

ДОДАТОК А

Слайди презентації



Дослідження ансамблевих методів машинного навчання для класифікації музичних жанрів у інформаційних системах



Мовчан А.С., ІПЗм-23-2
Керівник: професор каф. ПІ Смеляков С.В.

19 червня 2025

1

Рисунок А.1 – Слайд 1

Основні положення дослідження

- Об'єкт дослідження – процес автоматизованої класифікації музичних треків за жанрами в інформаційних системах;
- Напрямок дослідження – застосування ансамблевих методів машинного навчання для задач класифікації музичних жанрів;
- Контекст дослідження – зростання потреби в систематизації музичних даних у зв'язку з розвитком цифрових музичних сервісів.



2

Рисунок А.2 – Слайд 2

Аналіз предметної галузі

- Музичні архіви та стримінгові платформи швидко зростають;
- Потреба в автоматизованій класифікації та ефективній структуризації;
- Виклики: велика варіативність стилів, змішані жанри, суб'єктивність сприйняття.



3

Рисунок А.3 – Слайд 3

Аналіз предметної галузі

- Аудіоознаки: спектральні, ритмічні, тембральні характеристики мають ключове значення;
- Спектрограми дають уявлення про енергетичну структуру треків у часі;
- Роль ансамблів: дозволяють компенсувати слабкі сторони окремих моделей, підвищують точність.



4

Рисунок А.4 – Слайд 4

Огляд і аналіз літературних, наукових джерел

Джерело	Основні підходи	Методи
[3] Liu Z., De Mori J., Abayomi K.	Open Set Recognition (OSR) у контексті класифікації	Використання CNN з розпізнаванням «нових» жанрів за допомогою відкритого множинного підходу
[4] Pons J., Lidy T., Serra X.	«Музично мотивовані» Convolutional Neural Networks (CNN)	Застосування спеціалізованих фільтрів, що враховують ритмічні й гармонійні патерни
[5] Choi K., Fazekas G., Sandler M., Cho K.	Convolutional Recurrent Neural Networks (CRNN)	Поєднання CNN для вилучення локальних ознак і RNN для моделювання часових залежностей (Mel-спектрограми, MFCC)



5

Рисунок А.5 – Слайд 5

Огляд і аналіз літературних, наукових джерел

Джерело	Основні підходи	Методи
[6] Meng Y.	Порівняння CNN та XGBoost	Аналіз ефективності підходів Deep Learning проти Boosting із застосуванням Mel-спектрограм та MFCC
[7] Zhang J.	Гібридна модель ResNet + Bi-GRU	Глибинне вилучення ознак (ResNet) та двонапрямна GRU для врахування часових залежностей
[8] Ru G., Zhang X., Wang J., Cheng N., Xiao J.	Мультимодальний аналіз із урахуванням кореляцій між жанрами	Інтеграція аудіо- і текстових ознак, механізм крос-модальної уваги, врахування взаємозв'язків між жанровими категоріями



6

Рисунок А.6 – Слайд 6

Постановка задачі

Проблема: weak-learner моделі демонструють недостатню точність класифікації через складність жанрових меж, змішані стилі та високий рівень шуму в аудіоданих;

Очікуваний результат: побудова комбінованого ансамблю (stacking) з моделей Bagging, AdaBoost, Random Forest та SVC, який перевершує в точності кожен з базових моделей окремо та забезпечує стабільні результати на наборі даних FMA_small.



7

Рисунок А.7 – Слайд 7

Методологія та інструменти

Методи дослідження:

- Ансамблеві підходи: Bagging, AdaBoost, Random Forest, SVC, stacking;
- Байєсова оптимізація гіперпараметрів;
- Парето-аналіз.

Інструментарій та технології:

- Python 3.12 / Jupyter Notebook — середовище для експериментів;
- scikit-learn, joblib — реалізація моделей та паралельні обчислення;
- pandas, NumPy — обробка і аналіз даних;
- matplotlib, tqdm — візуалізація результатів та контроль прогресу.



8

Рисунок А.8 – Слайд 8

Зміст проведеного експерименту

Вхідні дані:

- Набір треків з FMA_small (6000 треків, 6 жанрів);
- 11 груп ознак: chroma, mfcc, spectral, zcr, tonnetz та інші.

Критерії оцінки:

- Точність класифікації (Accuracy);
- F-1 міра.



9

Рисунок А.9 – Слайд 9

Опис ознак з проведеного експерименту

Група	Інтуїтивна користь
mfcc	Ловить тембр і форманти – базовий дескриптор для жанрової сегментації
chroma_cqt / chroma_stft	Гармонічний контекст та акордові прогресії; CQT – логарифмічна шкала, STFT – лінійна
rmse	Енергетичний профіль треку
spectral_centroid / contrast	«Яскравість» і нерівномірність енергетичного спектра
tonnetz	Тональні центри, релевантні для гармонійно насиченої музики
zcr	Ритмічна інформація і рівень шумів, особливо для рок-/експериментальної сцени



10

Рисунок А.10 – Слайд 10

Зміст проведеного експерименту

Послідовність:

1. Попередня обробка ознак
2. Відбір ключових ознак
3. Побудова базових моделей
4. Отримання out-of-fold ймовірностей
5. Навчання мета-класифікатора (stacking)
6. Повторення кроків (3-5) на 5 фолдах (CV=5)

Вимірювання:

Для кожної моделі обчислюємо:

1. Середню точність
2. Стандартне відхилення на фолдах
3. Час тренування

Візуалізація:

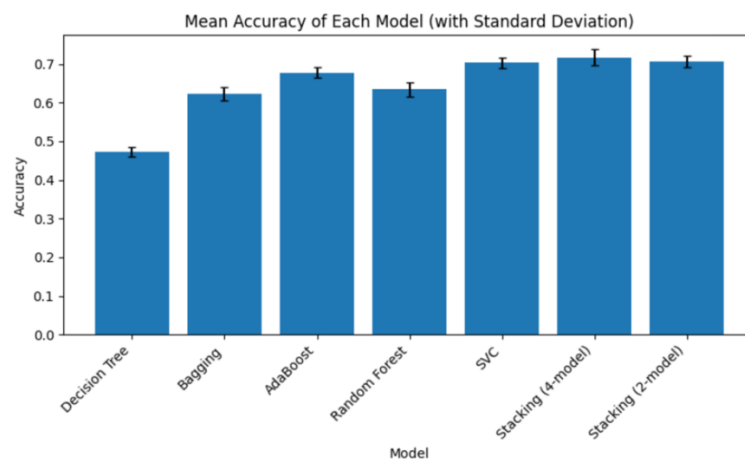
- Стовпчасті діаграми (accuracy)
- Графіки



11

Рисунок А.11 – Слайд 11

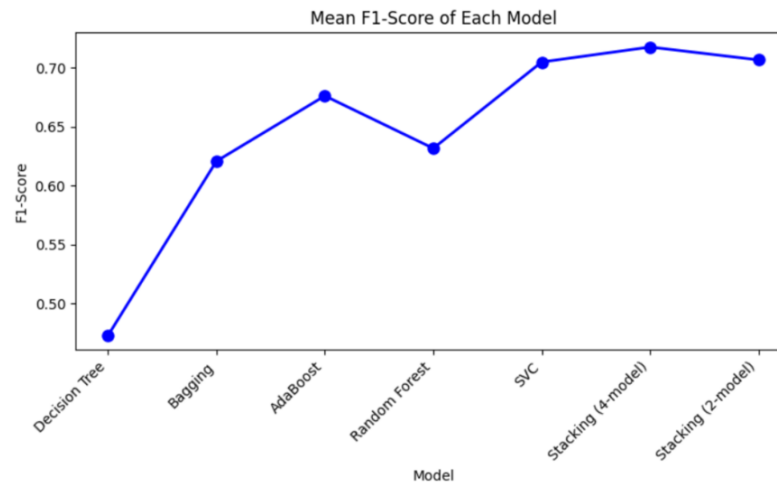
Результати експерименту



12

Рисунок А.12 – Слайд 12

Результати експерименту

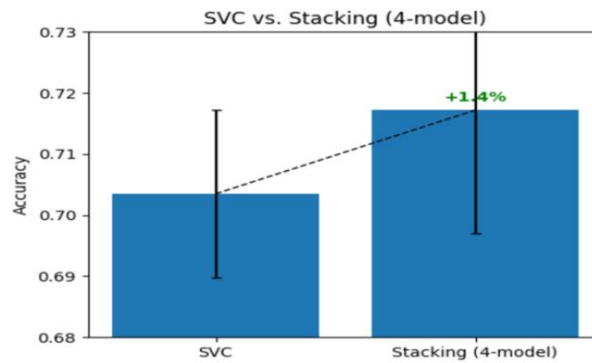


13

Рисунок А.13 – Слайд 13

Порівняння SVC та Stacking Classifier

- Standalone SVC досягає ~70% точності;
- Stacking ансамбль (SVC + Bagging + AdaBoost + RF) покращує точність до ~72%.



14

Рисунок А.14 – Слайд 14

Аналіз отриманих результатів

- Досягнуто мети: підвищення якості класифікації жанрів за допомогою ансамблевих методів;
- Підтверджено, що комбінування моделей через stacking ефективніше, ніж використання окремих моделей.

Інтерпретація результатів:

- SVC отримав низьку оцінку за формальними критеріями Парето, однак мав сильну емпіричну ефективність;
- Stacking дозволив поєднати сильні сторони моделей, компенсуючи слабкості різних підходів до навчання.



15

Рисунок А.15 – Слайд 15

Публікація результатів



16

Рисунок А.16 – Слайд 16

Підсумки

- Досягнуто стійкого підвищення точності класифікації жанрів шляхом поєднання моделей;
- Результати підтверджують практичну цінність застосування ансамблевих методів у реальних інформаційних системах.

Можливий напрям розвитку:

- Розширення ансамблю за рахунок методів глибокого навчання (наприклад, прості CNN чи CRNN) і мультимодальних дескрипторів (метадані, тексти тощо)



17

Рисунок А.17 – Слайд 17

ДОДАТОК Б

Апробація результатів роботи. Тези


 IOWORLD
PAPERS

ISSN 2710-3056

INTERNATIONAL SCIENTIFIC JOURNAL

GRAIL OF SCIENCE

№ **49** (February, 2025)

with the proceedings of the:

IV Correspondence International
Scientific and Practical Conference

**SCIENCE IN MOTION: CLASSIC AND
MODERN TOOLS AND METHODS IN
SCIENTIFIC INVESTIGATIONS**

held on February 21st, 2025 by

NGO European Scientific Platform
(Vinnytsia, Ukraine)
LLC International Centre Corporative
Management (Vienna, Austria)

МІЖНАРОДНИЙ НАУКОВИЙ ЖУРНАЛ

ГРААЛЬ НАУКИ

№ **49** (лютий, 2025)

за матеріалами:

IV Міжнародної науково-
практичної конференції

**НАУКА В РУСІ: КЛАСИЧНІ ТА
СУЧАСНІ ЗАСОБИ ТА МЕТОДИ
НАУКОВИХ ДОСЛІДЖЕНЬ**

що проводилася 21.02.2025

ГО «Європейська наукова
платформа» (Вінниця, Україна)
ТОВ «International Centre Corporative
Management» (Відень, Австрія)



Вінниця – Відень, 2025

Рисунок Б.1 – Перша сторінка збірника тез до обраної конференції



DOI 10.36074/grail-of-science.21.02.2025.084

ENSEMBLE-BASED APPROACHES FOR MUSIC TRACK CLASSIFICATION

Movchan Andrii Serhiiovych Student of the Faculty of Computer Science
*Kharkiv National University of Radio Electronics, Ukraine***Scientific adviser: Smelyakov Serhii Vacheslavovych** DSc,
Professor of the Department of Software Engineering
Kharkiv National University of Radio Electronics, Ukraine

In today's era of large-scale streaming services and vast music repositories, the ability to quickly and accurately classify music tracks by genre has become critical. This classification process underpins recommendation systems, user-centric navigation, and even musicological analysis. Yet, it remains challenging due to overlapping styles, subjective genre boundaries, and the rapid evolution of new subgenres. Ensemble techniques, such as Bagging, Random Forests, and AdaBoost, are recognized for their capacity to reduce variance and mitigate overfitting, while Support Vector Classifier (SVC) has proven to be a strong standalone model for many classification tasks. However, the question arises whether a carefully orchestrated stacking framework can leverage these diverse learners to provide improvements in stability and accuracy.

Music genre classification is a highly active area of investigation. Reference [1] presents an Open Set Recognition (OSR) approach, enabling classifiers to accommodate the continual emergence of previously unseen genres by permitting an "unknown" label. Similarly, [2] demonstrates that XGBoost can rival or even surpass deep CNNs in certain scenarios, highlighting how feature engineering and hyperparameter tuning remain crucial for complex audio data. In [3], "musically motivated" CNN architectures are explored, aligning neural filters with known harmonic and rhythmic patterns. These studies collectively reveal that, although deep learning has gained prominence, ensemble and boosting methods retain competitiveness, especially when harnessed to exploit domain-specific nuances. Nevertheless, the challenge of multi-class classification persists, as many audio recordings lie at the intersection of multiple genres or exhibit hybrid styles.

The objective is to develop and evaluate a suite of ensemble-based classification models on the *fma_small* dataset, focusing on both hyperparameter tuning and meta-ensemble integration. Our research specifically examines Bagging, AdaBoost, Random Forests, and SVC, then unifies them in a stacking model. We address the "white spots" surrounding how to optimally merge distinct algorithms under a single meta-learner. In addition, we aim to investigate whether maintaining



subtle correlations among audio features can ultimately boost classification. By selecting a subset of six genres (Experimental, Hip-Hop, Electronic, Instrumental, International, and Rock) out of the available eight in `fma_small`, we form a balanced corpus of 6000 samples (1000 per class), enabling a clear-cut experimental framework.

As a first step, we excluded features with correlations above 0.95 to eliminate redundancy. Despite exploring advanced feature reduction methods, such as PCA and RFE, results showed a decrease in accuracy, suggesting that the nuanced interrelationships within the original feature set might be beneficial for capturing genre-specific patterns. Each of the 6000 samples was labeled among the six chosen categories, ensuring balanced class distribution (1000 examples per genre).

Subsequently, both a Decision Tree and an SVC were tuned using Optuna, a hyperparameter optimization library. Unlike exhaustive grid searches, which evaluate a static set of configurations, Optuna iteratively refines its sampling to focus on promising parameter values, thus improving efficiency. The standalone decision tree, if unoptimized, obtained only 50% accuracy, but significantly improved within ensemble methods: Bagging and Random Forest each achieved 67%, while AdaBoost reached 69%. After adjusting kernel parameters, class weighting, and the regularization factor, the SVC alone scored 70%.

To further increase performance, a stacking ensemble was built that combined SVC, Bagging, AdaBoost, and Random Forest through a Logistic Regression meta-classifier. We generated out-of-fold (OOF) predictions with five folds for each base model, ensuring that the meta-learner was trained on predictions not derived from the same data on which the base models were fit (thus reducing overfitting). This integrated approach yielded 72% accuracy, notably outperforming any individual model, and exhibited more consistent results across folds, pointing to enhanced robustness. The diverse error patterns of the base classifiers served to offset each other under the stacking strategy, leading to this overall improvement.

Our experimentation affirms that ensemble methods benefit music genre classification tasks. Although a well-tuned SVC alone reached 70% accuracy, the stacking ensemble surpassed it with 72%, demonstrating stronger generalization. The `fma_small` dataset provided a balanced environment to evaluate these methods. We found that removing only extremely correlated features while retaining the rest preserved essential relationships that various algorithms could exploit.



For future work, we suggest exploring additional ensemble techniques, such as gradient boosting architectures, or incorporating multimodal data (e.g., textual metadata or user-generated tags). Investigating complex multi-label scenarios where multiple genres may coexist in a single track also remains compelling. Overall, the synergy of multiple base classifiers, particularly within a stacking framework, proves highly promising for constructing robust and scalable genre classification systems in large-scale music repositories.

References:

- [1] Liu, Z., De Mori, J., & Abayomi, K. (2022). Open Set Recognition for Music Genre Classification. Proceedings of the 23rd International Society for Music Information Retrieval Conference.


ДОДАТОК В

Звіт результатів перевірки на унікальність тексту в базі ХНУРЕ

Дата звіту 6/8/2025

Дата редагування ---



Звіт не був оцінений

Звіт подібності

метадані

Назва організації
Kharkiv National University of Radio Electronics


Заголовок
2025_М_ПІ_ІПЗ-23-2_Мовчан_А_С_скорочений

Автор Науковий керівник / Експерт
Мовчан Андрій СергійовичЄвген Кардаш


підрозділ
каф. ПІ

Обсяг знайдених подібностей

Коефіцієнт подібності визначає, який відсоток тексту по відношенню до загального обсягу тексту було знайдено в різних джерелах. Зверніть увагу, що високі значення коефіцієнта не автоматично означають плагіат. Звіт має аналізувати компетентна / уповноважена особа.



0.00%
0.00% КП 1



0.53%
0.53% КЦ

25

Довжина фрази для коефіцієнта подібності 2

8251

Кількість слів

64592

Кількість символів

Тривога

У цьому розділі ви знайдете інформацію щодо текстових спотворень. Ці спотворення в тексті можуть говорити про **МОЖЛИВІ** маніпуляції в тексті. Спотворення в тексті можуть мати навмисний характер, але частіше характер технічних помилок при конвертації документа та його збереженні, тому ми рекомендуємо вам підходити до аналізу цього модуля відповідально. У разі виникнення запитань, просимо звертатися до нашої служби підтримки.






Заміна букв		0
Інтервали		0
Мікропробіли		0
Білі знаки		0
Парафрази (SmartMarks)		0

Рисунок В.1 – Звіт подібності в системі StrikePlagiarism

ДОДАТОК Г

Експертний висновок результатів перевірки кваліфікаційної роботи на
відповідність вимогам ДСТУ 3008: 2015

Експертний висновок результатів перевірки кваліфікаційної роботи

студент
(посада)

програмної інженерії
(кафедра)

ІПЗМ-23-2
(група)

Мовчан Андрій Сергійович

(прізвище, ім'я, по батькові)

Зауваження

Пункт ДСТУ 3008-2015	Зміст пункту	Сторінка кваліфікаційної роботи
1	2	3
	7.1 Загальні положення	
	7.3 Нумерація сторінок звіту	
	7.4 Нумерація розділів, підрозділів, пунктів, підпунктів	
	7.5 Рисунки	
	7.6 Таблиці	
	7.7 Переліки	
	7.8 Примітки	
	7.9 Виноски	
	7.10 Формули та рівняння	
	7.11 Посилання	
	7.13 Список авторів	
	7.14 Скорочення та умовні позначки	
	7.15 Додатки	

зауважень немає

Експерт

(підпис)

Олена ОЛІЙНИК

(прізвище, ініціали)

09.06.2025

Рисунок Г.1 – Експертний висновок результатів перевірки роботи