

Лебьодкін Є. О., аспірант;

Варламов Є. М., канд. техн. наук, ст. наук. співр.;

Палагута О. А., канд. техн. наук;

Науково-дослідна установа, Український науково-дослідний інститут екологічних проблем, м. Харків, Україна

Лебьодкін О. І., керівник відділу автоматизованих систем моніторингу

Компанія Хімлаборреактив, м. Бровари, Київська область

МОДЕЛЮВАННЯ КОНЦЕНТРАЦІЙ PM_{10} У ЗИМОВИЙ ПЕРІОД У м. КРИВИЙ РІГ ЗА ДОПОМОГОЮ МОДЕЛІ RANDOM FOREST

Анотація. У цьому дослідженні розглянуто проблему прогнозування концентрацій дрібнодисперсних частинок PM_{10} у великому промисловому центрі України – місті Кривий Ріг – протягом зимового періоду 2024–2025 рр. З використанням даних спостережень за концентрацією PM_{10} та супутніми метеорологічними параметрами (температура повітря, відносна вологість, атмосферний тиск, швидкість і напрям вітру) побудовано модель регресії Random Forest для короткострокового прогнозування рівнів забруднення. Додатково враховано часові чинники (час доби, день тижня) та умовну змінну, що відображає інтенсивність викидів протягом доби. Перед побудовою моделі дані були нормалізовані та очищені від крайніх викидів.

Розроблена модель пояснює приблизно 72 % варіації концентрації PM_{10} на тестовій вибірці та демонструє значне зменшення середньої абсолютної похибки ($MAE \approx 8,5$ мкг/м³) порівняно з лінійною регресією. Графічний аналіз показує, що PM_{10} має виражену добову циклічність – найнижчі рівні в денні години, а найвищі – у вечірній та нічний час. Кореляційний аналіз підтверджує негативний зв'язок між PM_{10} та температурою, швидкістю вітру та позитивний із відносною вологістю. Визначено, що найбільший внесок у пояснення варіації вносять атмосферний тиск, температура та вологість, тоді як часові змінні мають менший вплив.

Ключові слова PM_{10} ; атмосферне повітря; прогнозування; метеорологічні чинники; машинне навчання; Random Forest; Кривий Ріг.

Вступ

Охорона атмосферного повітря є одним із ключових напрямів сучасної екологічної політики. Дрібнодисперсні частинки PM_{10} (аеродинамічний діаметр до 10

мкм) проникають у дихальні шляхи людини та асоціюються з підвищеним ризиком хронічних хвороб серцево-судинної та дихальної систем, а також передчасною смертністю [1]. Європейські та світові стандарти якості повітря будуються саме на граничних концентраціях PM_{10} та $PM_{2.5}$ [2]. З прийняттям нової Директиви (ЄС) 2024/2881 про якість атмосферного повітря та чистіше повітря для Європи, що набрала чинності 10 грудня 2024 р., Європейський Союз посилює вимоги до моніторингу та прогнозування забруднення. І хоча Україна тільки інтегрує ці вимоги у своє законодавство, актуальність оперативного прогнозу якості повітря в промислових містах є очевидною.

Кривий Ріг – один із найбільших промислових центрів України, де зосереджені металургійні комбінати, видобувні підприємства та значна транспортна інфраструктура. Викиди промисловості та транспорту разом із несприятливими метеорологічними умовами створюють умови для підвищених концентрацій твердих частинок, особливо в зимовий період. Традиційні методи прогнозування, такі як множинна лінійна регресія, часто виявляються недостатніми для відтворення складних нелінійних зв'язків між погодними факторами та забрудненням. Сучасні підходи машинного навчання, зокрема ансамблеві методи, мають потенціал покращити точність прогнозу, оскільки дозволяють виявляти приховані закономірності, недосяжні для лінійних моделей [3]. Random Forest – один із найпопулярніших ансамблевих методів машинного навчання. Він поєднує велику кількість дерев рішень, кожне з яких навчається на випадковій підмножині даних та ознак. Підсумковий прогноз отримують шляхом усереднення результатів окремих алгоритмів розгалуження даних (для регресії) або голосування (для класифікації) [3]. Такий підхід суттєво зменшує перенавчання, підвищує стійкість до шумів та забезпечує оцінку важливості кожної ознаки у моделі. Окремі дослідження показують, що моделі Random Forest здатні пояснити понад половину варіацій PM_{10} , використовуючи сукупність метеорологічних та часових факторів.

Мета даної роботи – побудувати та оцінити модель Random Forest для прогнозування концентрацій PM_{10} у місті Кривий Ріг у зимовий період. Робота базується на реальних даних сертифікованої в Україні станції спостережень «Cairnet» компанії ENVEA, що охоплюють період з грудня 2024 р. включно по березень 2025 р. Отримані результати порівнюються з висновками попередньої роботи, в якій застосовувалася множинна лінійна регресія. Докладний аналіз впливу

метеорологічних умов і часових змінних дозволяє запропонувати рекомендації щодо управління якістю повітря на місцевому рівні.

Матеріали та методи

Для аналізу використано набір даних, отриманий зі станції моніторингу якості повітря Cairnet, що встановлена у центральній частині м. Кривий Ріг. Станція сертифікована для контролю PM_{10} відповідно до державних стандартів, а також укомплектована метеорологічними датчиками. Вхідний файл містить 9220 вимірювань за період з 5 грудня 2024 р. по 31 березня 2025 р. Кожен запис характеризується такими показниками:

- **PM_{10} , $мкг/м^3$** – концентрація дрібнодисперсних частинок;
- **T, °C** – температура повітря на момент вимірювання;
- **RH, %** – відносна вологість;
- **P, мм рт. ст.** – барометричний (атмосферний) тиск;
- **SW, градуси** – швидкість вітру;
- **DW, м/с** – напрям вітру;
- **Data_Time** – мітка часу (дата та точний час спостереження).

Перед початком моделювання у таблицю було додано низку часових змінних, що відображають циклічну добову та тижневу динаміку:

- **Час доби** – категоріальна змінна з чотирма значеннями: ніч (0:00–06:00), ранок (06:00–12:00), день (12:00–18:00) та вечір (18:00–24:00). Такий поділ дозволяє врахувати зміну стабільності атмосфери, наявність інверсій та зміну активності джерел викидів;

- **Тип дня** – бінарна змінна (0 – робочий день, 1 – вихідний), що відображає зменшення транспортної та виробничої активності у вихідні;

- **Інтенсивність викидів** – числова змінна, що відповідає умовному рівню антропогенного навантаження протягом доби. Для ночі прийнято значення 1,00; ранку – 1,75; дня – 1,55; вечора – 2,15. Вибір коефіцієнтів ґрунтується на типових добових циклах активності у великих містах: максимум викидів припадає на вечірні години, мінімум – уночі.

Для виключення екстремальних сплесків, що пов'язані з аварійними ситуаціями або сильними пиловими бурями, із вибірки було вилучено 5 % найвищих значень PM_{10} (понад 95-й процентиль). Це дозволило зосередитися на нормальних умовах забруднення й уникнути сильного впливу викидів на навчання моделі.

Підготовка даних

- Обробка дат і часових ознак. Стовець Data_Time перетворено на тип datetime. З нього виділено годину та день тижня, на основі яких сформовано змінні «час доби» та «тип дня». Категоріальні значення закодовано методом бінарне кодування: Для кожного періоду доби сформовано окремі бінарні змінні (tod_night, tod_morning, tod_day, tod_evening), які набувають значення 1 у випадку належності спостереження до відповідної категорії та 0 – в іншому випадку..

- Нормалізація та заповнення пропусків. У числових ознак (T, RH, P, SW, DW, type_day, emission_intensity) траплялися пропуски, які було заповнено медіанним значенням відповідної ознаки на навчальній вибірці (метод Median Imputation). Після цього всі числові ознаки масштабовано до нульового середнього та одиничного стандартного відхилення за допомогою стандартного нормування (StandardScaler).

- Розділення на навчальну та тестову вибірки. Дані поділено у співвідношенні 70:30. Для забезпечення об'єктивної оцінки моделі дані змішано випадковим чином із фіксованим параметром random_state.

- Побудова моделі. Для прогнозування обрано Random Forest Regressor з 200 деревами рішень. Така кількість дерев забезпечує баланс між точністю прогнозу та швидкістю обчислень. Оптимізація гіперпараметрів (глибина дерев, мінімальний розмір листа) виконувалась експериментально з використанням крос-перевірки.

- Оцінка якості. Якість моделі оцінювали за метриками середньої абсолютної похибки (MAE), середньоквадратичної похибки (RMSE) та коефіцієнта детермінації R^2 на навчальній і тестовій вибірках. Крім того, виконано графічний аналіз: побудовано діаграми «факт – прогноз» та графік залишків (різниця між фактичними та прогнозованими значеннями).

Алгоритм Random Forest

Random Forest належить до ансамблевих методів **bagging**, у яких комбінування багатьох слабких моделей (дерев рішень) створює сильний узагальнюючий предиктор. Ключові особливості алгоритму[3]:

- Кожне дерево навчається на випадковій підвибірці навчальних даних, отриманій методом бутстреп (вибірка з поверненням). Це зменшує кореляцію між окремими деревами.

- Для кожного вузла дерева при виборі оптимального поділу розглядається випадкова підмножина ознак. Такий підхід запобігає домінуванню однієї сильної ознаки та підвищує різноманіття дерев.

- Для задач регресії підсумковий прогноз обчислюється як середнє арифметичне прогнозів усіх дерев. Для задач класифікації використовується більшість голосів (majority vote).

Завдяки паралельному навчанню дерев та усередненню результатів Random Forest демонструє високу стійкість до переобучення та шумів у даних. Метод дозволяє оцінити важливість ознак, що корисно для інтерпретації моделей екологічних процесів. У дослідженні, присвяченому нормалізації PM_{10} в Австралії, найкраща модель

Random Forest пояснила близько 59 % варіацій концентрації PM_{10} , використовуючи метеорологічні та часові фактори [3]. Це мотивує використання аналогічного підходу для українських умов. Загальний вигляд моделі:

$$f(x) = \frac{1}{B} \sum_{b=1}^B f_b(x);$$

де:

B - кількість дерев у моделі;

$f_b(x)$ - прогноз b -го дерева;

$f(x)$ - узагальнений прогноз Random Forest.

Результати

Для розуміння добової динаміки PM_{10} проаналізовано розподіли концентрацій залежно від часу доби (рис. 1). На графіку показано, як змінюється концентрація PM_{10} залежно від часу доби - вночі, вранці, вдень та ввечері. Середня лінія всередині прямокутника відображає медіану - типові значення. Сам прямокутник охоплює половину всіх даних. Вертикальні лінії, що виходять угору та вниз від прямокутника, показують діапазон нормальних значень. Окремі точки вище верхньої межі чи нижче нижньої межі - це рідкісні випадки, коли концентрації значно відхилялися від звичайних, наприклад під час смогу чи пилових бур. Таким чином:

- Ніч (00:00–06:00). Медіанне значення PM_{10} становить ~ 33 $\mu\text{кг}/\text{м}^3$. Дані характеризуються широким розкидом; верхній кuartиль сягає ~ 54 $\mu\text{кг}/\text{м}^3$. Високі значення (>80 $\mu\text{кг}/\text{м}^3$) трапляються рідко, але саме вночі фіксуються найвищі піки (до 120 $\mu\text{кг}/\text{м}^3$). Підвищення забруднення пов'язане зі слабким перемішуванням повітря та наземними інверсіями.

- Ранок (06:00–12:00). Розподіл схожий на нічний, але верхня межа дещо коротша. Медіана близько 35 $\mu\text{кг}/\text{м}^3$, верхній кuartиль ~ 53 $\mu\text{кг}/\text{м}^3$. Ранкові піки зумовлені підвищенням транспортної активності та накопиченням забруднювачів після

нічної інверсії.

- День (12:00–18:00). Найнижчі концентрації: медіана ≈ 27 мкг/м^3 , верхній кuartиль ≈ 39 мкг/м^3 . Під час денного прогрівання атмосфера стає турбулентною, відбувається активне вертикальне перемішування і розсіювання частинок, що пояснює зниження забруднення.

- Вечір (18:00–24:00). Рівень PM_{10} зростає: медіана ≈ 33 мкг/м^3 , IQR (середній діапазон значень) ≈ 22 – 51 мкг/м^3 . Вечірній максимум відображає пікові навантаження транспорту, побутових джерел та зменшення висоти змішування вночі

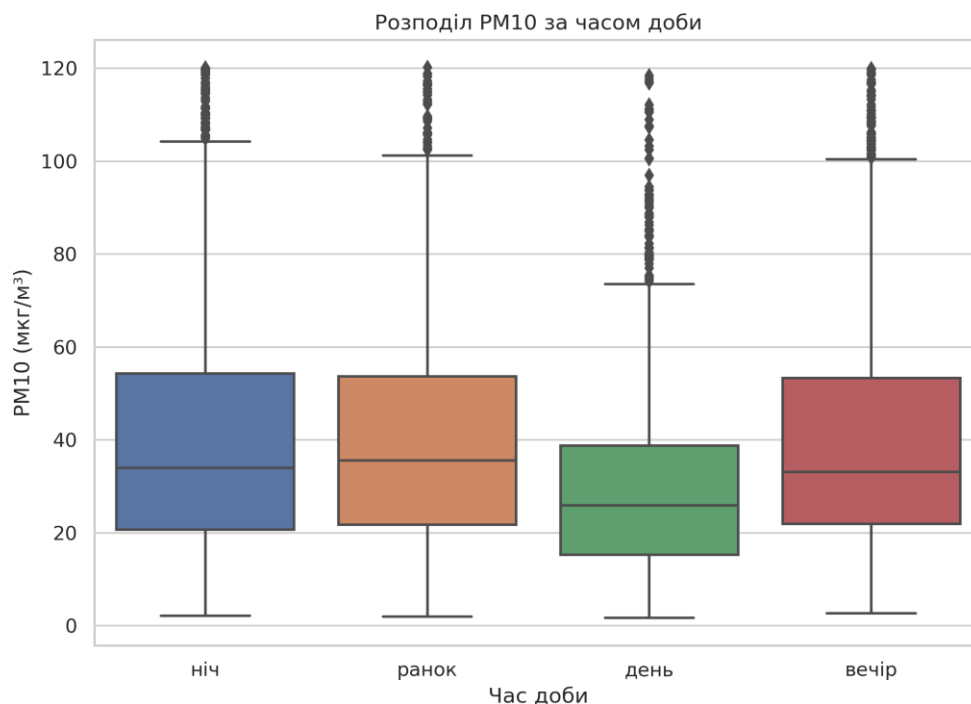


Рисунок 1 – Розподіл концентрацій PM_{10} для різних періодів доби. Рамкові діаграми показують медіани, кuartили та аномальні значення

Таким чином, добовий цикл характеризується мінімумом у денний час і максимумом у нічні та вечірні години. Це відповідає фізичним процесам турбулентності та добовій активності джерел.

Розподіл PM_{10} залежно від рівня умовної інтенсивності викидів наведено на рис. 2.

Порівняння груп підтверджує очікувану тенденцію: при найнижчій інтенсивності (1,00) медіана PM_{10} становить ~ 33 мкг/м^3 ; під час денного та ранкового періодів (1,55–1,75) медіани близькі до 27–35 мкг/м^3 , а у вечірні години (2,15) спостерігається помітне зростання до ~ 33 – 36 мкг/м^3 . Розподіли частково перекриваються, проте найвищі відхилення значень угору фіксуються саме для вечірнього та нічного часу, що свідчить

про вплив не лише антропогенних джерел, а й метеорологічних умов. Кореляція між PM_{10} та метеорологічними змінними

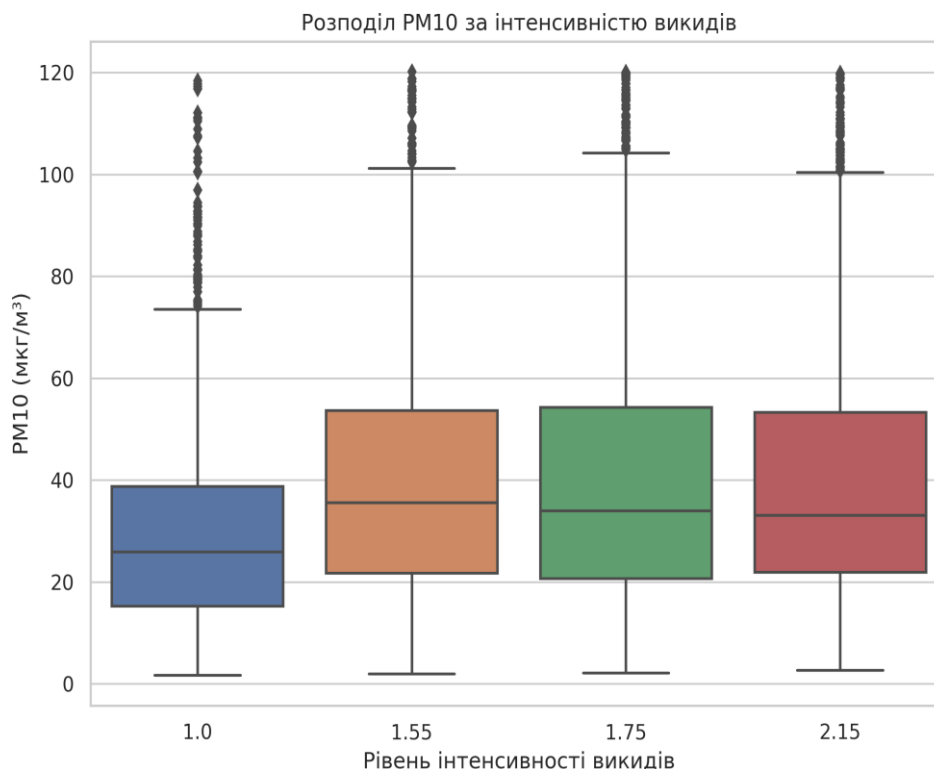


Рисунок 2 – Розподіл концентрацій PM_{10} для чотирьох рівнів інтенсивності викидів (1,00 – ніч; 1,55 – день; 1,75 – ранок; 2,15 – вечір).

Після відсікання 5 % найбільших значень PM_{10} було розраховано коефіцієнти лінійної кореляції між PM_{10} та незалежними ознаками. Найсуттєвіші зв'язки такі:

- Відносна вологість (RH): $r \approx 0,35$. Зростання вологості сприяє злипанню частинок та тривалішому перебуванню аерозолі у приземному шарі. Висока вологість часто супроводжується туманами та штилями, що сповільнюють розсіювання.
- Температура повітря (T): $r \approx -0,21$. Із підвищенням температури повітря концентрації PM_{10} зменшуються завдяки збільшенню висоти змішування та посиленню конвективних потоків. Похолодання, навпаки, сприяє накопиченню пилу.
- Напрямок вітру (DW): $r \approx -0,19$. Хоча у моделі напрям вітру представлено числовим значенням, цей зв'язок відображає переважання певних напрямків під час стійкої погоди. Оскільки напрямок є циклічною змінною, лінійна кореляція є умовною.
- Атмосферний тиск (P): $r \approx 0,08$. Спостерігається слабкий позитивний зв'язок. Високий тиск часто асоціюється зі спокійною погодою та температурними інверсіями, що сприяє акумуляції забруднювачів.

Інші фактори (швидкість вітру, тип дня, інтенсивність викидів) мають низьку лінійну кореляцію, але їхній вплив проявляється у нелінійних взаємодіях, які лінійна кореляція не відображає. Це ще раз підтверджує доцільність застосування нелінійних моделей.

Показники якості моделі Random Forest

Модель Random Forest була навчена на 6 117 спостереженнях (70 %) і перевірена на

2 622 спостереженнях (30 %). Результати оцінювання подано у таблиці 1.

Таблиця 1 – Показники точності моделі Random Forest (MAE – середня абсолютна похибка; RMSE – корінь середньоквадратичної похибки; R^2 – коефіцієнт детермінації).

Вибірка	MAE, мкг /м ³	RMSE, мкг/м ³	R^2
Навчальна	3,25	4,78	0,96
Тестова	8,51	12,63	0,72

Високе значення R^2 (0,96) на навчальній вибірці свідчить про здатність моделі майже ідеально відтворювати вихідні дані. На тестовій вибірці модель пояснює 72 % варіації PM_{10} , що значно перевищує результат лінійної регресії (≈ 27 %). Середня абсолютна похибка зросла з 3,25 до 8,5 мкг/м³. Це закономірно, оскільки на тренувальних даних модель відтворює вже відомі їй приклади й помиляється менше, тоді як на тестових даних вона стикається з новими умовами (іншими поєднаннями метеорологічних факторів та впливом додаткових джерел забруднення), що природно призводить до зростання похибки. Таким чином, більша похибка на незалежних даних відображає реальну здатність моделі до узагальнення. Однак навіть із такою похибкою модель залишається придатною для оцінки загального рівня забруднення й раннього попередження про перевищення нормативів.

Графічний аналіз прогнозів та залишків.

На рис. 3 зображено співвідношення між прогнозованими та фактичними значеннями PM_{10} для навчальної та тестової вибірок. Штрихова лінія показує ідеальну відповідність ($y=x$)

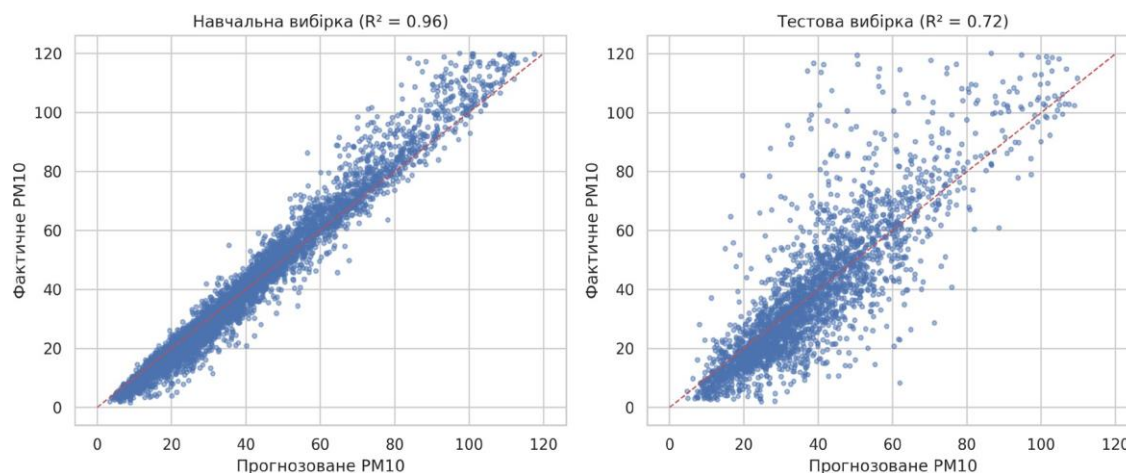


Рисунок 3 – Співвідношення фактичних та прогнозованих значень PM_{10} . Ліворуч – навчальна вибірка ($R^2=0,96$); праворуч – тестова вибірка ($R^2=0,72$).

На навчальній вибірці модель майже точно відтворює спостережені значення, що є природним результатом використання великої кількості алгоритмів, об'єднаних у єдину систему. На тестових даних розсіювання навколо діагоналі помітно більше, однак більшість точок концентруються поблизу лінії ідеальної відповідності. Модель дещо недооцінює високі концентрації (>80 $\mu\text{g}/\text{m}^3$), але загалом відтворює амплітуду коливань. Це свідчить про добру узагальнювальну здатність Random Forest щодо метеорологічних і часових факторів.

Для виявлення систематичних помилок побудовано графік залишків (фактичне значення мінус прогноз) у залежності від прогнозу (рис. 4).

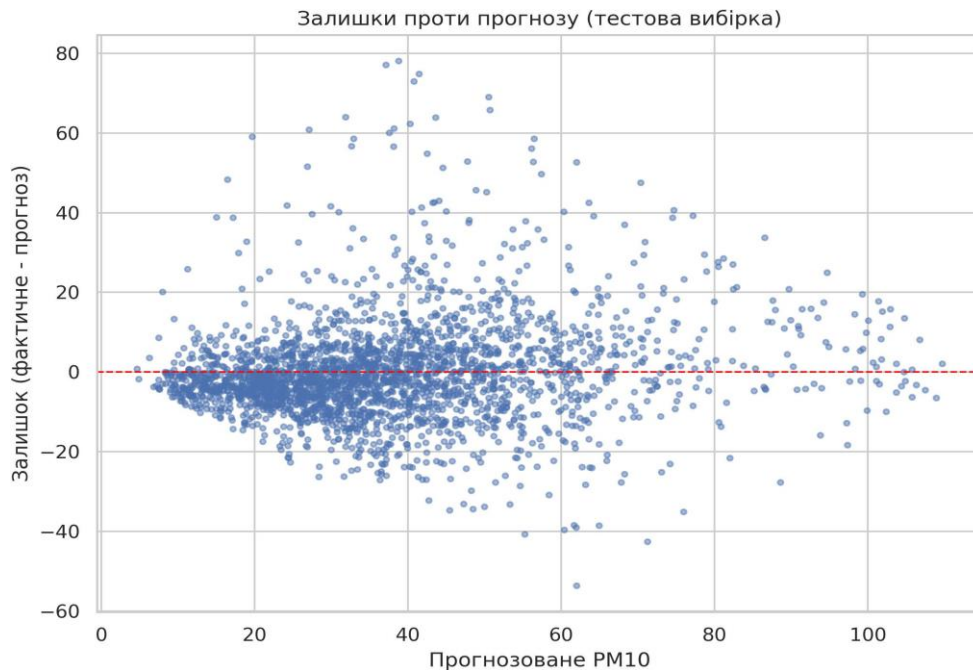


Рисунок 4 – Залишки (фактичне – прогнозоване) у залежності від прогнозованих значень PM_{10} на тестовій вибірці. Штрих лінія $y=0$ позначає ідеальний прогноз.

Графік демонструє, що залишки в середньому розподілені симетрично навколо нуля. У діапазоні прогнозів 0–40 $мкг/м^3$ спостерігається незначне заниження моделі – фактичні значення іноді вищі на 5–20 $мкг/м^3$. Це пояснюється впливом не врахованих факторів (наприклад, локальних джерел пилу або інверсій). При середніх прогнозах (40–70 $мкг/м^3$) залишки мають мінімальний розкид, що вказує на найкращу роботу моделі саме у цьому діапазоні. Для високих прогнозів (>70 $мкг/м^3$) з'являються як від'ємні, так і додатні залишки, але явних систематичних зсувів не спостерігається. Таким чином, Random Forest добре справляється з нелінійними взаємодіями, хоча деякі пікові випадки залишаються непередбачуваними.

Важливість ознак

Оцінка важливості ознак проводилася на основі відносного зниження середньої похибки при розщепленні дерев. На рис. 5 подано ранжування ознак.

Найбільший вклад у точність моделі вносять такі фактори:

- Атмосферний тиск (P), температура (T) та відносна вологість (RH) – кожна з цих змінних становить понад 25 % сумарної важливості. Це підтверджує, що синоптичні умови визначають рівень PM_{10} : підвищений тиск і низькі температури сприяють акумуляції частинок, а висока вологість посилює їхнє осідання та

агрегування.

- Швидкість вітру (SW) – близько 8 % важливості. Помірний вітер сприяє розсіюванню аерозолів, але водночас може переносити пил із джерел на метеостанцію.

- Тип дня (type_day) та напрям вітру (DW) – по 3–4 % важливості. Вихідні дні характеризуються дещо нижчими рівнями PM_{10} через зниження транспортної активності. Напрямок вітру відіграє меншу роль, ймовірно через циклічну природу змінної.

- Інтенсивність викидів і категорії часу доби мають порівняно невелику важливість (1–2 % кожна). Це пояснюється тим, що ці змінні частково корелюють із метеорологічними характеристиками та не додають суттєвої нової інформації для моделі.

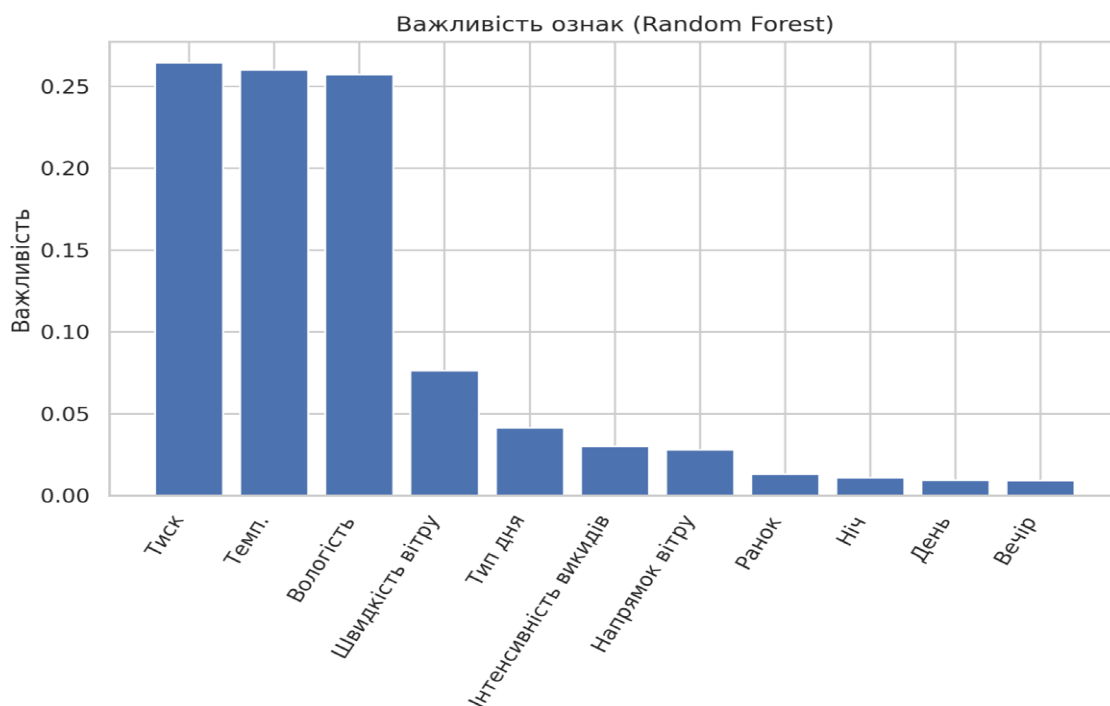


Рисунок 5 – Важливість незалежних змінних у моделі Random Forest. Висота стовпця пропорційна впливу ознаки на точність прогнозу.

Обговорення

Порівняно з моделлю множинної лінійної регресії, яка пояснювала лише 27 % варіації PM_{10} , модель Random Forest значно покращила точність прогнозу. Такий приріст обумовлений здатністю ансамблевого підходу відтворювати нелінійні зв'язки між метеорологічними умовами, часом доби та концентраціями забруднювачів. Лінійні моделі припускають сталий коефіцієнт впливу кожної змінної, але взаємодія факторів (наприклад, висока вологість при низькій температурі) має складніший характер.

Random Forest, навчаючи численні дерева на випадкових підвбірках, автоматично враховує такі взаємодії, що підвищує здатність узагальнення. Крім того, ансамблевий метод краще передбачає пікові значення, хоча певне заниження при дуже високих рівнях все ж зберігається.

Вплив метеорологічних умов

Отримані результати підтверджують, що основними регуляторами рівня PM_{10} у Кривому Розі взимку є температура, атмосферний тиск і відносна вологість. Високий тиск у поєднанні з низькими температурами часто відповідає антициклонним умовам, які формують приземні інверсії та зменшують висоту змішування. Це призводить до накопичення частинок у нижньому шарі атмосфери. Висока вологість може підсилювати утримання пилу завдяки конденсації на аерозолях, а також призводити до туманів, які обмежують вертикальне перемішування. Швидкість вітру відіграє подвійне значення: помірний вітер сприяє розсіюванню забруднення, але сильний може переносити пил із відкритих кар'єрів або промислових майданчиків. Температурний чинник має нелінійний характер впливу. При невеликих морозах ($<0^{\circ}C$) концентрації PM_{10} підвищуються через низьку висоту конвективного шару. Однак при ще більш низьких температурах пил може сильніше осідати або замерзати, що зменшує масову концентрацію. Random Forest успішно враховує такі нелінійні ефекти, чого не можна досягти за допомогою простих лінійних моделей.

Часові закономірності та інтенсивність викидів

Добова структура PM_{10} , виявлена за допомогою розподільчих діаграм, узгоджується з характерними змінами атмосферної стабільності. Мінімум забруднення фіксується вдень, коли поверхня землі добре прогрівається й формується розвинутий конвективний шар.

Увечері та вночі спостерігається стабілізація атмосфери, приземні інверсії та накопичення пилу. Ранкові концентрації відносно високі через активізацію транспорту й промислових джерел. Змінна «інтенсивність викидів» лише частково описує ці цикли, тому її важливість для моделі є помірною.

Зміна PM_{10} залежно від дня тижня виявилася незначною, що свідчить про домінування природних процесів та цілодобової роботи промислових підприємств. Однак невелике зниження концентрацій у вихідні підтверджує зменшення транспортного потоку та побутових джерел. Для більш детального врахування

людської активності доцільно включити дані про інтенсивність руху на ключових магістралях або виробничі графіки підприємств.

Обмеження та можливості для подальших досліджень

Не зважаючи на значне покращення прогнозу, модель Random Forest має певні обмеження:

- Обмежений набір факторів. У вибірці відсутні дані про кількість транспортних засобів, масивні викиди підприємств, опади (дощ, сніг), висоту граничного шару. Включення таких змінних може істотно збільшити точність прогнозу.
- Обробка циклічних змінних. Напрямок вітру подано у градусах від 0 до 360. Лінійне представлення не враховує циклічності ($0^\circ \approx 360^\circ$). Для кращого опису слід використовувати перетворення через синус і косинус. Аналогічно, час доби можна кодувати за допомогою тригонометричних функцій.
- Довгострокове прогнозування. Дана модель призначена для короткострокового прогнозу (1–3 години). Для довгострокових прогнозів доцільно застосовувати рекурентні нейронні мережі, що враховують часову залежність між спостереженнями.
- Інтерпретація нелінійних моделей. Хоча Random Forest забезпечує оцінку важливості ознак, інтерпретація конкретних залежностей залишається складною. Частково цю проблему вирішують методи SHAP або partial dependence plots, які аналізують внесок кожного фактора в прогноз, проте вони не входили до обсягу даної роботи.

Подальше дослідження може включати комбінування моделей (наприклад, градієнтний бустинг, екстремальні випадкові алгоритми), використання даних про перевалочні майданчики, трафік, опади, а також застосування багатомодальних підходів (спутникові знімки, дані з мережі спостережних станцій).

Висновки

- Дані та методи. На основі понад 9000 вимірювань концентрацій PM_{10} та метеорологічних параметрів у Кривому Розі за зимовий період 2024–2025 рр. було сформовано навчальну та тестову вибірки. Додані часові змінні дозволили врахувати добову циклічність і зниження активності у вихідні.
- Розподіли та кореляції. Розподіл PM_{10} виявив чітку добову динаміку: найнижчі концентрації спостерігаються вдень, найвищі – уночі та ввечері. Позитивна кореляція

із вологістю та негативна з температурою підтверджують залежність забруднення від синоптичних умов.

- Модель Random Forest. Побудований алгоритм забезпечує $R^2 \approx 0,72$ на тестовій вибірці та істотно знижує похибки прогнозу порівняно з лінійною регресією. Модель адекватно відтворює розподіл PM_{10} у діапазоні 0–100 мкг/м³ і лише частково недооцінює пікові значення.

- Важливість ознак. Найбільший вплив на рівні PM_{10} мають атмосферний тиск, температура та вологість. Часові змінні та інтенсивність викидів відіграють допоміжну роль. Це свідчить про те, що режими погоди є визначальними для якості повітря в зимовий період.

- Практичні рекомендації. Результати дослідження можуть бути використані органами місцевого самоврядування для оперативного прогнозу та попередження населення про високі концентрації PM_{10} . Запропонована модель застосована як основа для сценарного аналізу та планування заходів із зменшення викидів.

- Узагальнюючи, застосування моделі Random Forest до реальних даних моніторингу показало високу ефективність цього підходу для прогнозування рівнів PM_{10} у промислових містах. Надалі доцільно розширювати набір ознак, використовувати комбіновані методи та інтегрувати результати у системи екологічного менеджменту.

Література

1. European Environment Agency. What is particulate matter and what are its effects on *human health*? FAQ, ред. 20 серпня 2025 р. – [Офіційний вебсайт ЄС] [L55-L74].
2. European Environment Agency. *What is particulate matter and what are its effects on human health*? FAQ, ред. 20 серпня 2025 р. – [Офіційний вебсайт ЄС] [L60-L74].
3. Random Forest Algorithm in Machine Learning. GeeksforGeeks, 2025 р. – Опис принципів роботи, переваг та ключових особливостей Random Forest [L80-L110].