

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
Національний університет «Запорізька політехніка»

ФАКУЛЬТЕТ КОМП'ЮТЕРНИХ НАУК І ТЕХНОЛОГІЙ

(повне найменування інституту, факультету)

КАФЕДРА «СИСТЕМНИЙ АНАЛІЗ ТА ОБЧИСЛЮВАЛЬНА

МАТЕМАТИКА»

(повне найменування кафедри)

Пояснювальна записка

до дипломного проєкту (роботи)

магістр

(ступінь вищої освіти)

на тему **АНАЛІЗ РІВНЯ ВИКИДІВ ШКІДЛИВИХ РЕЧОВИН
ВІД АВІАЦІЙНИХ ГАЗОТУРБІННИХ ДВИГУНІВ
ANALYSIS OF THE LEVEL OF HAZARDOUS SUBSTANCES
EMISSIONS FROM AVIATION GAS TURBINE ENGINES**

Виконав: студент 2 курсу, групи КНТз-813м

Спеціальності 124-Системний аналіз

(код і найменування спеціальності)

Освітня програма (спеціалізація)

«Інтелектуальні технології та прийняття
рішень в складних системах»

ГУССВ В.М.

(прізвище та ініціали)

Керівник БАКУРОВА А.В.

(прізвище та ініціали)

Рецензент ЛОЗОВСЬКА Л.І.

(прізвище та ініціали)

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
Національний університет «Запорізька політехніка»
(повне найменування закладу вищої освіти)

Факультет Комп'ютерних наук і технологій

Кафедра «Системний аналіз та обчислювальна математика»

Ступінь вищої освіти магістр

Спеціальність 124 – Системний аналіз
(код і найменування)

Освітня програма

(спеціалізація) «Інтелектуальні технології та прийняття рішень в складних системах»

ЗАТВЕРДЖУЮ

В.о. завідувача кафедри

Терещенко Е.В.

“ 23 ” листопада 2024 року

З А В Д А Н Н Я

НА ДИПЛОМНИЙ ПРОЄКТ (РОБОТУ) СТУДЕНТА(КИ)

Гусєва Володимира Миколайовича

(прізвище, ім'я, по батькові)

1. Тема проєкту (роботи) аналіз рівня викидів шкідливих речовин від авіаційних газотурбінних двигунів. Analysis of the level of hazardous substances emissions from aviation gas turbine engines.

керівник проєкту (роботи) Бакурова Анна Володимирівна д.е.н., проф.,

(прізвище, ім'я, по батькові, науковий ступінь, вчене звання)

затвержені наказом закладу вищої освіти від «20» листопада 2024 року № 480

2. Строк подання студентом проєкту (роботи) «23» листопада 2024 року

3. Вихідні дані до проєкту (роботи) рекомендована література, технічне завдання, статистичні дані бази сертифікаційних випробувань авіаційних двигунів за 1971...2021 роки.

4. Зміст розрахунково-пояснювальної записки (перелік питань, які потрібно розробити) В першому розділі розглядається питання актуальності прогнозування викидів шкідливих речовин авіаційними двигунами в Україні та світі, а також технологічні цілі провідних країн світу, стимулюючих промисловість знижувати викиди шкідливих речовин. Показано наявність зв'язка основних етапів створення інноваційної технології спалювання та часу запровадження технології у виробництві, який обумовлює актуальність розробки методів прогнозування викидів шкідливих речовин NOx для українських підприємств. В другому розглянуто основні етапи побудови моделі для прогнозування даних методами машинного навчання та методика реалізації

дослідження. В третьому розділі приведено опис комп'ютерної моделі прогнозування викидів шкідливих речовин авіаційними двигунами та оцінку результатів роботи моделі.

5. Перелік графічного матеріалу (з точним зазначенням обов'язкових креслень)___

6. Консультанти розділів проєкту (роботи)

| Розділ | Прізвище, ініціали та посада консультанта | Підпис, дата | |
|---------------|---|----------------|---------------------------|
| | | завдання видав | прийняв виконане завдання |
| 1 | БАКУРОВА А.В., д-р екон. наук, проф. | 01.10.2014 | 01.11.2024 |
| 2 | БАКУРОВА А.В., д-р екон. наук, проф. | 14.10.2014 | 29.11.2024 |
| 3 | БАКУРОВА А.В., д-р екон. наук, проф. | 11.11.2024 | 10.12.2024 |
| Нормоконтроль | ШИРОКОРАД Д.В., к.ф.-м.н., доцент. | 19.01.2025 | 20.01.2025 |

7. Дата видачі завдання “01” жовтня 2024 року.

КАЛЕНДАРНИЙ ПЛАН

| № з/п | Назва етапів дипломного проєкту (роботи) | Строк виконання етапів проєкту (роботи) | Примітка |
|-------|--|---|----------|
| 1 | Сформулювати мету та основні завдання дипломної роботи | 01.10.2014 | |
| 2 | Опрацювати літературу та існуючі дослідження за темою роботи | 01.11.2014 | |
| 3 | Розробка програмної реалізації для вирішення задачі | 29.11.2024 | |
| 4 | Розрахунки та аналіз даних | 6.12.2024 | |
| 5 | Оформлення пояснювальної записки | 12.12.2024 | |
| 6 | Попередній захист дипломної роботи та отримання рецензій. | 16.12.2024 | |
| 7 | Захист дипломної роботи. | 23.01.2025 | |

Студент(ка) _____ ГУССВ В.М.
(підпис) (прізвище та ініціали)

Керівник проєкту (роботи) _____ БАКУРОВА А.В.
(підпис) (прізвище та ініціали)

РЕФЕРАТ

Пояснювальна записка до дипломної кваліфікаційної роботи магістра:
68 с., 7 табл., 24 рис., 24 джерел.

Об'єкт дослідження – викиди шкідливих речовин (NO_x) отримані в результаті сертифікаційних іспитів авіаційних двигунів.

Предмет дослідження – методи аналізу показників рівня викидів шкідливих речовин, які утворюються під час роботи авіаційних газотурбінних двигунів на режимах злітно-посадкового циклу літака.

Мета дослідження – розробка методики аналізу та прогнозування рівня викидів шкідливих речовин від авіаційних газотурбінних двигунів з використанням методу випадкового лісу.

Результати роботи представлені побудовою математичної моделі для прогнозування викидів шкідливих речовин авіаційними двигунами на основі аналізу і узагальнення вітчизняних та зарубіжних наукових праць, законодавчих актів та нормативних документів, методологічних підходів до визначення факторів, які необхідно враховувати для прогнозування рівня викидів шкідливих речовин від авіаційних газотурбінних двигунів на режимах злітно-посадкового циклу літака. Ця модель може бути використана для оцінки екологічних наслідків різних сценаріїв розвитку авіації, розробки заходів щодо зниження викидів та прогнозування обмежень на рівень викидів, що вводяться регулятором, залежно від зростання технічної досконалості двигунів провідних авіаційних компаній.

ПРОГНОЗУВАННЯ ВИКИДІВ ШКІДЛИВИХ РЕЧОВИН,
МАШИННЕ НАВЧАННЯ, ЗАБРУДНЕННЯ АТМОСФЕРИ, NO_x, RANDOM
FOREST

ABSTRACT

Explanatory note to the master's thesis: 68 pp., 24 tables, 13 figures, 24 sources.

The object of the study is the level of emissions of harmful substances (NO_x) obtained as a result of certification examinations of aircraft engines.

The subject of the study is methods for analyzing indicators of the level of emissions of harmful substances that are formed during the operating modes of aviation gas turbine engines in the aircraft take-off and landing cycle modes.

The purpose of the research is the development of a methodology for analyzing and forecasting the level of emissions of harmful substances from aircraft gas turbine engines using the random forest method.

The results of the work are presented by building a mathematical model for forecasting emissions of harmful substances by aircraft engines based on the analysis and generalization of domestic and foreign scientific works, legislative acts and regulatory documents, methodological approaches to determining factors, that must be taken into account for forecasting the level of emissions of harmful substances from aircraft gas turbine engines during the take-off and landing cycle modes of the aircraft. This model can be used to assess the environmental consequences of various aviation development scenarios, develop measures to reduce emissions, and predict emission limits imposed by the regulator, depending on the growth of technical perfection of the engines of leading aviation companies.

**FORECASTING OF HARMFUL EMISSIONS, MACHINE LEARNING,
AIR POLLUTION, NO_x, RANDOM FOREST**

ЗМІСТ

| | |
|--|----|
| ЗАВДАННЯ | 2 |
| РЕФЕРАТ | 4 |
| ПЕРЕЛІК СКОРОЧЕНЬ ТА УМОВНИХ ПОЗНАЧЕНЬ..... | 7 |
| ВСТУП | 8 |
| 1 ХАРАКТЕРИСТИКА ОБ'ЄКТУ ДОСЛІДЖЕННЯ..... | 13 |
| 1.1 Нормативні вимоги до викидів шкідливих речовин | 13 |
| 1.2 Дослідження методів прогнозування викидів шкідливих речовин..... | 20 |
| 2 МАШИННЕ НАВЧАННЯ ДЛЯ ПРОГНОЗУВАННЯ ВИКИДІВ ШКІДЛИВИХ РЕЧОВИН АВІАЦІЙНИМИ ДВИГУНАМИ | 23 |
| 2.1 Моделі штучного інтелекту для прогнозування викидів шкідливих речовин | 23 |
| 2.2 Основні етапи побудови моделі для прогнозування даних..... | 29 |
| 2.2.1 Формулювання завдання прогнозування викидів шкідливих речовин авіаційними двигунами..... | 31 |
| 2.2.2 Етап підготовки та дослідження даних | 32 |
| 2.3 Залежність якості роботи алгоритму Random Forest від структурних параметрів алгоритму | 43 |
| 3 КОМП'ЮТЕРНА МОДЕЛЬ ПРОГНОЗУ ВИКИДІВ РЕЧОВИН ВІД АВІАЦІЙНИХ ГАЗОТУРБІННИХ ДВИГУНІВ..... | 46 |
| 3.1 Аналіз і обґрунтування вибору середовища для програмної реалізації | 46 |
| 3.2 Програмна реалізація..... | 48 |
| 3.3 Процес прийняття рішень у методі Random Forest для прогнозування викидів шкідливих речовин | 53 |
| 3.4 Результати роботи програми | 61 |
| ВИСНОВКИ..... | 64 |
| ПЕРЕЛІК ПОСИЛАНЬ..... | 65 |

ПЕРЕЛІК СКОРОЧЕНЬ ТА УМОВНИХ ПОЗНАЧЕНЬ

- КЗ – камера згоряння;
- ШР – шкідливі речовини;
- NO_x – оксиди азоту;
- LTO – один цикл посадки та зльоту;
- CAEP – Комітет з охорони навколишнього середовища від авіації;
- ICAO – Міжнародна організація цивільної авіації;
- ISA – Міжнародна стандартна атмосфера.

ВСТУП

Однією з найважливіших глобальних проблем сучасного суспільства є посилення негативного антропогенного впливу на навколишнє середовище. Існуючі та передові технічні цілі ACARE (Flightpath 2050, Challenge 3, «Захист навколишнього середовища та енергопостачання»), Ціль 1, «У 2050 році доступні технології та процедури дозволять знизити викидів NO_x на 90%. Це відносно можливостей типових нових літаків 2000 року») [1].

Приблизно 81% загальної енергії, виробленої у світі, є хімічною енергією викопного палива, що виділяється як результат хімічної реакції з киснем повітря у вигляді тепла. Основними продуктами повного згоряння вуглеводневого палива в повітрі є вуглекислий газ і вода в суміші з азотом. У той же час, при спалюванні вуглеводневого палива утворюються речовини, небезпечні для здоров'я людей і навколишнього середовища – шкідливі речовини (ШР). З метою контролю за викидами ШР на повітряному транспорті Комітетом з охорони навколишнього середовища від авіації (CAEP) та Міжнародною організацією цивільної авіації (ICAO) введено в 1986 році перші Міжнародні стандарти, які обмежили викиди NO_x (оксиди азоту), CO, незгорілі вуглеводні HC та дим. Основною метою є зменшення забруднення повітря в районі аеропортів за один цикл посадки та зльоту (LTO) літака.

Міжнародні екологічні стандарти (ICAO та CAEP) переважно зосереджені на обмеженні викидів NO_x. Це пов'язано з їх значним внеском у зміну клімату та забруднення повітря. NO_x відіграє ключову роль процесів руйнування озонового шару у верхніх шарах атмосфери, що посилюється на великих висотах, де працюють авіаційні двигуни. Дим і CO відносно швидко розсіюються і мають локальний вплив, тоді як NO_x має тривалий вплив на атмосферні процеси. Рівень викидів CO та диму теж регулюються, але стандарти для NO_x суворіші через їхню важливість для екосистем. З 1986 року існує практика постійного посилення обмежень у міжнародних стандартах

ІСАО спрямованих на зменшення викидів NO_x від турбовентиляторних двигунів (CAEP/2 у 1996 р., CAEP/4 у 2004 р., CAEP/6 у 2008 р., CAEP/8 у 2014 р.), зберігаючи викиди інших ШР на тому ж рівні [2]. Контроль та зниження викидів NO_x потребує інноваційних підходів до проектування двигунів та покращення паливної ефективності. Ці аспекти мають першорядне значення для екологічної стійкості авіації.

Тенденція до підвищення чистого ККД турбовентиляторного двигуна з метою покращення його паливної ефективності призводить до істотного підвищення робочого тиску і температури на вході в турбіну сучасних турбовентиляторних двигунів і до значного підвищення швидкості реакції утворення NO_x в камері згоряння (КЗ), що ускладнює проблему забезпечення жорсткості майбутніх екологічних стандартів. Обмеження CAEP/8 було запроваджено з 2014 року. Вони передбачають зниження викидів NO_x на 15% порівняно з рівнем обмежень 2008 року (або на 50% до рівня обмежень 1986 року). Цільовий технологічний рівень параметра викидів NO_x має бути знижений на 60% відносно рівня обмежень 2008 року [3] у довгостроковій перспективі (до 2030 року).

Захист навколишнього середовища від наслідків людської діяльності став головним викликом і метою розробників авіаційних двигунів у останні роки, і тому такі міркування стали критично важливим компонентом інженерного проектування та функціонування, а також формулювання політики та законодавства. Щоб оцінити та контролювати викиди літаків, потрібен точний інструмент для їх надійного прогнозування. Багато методів прогнозування індексу викидів NO_x доступні у відкритих джерелах, тоді як деякі методи, які використовуються у галузі авіабудування, потребують секретної інформації. В статті [3] проаналізовано багато важливих методів прогнозування і проводиться систематичне дослідження їх точності та застосовності.

Метою роботи є розробка методики аналізу та прогнозування рівня викидів шкідливих речовин від авіаційних газотурбінних двигунів з

використанням методу випадкового лісу. Для досягнення мети потрібно виконати наступні задачі:

1. Збір та підготовка даних.
 - 1.1. Збір даних про характеристики двигунів (тип, модель, рік випуску), режими роботи, види палива, а також даних про фактичні викиди.
 - 1.2. Очищення даних від помилок та викидів.
 - 1.3. Нормалізація даних для забезпечення сумісності.
2. Вибір математичної моделі з найбільш поширених моделей для рішення подібних задач.
 - 2.1. Регресійний аналіз використовується для встановлення залежності між викидами та різними факторами (наприклад, лінійна, поліноміальна регресія).
 - 2.2. Нейронні мережі здатні моделювати складні нелінійні залежності між вхідними та вихідними даними.
 - 2.3. Методи часових рядів застосовуються для прогнозування викидів у часі, враховуючи сезонність та тренди.
 - 2.4. Фізико-хімічні моделі використовуються для детального опису процесів горіння та утворення шкідливих речовин у двигуні.
 - 2.5. Машинне навчання використовується для автоматизації аналізу даних, прогнозування, розпізнавання образів, створення рекомендацій і вирішення складних задач.
3. Навчання моделі.
 - 3.1. Поділ даних на навчальну та тестову вибірки.
 - 3.2. Навчання вибраної моделі на навчальній вибірці з метою мінімізації помилки прогнозування.
 - 3.3. Оцінка якості моделі на тестовій вибірці.
4. Прогнозування.
 - 4.1. Використання навченої моделі для прогнозування викидів за заданих умов.
 - 4.2. Аналіз чутливості прогнозу зміни вхідних параметрів.

5. Оцінка точності прогнозу.
6. Розрахунок метрик точності (наприклад, середньоквадратична помилка, коефіцієнт детермінації).
 - 6.1. Порівняння прогнозованих значень із фактичними даними.

Методологічною основою дослідження стали праці та публікації вітчизняних і зарубіжних досліджень з питань прогнозування рівня викидів ШР авіаційними двигунами, база результатів сертифікаційних випробувань авіаційних двигунів ІКАО.

Прогнозування викидів шкідливих речовин авіаційними двигунами має низку важливих економічних наслідків, як авіаційної галузі, так і для суспільства в цілому.

1. Для авіаційної галузі:
 - 1.1. Авіакомпанія може заздалегідь вжити заходів щодо зниження викидів, що дозволяє уникнути штрафів та інших фінансових санкцій за недотримання екологічних стандартів.
 - 1.2. Розуміння джерел викидів стимулює розвиток нових технологій та матеріалів для створення більш екологічних двигунів, що у довгостроковій перспективі може знизити витрати на обслуговування та ремонт.
 - 1.3. Демонстрація прихильності до екологічних стандартів підвищує репутацію авіакомпанії серед пасажирів та інвесторів.
2. Для суспільства:
 - 2.1. Точне прогнозування та подальші заходи щодо зниження викидів сприяють покращенню якості повітря та мінімізації негативного впливу авіації на навколишнє середовище.
 - 2.2. Інвестиції у дослідження та розробку в галузі екологічно чистих авіаційних технологій можуть створити нові робочі місця та стимулювати економічне зростання.

2.3. Країни, які активно займаються зниженням викидів авіаційної галузі, можуть відігравати більш активну роль у міжнародному співробітництві з охорони навколишнього середовища. [1].

1 ХАРАКТЕРИСТИКА ОБ'ЄКТУ ДОСЛІДЖЕННЯ

1.1 Нормативні вимоги до викидів шкідливих речовин

Прогнозування викидів шкідливих речовин авіаційними двигунами є актуальним завданням у зв'язку зі зростаючою стурбованістю екологічними проблемами та посиленням міжнародних норм у галузі авіації. Останні дослідження в цій галузі демонструють значний прогрес, пов'язаний з розвитком обчислювальних потужностей, появою нових методів аналізу даних та більш глибоким розумінням процесів горіння в авіаційних двигунах.

Протягом останніх десятиліть було виконано багато досліджень, які оцінювали вплив викидів шкідливих речовин авіаційних двигунів на місцеву та регіональну якість повітря поблизу аеропорту. Сумарні викиди шкідливих речовин комерційних літаків зазвичай обчислюються на основі сертифікованих емісійних індексів [4], які надаються виробниками двигунів і відображаються в базі даних Міжнародної організації цивільної авіації [5].

Детальна інформація про нормативні вимоги до викидів газотурбінних двигунів для дозвукових комерційних літаків щодо обмеження викидів диму, незгорілих вуглеводнів (СН), окису вуглецю (СО) та оксидів азоту (NO_x) наведено у таблиці 1.1. У таблиці 1.1 також наведено нормативні обмеження ІКАО на викиди NO_x [6]. Як випливає з таблиці 1.1, нормативні вимоги постійно посилюються, стимулюючи галузь авіадвигунобудування впроваджувати технологічні рішення, що забезпечують нижчий рівень викидів NO_x.

Існуючі перспективні технологічні цілі ACARE (SRIA1, SRIA2, Flighpath2050), NASA (HSCT, AST і UEET), а також цільові рівні програм (NEWAC, LEMCOTECH) стимулюють галузь авіадвигунобудування розробляти науково-технічні рішення для подальшого впровадження технологій, які забезпечують на 50–70%, а у перспективі 80–90% від нормативних вимог CAEP2 без погіршення інших характеристик камер

згоряння (вимог до запуску, запасів щодо меж бідного та багатого зриву, повноти згоряння, викидів шкідливих речовин CO та числа димності, нерівномірності поля температур на виході з КЗ). У таблицях 1.2 та 1.3 наводяться технологічні цільові рівні шкідливих речовин CAEP, NASA, SRIA та Flightpath 2050 [1, 4, 6–8].

Таблиця 1.1 – Нормативний рівень викидів NO_x відповідно до вимог ІКАО

| Нормативні вимоги ІКАО | |
|--|--|
| CAEP/2, після 31.12. 1999 $LTO\ NO_x \leq 32 + 1,6 \cdot OPR$ | |
| CAEP/4 після 31.12. 2003 <u>-16.25 % відносно рівня CAEP/2</u> $LTO\ NO_x \leq \begin{cases} 19 + 1,6 \cdot OPR; & OPR \leq 30, F_{00} > 89\text{ kN} \\ 37,572 + 1,6 \cdot OPR - 0,2087 \cdot F_{00}; & OPR \leq 30; 26,7\text{ kN} < F_{00} \leq 89\text{ kN} \\ 7 + 2,0 \cdot OPR; & OPR > 30, F_{00} > 89\text{ kN} \\ 42,71 + 1,4286 \cdot OPR - 0,4013 \cdot F_{00} + 0,00642 \cdot OPR \cdot F_{00}; & OPR > 30; 26,7\text{ kN} < F_{00} \leq 89\text{ kN} \\ 32 + 1,6 \cdot OPR; & OPR > 62,5 \end{cases}$ $LTO\ NO_x = 67\text{ g/kN, for ORP}=30$ | |
| CAEP/6 після 31.12. 2007 $LTO\ NO_x \leq \begin{cases} 16,72 + 1,4080 \cdot OPR; & OPR \leq 30, F_{00} > 89\text{ kN} \\ 38,5486 + 1,6823 \cdot OPR - 0,2453 \cdot F_{00} - 0,0308 \cdot OPR \cdot F_{00}; & OPR \leq 30; 26,7\text{ kN} < F_{00} \leq 89\text{ kN} \\ 7 + 2,0 \cdot OPR; & 30 < OPR \leq 82,6, F_{00} > 89\text{ kN} \\ 46,16 + 1,4286 \cdot OPR - 0,5303 \cdot F_{00} + 0,00642 \cdot OPR \cdot F_{00}; & 30 < OPR \leq 82,6; 26,7\text{ kN} < F_{00} \leq 89\text{ kN} \\ -1,04 + 2,0 \cdot OPR; & 30 < OPR \leq 82,6; F_{00} > 89\text{ kN} \\ 32 + 1,6 \cdot OPR; & OPR > 82,6 \end{cases}$ <u>12 % відносно рівня CAEP/4</u> $LTO\ NO_x = 59\text{ g/kN, for ORP}=30$ | |
| CAEP/8 після 31.12. 2013 <u>-15 % відносно рівня CAEP/6,</u> $LTO\ NO_x \leq \begin{cases} 7,88 + 1,408 \cdot OPR, & OPR \leq 30, F_{00} > 89\text{ kN} \\ -9,88 + 2,0 \cdot OPR, & OPR > 30, F_{00} > 89\text{ kN} \end{cases}$ $LTO\ NO_x = 50,1\text{ g/kN, for ORP}=30$ | |

Таблиця 1.2 – Перспективні технологічні рівні викидів NOx

| Технологічні цілі CAEP | | |
|--|---|-----|
| Середньострокові технологічні цілі CAEP 2016 | -45% of CAEP/6 | [6] |
| Довгострокові технологічні цілі CAEP 2016 | -60% of CAEP/6 | [6] |
| Технологічні цілі SRIA, Flightpath | | |
| SRIA1 (2020) | -80 of CAEP/2 engine: -65%CAEP6 | |
| Flightpath 2050 | -90 of CAEP/2 engine: | |
| Перспективні цілі ІКАО potential future regulations | | |
| CAEP Потенційні рівні обмеження викидів | $CAEP/10 = -17.401 + 2\pi;$ $CAEP/12 = -23.791 + 2\pi;$ $CAEP/14 = -29.223 + 2\pi;$ $CAEP/16 = -33.839 + 2\pi$ | [4] |
| CAEP /18 effective December 31, 2033 | $LTO NOx = -37.763 + 2\pi$ $LTO NOx = 22 \text{ g/kN, for ORP}=30$ | [4] |
| Технологічні цілі NASA | | |
| NASA N+1(2015г.) | 40% CAEP6 LTO -55% Cruise NOx Emissions relative to 2005 best in class | |

Продовження таблиці 1.2 – Перспективні технологічні рівні викидів NO_x

| | | | |
|---|-----|--|--|
| NASA (2020–2025г.) to achieve Technology Readiness Level (TRL) 6 by the year 2020 | N+2 | 25% CAEP6 LTO -70% Cruise NO _x Emissions relative to 2005 best in class | |
| NASA (2030–3035г.) | N+3 | 20% CAEP6 LTO respectively, with the corresponding 30 OPR LTO values 12 -80% Cruise NO _x Emissions relative to 2005 best in class | |
| ERA | | 25 % CAEP6 respectively, with the corresponding 30 OPR LTO values 15 | |
| AST | | 50% CAEP2 respectively, with the corresponding 30 OPR LTO values 40 | |
| UEET | | 30 % CAEP2 respectively, with the corresponding 30 OPR LTO values 24 | |
| Технологічні цілі ЄС | | | |
| NEWAC | | -60%...-70% <u>відносно рівня</u> CAEP/2 Досягнути рівня готовності технології TRL 5–6 | |
| Lemcotec | | -65...-70% <u>відносно рівня</u> CAEP/2 | |

Таблиця 1.3 – Порівняння довгострокових цілей за факторами впливу авіації на довкілля між Політикою ІКАО, програмою досліджень та розробок ЄС та США

| Фактори впливу на навколишнє середовище від авіації | Політика ІКАО [6] | Цілі EU ACARE (FP2050 до 2050 року) [1] | Цілі US FAA та NASA (NSTC2010 [7] та CLEEN II [8] до 2035 року) [9] |
|--|---|--|---|
| Викиди NO _x | Обмежити чи зменшити вплив авіаційних викидів на місцеву якість повітря | Знизити викиди NO _x на 90% | Скорочення викидів NO _x на 80% відносного стандарту ІКАО CAEP/6 |
| Викиди парникового газу та споживання паливо/енергія | Обмежити чи зменшити вплив викидів авіаційного парникового газу: скорочення авіаційних викидів CO ₂ на 50% до 2050 року по відношенню до рівня 2005 року | Скорочення викидів CO ₂ на 75% на пасажиро-кілометр | Зниження вартості паливо/енергія на 60% (CO ₂ викиди на пасажиро-кілометр) |

Аерокосмічні дослідження ЄС для досягнення цілей Flightpath 2050 [1] стикаються з низкою проблем. Охорона навколишнього середовища продовжуватиме залишатися ключовим фактором для авіаційної галузі в цілому. Проблема щодо навколишнього середовища полягає в тому, що постійно знижувати вплив на навколишнє середовище в умовах зростання попиту на авіацію та скорочення існуючих джерел енергії. Перед промисловістю стоїть задача завдяки розвитку технологій у 2050 знизити

викиди CO₂ на 75% на пасажиро-кілометр, а викиди NO_x на 90% відносно рівня типового нового літака в 2000 році.

На рис. 1.1 показано зв'язок основних етапів створення інноваційної технології спалювання, які характеризуються рівнем готовності технології (TRL – Technology readiness level) та часу запровадження технології у виробництві. Показник TRL використовуються для кількісного визначення статусу технологічної зрілості елемента відповідно до стандарту ISO-16290 [10]. Схема відображає процес розвитку нової технології двигуна від початкової ідеї (TRL 3) до введення в експлуатацію (TRL 9) схема побудована на підставі експертного досвіду ДП «Івченко-Прогрес». Показано, як технологія проходить через різні стадії розвитку, включаючи дослідження, розробку, випробування та сертифікацію. Кожна стадія пов'язана зі зростанням складності та вимагає інтеграції в промисловий продукт.



Рисунок 1.1 – Зв'язок основних етапів створення інноваційної технології спалювання TRL та часу на основі експертного досвіду ДП «Івченко-Прогрес»

Як показує досвід ДП «Івченко-Прогрес» та інших підприємств, після отримання ультранизьких викидів NO_x на одному або кількох розрахункових режимах при випробуваннях на відсіку КЗ при випробуваннях інноваційної технології спалювання (що відповідає рівню TRL 3–4) для доведення нової технології до рівня TRL 9 необхідно витратити щонайменше 10–15 років [4].

Це пояснюється великою кількістю вимог, що висуваються до камер згоряння авіаційних двигунів:

- відповідність викидів забруднюючих речовин NO_x, CO, HC нормативним обмеженням;
- обмеження на радіальну та максимальну нерівномірність поля температур на виході з камери згоряння;
- вимоги до діапазону висот та швидкостей польоту під час запуску двигуна;
- відсутність віброгоріння у всьому діапазоні режимів експлуатації;
- відсутність коксування палива;
- низькі температурні градієнти елементів жарової труби;
- міцнісні вимоги;
- високий ресурс.

На рис. 1.2 показано прогнозовані [11] дані характерного рівня викидів NO_x для наступних технологій спалювання палива: PERM EV (швидке змішування з попереднім випаровуванням), MSFI (багатоступеневе впорскування палива), LDI (збіднене пряме впорскування), які використовуються для оцінки рівня ефективності викидів NO_x 2030 року.

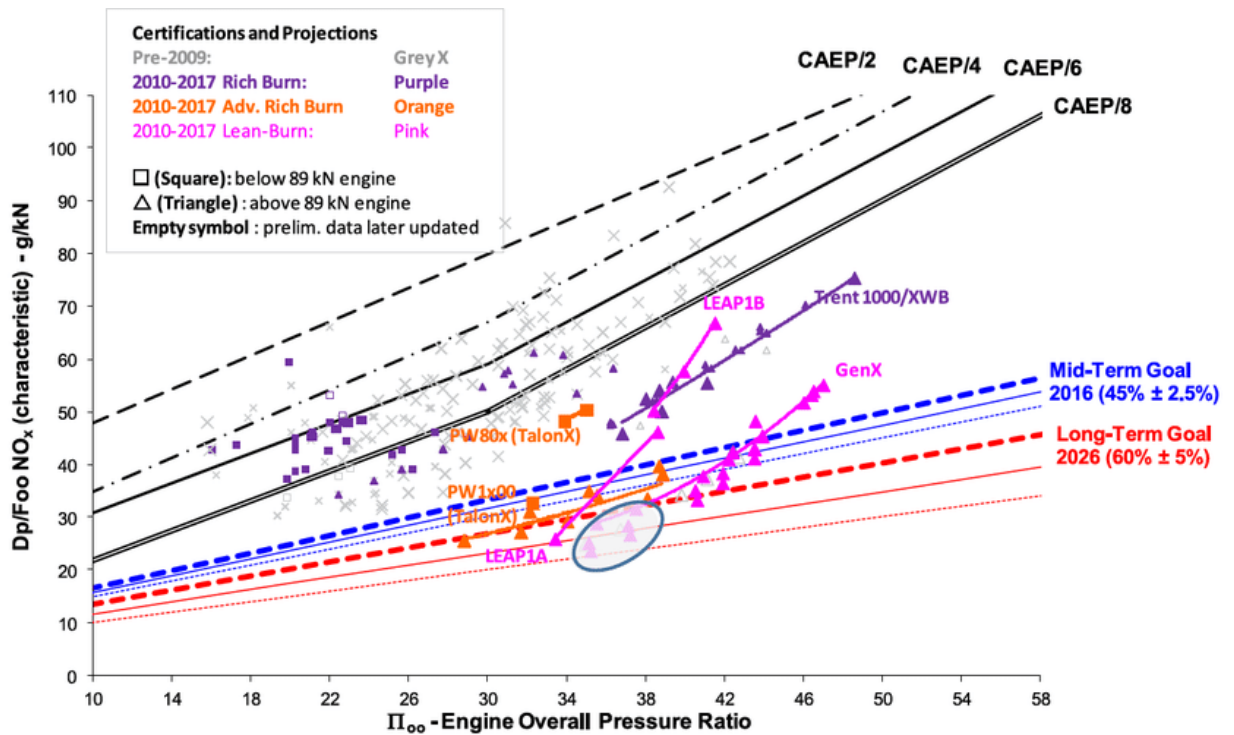


Рисунок 1.2 – Прогнозований рівень емісійної досконалості 2030 року в залежності від π_{Σ} в стандартних умовах на рівня моря (тиск = 1013.25 кПа, температура повітря = +15 °С)

1.2 Дослідження методів прогнозування викидів шкідливих речовин

Поточні дослідження процесів утворення шкідливих речовин про згоряння палива в камерах згоряння авіаційних двигунів охоплюють багато аспектів, таких як побудова моделей горіння, розробка підходів до моделювання горіння, удосконалення кінетичних схем для моделювання кінетики згоряння палива та оптимізація горіння в системах, розв'язанні рівняння масового переносу для різновидів NO_x з урахуванням конвекції, дифузії, виробництва та споживання NO_x та споріднених речовин. Багато експериментів і симуляцій генерують величезну кількість даних. Використання цих даних для кращого прогнозування викидів шкідливих речовин стало новим викликом і можливостями для дослідження. На щастя, машинне навчання (ML) надає передові методи для аналізу інформації з

масивних даних і допомагає виявити залежності утворення шкідливих речовин для камер згоряння в залежності від року сертифікації, який характеризує рівень технічної досконалості конструкції.

В роботі [4] наведено прогнозовані рівні викидів оксидів азоту (NO_x) для стандартного циклу зльоту-посадки. На основі результатів досліджень світового досвіду було розглянуто оціночний діапазон застосування перспективних схем спалювання палива з точки зору загального тиску в двигуні (OPR). Авторами було запропоновано використати методи регресійного аналізу, які дозволили встановити залежності між викидами та рядом факторів для невеликої кількості початкових даних. На підставі досвіду конструкторського бюро ДП «Івченко-Прогрес» у проектуванні камер згоряння було зроблено прогноз рівня викидів NO_x від авіаційних двигунів. Були визначені та показані орієнтовні темпи зниження рівня викидів NO_x для двигунів різних тягових класів до 2050 року. Показана ймовірність досягнення цілей програми «Flightpath 2050» [1], виклику 3, мети 1 для двигунів різних тягових діапазонів. К недолікам застосованих методів прогнозування слід віднести:

1. Обмежені можливості (через простоту моделей цей метод може не вловити складні взаємозв'язки між змінними).

2. Перенавчання (на малих вибірках існує ризик перенавчання моделі, коли вона надто добре підходить до навчальних даних, але погано узагальнюється на нові).

Запровадження методів аналізу даних за допомогою методів машинного навчання має наведені нижче переваги:

1. Великий обсяг даних (нейронні мережі дозволяють проаналізувати велику кількість даних, що дає можливість знайти складні закономірності).

2. Складні моделі (нейронні мережі можуть виявити та моделювати складні нелінійні залежності між змінними).

Для українських підприємств впровадження сучасних методів прогнозування викидів шкідливих речовин є актуальним завданням. Для таких

підприємств, як ДП «Івченко-Прогрес», точне прогнозування викидів NO_x має важливе значення для проектування та розробки двигунів, що відповідають екологічним стандартам.

Точне прогнозування може допомогти українським дослідникам визначити області, де можна удосконалити конструкцію та роботу двигунів для подальшого скорочення викидів NO_x.

На підставі аналізу публікацій на тему дослідження можна зробити такі висновки:

1. Завдяки використанню сучасних методів та великих даних можна значно підвищити точність прогнозування викидів шкідливих речовин авіаційними двигунами.

2. Викиди шкідливих речовин авіаційними двигунами залежать від багатьох факторів, і для точного прогнозування необхідно враховувати їхню комплексну взаємодію.

3. Незважаючи на досягнуті успіхи, залишається ще багато невирішених завдань, таких як прогнозування викидів нових типів двигунів.

2 МАШИННЕ НАВЧАННЯ ДЛЯ ПРОГНОЗУВАННЯ ВИКИДІВ ШКІДЛИВИХ РЕЧОВИН АВІАЦІЙНИМИ ДВИГУНАМИ

2.1 Моделі штучного інтелекту для прогнозування викидів шкідливих речовин

Машинне навчання (Machine Learning, ML) — це клас методів штучного інтелекту, що поєднує підходи та технології для створення алгоритмів, здатних самостійно навчатися на основі даних. На відміну від традиційних методів, при яких обчислювальні системи запрограмовані на виконання певних завдань, ML-моделі здатні покращувати свою роботу з часом, аналізуючи дані та виявляючи в них закономірності. Це дозволяє ML-системам вирішувати завдання, які складно чи неможливо запрограмувати вручну. Машинне навчання в галузі штучного інтелекту включає розробку моделей та алгоритмів, які дозволяють комп'ютерам вчитися на даних та приймати рішення чи прогнозувати рівень викидів в майбутньому [12].

На рис. 2.1 показані різні алгоритми машинного навчання різних додатків.

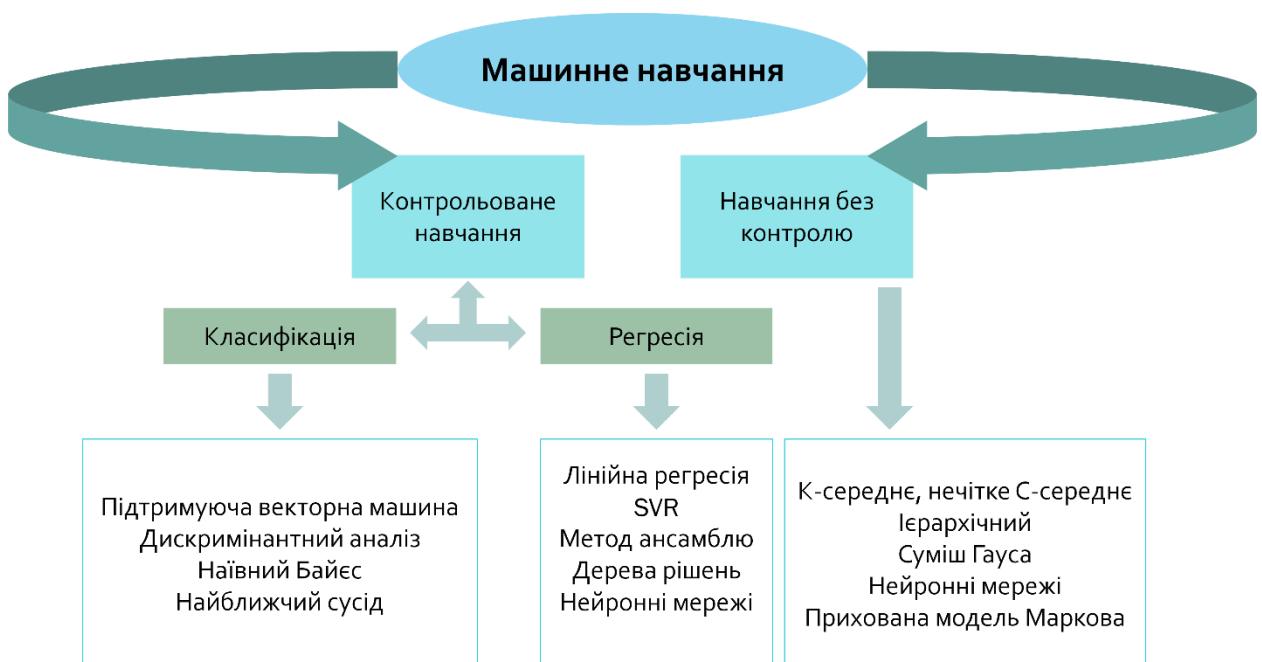


Рисунок 2.1 – Сучасні алгоритми машинного навчання [13]

Алгоритми ML будують модель на основі заданих даних, які походять зі статистики. Розвиток статистики прискорив розвиток ML. ML має на меті знайти закономірності та механізми серед величезної кількості даних, побудувати моделі на основі даних для прогнозів і допомогти вирішити проблеми. Загалом ML є дуже потужним інструментом для обробки складних і нелінійних фізичних і хімічних процесів, таких як горіння. Крім того слід зазначити, що високопродуктивні обчислення допомагають працювати з величезними обсягами даних і прискорюють процес моделювання фізичних явищ, інтелектуального аналізу даних та штучного інтелекту. Завдяки швидкому розвитку ML, його зручному розгортанню та покращенню продуктивності апаратного забезпечення, ML поширився на сферу дослідження горіння, і багато проблем горіння можуть бути вирішені шляхом прийняття методів ML [14]. ML надає різні моделі для обробки даних, які в основному можна класифікувати на три типи: контрольоване навчання, неконтрольоване навчання та навчання з підкріпленням, і існує багато конкретних моделей для кожного типу, як показано на рис. 2.2. У методах ML можна виділити три основні елементи: моделі, критерії навчання та оптимізацію.

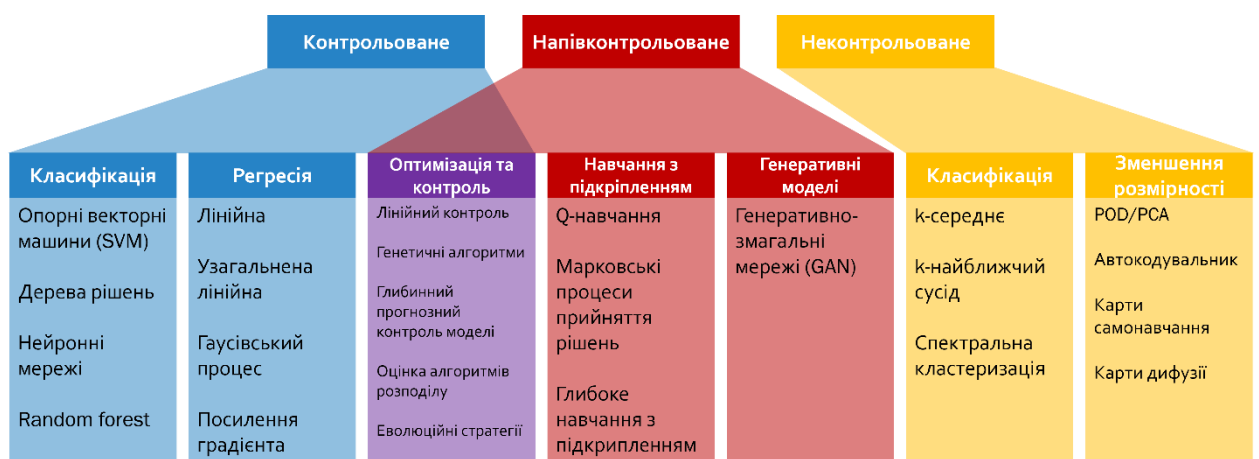


Рисунок 2.2 – Класифікація сучасних алгоритмів машинного навчання [14]

Алгоритми машинного навчання можна описати як навчання цільової функції f , яка найкраще співвідносить вхідні змінні X та вихідну змінну Y : $Y = f(X)$. Найбільш поширеним завданням у машинному навчанні є передбачення значень Y для нових значень X . Це називається прогностичним моделюванням, і мета вибору функції – зробити якомога точніше передбачення. У контрольованому навчанні вихідний простір часто містить дискретні значення для проблем класифікації або безперервні значення для задач регресії. У неконтрольованому навчанні результатом можуть бути кластери, низьковимірні вектори тощо.

Процес навчання ML триває на основі критеріїв або політики навчання. Якість моделі вимірюється за допомогою очікуваного ризику

$$R(\theta) = E_{(x,y) \sim p_r(x,y)}[L(y, \hat{y})] \quad (2.1)$$

де $p_r(x,y)$ – розподіл реальних даних;

$L(y, \hat{y})$ – функція втрат (loss function), яка кількісно визначає різницю між реальним вихідним значенням y і прогнозованим за моделлю значенням \hat{y} ;

Θ – параметри моделі;

y – реальні вихідні значення;

\hat{y} – прогнозовані значення.

Функція втрат може приймати різні форми [15], найбільш типовим вибором для задач регресії є:

– квадратична функція втрат MSE

$$L(y, \hat{y}) = (\hat{y} - y)^2; \quad (2.2)$$

– середня абсолютна помилка MAE

$$L(y, \hat{y}) = |\hat{y} - y|; \quad (2.3)$$

– середня відносна помилка MRE

$$L(y, \hat{y}) = \frac{|\hat{y} - y|}{y}; \quad (2.4)$$

– коефіцієнт детермінації R²

$$R^2(y, \hat{y}) = 1 - \frac{\sum(y_i - \hat{y}_i)^2}{\sum(y_i - \bar{y})^2}, \quad \bar{y} = \frac{\sum y_i}{n}. \quad (2.5)$$

Критерій навчання полягає в мінімізації очікуваного ризику для отримання моделі з хорошою точністю прогнозування. Якісно навчена модель ML повинна мати хорошу здатність до узагальнення, тобто здатність передбачати невідомі дані. Якщо навчена модель відповідає лише навченим даним, але погано працює на нетренованих тестових даних, вона, швидше за все, буде надмірною. Існує кілька підходів для покращення узагальнення моделі ML, таких як розділення наборів даних, перехресна перевірка, рання зупинка. Типова схема навчання моделі з використанням підходу розподілу даних наведена на рис. 2.3. Перед навчанням моделі ML набір даних буде розділено на дві групи, тобто навчальний набір і тестовий набір даних. Навчальний набір даних використовується для навчання моделі, а тестовий набір даних використовується для перевірки здатності моделі до узагальнення.

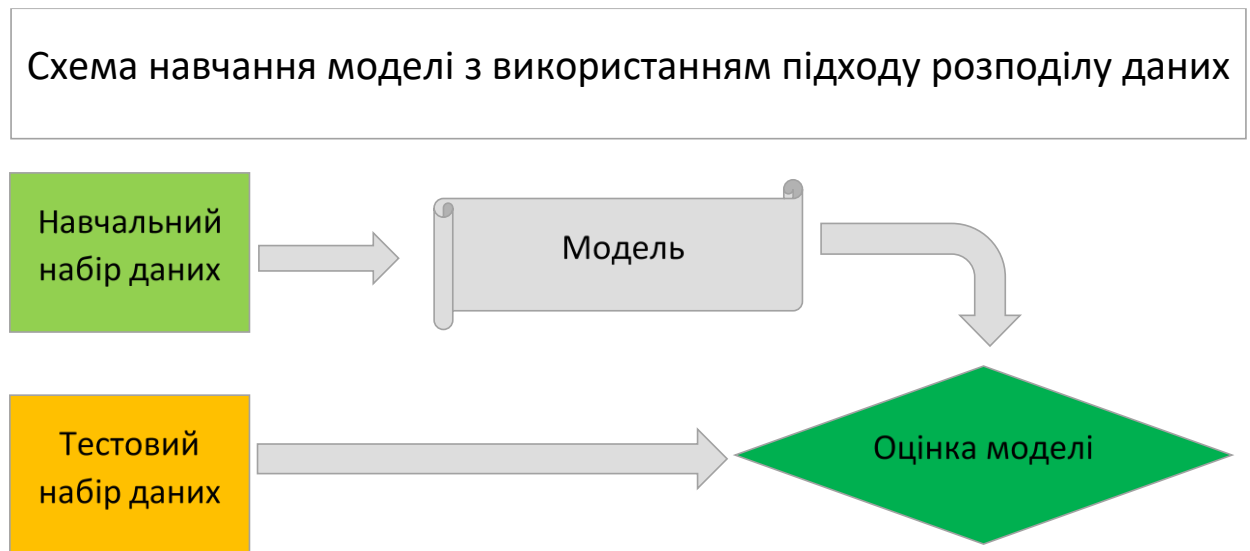


Рисунок 2.3 – Типова схема навчання моделі з використанням підходу розподілу даних

Існують різні види методів машинного навчання [16], такі як:

- контрольоване навчання – цей вид навчання передбачає поєднання вхідних даних із відповідним виходом після того, як алгоритм навчено на позначеному наборі даних. Модель має з’ясувати, як відображаються входи та виходи, щоб мати можливість передбачити нові, неспостережувані дані.

- Неконтрольоване навчання – це навчання, при якому алгоритм відповідає за розпізнавання шаблонів або структур у даних, які не позначені. Типові дії під час неконтрольованого навчання включають зменшення розмірності та кластеризацію даних.

- Напівконтрольоване навчання – цей тип навчання поєднує контрольовані та неконтрольовані методи. Щоб алгоритм навчався на основі позначених даних і узагальнював дані на немарковані дані, він навчається на наборі даних, який включає як позначені, так і непомічені зразки.

- Навчання з підкріпленням – за допомогою цього методу агента навчають максимізувати кумулятивний сигнал винагороди, приймаючи рішення в певному порядку всередині середовища. Він набуває досвіду, роблячи помилки та отримуючи зворотний зв’язок у вигляді заохочень або покарань за свої вчинки.

– Глибоке навчання – це одна з підсфер машинного навчання, яка називається «глибоким навчанням», що використовує мережу взаємопов'язаних «глибоких нейронних мереж», щоб зрозуміти, як ієрархічно представляти дані самостійно. Глибоке навчання було дуже успішним у багатьох сферах, включаючи обробку природної мови, розпізнавання звуку та ідентифікацію зображень.

Слід пам'ятати, що вибір архітектури нейронної мережі включає число і тип нейронів, а вибір алгоритму навчання може суттєво вплинути на його ефективність. Моделі та архітектура нейронної мережі повинні визначатися індивідуально для кожного випадку.

Надійне й точне прогнозування забруднення повітря вимагає ретельного розгляду вхідних параметрів. Різні типи моделей штучного інтелекту та машинного навчання можна використовувати для передбачення та прогнозування забруднювачів повітря зображено на рис. 2.4 [13].

У роботі [17] було дано початковий статистичний аналіз даних, досліджено параметри нейронної мережі, побудованої на основі Extreme Learning Machines, наведено результати, отримані для цієї моделі, та моделі на основі алгоритму Random Forest. Автори пропонують два показники якості моделей – модуль середнього відхилення (MAE) та коефіцієнт детермінації (R2).

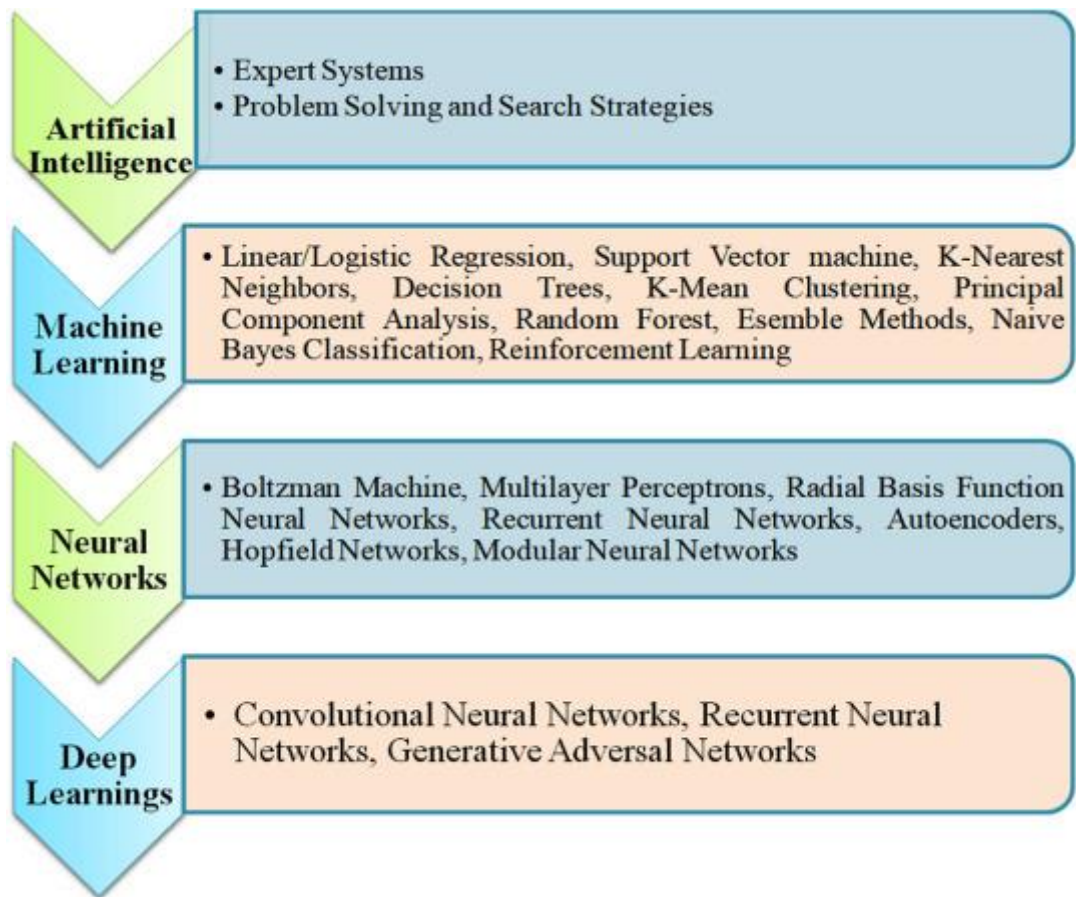


Рисунок 2.4 – Моделі штучного інтелекту для прогнозування викидів шкідливих речовин [13]

Найкращі з отриманих моделей у роботі [16] мають $MAE=0,97$, $R^2=0,56$ для викидів CO та $MAE=4,57$, $R^2=0,67$ для викидів NOx .

Представлені у роботі результати досліджень свідчать про доцільність використання методу Random Forest для прогнозування викидів шкідливих речовин з високими прогнозними якостями.

2.2 Основні етапи побудови моделі для прогнозування даних

Побудова моделі для прогнозування даних є частиною більш масштабного процесу, до якого входять усі завдання, від формулювання питань щодо даних та створення моделі для відповіді на ці питання до

розгортання моделі в робочому середовищі. Цей процес можна представити як послідовність наступних шести базових кроків [19]:

- формулювання завдання;
- підготовка даних;
- вивчення даних;
- побудова моделей;
- дослідження та перевірка моделей;
- розгортання та оновлення моделей.

На рис. 2.5 описані зв'язки між кожним етапом процесу побудови моделей та технологіями нейронних мереж, які можна використовувати для виконання кожного кроку [19].



Рисунок 2.5 – Схема зв'язків між кожним етапом процесу побудови моделей та технологіями

Процес представлений на рис 2.3 може вимагати виконати декілька послідовних циклів [18]. Наприклад на етапі вивчення даних, розробник може прийти до висновку, що існуючих даних недостатньо для створення побудови моделі, що потребує повернення до етапу підготовки даних та пошуку

додаткових джерел даних. Також на етапі побудови моделей може виникнути ситуація, коли виявиться, що усі розроблені моделі не дають адекватної відповіді на поставлене завдання, і тому необхідно виконати етап уточнення завдання. Може виникнути потреба в оновленні вже розгорнутих моделей за рахунок нових даних, отриманих після завершення етапу розробки моделі. Для розробки високоточної моделі може знадобитися багаторазово повторити кожний крок процесу [19].

2.2.1 Формулювання завдання прогнозування викидів шкідливих речовин авіаційними двигунами

Необхідно провести прогнозування розподілу максимальної концентрації забруднюючої речовини NO_x на режимах злітно-посадкового циклу літака.

В роботі використано dataset, який було сформовано на основі бази ІКАО сертифікаційних випробувань авіаційних двигунів за 1971..2021 роки [5]. Всього dataset містить 834 випробування двигунів. Розподіл компаній виробників наведено на рис. 2.6 та в таблиці 2.1.

Нейромережева модель буде складатися з чотирьох входів:

- вхід x_1 – ступень двоконтурності B/P Ratio;
- вхід x_2 – загальний ступінь стиснення PressureRatio;
- вхід x_3 – тяга двигуна RatedThrust;
- вхід x_4 – концентрація CO – CODpFooChar.

Та одного виходу:

- вихід y_1 – концентрація NO_x – NOxDpFooChar.

Таблиця 2.1 – Компанії виробники авіаційних двигунів які представлені в dataset

| Виробник | Кількість записів |
|----------------------------|--------------------------|
| General Electric Company | 216 |
| Pratt & Whitney | 176 |
| Rolls-Royce plc | 148 |
| CFM International | 142 |
| Rolls-Royce Corporation | 34 |
| Pratt & Whitney Canada | 30 |
| International Aero Engines | 27 |
| Rolls-Royce Deutschland | 17 |
| Honeywell | 16 |
| Aviadvigatel | 8 |
| ККВМ | 5 |
| PowerJet S.A. | 4 |
| Textron Lycoming | 4 |
| IVCHENKO PROGRESS ZMBK | 4 |
| Allied Signal | 2 |
| Engine Alliance | 1 |

2.2.2 Етап підготовки та дослідження даних

Другим етапу процесу побудови моделі є об'єднання, очищення та розподіл даних, визначених на етапі формування завдання.

Очищення даних – це не тільки видалення неприпустимих даних або інтерполяція відсутніх значень, але й пошук даних прихованих залежностей, визначення джерел найточніших даних і підбір стовпців, які найбільше підходять для використання в аналізі. Неповні дані, помилкові дані та вхідні

параметри, які виглядають як незалежні, але насправді мають міцний взаємозв'язок можуть непередбачуваним чином вплинути на результати моделі.

Кореляція Пірсона – це фундаментальний статистичний метод, який використовується для розуміння лінійних зв'язків між двома неперервними змінними (формула 2.1).

$$r = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{(\sum(x_i - \bar{x})^2)\sum(y_i - \bar{y})^2}}, \quad (2.7)$$

де r – коефіцієнт кореляції;

x_i – значення змінної x у вибірці;

\bar{x} – середнє значення змінної x ;

y_i – значення змінної y у вибірці;

\bar{y} – середнє значення змінної y .

На рис. 2.7 та в таблиці 2.2 наведено результати обчислення коефіцієнту кореляції Пірсона для dataset, який є мірою лінійної залежності між двома змінними. Значення коефіцієнта Пірсона варіюється від -1 до 1, де значення близькі до 1 або -1 вказують на сильну лінійну залежність, а значення близькі до 0 вказують на слабку або відсутню залежність.

Таблиця 2.2 – Значення коефіцієнта кореляції Пірсона для dataset

| | B/PRatio | PressureRatio | RatedThrust | CODpFooChar | NOxDpFooChar |
|----------------------|-----------------|----------------------|--------------------|--------------------|---------------------|
| B/PRatio | 1,000 | 1,000 | 0,754 | -0,580 | 0,311 |
| PressureRatio | 1,000 | 1,000 | 0,754 | -0,580 | 0,311 |
| RatedThrust | 0,754 | 0,754 | 1,000 | -0,373 | 0,534 |
| CODpFooChar | -0,580 | -0,580 | -0,373 | 1,000 | -0,122 |
| NOxDpFooChar | 0,311 | 0,311 | 0,534 | -0,122 | 1,000 |

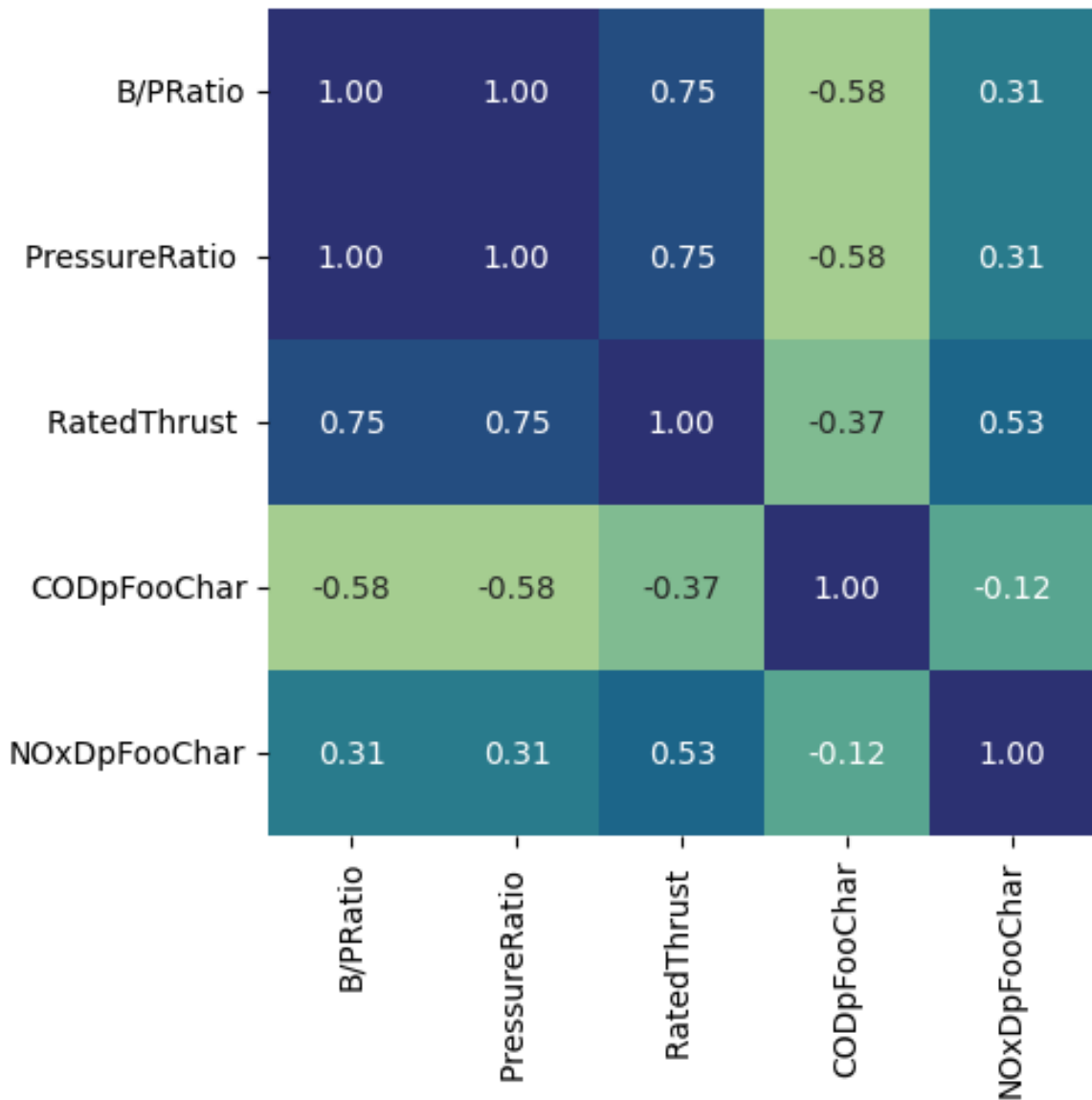


Рисунок 2.7 – Значення коефіцієнта кореляції Пірсона для dataset

При аналізі даних корисно переглядати всі діаграми розсіювання серед групи змінних, тобто побудувати парні графіки чи матрицю діаграм розсіювання [20]. У даній роботі використана функція *pairplot* бібліотеки *seaborn*, яка, що важливо, підтримує розміщення гістограм або оцінок щільності кожної змінної діагоналі. Наведена на рис. 2.8 матриця діаграм розсіювання дозволяє визначити вид та тісноту зв'язку між парами відповідних змінних. Для прийняття правильних рішень при побудові моделі необхідно проаналізувати дані. Методи дослідження даних включають розрахунок мінімальних і максимальних значень, обчислення середньо

ймовірного і стандартного відхилення і вивчення розподілу даних. Велика величина стандартного відхилення може свідчити, що додавання нових даних допоможе вдосконалити модель. Дані, які сильно відхиляються від стандартного розподілу, можуть бути спотвореними або представляти точну картину реальної проблеми, яка робить складним підбір відповідної моделі для даних [20, 21].

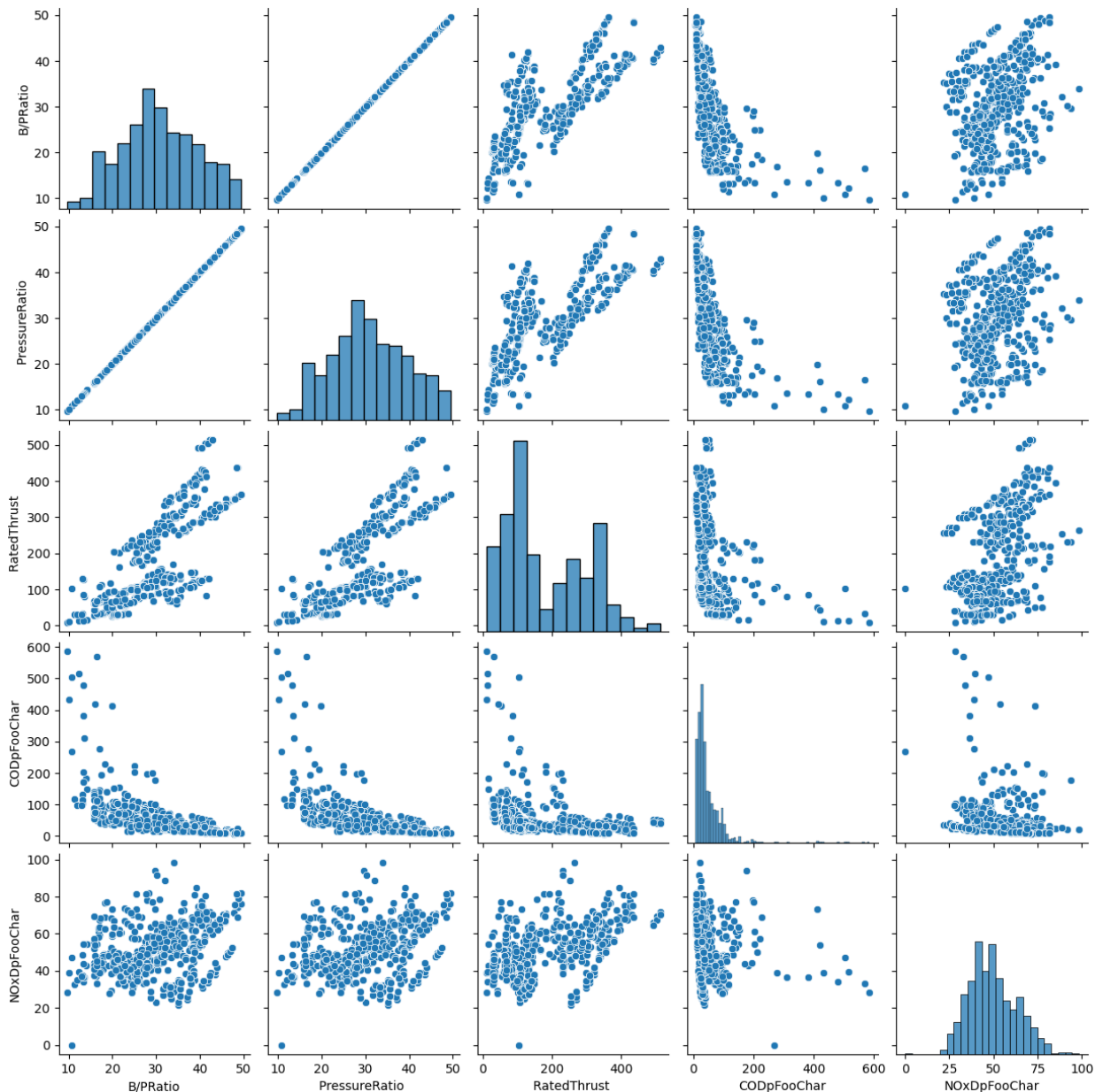


Рисунок 2.8 – Матриця діаграм розсіювання для dataset

Розподіл даних на навчальні та перевіірочні набори є важливою частиною оцінки моделей інтелектуального аналізу даних. Зазвичай під час поділу набору даних на навчальний і перевіірочний набори більшість даних використовують для навчання, а менша — для перевірки. В роботі був використаний метод Scikit-learn *train_test_split*, який формує вибірку даних випадковим чином, щоб переконатися, що набори для тестування та навчання схожі. Використання однакових даних для навчання та перевірки дозволяє звести до мінімуму вплив невідповідності даних та краще зрозуміти характеристики моделі. Після обробки моделі на основі навчального набору здійснюється її перевірка шляхом формування прогнозів на основі набору перевірок. Оскільки дані в перевіірочному наборі містять відомі значення для прогнозованого атрибута, дуже просто визначити, чи правильні висновки були отримані моделлю. На рис. 2.9 показано найпоширеніші варіанти розподілу даних.

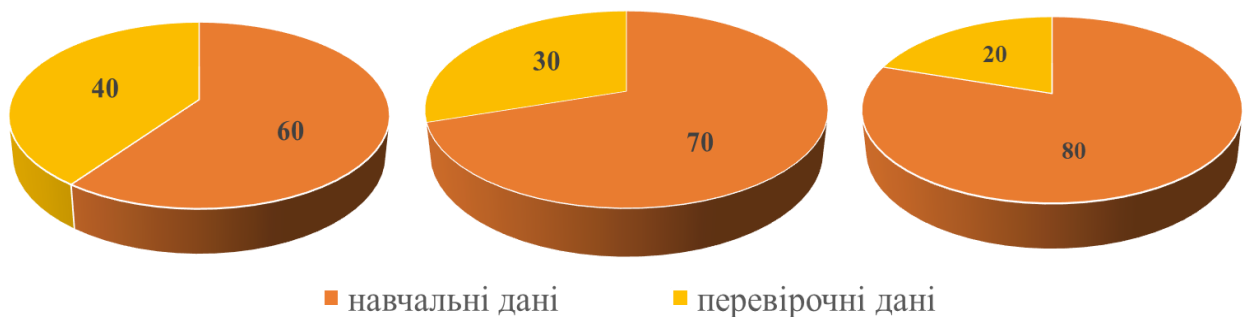


Рисунок 2.9 – Найпоширеніші варіанти розподілу даних

Не існує простого способу визначити, якого обсягу набору даних буде достатньо для повноцінного навчання і наскільки розрідженим може бути навчальний набір, щоб не викликати появи хибних взаємозв'язків. У більшості випадків варіанти розподілу даних, що приведені на рис. 2.7 забезпечують добрий баланс між навчальними та перевіірочними даними. Після побудови моделі доцільно виконати перехресну перевірку, щоб оцінити набір даних

щодо конкретної моделі. В розробленому проєкті було використано значення розподілу між навчальними та перевірочними даними – 70% та 30% відповідно.

2.2.3 Етап побудови моделей прогнозування викидів шкідливих речовин авіаційними двигунами

Побудова моделей прогнозування викидів шкідливих речовин авіаційними двигунами – це складний процес, який вимагає знань в області машинного навчання, статистики та авіаційної техніки.

Побудова моделі дозволяє досліджувати ступінь впливу різних факторів на ці процеси, як а також підготувати прогнози щодо очікуваних рівнів викидів шкідливих речовин у майбутньому спираючись на рівень TRL.

В роботі була використана бібліотека Scikit-learn – це бібліотека машинного навчання з відкритим кодом, написана мовою Python. Вона надає зручний інтерфейс для виконання завдань класифікації, регресії, кластеризації, зниження розмірності, обробки даних та валідації моделей. Бібліотека заснована на інших популярних бібліотеках Python, таких як NumPy, SciPy та Matplotlib, що забезпечує її продуктивність та широкі можливості для візуалізації. Scikit-learn має моделі, які можуть реагувати на більш тонкі особливості даних. Програмна реалізація бібліотеки Scikit-learn на мові Python відрізняється модульною архітектурою, зручним інтерфейсом і високою продуктивністю.

Ядро бібліотеки реалізовано на Python із використанням NumPy для роботи з масивами та SciPy для наукових обчислень. Важливо, що більшість обчислювально-витратних операцій (наприклад, розрахунок матриць або лінійної алгебри) виконується на рівні низькорівневих оптимізованих бібліотек (наприклад, BLAS, LAPACK). Scikit-learn надає зручний та

уніфікований інтерфейс програмування (API), який робить роботу з бібліотекою інтуїтивно зрозумілою для користувачів, що досягається завдяки реалізацію усіх моделей та методів за допомогою методів об'єктно-орієнтоване програмування.

Алгоритм випадкового лісу (Random Forest) [14] – універсальний алгоритм машинного навчання, суть якого полягає у використанні ансамблю вирішальних дерев. Саме по собі вирішальне дерево надає вкрай невисоку якість класифікації, але через велику їх кількість результат значно покращується. Також це один із небагатьох алгоритмів, який можна використовувати в абсолютній більшості завдань.

Порядок дій алгоритму:

1. Завантаження даних.
2. У заданому наборі даних визначити довільну вибірку.
3. Алгоритм будує на дану вибірку дерево рішень.
4. Дерево буде продовжувати будуватися, поки в кожному листку не буде більше n об'єктів, або поки не буде досягнуто певної заданої висоти.
5. Отримаємо результат прогнозування кожного дерева рішень.
6. Проводиться етап голосування для кожного прогнозованого результату: для кожної обраної ознаки робиться нове розбиття у дереву та виконується циклічне повторення голосування до вичерпання усієї вибірки.
7. Вибір результату прогнозу із найбільшою кількістю голосів – це і буде остаточним результатом прогнозування.

Принципову схему роботи алгоритму Random Forest наведено на рис. 2.10.

Алгоритм Random Forest не є нейронною мережею і, відповідно, не має архітектури, характерної для нейромереж. Натомість він відноситься до методів ансамблевого навчання і базується на комбінації кількох дерев рішень (decision trees). Порівняння архітектури алгоритму Random Forest з архітектурою нейронної мережи наведено у таблиці 2.3.

МОДЕЛЬ RANDOM FOREST CLASSIFIER

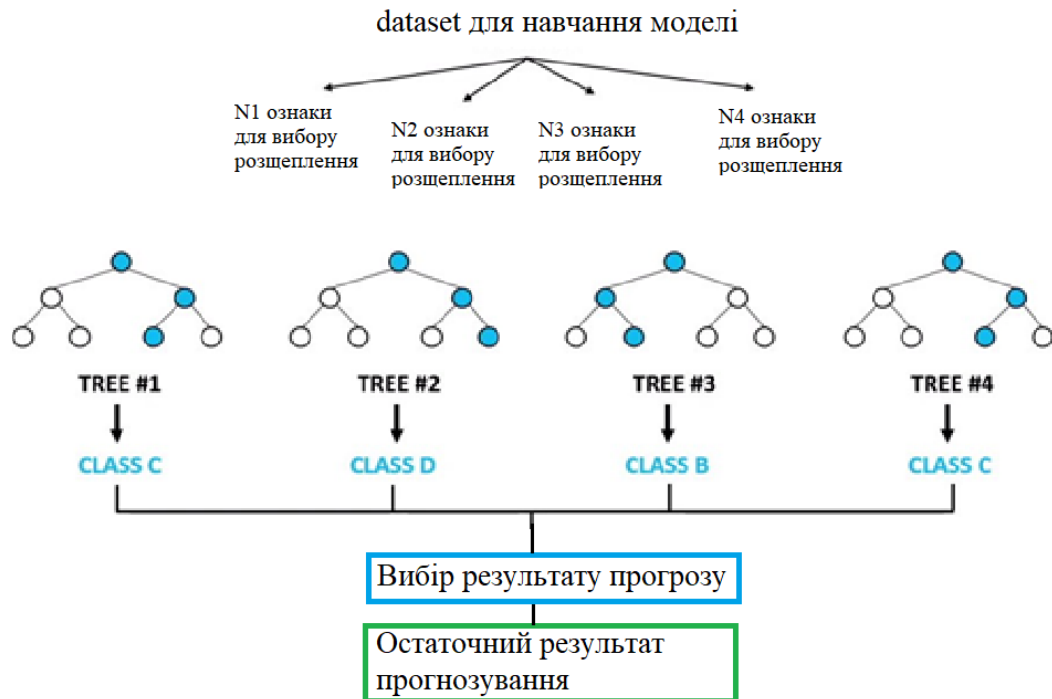


Рисунок 2.10 Принципова схема роботи алгоритму Random Forest [22]

Таблиця 2.3 Порівняння архітектури алгоритму Random Forest та нейронної мережи

| Особливість | Random Forest | Нейронна мережа |
|---------------------|--|--|
| Моделі | Безліч незалежних дерев рішень | Мережа шарів (вхідний, приховані та вихідний) |
| Взаємодія елементів | Дерева працюють паралельно та незалежно | Нейрони пов'язані та передають інформацію через шари |
| Навчання | Кожне дерево навчається на підмножині даних | Гرادієнтний спуск, заснований на обчисленні функції помилки |
| Інтерпретація | Проста (шлях ухвалення рішення можна відстежити) | Менш прозора, складна інтерпретація |
| Випадковість | Використовує випадкові вибірки даних та ознак | Немає випадкового семплювання (зазвичай робота з повним набором) |

Для об'єкту x алгоритм випадкового лісу робить прогноз таким чином (формула 2.2)

$$RF(x) = \sum_{i=1}^N \frac{a_i(x)}{N}, \quad (2.8)$$

де N – кількість вирішальних дерев, що необхідно підбирати вручну;
 $a_i(x)$ – вирішальне дерево.

Необхідні параметри алгоритму [23]:

1. Число дерев – $n_estimators$ (чим більше дерев, тим краща якість. Варто зазначити, що час налаштування та роботи Random Forest пропорційно збільшуватиметься, що може позначитися на продуктивності). Часто при великому збільшенні $n_estimators$ якість на навчальній вибірці може навіть сягати 100%, у той час як якість на тесті виходить на асимптоту, що сигналізує про перенавчання нашої моделі. Найкращий спосіб уникнути цього – прикинути, скільки дерев вам достатньо, зафіксувавши момент, коли якість тесту ще не стає стабільно незмінною.

2. Критерій розщеплення – $criterion$ (також один із найважливіших параметрів для побудови, але без значної можливості вибору. У бібліотеці *sklearn* для класифікаційних завдань реалізовані критерії *gini* та *entropy*. Вони відповідають класичним критеріям розщеплення). Для завдань регресії реалізовано два критерії (*mse* та *mae*), які є функціями помилок Mean Square Error та Mean Absolute Error відповідно.

3. Число ознак для вибору розщеплення – $max_features$ (при збільшенні $max_features$ збільшується час побудови лісу, а дерева стають схожими одне на одне. У задачах класифікації він за замовчуванням дорівнює \sqrt{n} , завдання регресії – . Є одним із найважливіших параметрів в алгоритмі. Він налаштовується насамперед після того, як визначено тип завдання).

4. Мінімальна кількість об'єктів для розщеплення – *min_samples_split* (другорядний за своїм значенням параметр, його можна залишити в стані за замовчуванням).

5. Обмеження кількості об'єктів у листі – *min_samples_leaf* (аналогічно з *min_samples_split*, але зі збільшенням даного параметра якість моделі навчання падає, тоді як час побудови моделі скорочується).

6. Максимальна глибина дерев – *max_depth* (чим менша максимальна глибина, тим швидше будується і працює алгоритм випадкового дерева. При збільшенні глибини різко зростає якість як навчання моделі, так і її тестування). Рекомендується використовувати максимальне значення даного параметра, але ще суттєво збільшить час побудови глибоких дерев. Зазначимо, що неглибокі дерева рекомендується використовувати в задачах зі значною кількістю шумових об'єктів (викидів).

Переваги даного алгоритму Random Forest [23]:

- має високу точність прогнозування, прогностичні якості якої східни з характеристиками методу градієнтного бустингу.

- Для багатьох класів задач не потребує ретельного налаштування параметрів, добре працює з параметрами за замовчуванням.

- Низька чутливість до масштабування та до інших монотонних перетворень характеристик моделі.

- Низькі показники перенавчання, з ростом кількості дерев відмічається тенденція до покращення якості прогнозу.

- Якщо було виявлено проблему перенавчання, то рекомендовано виконати усереднення чи об'єднання результатів прогнозування для різних дерев рішень.

- Здатний ефективно будувати прогностичну модель використовуючи дані з великою кількістю ознак.

- Характеризується прийнятною точністю за наявності пропущених даних.

- Однаково добре обробляє як безперервні, і дискретні ознаки.

- Висока паралелізованість та масштабованість.

Недоліки алгоритму Random Forest [23]:

- Для обчислення моделі алгоритму випадкового лісу потрібно мати значний обсяг обчислювальних ресурсів, проте цей недолік може бути вирішений шляхом використання сучасних промислових кластерних рішень або виконанням коду на віддалених серверах.

- Побудова випадкового лісу забирає більше часу, ніж дерева рішень чи лінійних алгоритмів, проте цей недолік може бути вирішений шляхом використання сучасних промислових кластерних рішень або виконання коду на віддалених серверах.

- Для зниження впливу цього недоліку необхідно ретельно перевіряти початкові дані на етапі підготовки даних.

- На відміну від простіших алгоритмів, результати випадкового лісу складніше інтерпретувати.

- Процес прогнозування з допомогою випадкових лісів дуже трудомісткий, на відміну від інших алгоритмів.

2.2.4 Етап дослідження та перевірки моделей

Перед запровадженням моделі у процес проектування двигунів необхідно виконати оцінку ефективності роботи моделі. Крім того, під час побудови моделі зазвичай створюється кілька моделей з різною конфігурацією, а потім перевіряються оцінки якості прогнозування всіма моделями, щоб визначити, яка з них забезпечує найкращі результати для поставленого завдання та наявних даних. Точність прогнозів, створених моделями, можна перевірити за допомогою засобів конструкторів, таких як діаграми точності прогнозів і матриці класифікації. Щоб перевірити, чи обмежена застосовність моделі наявними даними або вона може

використовуватися для висновків щодо генеральної сукупності, можна застосувати статистичний метод, званий перехресною перевіркою, щоб автоматично створити підмножини даних і перевірити модель на кожній підмножині.

2.3 Залежність якості роботи алгоритму Random Forest від структурних параметрів алгоритму

Перш ніж приступити до моделювання, важливо визначити відповідні вхідні параметри, які можна безпосередньо пов'язати із викидами.

Наступний крок – налаштування моделі. Якість роботи будь-якого алгоритму на тестовій вибірці сильно залежить від гіперпараметрів, що використовуються при його налаштуванні, тому важливим аспектом цього процесу є проведення наступних експериментів:

- проведення аналізу різних роздільних критеріїв, що використовуються для дроблення поточного завдання, що розглядається у вузлі дерева, на дрібніші підзавдання;

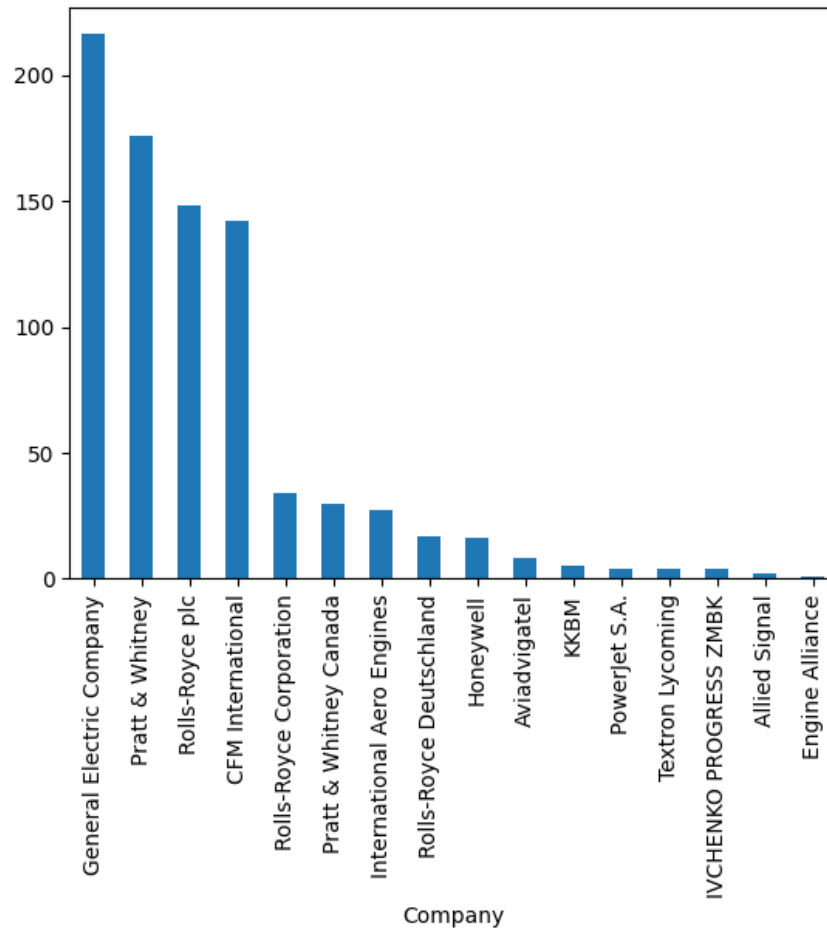


Рисунок 2.6 – Кількість експериментальних даних в залежності від виробника авіаційних двигунів, які представлені в dataset

- дослідження залежності якості роботи алгоритму від різної кількості ознак, що використовуються при побудові вузла;
- аналіз зв'язку якості розв'язання задачі з порогом на кількість об'єктів, що потрапляють до листа дерева.

Параметри алгоритму машинного навчання Random Forest [23], вплив яких досліджувався при розробці моделі:

- *#n_estimators* – число «дерев» у «випадковому лісі»;
- *#max_features* – число ознак для вибору розщеплення;
- *#max_depth* – максимальна глибина дерев;
- *#min_samples_split* – мінімальна кількість об'єктів, необхідна для того, щоб вузол дерева міг би розщепитися;
- *#min_samples_leaf* – мінімальна кількість об'єктів у листі;

– *#bootstrap* – використання для побудови дерев підвибірки із поверненням.

3 КОМП'ЮТЕРНА МОДЕЛЬ ПРОГНОЗУ ВИКИДІВ РЕЧОВИН ВІД АВІАЦІЙНИХ ГАЗОТУРБІННИХ ДВИГУНІВ

3.1 Аналіз і обґрунтування вибору середовища для програмної реалізації

У рамках кваліфікаційної роботи була розроблена модель прогнозу викидів шкідливих речовин від авіаційних газотурбінних двигунів.

Розробка моделі здійснювалася в інтерактивному середовищі Google Colaboratory, що забезпечує зручний доступ до обчислювальних ресурсів та популярних бібліотек Python.

Google Colab – це хмарний сервіс, який забезпечує виконання коду відбувається на віддалених серверах. Це дозволяє користувачам використовувати обчислювальну потужність Google без необхідності використання потужних локальних комп'ютерів. Обчислення проводилися на віртуальних машинах Google Colaboratory, що дозволило масштабувати обчислювальні ресурси відповідно до потреб проекту.

Google Colab забезпечує інтеграцію з іншими сервісами Google, такими як Google Drive та Google Cloud Storage. Це дозволяє користувачам легко отримувати доступ та зберігати дані, моделі та інші ресурси.

Як інструмент для розробки було обрано Jupyter Notebook, що дозволяє поєднувати код, текст та візуалізації в єдиному документі. Для реалізації моделі були використані бібліотеки Scikit-learn та Graphviz, що надають широкий набір інструментів для глибокого навчання та візуалізації результатів.

За допомогою програмного середовища Google Colab можна розробити себе якісний код, вбудовані алгоритми здійснюють перевірку коду на наявність синтаксичних помилок та дає рекомендації по їх виправленню.

На рис. 3.1 зображено середовище розробки програмного застосунку Google Colab за допомогою мови програмування Python.

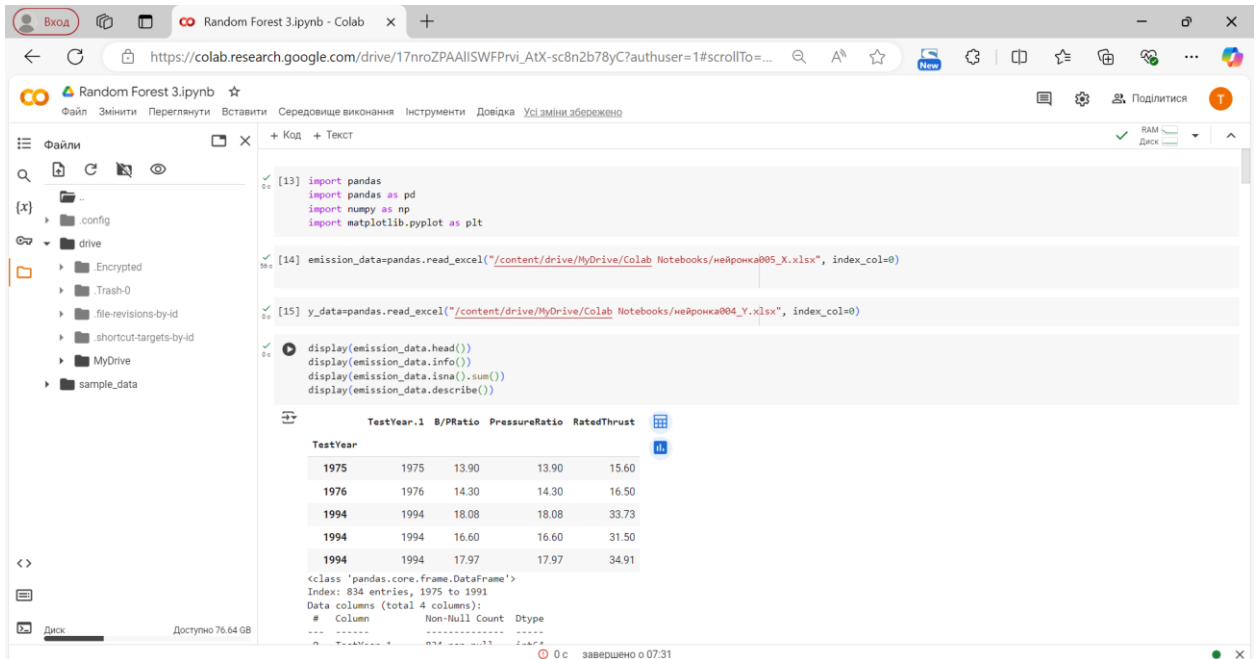


Рисунок 3.1 – Інтерфейс програмного середовища Google Colab

На рис. 3.2 зображений приклад візуального представлення результатів моделювання моделі RandomForest у середовищі Google Colab

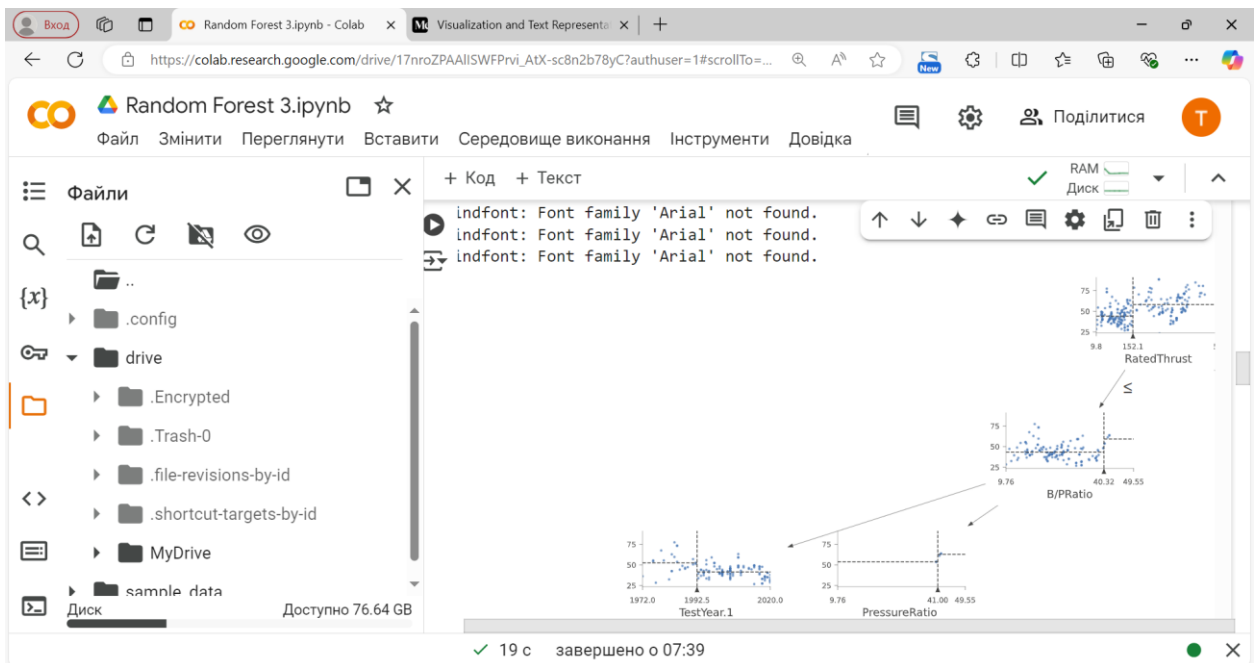


Рисунок 3.2 – приклад візуального представлення результатів моделювання за допомогою середовища Google Colab

На підставі вищевикладеного можна зробити висновок, що середовище Google Colab – це найзручніша середа для роботи з великими dataset мовою програмування Python.

3.2 Програмна реалізація

Для реалізації алгоритму прогнозування викидів шкідливих речовин від авіаційних газотурбінних двигунів потрібно імпортувати у середовище Google Colab необхідні бібліотеки Python. Розглянемо призначення обраних бібліотек [24]:

- бібліотека pandas (<http://pandas.pydata.org>) – бібліотека, яка базується на NumPy і містить високопродуктивні та зручні у використанні структури даних та інструменти обробки і аналізу даних;

- бібліотека NumPy (<http://www.numpy.org>, <http://scipy.org>) – вільна бібліотека Python для високопродуктивних операцій з багатовимірними масивами і матрицями;

- бібліотека scikit-learn (<http://scikit-learn.org>) містить зручні для використання алгоритми машинного навчання з учителем (регресія, класифікація) і без учителя (кластеризація, зменшення розмірності), а також засоби для підготовки даних і вибору найкращої моделі даних

- бібліотека seaborn (<https://seaborn.pydata.org/>) – бібліотека мови програмування Python, яка використовується для побудови статичних графіків.

Після імпортування необхідних бібліотек виконується етап завантаження початкових даних з Google Drive, зображений на рис. 3.3.

В dataset на першому етапі робіт було добавлено 10 змінних, які містять як параметри двигуна, так і характеристики палива, які безпосередньо впливають на утворення викидів шкідливих речовин, а також умови

зовнішнього середовища під час випробувань. Для зниження витрат часу та ресурсів під час розробки моделі, потрібно визначити які самі змінні потрібно використовувати. Також необхідно перевірити початкові дані на наявність невизначених значень, що може погіршити кінцевий прогноз. Програмний код перевірки на наявність невизначених значень показано на рис. 3.4.

The screenshot shows a Google Colab notebook with the following code cells:

```
[1] import pandas
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import StandardScaler
from sklearn.ensemble import RandomForestRegressor
from sklearn.model_selection import RandomizedSearchCV
from sklearn.metrics import recall_score
from sklearn.metrics import mean_squared_error, r2_score
from sklearn.model_selection import train_test_split
from sklearn.tree import export_graphviz
import seaborn as sns
```

```
[2] !pip install dtreeviz
from dtreeviz import model
```

Показати прихований результат

```
[3] emission_data=pandas.read_excel("/content/drive/MyDrive/Colab Notebooks/нейронка005_X.xlsx", index_col=0)
```

```
[4] y_data=pandas.read_excel("/content/drive/MyDrive/Colab Notebooks/нейронка004_Y.xlsx", index_col=0)
```

Рисунок 3.3 – Лістинг коду, який ілюструє підключення бібліотек Python та завантаження початкових даних з Google Drive

The screenshot shows a Google Colab notebook with the following code cells:

```
[8] y_data=pandas.read_excel("/content/drive/MyDrive/Colab Notebooks/нейронка004_Y.xlsx", index_col=0)
```

```
[9] display(emission_data.head())
display(emission_data.info())
display(emission_data.isna().sum())
display(emission_data.describe())
```

The output of the code shows the following information:

```
<class 'pandas.core.frame.DataFrame'>
Index: 834 entries, 1975 to 1991
Data columns (total 10 columns):
#   Column                Non-Null Count  Dtype
---  -
0   TestYear.1            834 non-null   int64
1   B/PRatio              834 non-null   float64
2   PressureRatio         834 non-null   float64
3   RatedThrust           834 non-null   float64
4   NOx Number Eng        834 non-null   int64
5   Fuel H/C Ratio Max    834 non-null   float64
6   Fuel Arom Min (%)     834 non-null   float64
7   Ambient Baro (kPa)    834 non-null   float64
8   Ambient Temp (K)     834 non-null   float64
9   Manufacturer           834 non-null   object
dtypes: float64(7), int64(2), object(1)
memory usage: 71.7+ KB
None
```

```
0
TestYear.1    0
B/PRatio     0
PressureRatio 0
```

Рисунок 3.4 – Лістинг коду, який ілюструє перевірку на наявність невизначених значень

Наступним кроком при виконанні при розробці методики прогнозування викидів шкідливих речовин від авіаційних газотурбінних двигунів програмного застосунку було виконано аналіз даних. Основною ціллю аналізу даних було виявлення основних зв'язків між змінними. Для цього потрібно було побудовано кореляційну матрицю. На рис. 3.5 та 3.6 представлені результати кореляційного аналізу у вигляді таблиці кореляційної матриці для всіх змінних та приклад візуалізації кореляційної матриці по всіх змінних у вигляді графіків.



Рисунок 3.5 – Лістинг коду, який ілюструє побудову кореляційної матриці

На рис. 3.6–3.7 можна побачити детальні графіки взаємодії цільової змінної та змінних, що характеризують конструкцію двигуна та характеристики палива. За допомогою отриманих графіків було виконано

аналіз цільової змінної з первинним набором параметрів, виявлено приховані взаємозв'язки між змінними. На рис. 3.8 наведено приклад обчислення та візуалізації параметру feature importance для досліджуваного dataset. По результатам аналізу було сформовано перелік параметрів, які будуть використані при формуванні методики прогнозування викидів шкідливих речовин від авіаційних газотурбінних двигунів.



Рисунок 3.6 – Графіки аналізу взаємодії змінних

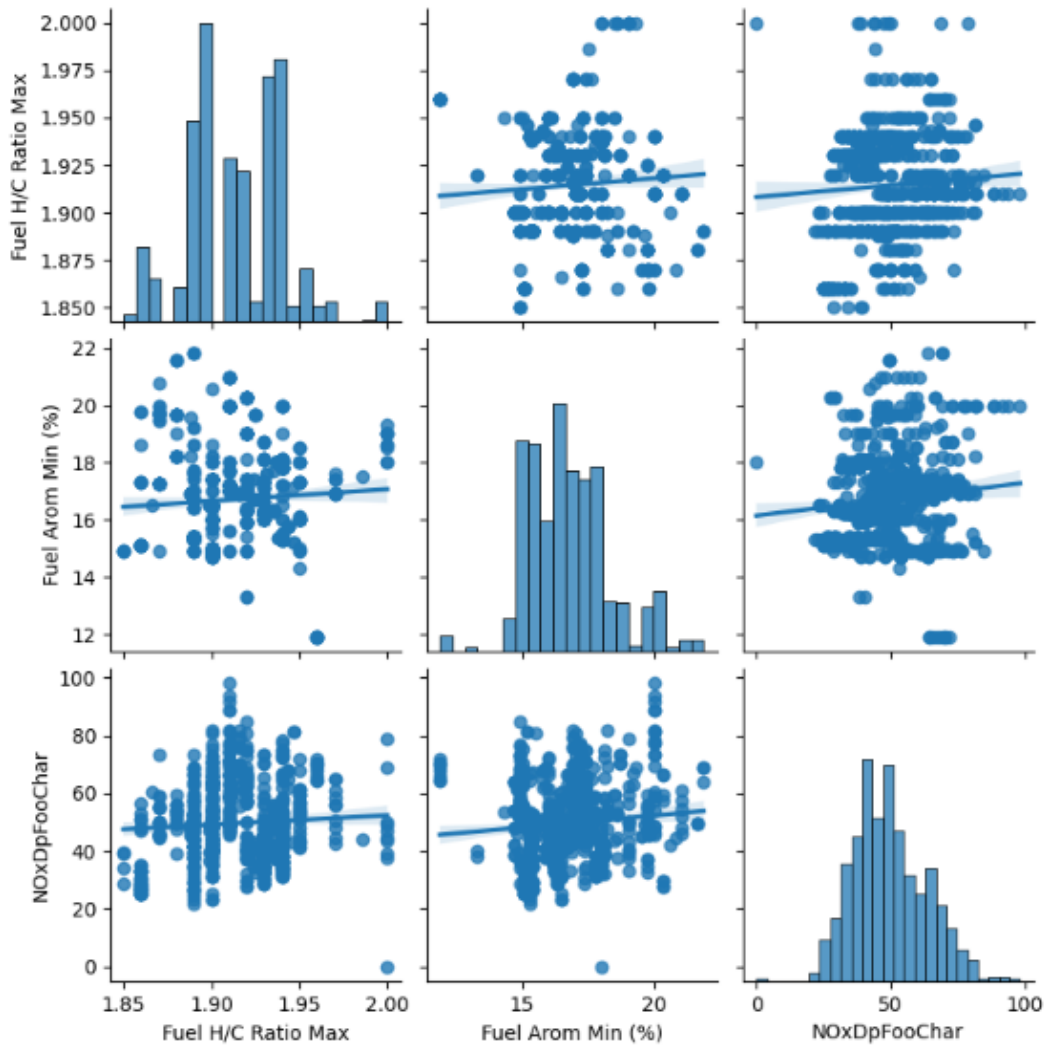


Рисунок 3.7 – Графіки аналізу взаємодії змінних

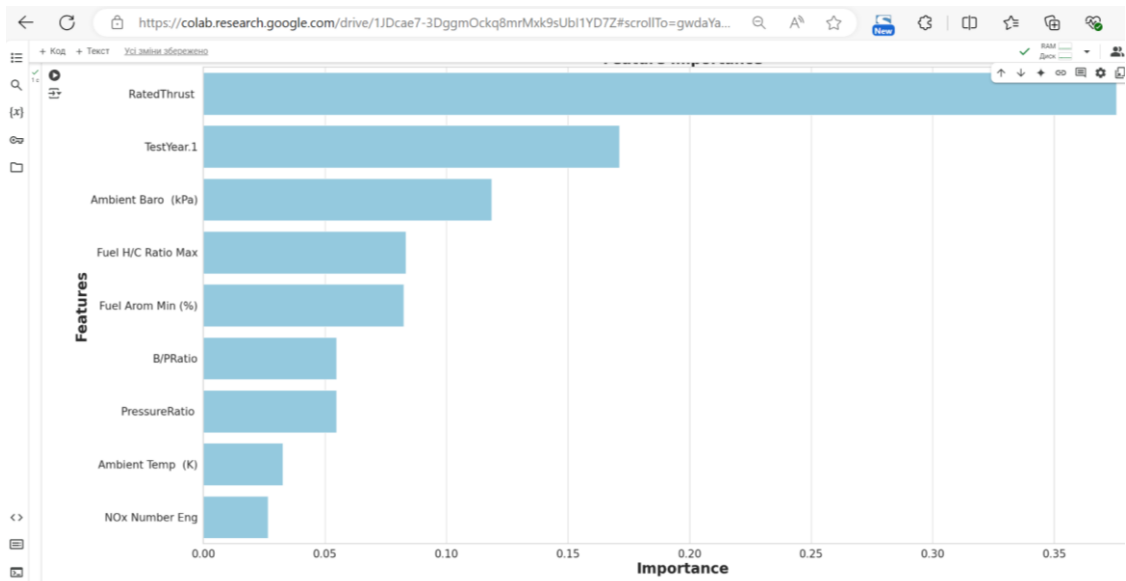


Рисунок 3.8 – Візуалізація параметру feature importance для досліджуваного dataset

На рис. 3.9 показана програмна реалізація розділення dataset згідно варіанту розподілу приведеному у розділі 2.2.2.

```
[69] #розподіляємо модель на модель для навчання та тестову частину
      from sklearn.model_selection import train_test_split
      X_train1, X_test1, y_train1, y_test1 = train_test_split(
          X, y, test_size=0.7, random_state=42)
      X_train = X_train1.values
      y_train = y_train1.values
      X_test = X_test1.values
      y_test = y_test1.values
```

Рисунок 3.9 – Програмна реалізація розділення dataset

3.3 Процес прийняття рішень у методі Random Forest для прогнозування викидів шкідливих речовин

Метод Random Forest формує набір дерев рішень, кожне дерево робить свій внесок у результат. Візуалізація окремих дерев може дати уявлення про процес прийняття рішень на детальному рівні. На рис. 3.10–3.14 показано ряд дерев, які було сформовано в залежності від значення параметра *max_depth*.

Аналіз рис. 3.10 демонструє приклад сильного недонавчання моделі при *max_depth* = 1. Дерево занадто просте, щоб вловити складні залежності даних. Цей варіант підходить для грубої класифікації чи регресії, але якість моделі на тестових даних буде низькою.

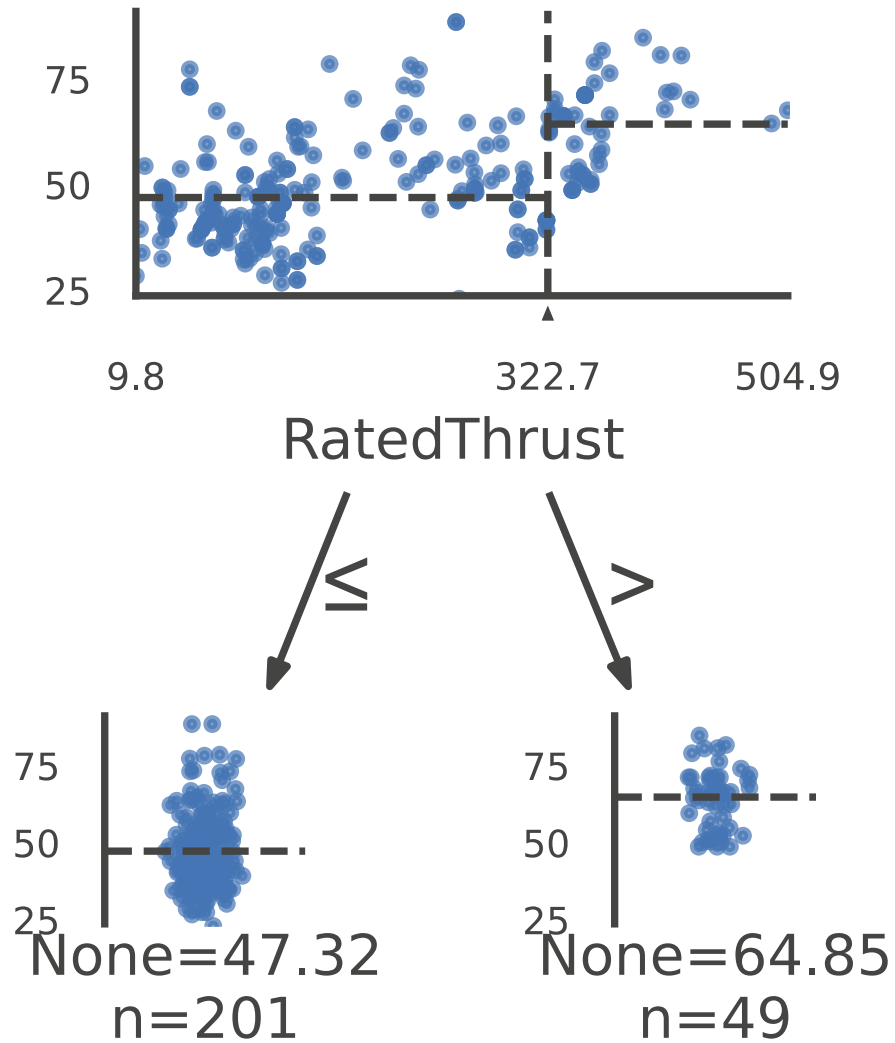
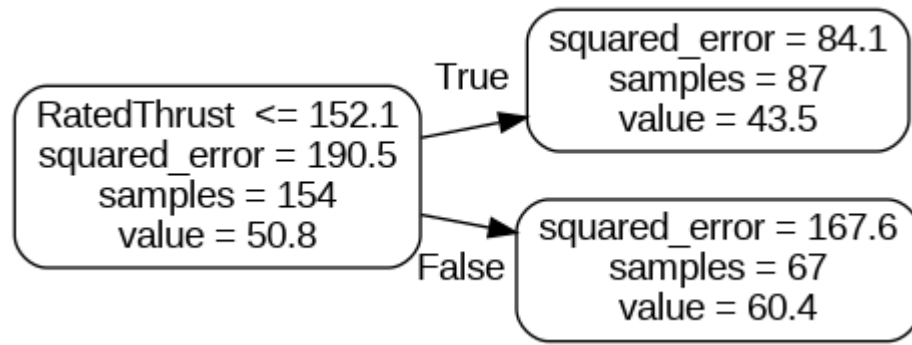


Рисунок 3.10 – візуалізація дерев методу Random Forest моделі прогнозування викидів шкідливих речовин при значенні $max_depth=1$

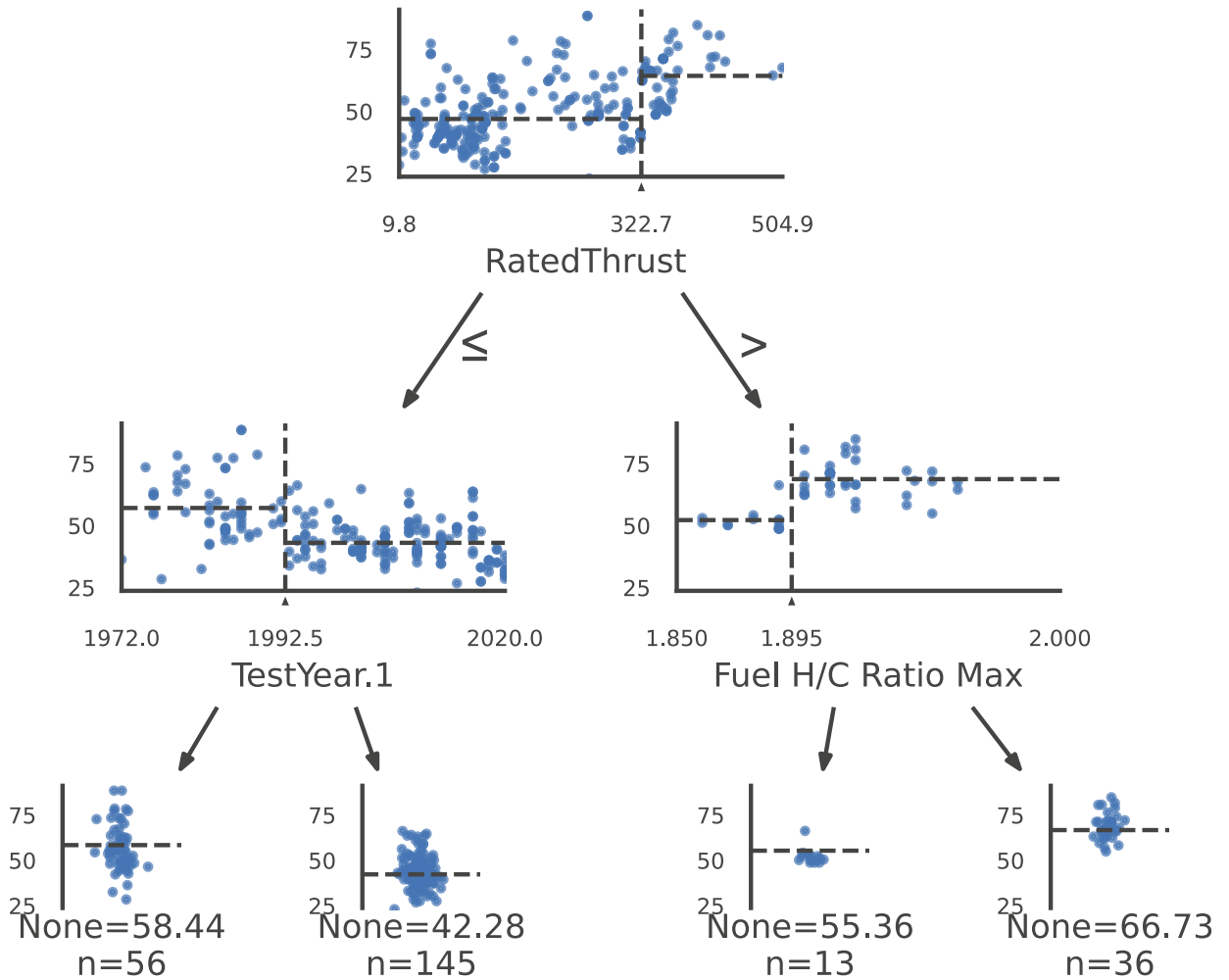
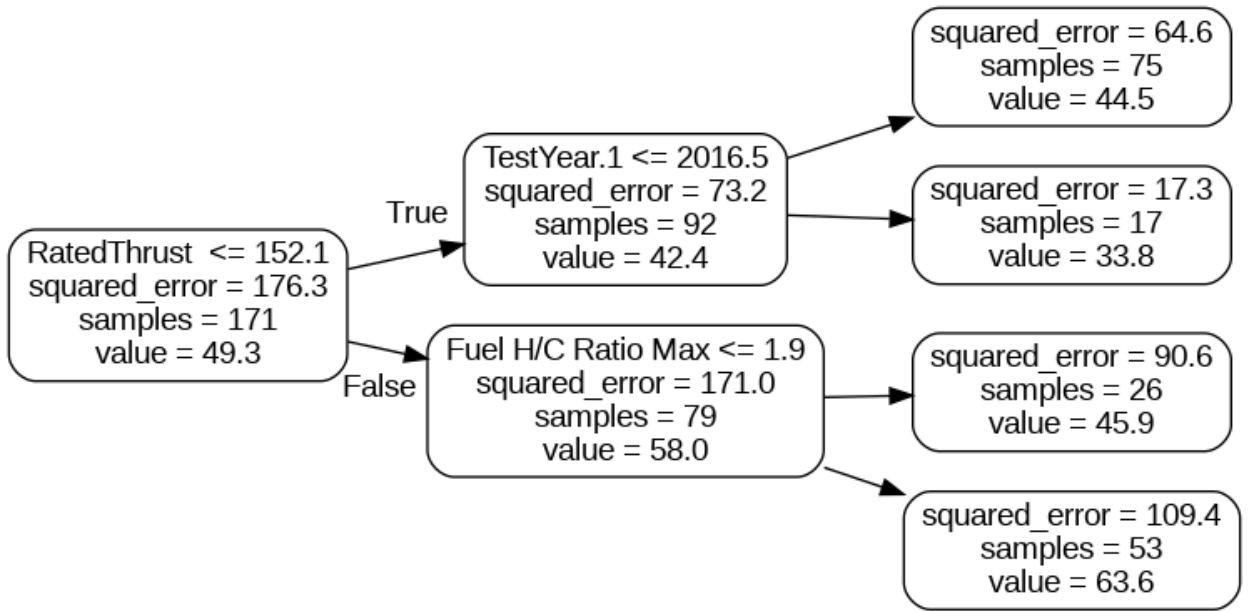


Рисунок 3.11 – візуалізація дерев методу Random Forest моделі прогнозування викидів шкідливих речовин при значенні $max_depth=2$

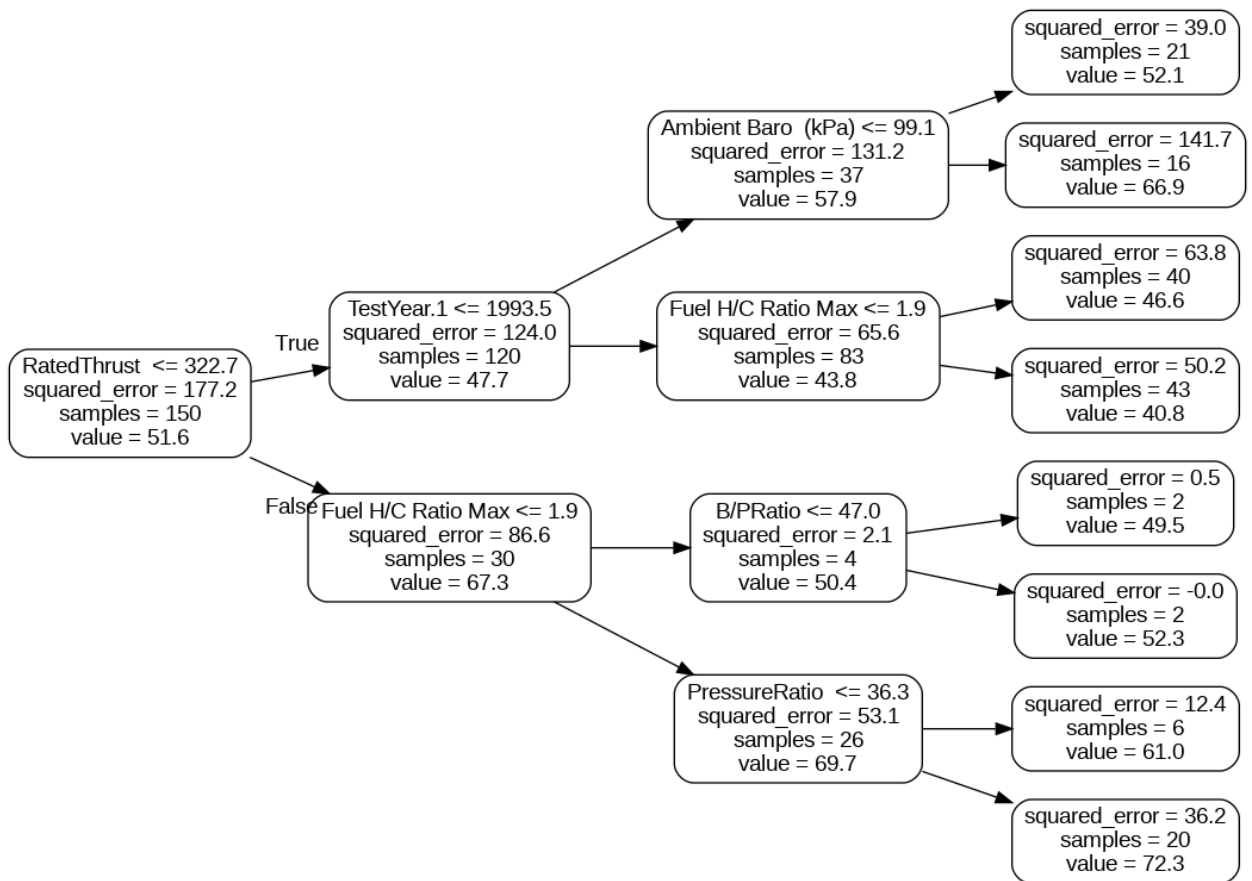


Рисунок 3.12 – візуалізація дерев методу Random Forest моделі прогнозування викидів шкідливих речовин при значенні $max_depth=3$

Дерево показане на рис. 3.11 має два рівні після кореневого вузла. Після першого поділу RatedThrust додається ще один критерій поділу (наприклад, TestYear.1 або інші ознаки). У кожному вузлі спостерігається зниження помилки (squared error), що говорить про точнішу сегментацію даних. Дерево стає більш здатним до навчання, але ще недостатньо складним для подання складних залежностей. Модель здатна краще захоплювати структуру даних у порівнянні з $max_depth = 1$, але все ще може страждати від недонавчання.

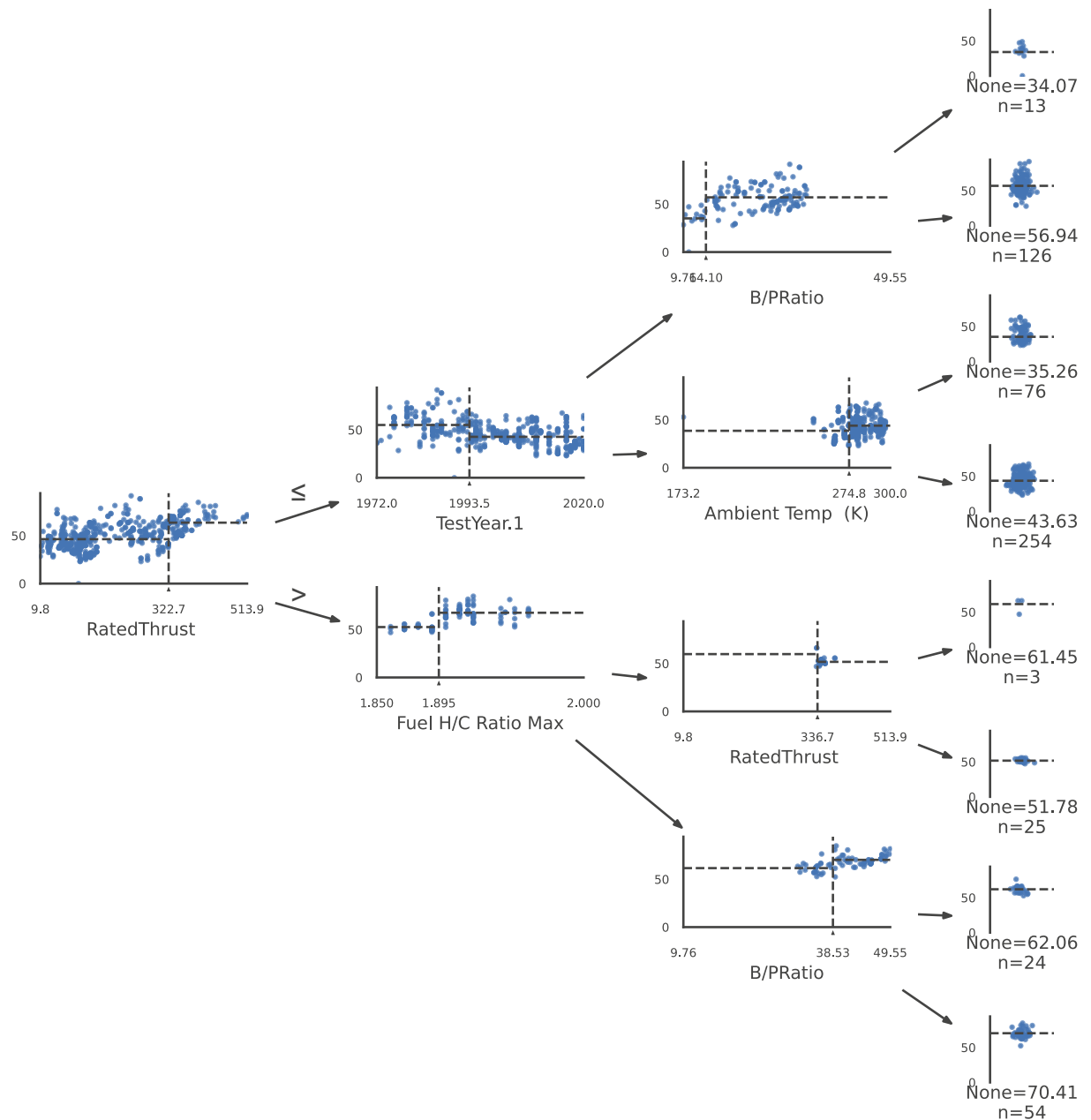


Рисунок 3.13 – візуалізація дерев методу Random Forest моделі прогнозування викидів шкідливих речовин при значенні $max_depth=3$

Дерево показане на рис. 3.12–3.13 має три рівні після кореневого вузла. З'являються додаткові критерії поділу (наприклад, PressureRatio, B/PRatio, TestYear.1), що робить сегментацію більш детальною. Помилки (squared error) у листових вузлах стають мінімальними, оскільки кожна група даних тепер однорідніша. Дерево виглядає добре збалансованим. Це показує, що збільшення глибини дозволяє краще адаптуватись до даних. Подібна глибина

часто виявляється оптимальною для моделей Random Forest, тому що дозволяє захоплювати складні закономірності без сильного перенавчання.

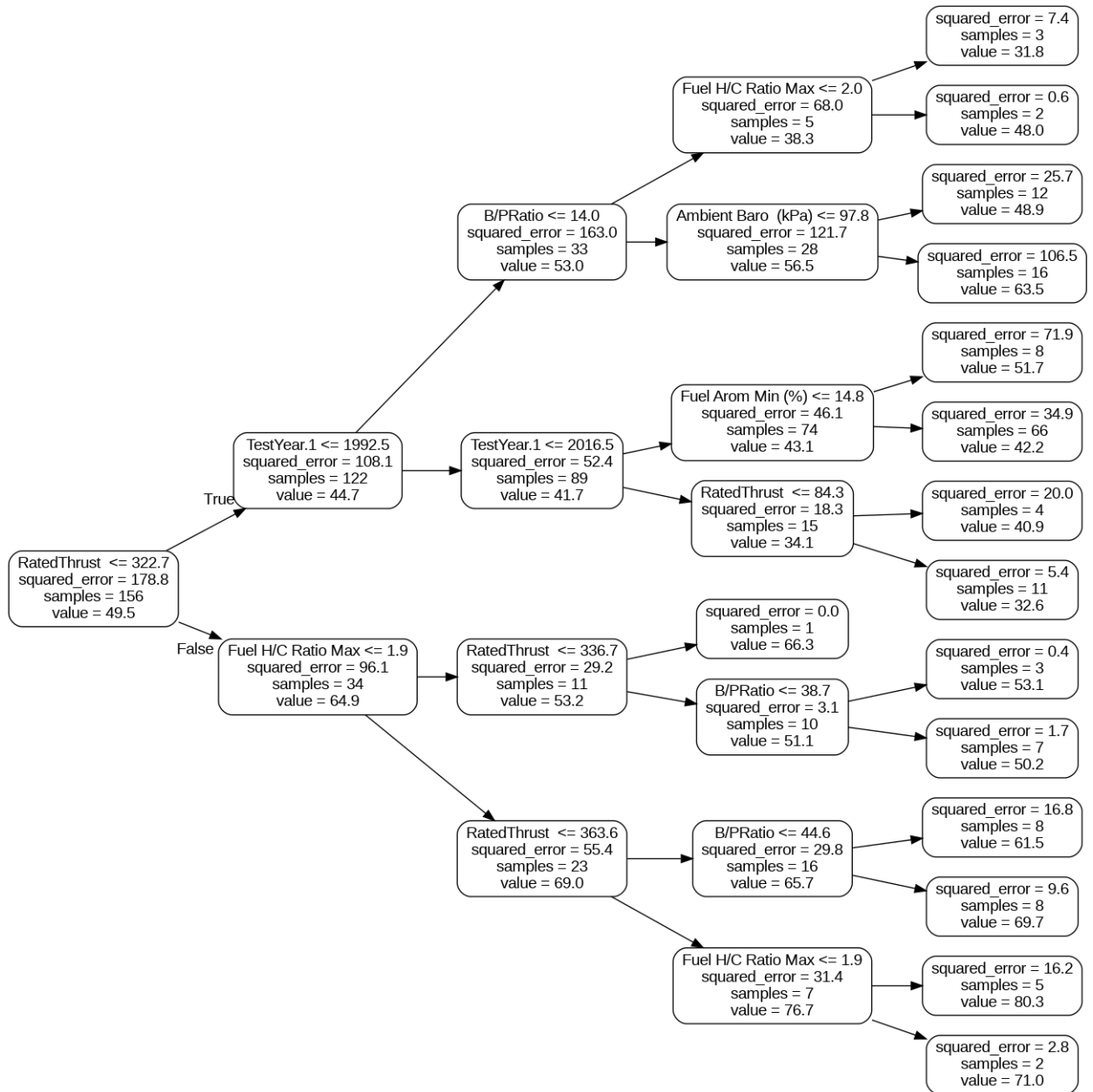
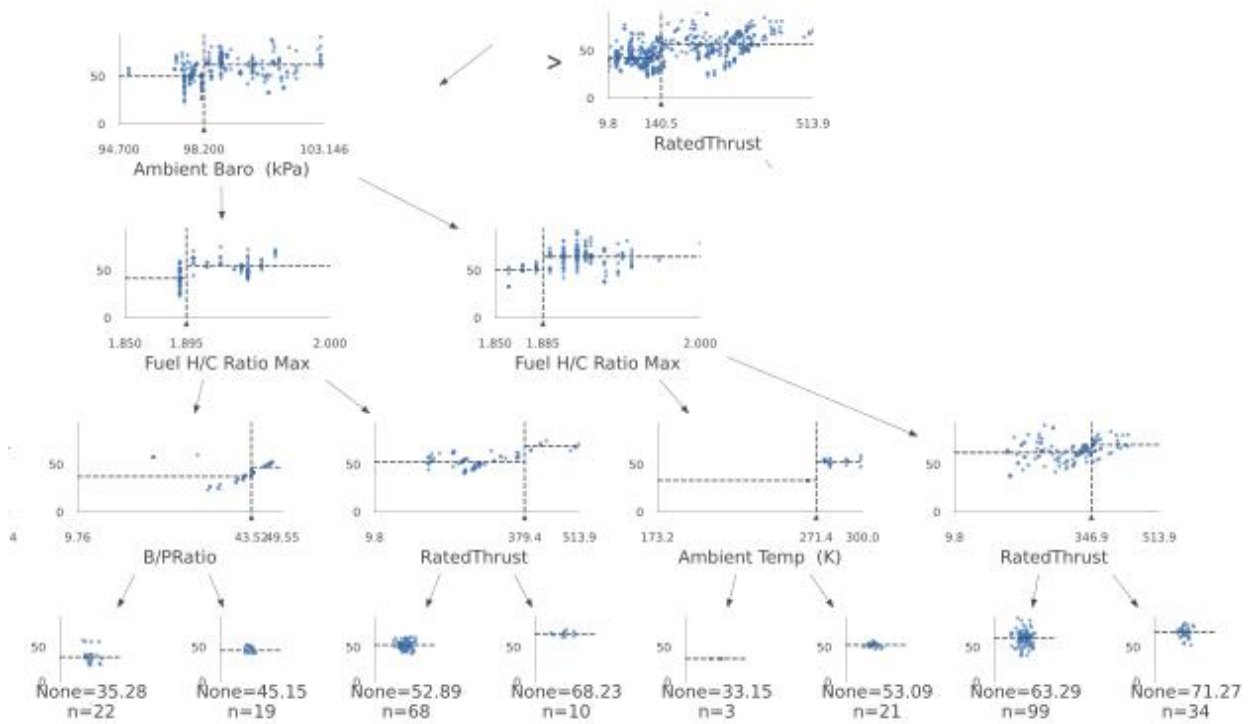
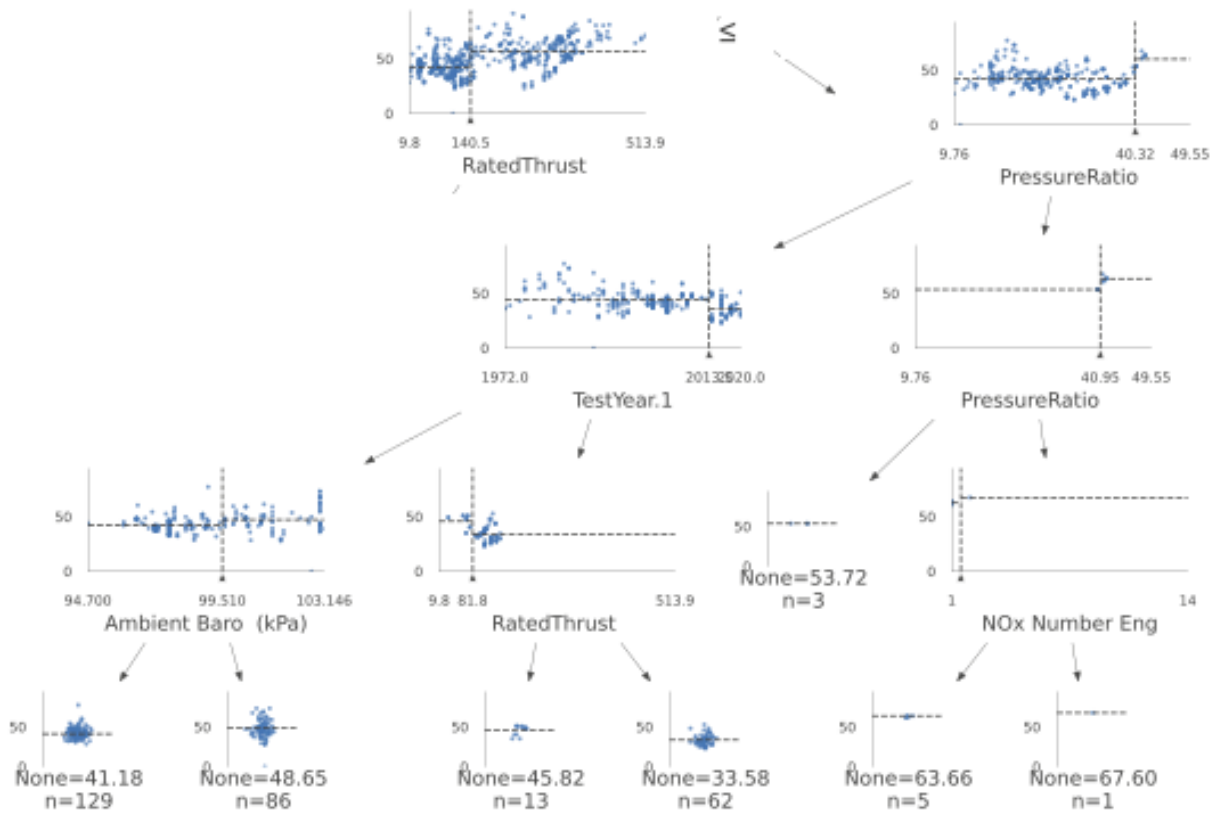


Рисунок 3.14 – візуалізація дерев методу Random Forest моделі прогнозування викидів шкідливих речовин при значенні $max_depth=4$



a) $RatedThrust > 140.5$



б) $RatedThrust \leq 140.5$

Рисунок 3.15 – візуалізація дерев методу Random Forest моделі прогнозування викидів шкідливих речовин при значенні $max_depth=4$

Дерева, що показані на рис. 3.14–3.15 мають чотири рівні. Після кореневого вузла дерево глибше, ніж друге. Воно має безліч вузлів та листя, що дозволяє більш точно розділяти дані. Використовуються параметри: *RatedThrust*, *PressureRatio*, *B/PRatio*, *TestYear.1*. Модель робить більш точні прогнози, оскільки кожне розбиття зменшує середньоквадратичну помилку (*squared_error*). Багато листових вузлів (кінцеві вузли дерева), що вказує на високий ступінь деталізації та увагу до даних. Є ризик перенавчання (*overfitting*), тому що дерево може надто сильно підлаштовуватися під дані.

У моделі випадкового лісу (*Random Forest*) кожне дерево є ієрархічною структурою, де листя — це кінцеві вузли, в яких приймається рішення про прогноз. Для оцінки складності моделі та перевірки її стійкості була використана функція *leaf_sizes()*, яка повертає список, що містить кількість листя у кожному дереві лісу. Даний аналіз дозволяє дізнатися, наскільки "розгалужені" дерева у прогностичній моделі. Залежність кількості листя у розробленій моделі випадкового лісу від параметру *max_depth* показана на рис. 3.16 та у таблиці 3.1.

Таблиця 3.1 – Порівняльний аналіз впливу параметру *max_depth*

| Характеристика | Дерево с <i>max_depth</i> =[1;3] | Дерево с <i>max_depth</i> ≥4 |
|----------------------|-------------------------------------|------------------------------|
| Точність передбачень | Низька | Висока |
| Ризик перенавчання | Низький | Високий |
| Інтерпретованість | Проста | Складна |
| Обробка шумів | Більш стійка | Менш стійка |

Наявність великої кількості унікальних ідентифікаторів листя свідчить, що модель досить добре розділяє дані, не схильна до перенавчання і може виділити безліч різних класів чи підкласів. Як видно, збільшення максимальної глибини дерева призводить до збільшення середньої кількості листя та більшої варіативності цього показника. Це пов'язано з тим, що

глибокі дерева мають більше можливостей для розгалуження. Однак надмірне збільшення глибини може призвести до перенавчання моделі. Для розглянутої моделі оптимальне значення max_depth знаходиться в діапазоні $[5, 10]$, так як при цьому досягається баланс між складністю моделі та її узагальнюючою здатністю.

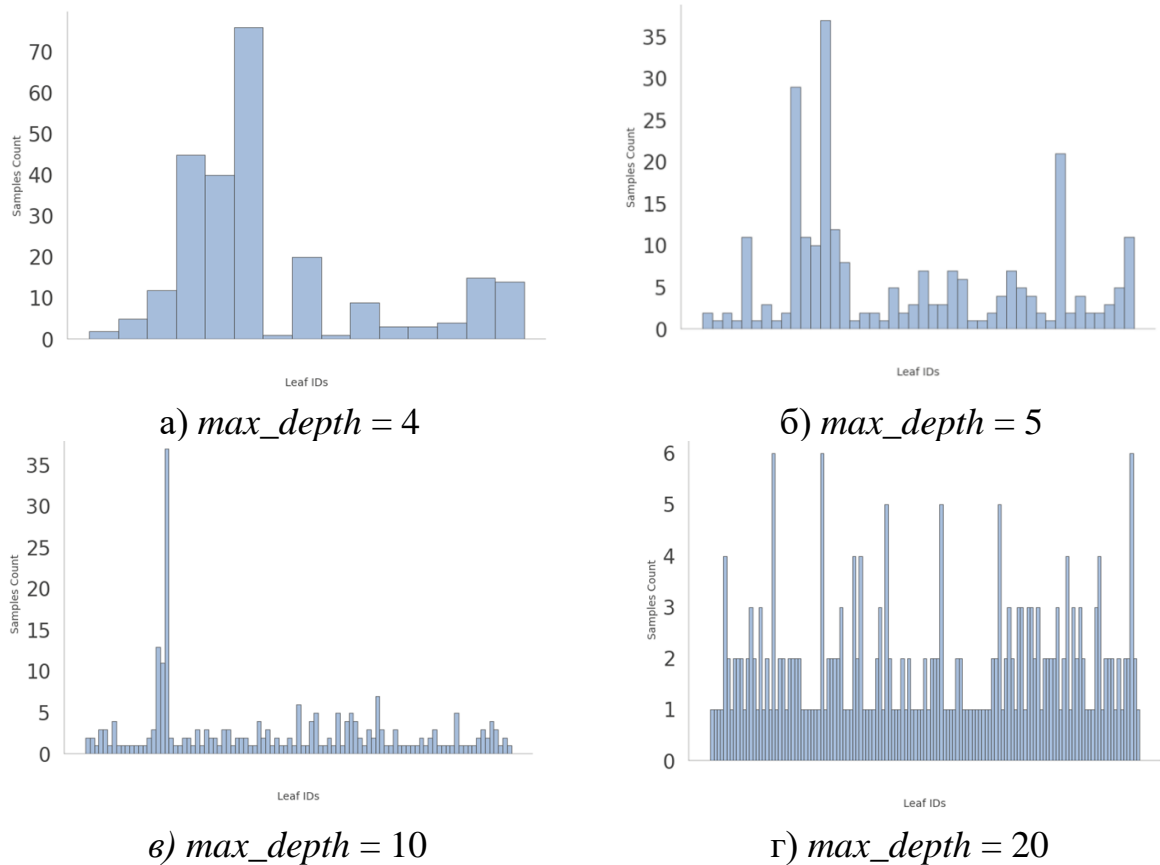


Рисунок 3.16 – Залежність кількості листя у розробленій моделі випадкового лісу від max_depth

3.4 Результати роботи програми

Результат роботи програми для тестового dataset у вигляді графіку на якому зображені лінія ідеального прогнозування, рівень викидів, який був цільовою змінною, та прогнозований рівень викидів зображено на рис. 3.17–3.18.

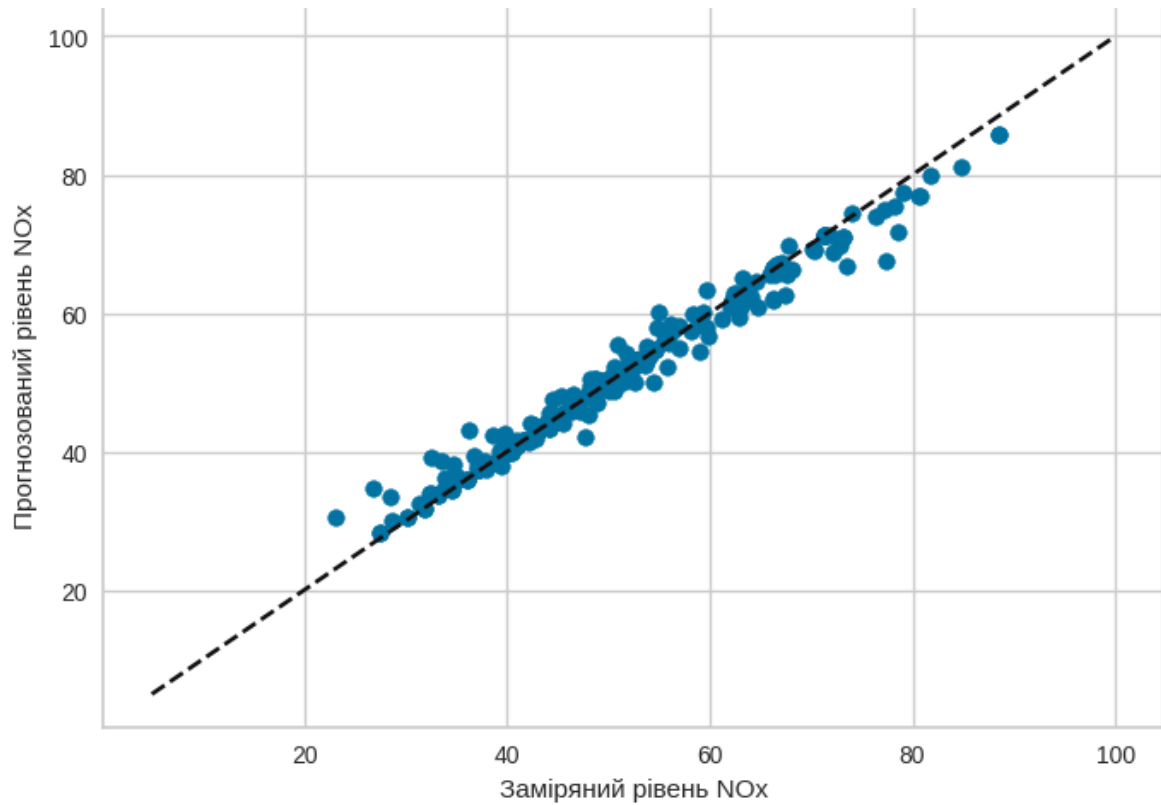
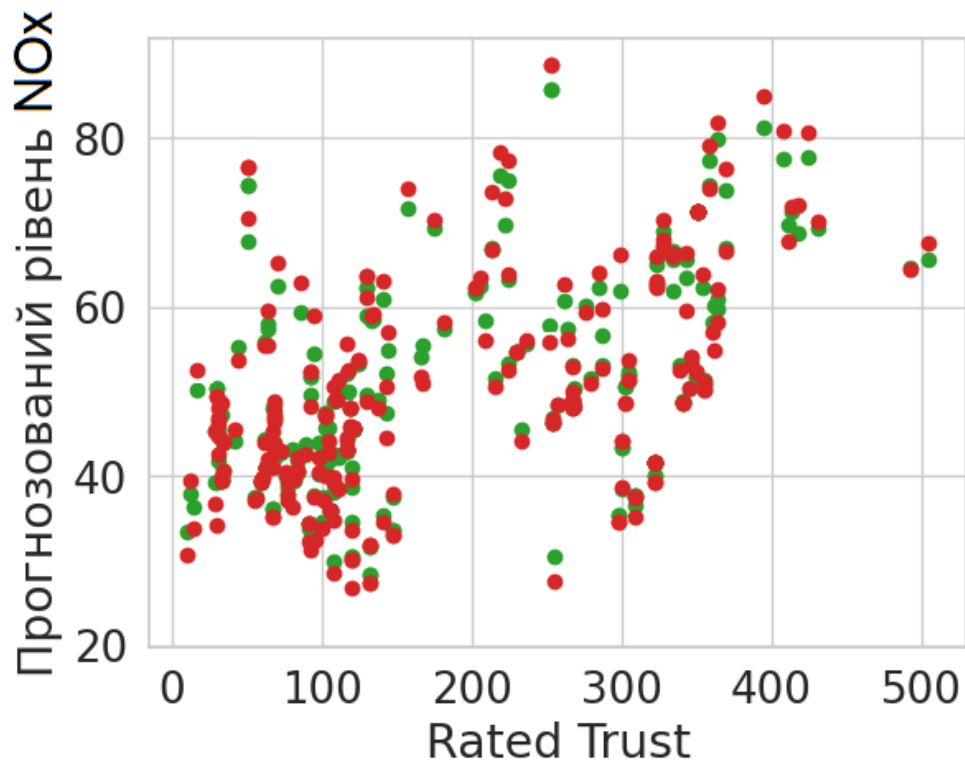


Рисунок 3.17 – Результат роботи програми для тестового dataset



- – в рівень викидів
- – прогнозований рівень викидів

Рисунок 3.17 – Залежність прогнозованого та виміряного рівня викидів від параметра RatedTrust для тестового dataset

Результати регресійного аналізу показали високий рівень відповідності розробленої моделі прогнозу викидів речовин від авіаційних газотурбінних двигунів емпіричним даним. Коефіцієнт детермінації становив 0.953, що означає, що модель пояснює 95.3% варіації результатів сертифікаційних випробувань ІКАО. Це дуже високий показник, що свідчить про те, що модель добре підходить для прогнозування рівня викидів від авіаційних газотурбінних двигунів. Розроблена модель демонструє високу здатність передбачати значення залежної змінної на основі незалежних змінних. Модель добре пояснює взаємозв'язки між змінними, що дозволяє глибше зрозуміти досліджуваний процес.

ВИСНОВКИ

У рамках кваліфікаційної роботи була розроблена модель прогнозу викидів шкідливих речовин від авіаційних газотурбінних двигунів.

У ході роботи були вирішені наступні завдання:

- проаналізовано методи класифікації, які виконують завдання прогнозування даних;

- проаналізовано метод класифікації даних Random Forest ;

- проаналізована математична модель методу класифікації Random Forest

- зроблено вибір вхідних та вихідних даних із чіткою залежністю між ними що дозволило отримати достовірні результати

- розроблена модель для прогнозу викидів речовин від авіаційних газотурбінних двигунів на базі результатів сертифікаційних випробувань ІКАО;

- виконана комп'ютерна модель алгоритму прогнозування викидів шкідливих речовин від авіаційних газотурбінних двигунів з використанням платформи Google Colab та мови програмування Python;

- зроблено висновки, щодо виконаної роботи.

Коефіцієнт детермінації розробленої моделі становив 0.953, що свідчить про те, що модель добре підходить для прогнозування рівня викидів від авіаційних газотурбінних двигунів. Розроблена модель демонструє високу здатність передбачати значення залежної змінної на основі незалежних змінних.

ПЕРЕЛІК ПОСИЛАНЬ

1. Flightpath 2050. Europe's Vision for Aviation Report of the High Level Group. Policy. EU Directorate-General for Research and Innovation, Directorate General for Mobility and Transport on Aviation Research [Electronic resource]. — Mode of access: https://www.arcs.aero/sites/default/files/downloads/Bericht_Flightpath_2050.pdf (date of access: 04.10.2024). — Title from screen.

2. ICAO Resolution A39-1. Consolidated statement of continuing ICAO policies and practices related to environmental protection – General provisions, noise and local air quality. Resolutions Adopted By The Assembly 39th SESSION, Montréal, 27 September—6 October 2016, provisional edition, October 2016. [Electronic resource]. — Mode of access: https://www.icao.int/environmental-protection/Documents/Resolution_A39_1.PDF (date of access: 04.10.2024). — Title from screen.

3. Mongia H.C. Future Trends in Commercial Aviation Engines' Combustion. / H.C. Mongia // Novel Combustion Concepts for Sustainable Energy Development /ed. by Avinash K Agarwal, Ashok Pandey, Ashwani K. Gupta, Suresh K. Aggarwal, Abhijit Kushari. – Springer, New Delhi, 2014. – P. 113-176.

4. Dmytriiev S. Analysis and forecast on changes in the level of harmful substance emissions from aviation gas turbine engines [Electronic resource] / S. Dmytriiev, V. Loginov, V. Gusev. // 8th European Conference for Aeronautics and Space Sciences. – 2019. – Mode of access: <https://www.eucass.eu/doi/EUCASS2019-1056.pdf> (date of access 04.10.2024). — Title from screen.

5. Emissions Databank (07/2024) [Electronic resource]. – Mode of access: <https://www.easa.europa.eu/en/domains/environment/icao-aircraft-engine-emissions-databank> (date of access 04.10.2024). — Title from screen.

6. Annex 16. Environmental Protection. Volume II. Aircraft Engine Emissions. 5th Edition. — Montréal : International Civil Aviation Organization, 2023. — 150 p.

7. ICAO/CAEP Report of the Independent Experts on the NO_x review and medium and longterm 14 technology goals for NO_x. (Doc 9887) / International Civil Aviation Organization. — Montreal, 2008. —120 p.

8. National Science and Technology Council: National aeronautics research and development plan // National science and technology council. —Washington, D.C., 2010. —56 p.

9. ASCENT Project 037 CLEEN II Technology Modeling and Assessment. Continuous Lower Energy, Emissions, and Noise (CLEEN) Program: program // Institute of Technology. — Georgia Institute of Technology, 2015. — 172 p.

10. ДСТУ EN 16603-11:2022 Космічна інженерія. Визначення рівнів технологічної готовності (TRL) та критерії оцінювання (EN 16603-11:2019, IDT; ISO 16290:2013, MOD). — Київ : УкрНДНЦ, 2022. - 22 с.

11. Vamsi K. A preliminary design of a schematic evaluation of low emission turbofan combustors [Electronic resource] / K. Vamsi, N. Mitrokhov // Journal of Physics: Conference Series. — 2021. — V. 1891. — 012026. — Mode of access: <https://iopscience.iop.org/article/10.1088/1742-6596/1891/1/012026/pdf> (date of access 14.10.2024). — Title from screen.

12. Що таке Machine Learning? [Електронний ресурс] // — Режим доступу: <https://denovo.ua/resources/what-is-machine-learning> (дата звернення: 10.12.2024). — Назва з екрана.

13. Zhou L. Machine learning for combustion [Electronic resource] / L. Zhou, Y. Song, W. Ji, H. Wei // Energy and AI — 2022 — Vol. 7. — Mode of access: <https://www.sciencedirect.com/science/article/pii/S2666546821000756> (date of access 14.10.2024). — Title from screen.

14. Yadav V. Artificial neural network an innovative approach in air pollutant prediction for environmental applications: A review [Electronic resource] / V. Yadav, A. Kumar, V. Singh, T. Singh // Results in Engineering. — 2024. —

Volume 22. — Mode of access: <https://www.sciencedirect.com/science/article/pii/S2590123024005607> (date of access 14.10.2024). — Title from screen.

15. Харченко В. О. Основи машинного навчання : навч. посіб. / В. О. Харченко. — Суми: Сумський державний університет, 2023. — 264 с.

16. Breiman, L. Random forests / Leo Breiman // Machine Learning. — 2001. — 45(1). — P. 5–32.

17. Kaya H. Predicting CO and NO_x emissions from gas turbines: Novel data and a benchmark PEMS / H. Kaya, P. Tüfekci, E. Uzun // Turkish journal of electrical engineering and computer sciences. — 2019. — № 27(6). — P. 4783 – 4796.

18. Лобода Ю.Г. Процес побудови моделі інтелектуального аналізу даних та оцінка продуктивності [Електронний ресурс] / Ю.Г. Лобода, Р.Є. Ключник // V International Scientific and Theoretical Conference «The process and dynamics of the scientific path» February 9, 2024; Athens, Greece – 2024 – Greece. Режим доступу: <https://previous.scientia.report/index.php/archive/article/view/1598/1611> (дата звернення: 10.12.2024). — Назва з екрана.

19. Data Mining Concepts | Microsoft Learn. [Електронний ресурс] – Mode of access: <https://learn.microsoft.com/en-us/analysis-services/data-mining/data-mining-concepts?view=asallproducts-allversions> (date of access 20.10.2024). — Title from screen.

20. Інтелектуальний аналіз даних: конспект лекцій для здобувачів ступеня вищої освіти бакалавра спеціальності «Економіка» освітньо-професійної програми «Економічна кібернетика» / Укладачі: Іванов С.М., Максишко Н.К., Бречко Д.О. — Запоріжжя: ЗНУ, 2020. — 156 с.

21. Kotronoulas G. An Overview of the Fundamentals of Data Management, Analysis, and Interpretation in Quantitative Research. / Grigorios Kotronoulas, Susana Miguel, Maura Dowling, Paz Fernández-Ortega // Seminars in Oncology Nursing Glasgow –Volume 39, Issue 2 –2023 Mode of access: <https://www.sciencedirect.com/science/article/pii/S0749208123000293> (date of access 20.10.2024). — Title from screen.

22. Random Forest Classifier Tutorial. [Електронний ресурс] – Mode of access: <https://www.kaggle.com/code/prashant111/random-forest-classifier-tutorial> (date of access 20.10.2024). — Title from screen.

23. RandomForestRegressor — scikit-learn 1.6.1 documentation. Mode of access: <https://scikit-learn.org/1.6/modules/generated/sklearn.ensemble.RandomForestRegressor.html> (date of access 20.10.2024). — Title from screen.

24. Копей В. Б. Мова програмування Python для інженерів і науковців : навчальний посібник / В. Б. Копей. – Івано-Франківськ : ІФНТУНГ, 2019. – 272 с.