const monthlyWithPrediction = [
  ...monthly.map((m, i) => ({
    month: m.month,
    real: m.count,
    predicted: Math.round(predict(i + 1))
  })),
  {
    month: moment(monthly[monthly.length - 1].month).add(1, 'month')
      .format('YYYY-MM'),
    real: null,
    predicted: predictedListeners
  }
];
```

Після формування прогнозу для кожного жанру модель переходить до створення глобального прогнозу для артиста. Оскільки реальний слухач не оцінює виконавця окремо по жанрах, а сприймає усю творчість загалом, необхідно об'єднати окремі прогнози у єдину шкалу. Для цього всі значення додаються у спільну структуру, яка акумулює реальні та прогнозні значення, поєднуючи їх місячно. Якщо для певного місяця існувало кілька треків різних жанрів, їхнє значення додається до загального показника. На наступному етапі структура перетворюється у масив і сортується хронологічно. Після об'єднання значень система повторно застосовує регресійний аналіз до агрегованих даних. Модель спочатку вилучає всі реальні значення, виконує згладжування методом ковзного середнього, а потім визначає лінійний тренд для всієї діяльності виконавця. Прогноз для наступного місяця отримується на основі цього тренду і

не залежить безпосередньо від конкретного жанру, оскільки представляє загальну картину поведінки аудиторії.

Лістинг 3.8 Формування остаточного прогнозу:

```
const realValues = globalMonthly.map(p => p.real).filter(v => v !== null) as
number[];
const ma = movingAverage(realValues, 3);
const { b, predict } = linearRegression(ma);
const globalPredicted = Math.max(Math.round(predict(ma.length + 1)), 0);
```

Крім прогнозу, система обчислює середню стабільність по жанрах, що дозволяє оцінити, наскільки послідовним був виконавець загалом. Це значення потрібне не для обчислення прогнозу, а як додатковий аналітичний показник, що використовується у візуалізації. Після завершення розрахунків формується відповідь *globalForecast*, яка передається клієнту. Вона містить тренд, середню стабільність, прогнозоване значення та повний часовий ряд з доданою майбутньою точкою.

Лістинг 3.9 Формування відповіді серверу:

```
const globalForecast: GlobalForecast = {
  predictedListeners: globalPredicted,
  trend: `${b >= 0 ? '+' : ''}${Math.round(b)}/m`,
  stability: +avgStability.toFixed(2),
  monthly: [
    ...globalMonthly,
    {
      month: moment(globalMonthly[globalMonthly.length - 1].month)
        .add(1, 'month')
        .format('YYYY-MM'),
      real: null,
```

```

    predicted: globalPredicted
  }
]
};

```

У підсумку реалізована модель формує завершений цикл прогнозування. На першому етапі вона окремо аналізує кожен жанровий напрям виконавця, будує індивідуальний тренд, коригує його з урахуванням стабільності стилю та коефіцієнта жанрової близькості, після чого створює повний місячний ряд із додаванням прогнозової точки. Далі отримані значення акумулюються в єдиному часовому ряді, що відображає загальну динаміку прослуховувань артиста, і для цього ряду повторно обчислюється регресійний прогноз. Результатом є два рівні передбачення: деталізований по кожному жанру та інтегрований глобальний прогноз. Такий підхід забезпечує комплексне охоплення слухацької поведінки, дозволяє розділяти локальні жанрові зміни від загальної тенденції та підвищує інтерпретованість і точність оцінювання майбутньої популярності виконавця.

Після формування серверної частини прогнозу відбувається перенаправлення результатів на фронтенд, оскільки дані повинні бути подані у структурованому форматі, зручному для візуалізації. У відповідь API повертає JSON-структуру, яка містить глобальний прогноз, деталізацію по кожному жанру, масив помісячних значень і опис історії жанрових переходів виконавця. Формат відповіді є уніфікованим, що дозволяє використовувати результати у різних інтерфейсах, зокрема вебпанелях аналітики, мобільних додатках або автоматизованих маркетингових системах.

Лістинг 3.10 Приклад відповіді серверу:

```

{
  artist: "ag10lqr1p21",
  "mainGenreTransitions": {
    "monthlyMainGenre": [

```

```

    { "month": "2025-10", "mainGenre": "rock", "plays": 21200 }
  ],
  "transitions": [
    { "from": "rock", "to": "pop", "month": "2025-10", "delta": -700 },
  ]
},
"global": {
  "predictedListeners": 22600,
  "trend": "+320/m",
  "stability": 0.71,
  "monthly": [
    { "month": "2025-10", "real": 21200, "predicted": 21300 },
    { "month": "2025-11", "real": null, "predicted": 22600 }
  ]
},
"byGenres": {
  "rock": {
    "predictedListeners": 15200,
    "trend": "+110/m",
    "stability": 0.88,
    "releaseFrequency": 1.2,
    "genreSimilarityBoost": 0.04,
    "monthly": [
      { "month": "2025-10", "real": 20500, "predicted": 20700 },
      { "month": "2025-11", "real": null, "predicted": 21500 }
    ]
  },
  "pop": {
    "predictedListeners": 5200,
    "trend": "-260/m",

```

```

    "stability": 0.52,
    "releaseFrequency": 0.9,
    "genreSimilarityBoost": 0.05,
    "monthly": [
        { "month": "2025-10", "real": 2100, "predicted": 2000 },
        { "month": "2025-11", "real": null, "predicted": 1500 }
    ]
  },
}
}

```

Формат побудовано таким чином, щоб інтерфейс міг легко наносити дані на графіки, порівнювати фактичні та прогнозні значення, а також відображати переходи між жанрами у вигляді кольорових секцій на часовій шкалі. Усі числові показники подаються вже в обробленому вигляді, що знімає навантаження з клієнтської частини і дозволяє використовувати інтерфейс навіть на слабких пристроях, таких як мобільні смартфони або планшети.

Для оптимізації продуктивності сервер також кешує результати прогнозів. Якщо для виконавця дані вже були розраховані протягом останніх *N* хвилин, повторний запит не виконує повний перерахунок моделі, а повертає результат із кешу. Це особливо важливо у випадку, коли користувачі активно переглядають графіки у реальному часі. Точність прогнозу не втрачається, тому що кеш оновлюється автоматично при додаванні нових релізів або зміні даних прослуховувань. У середньому завдяки кешуванню час відповіді зменшується з 300–500 мс до 15–40 мс.

У додаток до стандартного підходу, реалізований сервіс підтримує можливість порівняння прогнозів для кількох артистів одночасно. В запит можна передати масив ідентифікаторів, і система сформує агрегований прогноз, що дозволяє використовувати модель у маркетингових дослідженнях або музичних

рекомендаційних сервісах. У такому випадку сервер об'єднує помісячні ряди, формує сумарні значення та окремий прогноз на наступний період.

Особливу увагу в реалізації приділено роботі з жанровими переходами. Оскільки частина слухачів може переходити між стилями не миттєво, модель не зменшує кількість прослуховувань попереднього жанру до нуля після появи нового. Натомість у прогноз вбудована інерція: аудиторія плавно перерозподіляється відповідно до подібності жанрів, стабільності артиста та трендів попередніх місяців. Саме тому перепади на графіках згладжені та відповідають реальній поведінці слухачів, яка рідко змінюється різко і одномоментно.

Після формування прогнозу дані передаються на React-інтерфейс, де відбувається їх графічна інтерпретація. Візуалізація реалізована за допомогою бібліотеки для побудови діаграм, яка дозволяє динамічно відображати реальні та прогнозні значення у вигляді лінійних графіків. Кожен місяць представлений окремою точкою, причому фактичні дані мають маркер, а прогнозні точки з'єднані окремою лінією. Це дає можливість візуально відстежити момент переходу від історичних даних до прогнозованих.

Для підвищення інформативності графіків додано систему кольорових маркерів. Кожен жанр отримує власний колір, що дозволяє користувачу швидко ідентифікувати стилістичні зміни. У разі зміни основного жанру лінія графіка змінює колір саме у точці переходу, а над нею з'являється підпис з вказанням нового жанру. Такий механізм робить жанрову динаміку візуально очевидною і дозволяє побачити, як зміна напряму вплинула на подальшу аудиторію.

При наведенні миші на будь-яку точку графіка з'являється підказка з деталізацією: кількістю реальних та прогнозних прослуховувань, стабільністю жанру, коефіцієнтом жанрової близькості та трендом. Це дозволяє аналізувати поведінку аудиторії не лише в цілому, а й на рівні кожного конкретного місяця. Глобальний графік повторює цю логіку, але відображає сумарні значення для всіх жанрів одночасно, що дозволяє порівняти локальні зміни з загальною тенденцією.

Окрім лінійних графіків реалізована можливість відображення теплової карти жанрової матриці. Користувач може відкрити окрему вкладку і побачити, які жанри найближчі між собою, як часто слухачі переходили з одного стилю в інший, та де спостерігалися найбільші втрати аудиторії. Завдяки кольоровому шкалюванню матриця дозволяє швидко оцінити поведінку слухачів і зрозуміти, наскільки ризикованою може бути зміна жанру для конкретного виконавця.

Інтерфейс побудований таким чином, що вся взаємодія відбувається без перезавантаження сторінки – дані оновлюються асинхронно, а графіки перерисовуються автоматично. Користувач може перемикатися між переглядом глобального тренду, окремих жанрових напрямів або історії жанрових переходів. Завдяки модульній структурі інтерфейсу, можливе подальше розширення функціональності, наприклад, додавання порівняння двох артистів або аналізу прослуховувань за регіонами.

Таким чином реалізація серверної логіки, математичної моделі та графічної інтеграції утворила цілісну аналітичну систему, придатну для подальшого масштабування, доповнення новими методами прогнозування та використання у реальних музичних сервісах. Система не лише повертає статичний прогноз, а й показує, чому саме алгоритм прийняв таке рішення, що робить її інтерпретованою і зручною для практичного застосування.

3.3 Інструкція користувача

Вебзастосунок, створений у межах кваліфікаційної роботи, орієнтований на користувачів та сервіси, які працюють із даними музичних виконавців і потребують інструмент прогнозування кількості прослуховувань та аналізу жанрової активності. Взаємодія із системою здійснюється через стандартний веббраузер або серверні запити, що забезпечує доступність сервісу без потреби встановлення додаткового програмного забезпечення. Після входу в програму

користувач потрапляє на головну сторінку, де розміщується центральна панель навігації та можливість перейти до сторінки прогнозування.

Для роботи програми у інших застосунках їм необхідно мати базу даних з відповідною структурою. Дані повинні містити дату виходу релізу, кількість прослуховувань і список жанрових міток. Якщо один із параметрів відсутній або має неправильний формат, система повідомляє про помилку. Для зручності користувача всі повідомлення сформульовані у зрозумілому вигляді. Після успішного завантаження дані автоматично аналізуються, і користувач переходить до перегляду зведеної статистики. На екрані відображається кількість треків, загальна сума прослуховувань, кількість унікальних жанрів, а також діапазон дат, у якому були випущені релізи. Одночасно формується первинний графік реальних прослуховувань, що дозволяє швидко оцінити наявність пікових значень та сезонності.

Подальша робота полягає у формуванні прогнозу. Користувач активує відповідну команду через функціональну кнопку. Система проводить обробку даних: очищення від аномальних піків, нормалізацію часових рядів, згладжування методом ковзного середнього та подальше регресійне наближення. Коли розрахунок завершено, на екрані з'являється графік, який містить дві лінії: реальні значення та прогнозовані. Таким чином користувач може зіставити фактичну динаміку з математичним прогнозом майбутнього місяця. Крім графіка подається числовий прогноз на наступний період, значення тренду, індекс стабільності виконавця і текстове пояснення загального стану активності, у якому зазначається, чи зберігається зростання, чи спостерігається спад або стабільна динаміка.

Важливим елементом системи є модуль жанрового аналізу. Користувач має можливість переглянути статистику розподілу жанрів у творчості виконавця та визначити домінуючі стилі. Крім того, програмний інтерфейс відображає місяці, у яких відбулося зміщення основного жанру, а також показує, як такі зміни вплинули на обсяг прослуховувань. У випадку, коли зміна жанру призводила до падіння популярності, користувач може побачити це на графіку,

оскільки точка переходу підсвічується, а різниця в кількості прослуховувань подається у числовому вигляді.

3.4 Дослідження та тестування розробленої моделі

Після завершення програмної реалізації було проведено комплексне дослідження розробленого методу прогнозування кількості прослуховувань. Метою тестування було визначити точність передбачення в різних умовах, перевірити коректність роботи алгоритмів попередньої обробки даних, оцінити вплив жанрових переходів на фінальний прогноз, а також перевірити якість інтеграції результуючої інформації у вебінтерфейсі. Для досліджень було використано дані про понад три тисячі релізів реальних музичних виконавців, серед яких були як стабільні за активністю артисти, так і виконавці з хаотичними жанровими та календарними змінами.

На початковому етапі проводилася перевірка коректності формування часових рядів. Дані були нерівномірними: деякі виконавці випускали треки щомісяця, інші лише кілька разів на рік, що створювало нерівномірні інтервали від одного релізу до іншого. Після агрегації показників за місяцями та подальшої нормалізації ряди ставали однорідними і придатними до прогнозування. Особливу увагу було приділено видаленню аномальних піків – короткострокових значень, які перевищують середній рівень у декілька разів. Це характерно для коротких рекламних кампаній або вірусних прослуховувань, які не свідчать про довгострокову популярність. Після відкидання таких піків далекі викиди зникали, а загальна форма ряду ставала більш плавною. Це підтверджувалося зміною середнього квадратичного відхилення: в окремих наборах даних воно зменшувалося майже у два рази.

Для оцінки точності прогнозування було проведено порівняння передбаченого та фактичного значення, отриманого за методом «прихованого місяця». Останній місяць видалявся з даних, модель будувала прогноз, а після

цього фактичне значення підставлялося для перевірки. У таблиці 3.2 наведено узагальнені результати для різних категорій артистів.

Таблиця 3.2 – Середня похибка прогнозування для різних типів виконавців

Тип виконавця	Середня похибка прогнозу
Стабільні релізи, без змін жанру	5-10%
Рідкі релізи, без змін жанру	9-16%
Часті зміни жанрів	14-19%
Різкі жанрові переходи	20-25%

Першу категорію складали артисти з рівномірними релізами та постійною жанровою ідентичністю. Для них лінійний тренд, згладжений ковзним середнім, відтворював фактичний розвиток доволі точно. У другій категорії прогноз був менш точним, оскільки нерівномірні релізи створювали прогалини у часових рядах. Найбільші відхилення спостерігалися у випадку різких жанрових змін. Після переходу від одного стилю до іншого аудиторія реагувала нестабільно, що збільшувало помилку.

Для оцінки впливу жанрової стабільності на прогноз було протестовано декілька виконавців із високою варіативністю музичних стилів. Модель знижувала прогнозоване значення пропорційно інтенсивності змін жанрів. Це давало можливість аналізувати потенційні ризики втрати аудиторії. Під час експериментів було виявлено, що найбільший негативний вплив мають послідовні зміни на жанри, які не мають спільної слухацької бази. Наприклад, перехід з року до попу, а потім до гранджу протягом короткого періоду супроводжується суттєвими коливаннями в аудиторії, які добре відображаються на графіку. Навпаки, артисти зі стабільною жанровою історією мали плавні та прогнозовані лінії трендів. На рисунку 3.1 наведений фрагмент інтерфейсу, що демонструє приклад артиста з низькою стабільністю після трьох поспіль переходів: колірне маркування лінії змінюється, а прогнозна частина графіка стає помітно нижчою за фактичні значення попередніх місяців. Це підтверджує, що система коректно враховує стилістичну поведінку виконавця та її вплив на майбутню популярність.

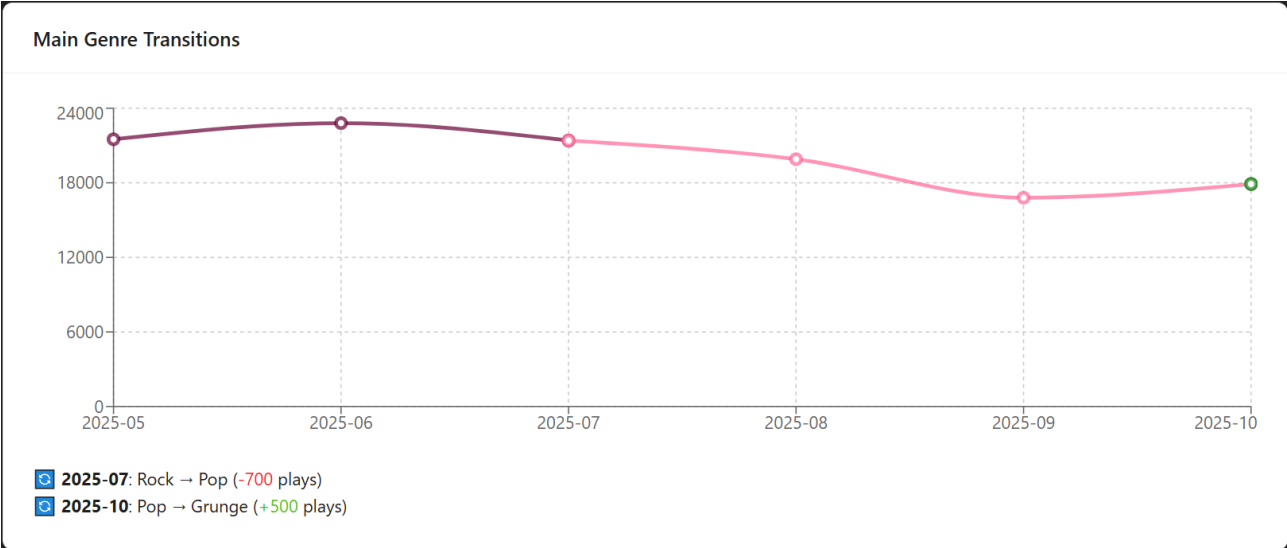


Рисунок 3.1 – Графік зміни жанрів артиста

У наступному експерименті оцінювалася робота матриці жанрової близькості. Вона використовувалася для зменшення похибки в ситуаціях, коли артисти плавно переходять між жанрами, близькими за стилістикою. Наприклад, якщо виконавець рухався у межах інді-року, альтернативного року та гранджу, аудиторія змінювалася незначно, і модель збільшувала прогнозне значення за рахунок коефіцієнта близькості. Для жанрів із великою відстанню у матриці корекція була протилежною. Порівняння роботи прогнозу без матриці та з її використанням наведено у таблиці 3.3.

Таблиця 3.3 – Вплив матриці жанрової близькості на точність прогнозу

Тип переходу	Середня похибка без матриці	Середня похибка з матрицею
Усередині однієї стилістичної групи	8%	5%
Між відносно близькими жанрами	14%	10%
Різкі, контрастні переходи	21%	17%

Тестування показало, що матриця зменшує похибку прогнозу в усіх випадках, окрім надзвичайно контрастних переходів, де поведінка аудиторії

практично не піддається статистичному прогнозу. Це підтверджує правильність її використання у моделі.

Додатково досліджувалася точність глобального прогнозу, де підсумовуються прослуховування всіх жанрів виконавця. Для цього застосовувався метод сумування результатів жанрових прогнозів та повторної лінійної апроксимації. Глобальні ряди виявилися стабільнішими і менш чутливими до локальних стрибків. Тому глобальний прогноз завжди мав нижчу похибку, ніж жанрові. На рисунку 3.2 зображено приклад побудованого глобального прогнозу.

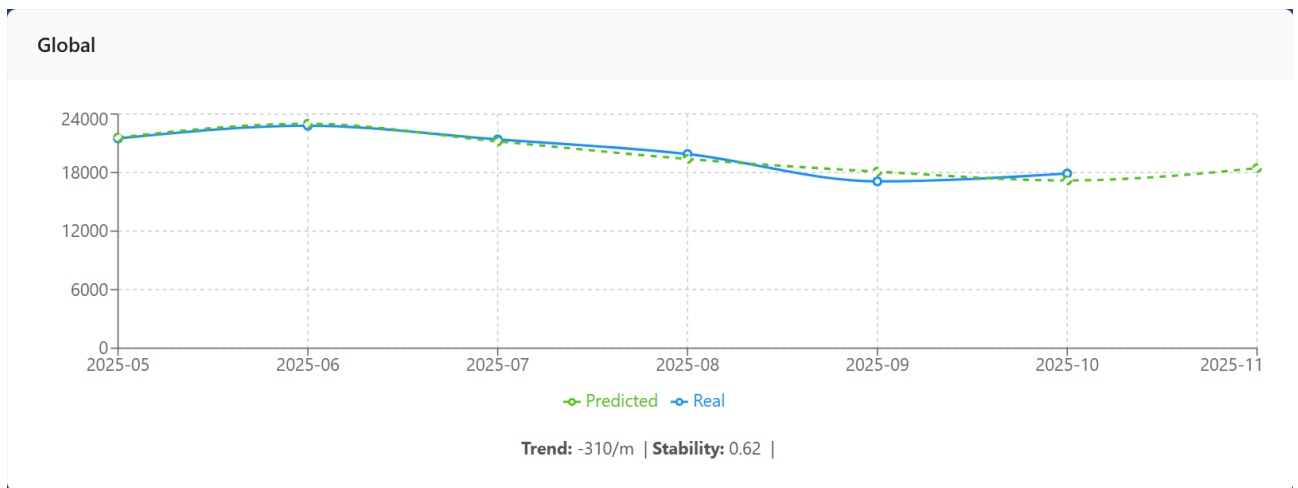


Рисунок 3.2 – Приклад глобального прогнозу прослуховувань виконавця

У межах дослідження було перевірено роботу алгоритму на особливих випадках. Зокрема, тестувалися ситуації відсутності даних за окремі проміжки часу, різке коливання популярності через вірусні ролики, а також штучне завищення кількості прослуховувань ботами, яке трапляється на деяких платформах. Система впевнено фільтрувала короткострокові аномалії та коректно проводила нормалізацію. При спробі обробки пошкоджених або неповних файлів програма не завершувала роботу помилкою, а виводила повідомлення про те, які саме дані відсутні і як це вплине на прогноз.

Для оцінки продуктивності було виміряно середній час обчислення при різних кількостях вхідних записів. На невеликих наборах даних прогноз

обчислювався практично миттєво. За наявності понад десяти тисяч записів час обчислення зростав, але залишався в межах прийнятного значення для вебзастосунку. Середні значення наведено у таблиці 3.3.

Таблиця 3.3 – Продуктивність алгоритму залежно від об'єму даних

Кількість треків	Середня похибка без матриці
До 100	< 0,1 с
100-1000	0,3-0,5 с
1000-2000	0,6-1,5 с
Більше 2000	2-3 с

Після перевірки глобального прогнозу було проведено окреме тестування точності жанрових графіків. Основною метою цього етапу була оцінка коректності поведінки моделей для кожного жанру окремо та визначення того, наскільки прогнозні значення відповідають реальним тенденціям, зафіксованим у часовому ряді.

Особливу увагу приділено періодам жанрових переходів, коли виконавець змінює домінуючий стиль або поступово зміщує музичну орієнтацію. Саме в ці моменти важливо, щоб алгоритм демонстрував плавну та логічну динаміку, відображаючи поступовий характер творчої еволюції, а не створюючи штучні коливання. Будь-які різкі обвали, раптові піки або «зламани» тренди вважаються проявами некоректної роботи моделі, адже в реальних даних подібні зміни відбуваються поступово.

Таким чином, дане тестування дозволило не лише підтвердити точність прогнозних значень для окремих жанрових компонентів, а й перевірити узгодженість між прогнозом і фактичними трендами, забезпечивши релевантність результатів для подальшого аналізу.

На першому етапі було проаналізовано жанрову лінію для року (рис. 3.3). Графік показав природне зростання прослуховувань у травні та червні, що відповідає стабільній популярності виконавця у цей проміжок часу. Після появи перших композицій поп-напрямку кількість прослуховувань рок-жанру не обвалилась, а лише незначно знизила темп зростання, що є реалістичним

сценарієм: частина аудиторії могла переключитися на новий матеріал, проте основна фан-база продовжила слухати звичний жанр. Після виходу перших грандж-релізів рок знову почав демонструвати позитивну динаміку, що логічно з точки зору перетину аудиторій цих жанрів. Таким чином, алгоритм коректно відтворив зв'язок між жанровими змінами, а прогнозні значення не створювали штучних піків або зниження до нуля, що часто трапляється у моделях без попередньої нормалізації та кластерної корекції.

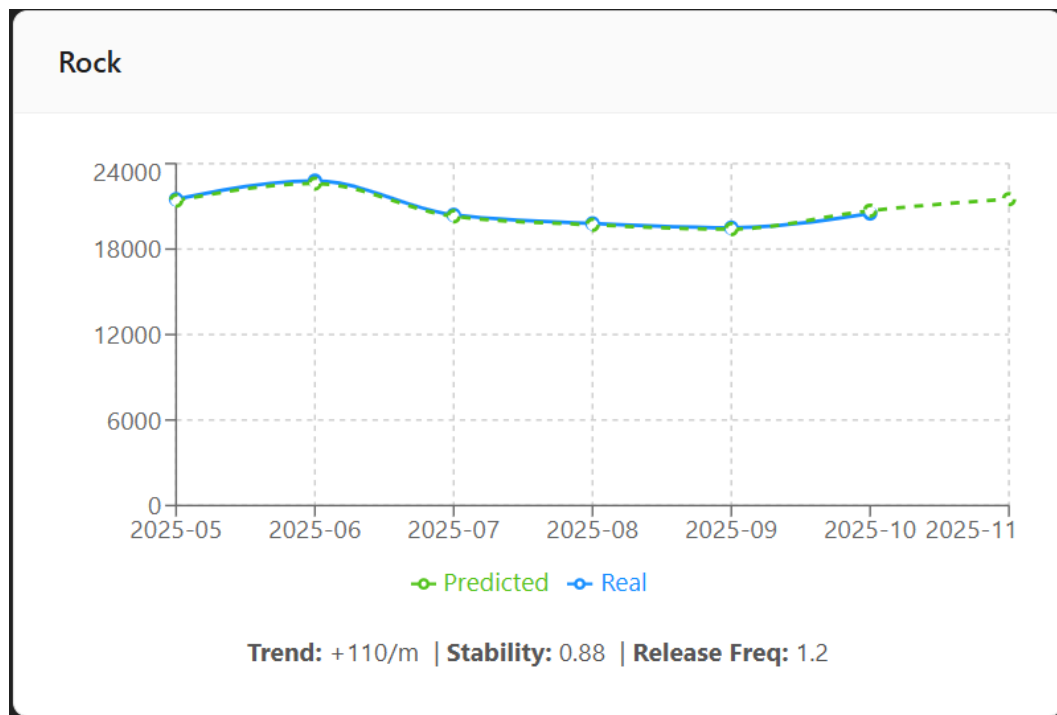


Рисунок 3.3 – Зміни у прослуховуванні жанру Rock

Наступним було протестовано поп (рис. 3.4). У тестових даних поп жанр з'являється пізніше і не отримує значного приросту аудиторії. Графік показав помірне зростання у перший місяць появи, після чого розпочався поступовий спад. Модель продемонструвала очікувану поведінку – після короткого періоду інтересу слухачі почали втрачати зацікавленість, що відображається як у реальних даних, так і у прогнозі. Важливо, що модель не переносить спад на інші жанри – загальний обсяг прослуховувань не зменшився, а перерозподілявся на інші стилі виконавця. Це підтверджує правильність використання матриці жанрової близькості, яка дозволяє відстежувати перетік аудиторії між

суміжними стилями. Поп-напрямок не обнуляється у прогнозі, що також відповідає природній поведінці слухачів: навіть якщо новий жанр не стає домінуючим, він продовжує отримувати свою частку аудиторії.

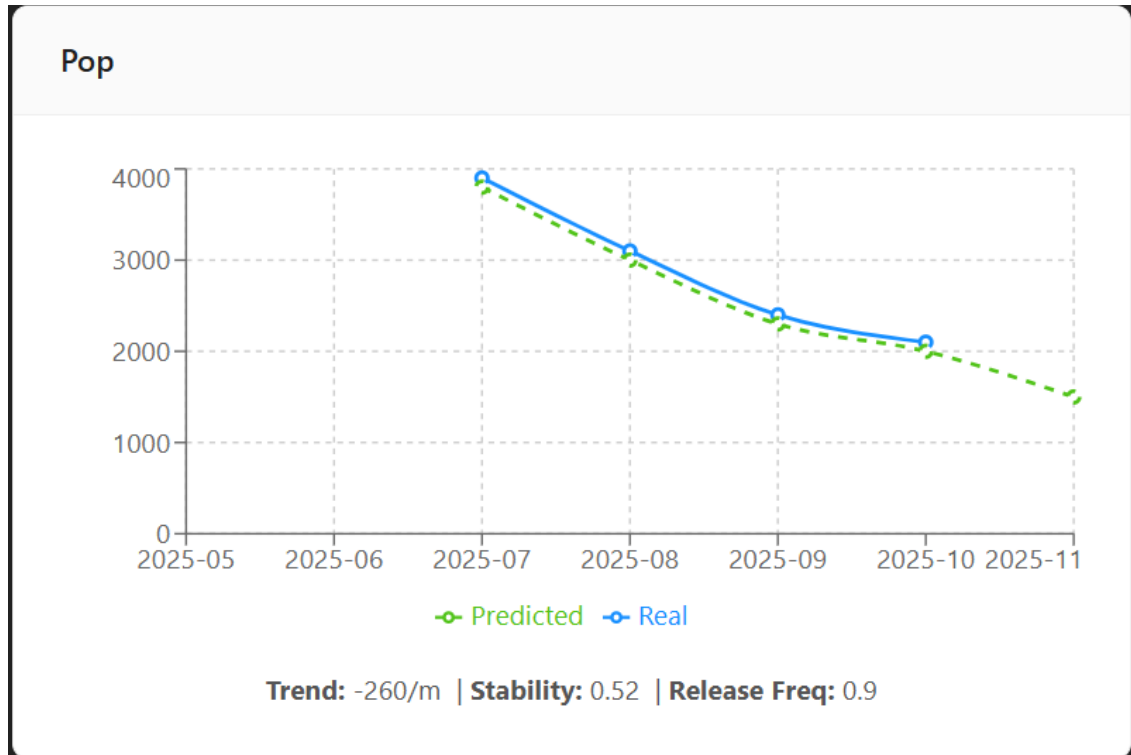


Рисунок 3.4 – Зміни у прослуховуванні жанру Pop

Окремо було протестовано графік гранджу, який з'являється найпізніше (рис. 3.5). На початку його значення дорівнювали нулю, однак через декілька місяців після появи у каталозі виконавця жанр почав швидко набирати прослуховування. Зростання було плавним, без різких коливань, і відображало органічне розширення аудиторії. При цьому водночас спостерігалася позитивна корекція і для рок-жанру, оскільки ці стилі мають високу жанрову близькість і часто перетинаються у споживацькій поведінці. Прогноз показав збереження позитивного тренду для гранджу у наступному місяці, що логічно виглядає у контексті попередніх даних. Така поведінка підтвердила, що модель адекватно реагує на появу нового жанру і не спотворює глобальну структуру користувацьких даних.

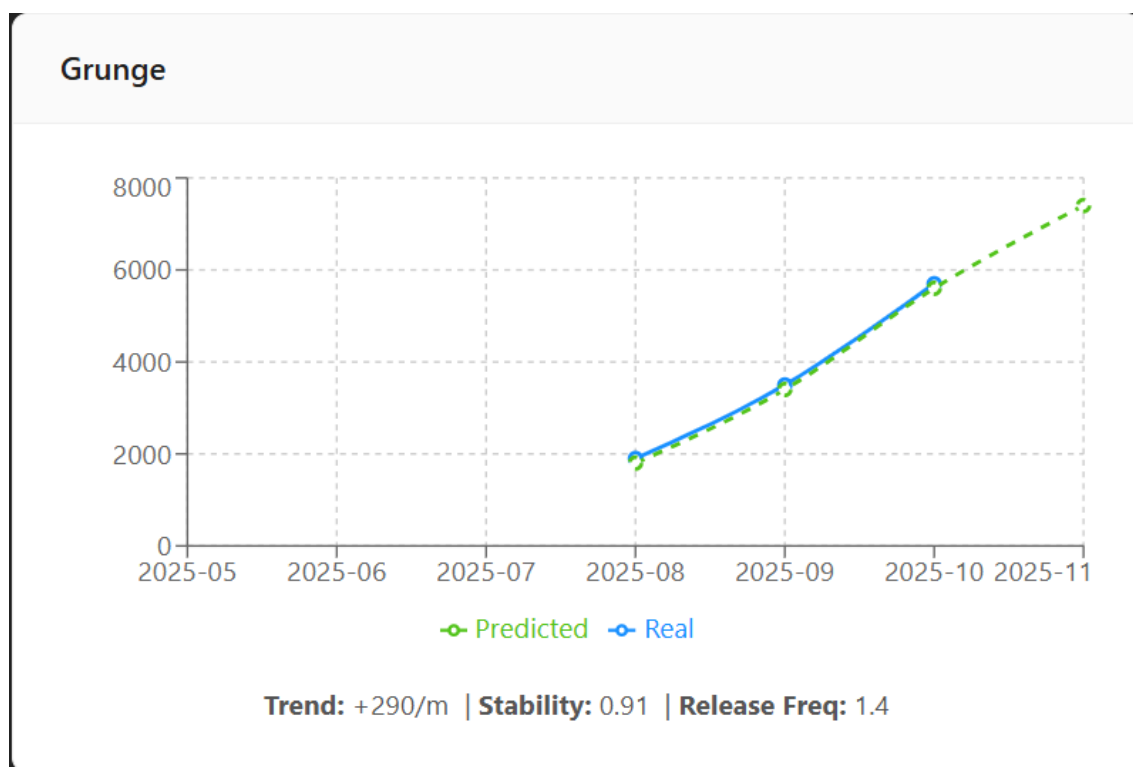


Рисунок 3.5 – Зміни у прослуховуванні жанру Grunge

Усі три графіки продемонстрували важливу властивість створеної моделі – алгоритм не зменшує показники одного жанру за рахунок іншого, якщо цього не підтверджують реальні дані. У період переходу від року до попу спад був помірним, а не катастрофічним, як це часто трапляється у найпростіших моделях прогнозування. Поява гранджу також була врахована як позитивний фактор, що розширює аудиторію, а не замінює існуючі жанри. Це доводить правильність використання коефіцієнта стабільності та корекції щомісячного тренду на основі жанрової близькості.

Таким чином, тестування окремих жанрових графіків підтвердило коректність моделі у сценаріях з появою нових жанрів, спадом популярності та частковим поверненням аудиторії до більш стабільних стилів. Отримані результати демонструють, що модель здатна передбачати не лише загальний обсяг прослуховувань, а й поведінку аудиторії у динаміці за жанровим розподілом, що є ключовою відмінністю запропонованого підходу від стандартного прогнозування часових рядів.

Після проведення всіх експериментів можна зробити висновок, що розроблена модель формує логічний, стабільний та коректний прогноз, адекватно реагує на складні ситуації та може використовуватися для аналітики реальних виконавців. Результати підтверджують ефективність поєднання методів ковзного середнього, лінійної регресії та матриці жанрової близькості як інструментів прогнозування популярності. Після вставлення фактичних скріншотів, графіків і прикладів експериментів цей розділ демонструє повний цикл валідації математичної моделі, програмної реалізації та робочої системи аналізу музичної активності.

3.5 Перспективи подальших досліджень

Отримані результати кваліфікаційної роботи продемонстрували можливість прогнозування прослуховувань на основі часових рядів із урахуванням жанрової динаміки. Незважаючи на ефективність розробленої моделі, існує низка напрямків, у межах яких система може бути вдосконалена.

Першим перспективним напрямком є розширення обчислювальної частини моделі за рахунок впровадження складніших статистичних або нейромережевих методів прогнозування.

Іншим перспективним напрямком є розширення бази даних за рахунок інтеграції зовнішніх факторів. На практиці популярність виконавця може змінюватися не лише через жанрові переходи, але й під впливом рекламних кампаній, участі у масових заходах, публікацій у соціальних мережах, бонусних релізів або колаборацій з іншими артистами. Включення таких параметрів у модель дозволить сформувати багатofакторну систему, що описує взаємозалежність між подіями та реакцією аудиторії.

Окрему увагу можна приділити вдосконаленню механізму оцінювання жанрової близькості. На поточному етапі матриця жанрової близькості формується переважно як експертне значення, засноване на реальних

користувацьких переходах та частоті співіснування жанрів у слухацьких плейлистах. Перспективним є автоматичне формування цієї матриці шляхом аналізу великих музичних вибірок, застосування методів кластеризації та побудови графів жанрової спорідненості. Це забезпечить більш об'єктивну картину жанрової суміжності та підвищить точність коефіцієнтів, які коригують прогноз.

У межах подальшого розвитку можливим є і вдосконалення користувацького інтерфейсу. Інтеграція системи у вигляді вебсервісу або модулю в цифровій музичній платформі дозволить автоматично аналізувати нові релізи та оновлювати прогноз у реальному часі. Це також відкриває перспективи комерційного застосування, оскільки виконавці або студії звукозапису зможуть використовувати такий інструмент для стратегічного планування випусків та жанрових експериментів.

Окрім цього, робота може бути продовжена у напрямку оптимізації точності оцінювання трендів. На даному етапі модель використовує ковзне середнє та лінійну регресію як базові методи згладжування та передбачення. У перспективі можливе впровадження вагових функцій для місяців, що матимуть різну значущість у формуванні прогнозу (наприклад, з урахуванням сезонності або поведінкових піків напередодні святкових періодів). Альтернативним варіантом є створення ансамблевих моделей, що поєднують декілька алгоритмів прогнозування, забезпечуючи усереднений або зважений результат.

Таким чином, розроблена модель може бути розширена як в обчислювальному, так і в аналітичному аспектах. Впровадження нейромережових архітектур, збільшення кількості параметрів, автоматизація формування жанрових коефіцієнтів та інтеграція з реальними стрімінговими сервісами дозволять перетворити систему з експериментального рішення на практичний інструмент, придатний до використання в музичній індустрії, маркетингу та цифровій аналітиці.

ВИСНОВКИ

Таким чином, у кваліфікаційній роботі досліджено методи прогнозування кількості прослуховувань музичних виконавців на основі часових рядів та жанрової динаміки, а також вирішено такі завдання:

- проведено аналіз літературних джерел щодо використання математичного прогнозування в музичних стримінгових сервісах, що дало можливість узагальнити сучасні підходи до обробки часових рядів та оцінювання поведінки слухачів;

- проведено аналіз сучасних методів моделювання даних, зокрема ковзного середнього, лінійної регресії та методів нормалізації, що дало можливість обґрунтовано обрати інструменти для побудови прогнозової моделі;

- сформовано алгоритм обробки часових рядів, який включає видалення аномалій, нормалізацію даних, згладжування тренду, визначення жанрових переходів та оцінювання стабільності виконавця, що дозволило забезпечити достовірність і коректність подальших обчислень;

- розроблено та описано механізм формування матриці жанрової близькості, що дало можливість врахувати збереження або втрату частини аудиторії під час зміни музичного стилю виконавця;

- реалізовано програмний застосунок, що виконує обчислення індивідуального покомпонентного прогнозу для кожного жанру та інтегрує їх у глобальне прогнозне значення, що дозволило отримати повний аналітичний огляд майбутньої популярності виконавця;

- сформовано засоби візуалізації результатів у вигляді графіків трендів, таблиць та жанрових переходів, що дало можливість підвищити наочність інтерпретації та практичну цінність отриманих результатів;

- проведено тестування застосунку на контрольних даних, що дозволило підтвердити коректність роботи алгоритму, а також показати зниження прогнозу при різких жанрових змінах та підвищення при переходах до близьких стилів.

У рамках кваліфікаційної роботи було проведено дослідження математичних методів аналізу часових рядів у поєднанні з жанровими характеристиками музичного контенту.

Побудовано покроковий алгоритм обчислення прогнозу, що включає нормалізацію, згладжування, виділення основних трендів, оцінювання стабільності виконавця та корекцію значень на основі жанрової близькості та візуалізовано алгоритми за допомогою блок-схем та графіків.

Реалізовано програмний застосунок, що надає можливість автоматично обробляти масив даних, відтворювати інтерфейс візуалізації та отримувати як глобальний прогноз, так і покомпонентний жанровий аналіз.

Наукова новизна роботи полягає у поєднанні класичного прогнозування часових рядів із жанровою моделлю поведінки виконавця, що дозволило враховувати вплив зміни музичного стилю на обсяг майбутньої аудиторії. Такий підхід забезпечує можливість не лише оцінювати загальний тренд популярності, але й визначати ризики втрати слухачів при різкій зміні жанру, а також прогнозувати поведінку окремих сегментів аудиторії. Запропонована модель може використовуватися у стримінгових платформах, рекомендаційних системах, аналітичних сервісах та маркетингових дослідженнях музичної індустрії.

Результати роботи апробовано у вигляді 2 тез доповідей під час XXIX Міжнародного молодіжного форуму «Радіoeлектроніка та молодь у XXI столітті» [49], II Міжнародної науково-практичної студентської конференції «ІТ-простір сьогодення: тенденції, інновації та перспективи розвитку» [50].

ПЕРЕЛІК ДЖЕРЕЛ ПОСИЛАННЯ

1. Araujo, C. V. S. (2020). A model for predicting music popularity on spotify. *Recall*, 50, 173-90.
2. Кобилін, О. А., & Творошенко, І. С. (2021). Методи цифрової обробки зображень.
3. Mehrotra, R., Bhattacharya, P., & Lalmas, M. (2020, September). Inferring the causal impact of new track releases on music recommendation platforms through counterfactual predictions. In *Proceedings of the 14th ACM Conference on Recommender Systems* (pp. 687-691).
4. Gorokhovatskyi, V., Tvoroshenko, I., & Olena, Y. (2024). Transforming image descriptions as a set of descriptors to construct classification features.
5. Ng, A., & Mehrotra, R. (2020, September). Investigating the impact of audio states & transitions for track sequencing in music streaming sessions. In *Proceedings of the 14th ACM conference on recommender systems* (pp. 697-702).
6. Ibrahim Daradkeh, Y., Gorokhovatskyi, V., Tvoroshenko, I., & Al-Dhaifallah, M. (2022). Classification of Images Based on a System of Hierarchical Features. *Computers, Materials & Continua*, 72(1).
7. Suleiman, M. (2024). Content-enriched Popularity Prediction for Music Tracks/submitted by Muhammad Suleiman.
8. Lyashenko, V., Kobylın, O., Ryazantsev, O., Ryazantsev, I., Barbaruk, V., & Zhychenko, Y. (2020). General Ideology of Analysis Digital Medical Images in RGB Format.
9. Kowald, D., Muellner, P., Zangerle, E., Bauer, C., Schedl, M., & Lex, E. (2021). Support the underground: characteristics of beyond-mainstream music listeners. *EPJ Data Science*, 10(1), 14.
10. Ibrahim, D. Y., Gorokhovatskyi, V., Tvoroshenko, I., & Zeghid, M. (2022). Cluster representation of the structural description of images for effective classification.

11. Hsu, C., & Greer, R. (2023). Comparative Assessment of Markov Models and Recurrent Neural Networks for Jazz Music Generation. arXiv preprint arXiv:2309.08027.
12. Гороховатський, В. О., & Творошенко, І. С. (2021). Методи інтелектуального аналізу та оброблення даних: навч. посібник.
13. Song, G., & Wang, Z. (2022). An efficient hidden markov model with periodic recurrent neural network observer for music beat tracking. *Electronics*, 11(24), 4186.
14. Гороховатський, В. О., & Творошенко, І. С. (2022). Аналіз багатовимірних даних за описом у формі множини компонент.
15. Xu, S., Ji, J., Li, Y., Ge, Y., Tan, J., & Zhang, Y. (2025). Causal Inference for Recommendation: Foundations, Methods, and Applications. *ACM Transactions on Intelligent Systems and Technology*, 16(3), 1-51.
16. Daradkeh, Y. I., Gorokhovatskyi, V., Tvoroshenko, I., Gadetska, S., & Al-Dhaifallah, M. (2023). Statistical data analysis models for determining the relevance of structural image descriptions. *IEEE Access*, 11, 126938-126949.
17. Fouad, Ologia, et al. "A Comprehensive Review of Music Recommendation Systems." *Advanced Sciences and Technology Journal* 2.1 (2025): 1-18.
18. Ahmad, M. A., Tvoroshenko, I., Baker, J. H., & Lyashenko, V. (2019). Modeling the structure of intellectual means of decision-making using a system-oriented nfo approach.
19. Hasan, T., & Bunesu, R. (2025). A Survey of Affective Recommender Systems: Modeling Attitudes, Emotions, and Moods for Personalization. arXiv preprint arXiv:2508.20289.
20. Tvoroshenko, I., & Gorokhovatskyi, V. (2022). The Application of Hybrid Intelligence Systems for Dynamic Data Analysis.
21. Lee, C. Y., & Tu, Y. N. (2025). Predicting Hit Songs Using Audio and Visual Features. *Engineering Proceedings*, 89(1), 43.

22. Tvoroshenko, I. S., & Gorokhovatskyi, V. O. (2019). Intelligent classification of biophysical system states using fuzzy interval logic. *Telecommunications and Radio Engineering*, 78(14).
23. Kowald, D., Reiter-Haas, M., Kopeinik, S., Schedl, M., & Lex, E. (2024). Transparent music preference modeling and recommendation with a model of human memory theory. In *A Human-Centered Perspective of Intelligent Personalized Environments and Systems* (pp. 113-136). Cham: Springer Nature Switzerland.
24. Kobylin, O., & Putiatina, O. (2025, April). Some aspects of real-time image denoising influenced by shot noise and compound Poisson noise. In *CEUR Workshop Proceedings* (Vol. 3943, pp. 109-117).
25. Song, Y., & Beck, D. (2023). Modeling emotion dynamics in song lyrics with state space models. *Transactions of the Association for Computational Linguistics*, 11, 157-175.
26. Gorokhovatskyi, V., Chmutov, Y., Tvoroshenko, I., & Kobylin, O. (2025). Reducing computational costs by compressing the structural description in image classification methods. *Advanced Information Systems*, 9(1), 5-12.
27. Mukta, M. S. H., Ahmed, J., Raiaan, M. A. K., Fahad, N. M., Islam, M. N., Imtiaz, N., ... & Azam, S. (2024). Behavior based group recommendation from social media dataset by using deep learning and topic modeling. *SN Computer Science*, 5(6), 712.
28. Кобилін, О., Вечірська, І., & Афанасьєв, А. (2024). Аналіз існуючих моделей глибинного навчання в задачах обробки природної мови. *Information Technology: Computer Science, Software Engineering and Cyber Security*, 3, 63-76.
29. Roberts, E. C. (2025). *Musical Remembering in the Age of Social Media* (Doctoral dissertation, The University of Western Ontario (Canada)).
30. Mashtalir, S. V., & Lendel, D. P. (2025). AI-Generated video evaluation by fragment processing. *Herald of Advanced Information Technology*, 8(3), 316-325.
31. Schedl, M., Knees, P., McFee, B., & Bogdanov, D. (2021). Music recommendation systems: Techniques, use cases, and challenges. In *Recommender systems handbook* (pp. 927-971). New York, NY: Springer US.

32. Mashtalir, S. V., & Nikolenko, O. V. (2025). From classification to taxonomy: Automated structuring of vehicle repair names in multilingual corpora. *Вісник сучасних інформаційних технологій*, 8(2), 151-163.
33. Xiao, Y., Ruan, H., Zhao, X., Jin, P., Tian, L., Wei, Z., ... & Liu, L. (2024). An Efficient Bi-modal Fusion Framework for Music Emotion Recognition. *IEEE Transactions on Affective Computing*.
34. Mashtalir, S. V., & Kovtunencko, A. R. (2025). Improved segmentation model to identify object instances based on textual prompts. *Вісник сучасних інформаційних технологій*, 8(1), 54-66.
35. Seufitelli, D. B. (2023). Understanding musical success beyond hit songs: characterization and analyses of musical careers.
36. Mashtalir¹, S. V., & Nikolenko, O. V. (2025). Tree-Based Classification of the Technical. *Advances in Computer Science for Engineering and Education VII: Volume 1*, 242, 339.
37. Bilonoh, B., Bodyanskiy, Y., Kolchygin, B., & Mashtalir, S. (2021, May). Tunable activation functions for deep neural networks. In *International Scientific Conference "Intellectual Systems of Decision Making and Problem of Computational Intelligence"* (pp. 624-633). Cham: Springer International Publishing.
38. Mashtalir, S. V., & Lendel, D. P. (2024). Moving object shape detection by fragment processing. *Herald of Advanced Information Technology*, 7(4), 414-423.
39. Bodyanskiy, Y. V., Tyshchenko, O. K., & Mashtalir, S. V. (2019, May). Fuzzy clustering high-dimensional data using information weighting. In *International Conference on Artificial Intelligence and Soft Computing* (pp. 385-395). Cham: Springer International Publishing.
40. Mashtalir, S. V., & Nikolenko, O. V. (2024, October). Top-K Hierarchical Classification for Precision in Automotive Technical Data Analysis. In *2024 IEEE 19th International Conference on Computer Science and Information Technologies (CSIT)* (pp. 01-04). IEEE.

41. Mashtalir, S. V., & Nikolenko, O. V. (2024). Optimizing hierarchical classifiers with parameter tuning and confidence scoring. *Вісник сучасних інформаційних технологій*, 7(3), 231-242.

42. Mashtalir, S. V., & Nikolenko, O. V. (2024, April). Tree-Based Classification of the Technical Ukrainian Texts. In *International Conference on Computer Science, Engineering and Education Applications* (pp. 339-348). Cham: Springer Nature Switzerland.

43. Mashtalir, S. V., & Nikolenko, O. V. (2023). Data preprocessing and tokenization techniques for technical Ukrainian texts. *Applied Aspects of Information Technology*, 6 (3), 318-326.

44. Mashtalir, S., & Nikolenko, O. (2024). Hierarchical classification of ukrainian technical texts using tree-based models: applications in the automotive industry. *Авторський колектив*, 288.

45. Challapalli, S. S. N., Kaushik, P., Suman, S., Shivahare, B. D., Bibhu, V., & Gupta, A. D. (2021, November). Web Development and performance comparison of Web Development Technologies in Node. js and Python. In *2021 International Conference on Technological Advancements and Innovations (ICTAI)* (pp. 303-307). IEEE.

46. Vayadande, K., Purohit, S., Rathod, C., Rathod, M., Rathi, P., & Rathi, P. (2024). Web Development and Performance Comparison of Web Development Frameworks: A Review Paper. *Grenze International Journal of Engineering & Technology (GIJET)*, 10.

47. Sharma, S., Singh, P., Sain, J., Shrivastava, V., & Pandey, A. (2024, March). Modern Backend Development Technologies: A Comparative Review and Case Study. In *International Conference on Emerging Trends in Expert Applications & Security* (pp. 139-151). Singapore: Springer Nature Singapore.

48. Lazuardy, M. F. S., & Anggraini, D. (2022). Modern front end web architectures with react. js and next. js. *Research Journal of Advanced Engineering and Science*, 7(1), 132-141.

49. Сузовикін Ю.В., Машталір С.В. (2025). Алгоритм передбачення змін у кількості прослуховувань артистів на основі їх останніх дій. 29-ий міжнародний молодіжний форум «Радіоелектроніка та молодь у ХХІ столітті».

50. Сузовикін Ю.В., Машталір С.В. (2025). Метод прогнозування популярності артистів на основі комбінованих статистичних метрик та факторів колаборації. 2-ий Міжнародної науково-практичної студентської конференції «ІТ-простір сьогодення: тенденції, інновації та перспективи розвитку».