

УДК 621.317.38:519.254]:621.311.243

Алгоритм підвищення точності прогнозування вихідної потужності сонячної електростанції при застосуванні методу найближчих сусідів

Седлярів Є. В., Клен К. С.

Національний технічний університет України "Київський політехнічний інститут імені Ігоря Сікорського", м. Київ, Україна

E-mail: e.sedlyarov-ds21mn@iit.kpi.ua, ekateryna.osypenko@gmail.com

У статті наведено метод прогнозування вихідної потужності сонячної електростанції на основі методу регресії найближчих сусідів. Для створення прогнозу вихідної потужності використано дані з сонячної електростанції в Китаї потужністю 50 МВт, що має дані значень сонячної радіації, навколишньої температури, тиску та вологості і відповідні їм значення вихідної потужності в період з 01.01.2019 до 31.12.2020 з дискретністю 15 хвилин. Перед використанням даних проведено їх очищення від викидів методом стандартного міжквартильного розмаху, розділення точок даних на тестову та тренувальну групи і застосовано масштабування ознак методом стандартизації для коректного обчислення Евклідової дистанції між точками даних. Запропоновано застосування кластеризації погодних параметрів методом k-середніх, що дозволяє провести індивідуальний вибір кількості сусідів для кожного кластеру і виключити вплив точок одного кластеру на кількість сусідів іншого кластеру. Кількість кластерів обирається за допомогою визначення коефіцієнта силуету, тренувальна група погодних параметрів з відповідними їм значеннями вихідної потужності розподіляється по кластерам на основі Евклідової відстані до центрів кластерів. Тестова група погодних параметрів розподіляється по кластерам, після чого прогнозування методом найближчих сусідів відбувається в межах кожного кластеру окремо. Вихідна потужність сонячної електростанції розраховується як середня арифметична зважена потужностей сусідів кожної точки тестової групи. В кінці алгоритму послідовність точок погодних параметрів тестової групи відновлюється і створюється часовий ряд прогнозу вихідної потужності. Запропонований алгоритм дозволив зменшити похибки прогнозу MSE, RMSE, MAPE, MAE на 1 день на 0,5348, 0,2265, 0,38%, 0,1448 відповідно, на 7 днів похибки стали меншими на 0,1992, 0,0384, 0,1%, 0,0193 відповідно. В результаті відносна похибка при прогнозуванні на 24 години складає 4,22%.

Ключові слова: метод найближчих сусідів; прогнозування потужності; кластеризація; сонячна енергія; сонячна радіація; регресія

DOI: [10.20535/RADAP.2024.95.39-46](https://doi.org/10.20535/RADAP.2024.95.39-46)

Вступ

Серед відновлюваних джерел енергії найбільшу популярність мають сонячні електростанції, загальна встановлена потужність яких у 2023 році склала 1046,6 ГВт [1–3]. Сонячна енергія складає вагомую частку в виробництві електроенергії відновлюваними джерелами енергії через переваги сонячних панелей, такі як низька вартість обслуговування, модульність, довгий час експлуатації, доступність сонячної енергії та практично нульовий викид парникових газів [4,5]. Реалізація прогнозування обсягу потужності на виході сонячних електростанцій з найменшою похибкою є необхідною для інтеграції сонячної енергетики у загальну мережу, ефективного планування та менеджменту енергії [6,7]. На генерацію електроенергії сонячними панелями впливають багато чинників, серед яких основними

є: кут нахилу панелі відносно Сонця, температура панелі, хмарність, опади, пил, швидкість вітру та інші [4,8–10]. Створення точної моделі прогнозування для відновлюваних джерел енергії є складною та комплексною проблемою, вирішення якої дозволяє створювати плани вироблення електроенергії, технічного обслуговування, ефективно використовувати енергетичні ресурси та знизити вартість електроенергії. Згідно з Законом України «Про внесення змін до деяких законів України щодо відновлення та "зеленої" трансформації енергетичної системи України», відхилення фактичних погодинних обсягів відпуску електричної енергії суб'єкта господарювання, що використовує енергію сонячного випромінювання має бути не більше 5% [11]. В Україні працює "ринку на добу вперед" на підставі Закону України «Про ринок електричної енергії», що визначає горизонт прогнозу в 24 години [12]. Тому метою даної

роботи є розробка моделі прогнозування електроенергії на виході сонячної електростанції з горизонтом прогнозу в 24 години та з відносною похибкою, що не перевищує 5%.

1 Огляд літератури

Класифікацію методів прогнозування потужності на виході відновлюваних джерел показано на Рис. 1 [4–6, 13]. Персистентні методи, які припускають, що генерація електроенергії відновлюваними джерелами енергії залишається постійною з часом, є занадто спрощеними і використовуються як референсні для визначення ефективності запропонованих моделей прогнозування. Перевагою методів авторегресійної інтегрованої ковзної середньої (англ. ARIMA), в яких враховується сезонність та екзогенні змінні для прогнозування часових рядів, є використання лише часового ряду виробленої електроенергії відновлюваними джерелами енергії, здатність враховувати тренди та сезонність часових рядів, а також розповсюдженість програмних пакетів для реалізації цього методу. Основними перевагами методу найближчих сусідів (англ. k-nearest neighbor method або KNN) є можливість використання непослідовних часових рядів, відсутність потреби в навчанні моделі при появі нових даних та простота алгоритму. Прогнозування за допомогою лінійної регресії має низьку точність і сильно залежить від впливу викидів у погодних даних [14].

Перевагами нейронних мереж є вміння використовувати нелінійні залежності між погодними даними та виробництвом електроенергії відновлюваними джерелами енергії, а також можна відмітити адаптивність та надійність таких моделей [4–6, 15–17].

Фізичні моделі використовують сонячну радіацію в різних проекціях, локацію, кут нахилу сонця, швидкість вітру, вологість та інші погодні параметри для вираховування напруги та струму, що генерують сонячні панелі. Якість прогнозу таких моделей сильно залежить від стабільності погодних умов [5, 11]. Крім цього, на прогноз фізичних моделей можуть впливати особливості обладнання, такі як мікротріщини в сонячних панелях, пил на панелях, температура поверхні панелі [18].

Гібридними є моделі, що поєднують в собі як статистичні, так і фізичні моделі, що підвищує точність прогнозу [19]. Недоліком такого поєднання можна назвати підвищення складності моделі і необхідність залучення більш обчислювальних ресурсів [4].

Вибір прогнозувальної моделі також залежить від дальності горизонту прогнозу [4, 5]. Більшість статистичних методів найбільше підходять для короткострокового та дуже короткострокового прогнозу, персистентні методи використовуються у дуже короткостроковому прогнозі, середньостроковим та довгостроковим прогнозами займаються центри прогнозів, що враховують багато параметрів навколишнього середовища [13].

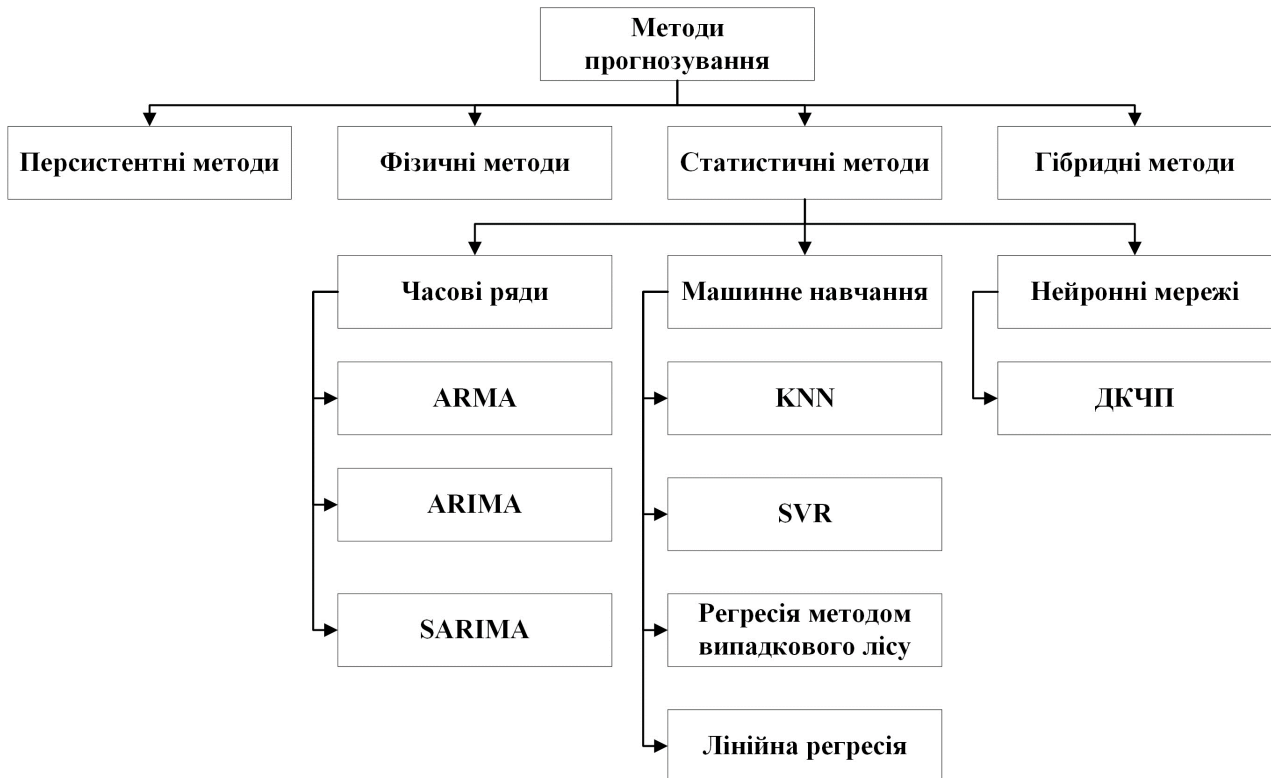


Рис. 1. Класифікація методів прогнозування потужності відновлюваних джерел енергії

Вхідними даними для прогнозової моделі слугують різноманітні параметри навколишнього середовища, такі як: сонячна радіація, температура навколишнього середовища, температура сонячних панелей, аерозольний індекс, вологість, швидкість та напрям вітру. Наявність хмарного покриву спричиняє несподівані просідання або флуктуації генерації електроенергії, і в залежності від часових рамок таких змін зростає необхідність в регульованому обладнанні для електромережі, падає стабільність та якість електроенергії, зростають збитки та витрати на технічне обслуговування [9]. Для врахування впливу хмарності існують такі підходи, як фотографування місцевості з супутника, що також дає змогу коректувати значення сонячної радіації [20], фотографування неба з Землі [19], визначення класу хмар, їх впливу на сонячні промені та проєкції цього впливу на земну поверхню [21], визначення фрактальної розмірності та передавальної функції хмар [22].

Серед оглянутих методів прогнозу виділяється метод найближчих сусідів, оскільки він не потребує значних обчислювальних ресурсів, є простим у використанні та може працювати з непослідовними часовими рядами, що є суттєвими перевагами перед популярними в останній час нейронними мережами та прогнозними моделями групи ARMA.

2 Опис та підготовка даних

При збиранні даних можуть виникнути проблеми, які негативно вплинуть на точність прогнозу, такі як невірно сприйняті датчиками погодні умови чи помилка при записі вихідної потужності сонячної електростанції. Щоб уникнути хибно навченої моделі, перед використанням даних у моделях їх необхідно підготувати, а саме очистити від викидів, дублікатів, провести масштабування ознак для коректного розрахунку відстаней між точками та розділити дані на тестову та тренувальну групи.

Моделі створені на основі даних про погодні умови та відповідні значення вихідної потужності з сонячної електростанції в Китаї [23]. Ці дані містять параметри навколишнього середовища, такі як загальна сонячна радіація (англ. total solar irradiance), пряма нормальна сонячна радіація (англ. direct normal irradiance), глобальна горизонтальна сонячна радіація (англ. global horizontal irradiance), температура зовнішнього середовища, тиск та вологість повітря і відповідні цим умовам значення вихідної потужності у МВт. У [23] наведені дані про 8 сонячних електростанцій та 6 вітрових електростанцій за період з 01.01.2019 до 31.12.2020 з дискретністю 15 хвилин. Для створення моделей обрана сонячна електростанція потужністю 50 МВт, що має найменшу кількість втрачених даних. Загалом, перша сонячна електростанція містить 70176 точок

погодних даних та відповідних значень виробленої електроенергії.

Для всіх моделей важливим етапом є очищення даних від точок, що сильно відрізняються від загальної вибірки, які ще називають викидами. Такі точки будуть перешкоджати ефективному навчанню моделей та знижувати їхню точність прогнозування. У статті використані вже оброблені дані з [23], де застосовувався стандартний міжквартильний розмах, де всі дані, що знаходяться вище за 1,5 міжквартильних розмахи від 75-го перцентилію та нижче за 1,5 міжквартильних розмахи від 25-го перцентилію, вважаються викидами.

Перед використанням даних у моделі необхідно провести їх попередню обробку. Для методів, що базуються на відстанях між точками, такими як метод найближчих сусідів та метод k-середніх, необхідно застосувати масштабування вхідних параметрів для коректного розрахунку відстаней між точками. Масштабування буде проводитися методом стандартизації, що приводить дані до діапазону від -1 до 1, де середнє значення даних прийняте за 0 і середнє квадратичне відхилення має значення 1. Стандартизація проводиться з використанням формули:

$$x'_i = \frac{(x_i - \mu)}{\sigma}, \quad (1)$$

де x – параметри навколишнього середовища та вихідна потужність сонячної електростанції перед масштабуванням, x' – параметри навколишнього середовища та вихідна потужність сонячної електростанції після масштабування, μ – середнє значення даних x перед масштабуванням, σ – середнє квадратичне відхилення даних x перед масштабуванням.

Для оцінки точності прогнозів моделей необхідно розділити дані на тренувальні та тестові групи, де тренувальні мають більший об'єм і використовуються для навчання моделей, а тестові мають невеликий об'єм та використовуються для перевірки моделей. У цій статті тестові групи мають кількість даних на 3 дні, 7 днів та 70 днів для оцінки можливостей прогнозування на короткостроковому та середньостроковому горизонті прогнозу.

Таким чином, на основі попередньо оброблених даних проводиться створення прогнозової моделі вихідної потужності сонячної електростанції методом найближчих сусідів.

3 Метод найближчих сусідів

Метод найближчих сусідів є непараметричним методом керованого навчання, що використовує відстані між точками тестової групи та найближчими до них точками тренувальної групи для вирішення задач класифікації чи регресії. Недоліком цього методу є значення k , що визначає кількість сусідів, тобто кількості точок тренувальної групи, до

якої розраховуються відстані. Зазвичай цей параметр визначається емпірично або обирається таким, щоб для перевірконої групи значення похибки було мінімальним, і це значення використовується для подальшої роботи методу. Це призводить до неточної моделі прогнозу, бо кількість сусідів для різних груп тестових точок буде різною. Проблема вибору значення k полягає у неможливості визначити похибку прогнозування без наявних реальних та прогнозованих значень. Для вирішення проблеми вибору значення k , при якому похибка прогнозування мінімальна, у цьому розділі статті наведено алгоритм вибору кількості сусідів, що базується на визначенні зміни середнього значення відстаней між точками тестової та тренувальної груп, а також наведено метод зменшення похибки прогнозування за допомогою попередньої кластеризації тренувальних та тестових даних методом k -середніх (англ. k -means clustering).

3.1 Класичний метод KNN

Для порівняння класичного методу KNN з запропонованим методом об'єднання KNN та кластеризації k -середніх значення k , тобто кількість сусідів, буде обрано таким, що похибка прогнозу буде мінімальною.

Спочатку визначається відстань між точками тестової групи та точками тренувальної групи за допомогою формули Евклідової відстані, що наведена нижче:

$$d(x', y') = \sqrt{\sum_{i=1}^n (y'_i - x'_i)^2}, \quad (2)$$

де x' – масштабована точка з тренувальної групи, y' – масштабована точка з тестової групи, n – число вимірів точки даних. Евклідова відстань дозволяє визначити найкоротший шлях між двома точками, тим самим визначаючи схожість двох точок між собою. Розрахунок залежної змінної, яка у нашому випадку є змінною вихідної потужності сонячної електростанції, проводиться з використанням формули середньої арифметичної зваженої значень потужності сусідів тестової точки. Вага обчислюється за формулою:

$$\omega_i = 1/d_i, \quad (3)$$

де d_i – Евклідова відстань між тестовою точкою і її сусідом. Тоді формула середньої арифметичної зваженої набуває вигляду:

$$\bar{P} = \frac{\sum_{i=1}^n \omega_i P_i}{\sum_{i=1}^n \omega_i}, \quad (4)$$

де P_i – значення вихідної потужності сонячної електростанції сусіда. Таким чином, чим далі сусід знаходиться від тестової точки, тим менше його внесок в обчислення залежної змінної тестової точки.

На Рис. 2 зображено приклад визначення вихідної потужності точки A:

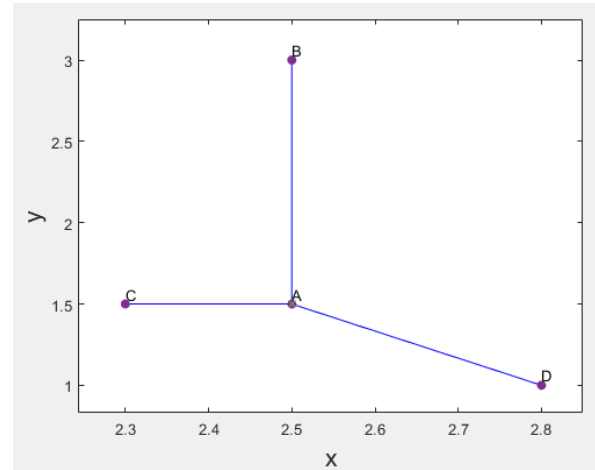


Рис. 2. Приклад визначення вихідної потужності точки A

Точка A є точкою погодних параметрів з тестової групи з кількістю сусідів 3, точки B, C, D є сусідами і точками погодних параметрів з тренувальної групи. Сині лінії між точками позначають Евклідову відстань між точками. Значення вихідної потужності точки A записується як:

$$P_A = \frac{P_B \omega_B + P_C \omega_C + P_D \omega_D}{\omega_B + \omega_C + \omega_D}. \quad (5)$$

Модель приймає значення погодних параметрів для визначення вихідної потужності електростанції.

3.2 Запропонований метод об'єднання KNN та кластеризації k -середніх

Для запропонованого методу об'єднання KNN та кластеризації k -середніх та класичного методу KNN значення k , тобто кількість сусідів, буде обрано таким, що похибка прогнозу буде мінімальною.

Для зменшення похибки прогнозу також застосовано кластеризацію методом k -середніх, що дозволяє створити кластери тренувальної групи, точки яких схожі за своїми погодними параметрами. Алгоритм роботи методу k -середніх наведено на Рис. 3.

Точки тестової групи розподіляються по цим кластерам за допомогою визначення Евклідової відстані від тестової точки до кожної з центроїдів кластерів, точка розподіляється у кластер, відстань до центроїди якого є найменшою. Далі проводиться алгоритм пошуку оптимальних значень кількості сусідів з мінімальними похибками прогнозу в рамках кожного кластеру, що дозволяє провести індивідуальний вибір кількості сусідів для кожного кластеру і виключити вплив точок одного кластеру на кількість сусідів іншого кластеру. Регресія методом найближчих сусідів проводиться в рамках кожного кластеру з подальшим відновленням часового ряду тестової групи точок і створенням прогнозу залежної змінної за алгоритмом класичного методу KNN.

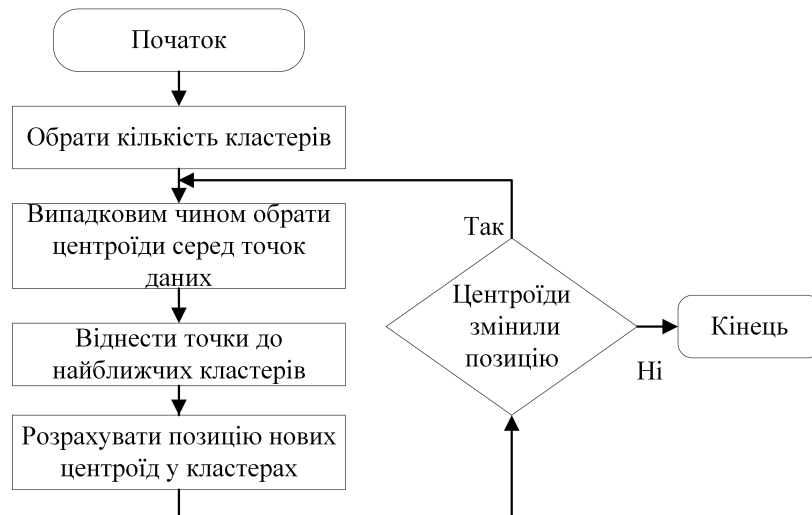


Рис. 3. Алгоритм роботи методу k-середніх

Кластер обирається за допомогою коефіцієнта силуету для визначення якості проведеної кластеризації. Чим вищий коефіцієнт силуету, тим більше точка даних схожа на свій кластер в порівнянні з іншими кластерами. Значення коефіцієнта силуету знаходяться в діапазоні від -1 до 1, де 1 означає, що об'єкт добре підходить під свій кластер, -1 означає, що взагалі не підходить в свій кластер. Формула коефіцієнта силуету:

$$s(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}}, \quad (6)$$

де i – номер точки даних, a_i – середня відстань між точкою даних та іншими точками у власному кластері, b_i – середня відстань від точки даних до точок в іншому найближчому кластері.

На Рис. 4 зображено результат обчислень коефіцієнта силуету для шести кластерів:

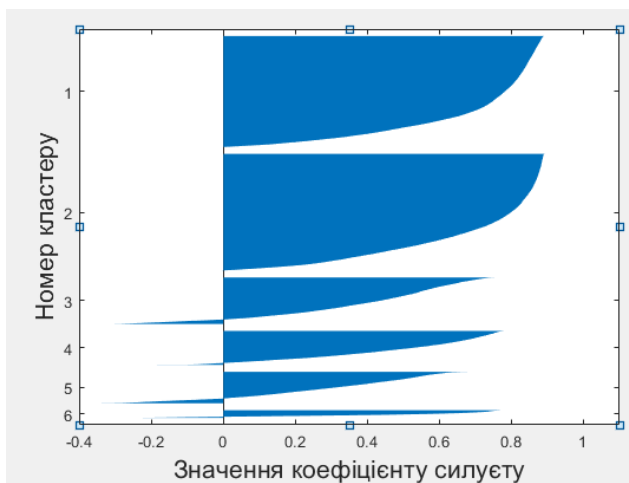


Рис. 4. Значення коефіцієнта силуету для кожного кластеру погодних параметрів

Шість кластерів мають найбільше середнє значення коефіцієнта силуету, тому така кількість кластерів обрана для проведення кластеризації.

4 Експериментальний аналіз

Для оцінки запропонованого методу найближчих сусідів з використанням кластеризації методом k-середніх буде наведено порівняння його похибки прогнозу з похибкою прогнозу методу найближчих сусідів.

Горизонт прогнозу був обраний як перші 1, 3 та 7 днів, що виділені з загального набору даних в якості тестової групи, всі інші дані використані як тренувальна група.

У якості похибок прогнозу використано корінь з середньоквадратичної похибки (англ. RMSE), середньоквадратичну похибку (англ. MSE), середню абсолютну похибку (англ. MAE) та середню відносну похибку (англ. MAPE), що часто використовуються для оцінки точності прогнозу [4–6]. Формули для їх обчислення такі:

$$MSE = \frac{1}{N} \sum_{i=1}^N (A_i - F_i)^2, \quad (7)$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (A_i - F_i)^2}, \quad (8)$$

$$MAE = \frac{1}{N} \sum_{i=1}^N |A_i - F_i|, \quad (9)$$

$$MAPE = \frac{1}{N} \sum_{i=1}^N \frac{|A_i - F_i|}{A_i} \times 100\%, \quad (10)$$

де A_i – реальні значення, F_i – прогнозовані значення, N – кількість точок прогнозу.

Кількість сусідів обиралась такою, щоб похибка MAPE була мінімальною, усі інші похибки розраховувались при цьому ж значенні кількості сусідів.

Результати розрахунку похибок наведені у Таблиці 1.

На Рис. 5 а, б, в наведено результати прогнозу методом KNN та методом KNN+k-середніх для горизонтів прогнозу 1, 3 та 7 днів відповідно.

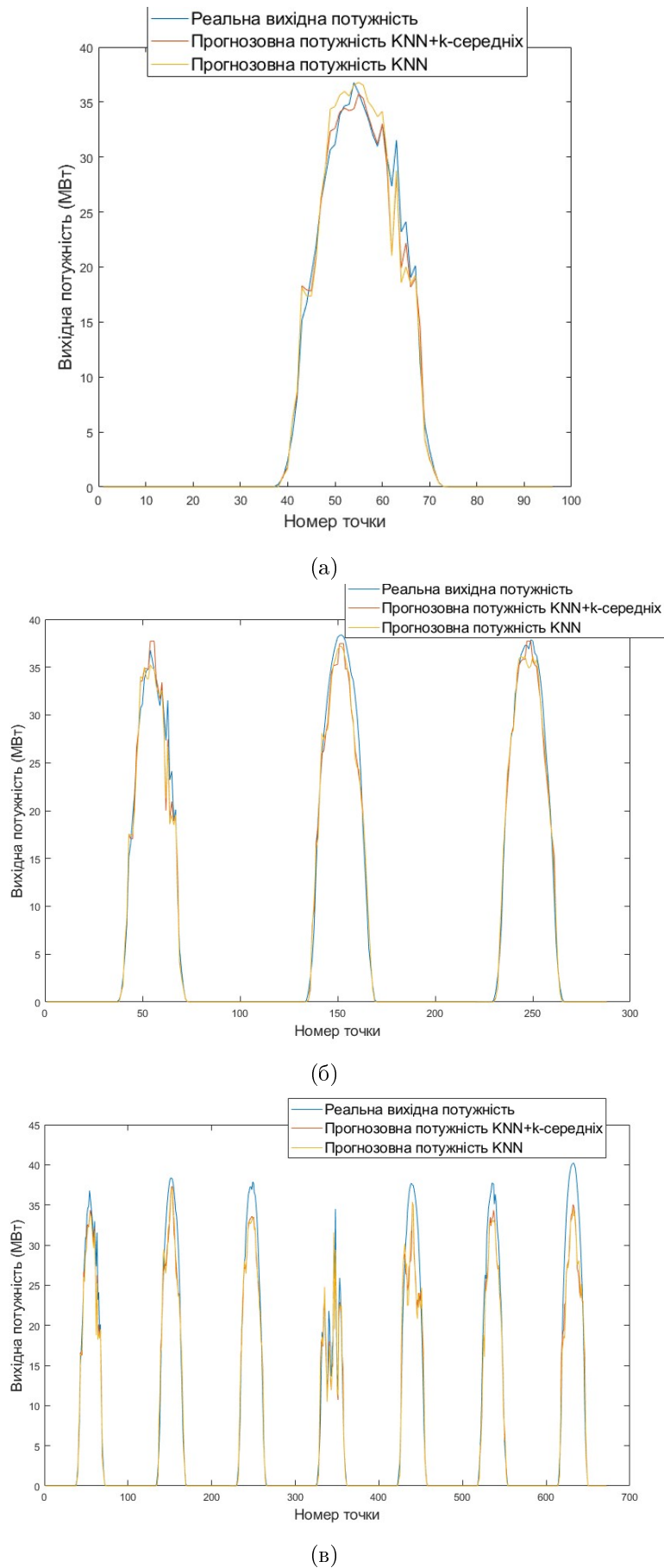


Рис. 5. Результати прогнозування методом KNN та методом KNN+k-середніх для одного дня (а); для трьох днів (б); для семи днів (в)

Табл. 1 Результати розрахунку похибок

| Метод | Кількість днів прогнозу | MSE | RMSE | MAPE, % | MAE |
|------------------|-------------------------|--------|--------|---------|--------|
| KNN | 1 | 1,6744 | 1,2940 | 4,6% | 0,5780 |
| | 3 | 1,4946 | 1,2225 | 5,80% | 0,5682 |
| | 7 | 6,8202 | 2,6115 | 8,14% | 1,2148 |
| KNN + k-середніх | 1 | 1,1396 | 1,0675 | 4,22% | 0,4332 |
| | 3 | 1,5028 | 1,2259 | 5,75% | 0,5562 |
| | 7 | 6,6210 | 2,5731 | 8,04% | 1,1955 |

За допомогою методу KNN+k-середніх вдалося досягти більшої точності порівняно з методом KNN, а саме похибки прогнозу MSE, RMSE, MAPE, MAE на 1 день стали меншими на 0,5348, 0,2265, 0,38%, 0,1448 відповідно, на 7 днів стали меншими на 0,1992, 0,0384, 0,1%, 0,0193 відповідно. При прогнозі на 3 дні меншими стали похибки MAPE на 0,05% та MAE на 0,012, похибки MSE та RMSE більші за звичайний метод KNN на 0,0082 та 0,0034, що виникає внаслідок того, що похибки MSE та RMSE більше виділяють великі значення різниці між прогнозованими та реальними даними. Тобто у запропонованому методі при прогнозі на 3 дні виникла похибка в одній із точок, абсолютне значення якої трохи збільшилось в порівнянні з звичайним методом KNN, на що вказали похибки MSE та RMSE, хоча похибка MAE і стала меншою.

Висновки

У цій статті запропоновано використання кластеризації методом k-середніх тренувальних та тестових груп даних, що дозволило підвищити точність прогнозу вихідної потужності сонячної електростанції регресійним методом найближчих сусідів. Для визначення кількості кластерів використано коефіцієнт силуету. Для створення прогнозу вихідної потужності використовується Евклідова відстань між погодними параметрами тренувальної та тестової груп, що обрані з набору даних сонячної електростанції в Китаї потужністю 50 МВт. Вплив значення вихідної потужності тренувальної групи визначається відстанню до точок тестової групи і розраховується за допомогою формули середньої арифметичної зваженої обраної кількості сусідів кожної точки тестової групи. Застосування кластеризації дозволило зменшити похибки MSE, RMSE, MAPE, MAE на 0,5348, 0,2265, 0,38%, 0,1448 відповідно при прогнозі на один день. В результаті отримано модель прогнозу, яка має значення похибок MSE — 1,1396, RMSE — 1,0675, MAPE — 4,22%, MAE — 0,4332 при горизонті прогнозу 24 години.

References

- [1] Paris Agreement. Treaty Series (2015) *United Nations*, Vol. 3156, Registration: 4 Nov. 2016, № 54113. ENVIRONMENT, Cham. XXVII 7 D, pp. 1-8.
- [2] European Commission, Secretariat-General (2019). Communication from the Commission to the European parliament, the European council, the council, the European economic and social committee and the committee of the regions. The European Green Deal. *An official website of the European Union*, Document 52019DC0640.
- [3] Renewable capacity statistics (2023) *International Renewable Energy Agency*.
- [4] Iheanetu K. J. (2022) Solar Photovoltaic Power Forecasting: A Review. *Sustainability*, Vol. 14, Iss. 24, 17005. doi:10.3390/su142417005.
- [5] Raza M. Q., Nadarajah M., Ekanayake C. (2016) On recent advances in PV output power forecast. *Solar Energy*, Vol. 136, pp. 125-144. doi:10.1016/j.solener.2016.06.073.
- [6] Antonanzas J., Osorio N., Escobar R., Urraca R., Martinez-de-Pison F. J., Antonanzas-Torres F. (2016) Review of photovoltaic power forecasting. *Solar Energy*, Vol. 136, pp. 78-111. doi:10.1016/j.solener.2016.06.069.
- [7] Klen K. S., Zhuikov V. Y. (2020) Vplyv stokhastychnoho kharakteru enerhii v systemakh rozoseredzhenoї heneratsii na yikhniu stiikist [Influence of stochastic nature of energy in distributed generation systems on their stability]. *Technical Electrodynamics*, Vol. 3, pp. 62-68. doi:10.15407/techned2020.03.062.
- [8] Santiago I., Trillo-Montero D., Moreno-Garcia I. M., Pallarés-López V., Luna-Rodríguez J. J. (2018) Modeling of photovoltaic cell temperature losses: A review and a practice case in South Spain. *Renewable and Sustainable Energy Reviews*, Vol. 90, pp. 70-89. doi:10.1016/j.rser.2018.03.054.
- [9] Poddar S., Evans J. P., Kay M., Prasad A., Bremner S. (2023). Assessing Australia's future solar power ramps with climate projections. *Scientific Reports*, Vol. 13, 11503. doi:10.1038/s41598-023-38566-z.
- [10] Welby S. J., Engerer N. A. (2016) Categorizing the Meteorological Origins of Critical Ramp Events in Collective Photovoltaic Array Output. *Journal of Applied Meteorology and Climatology*, Vol. 55, Iss. 6, pp. 1323-1344. doi:10.1175/JAMC-D-15-0107.1.
- [11] Zakon Ukrainy «Pro vnesennia zmin do deiakykh zakoniv Ukrainy shchodo vidnovlennia ta "zelenoi" transformatsii enerhetychnoi systemy Ukrainy» vid 30.06.2023 (№ 3220-IX) [Law of Ukraine On Amendments to Certain Laws of Ukraine on Restoration and Green Transformation of the Energy System of Ukraine]. *Verkhovna Rada Ukrainy [Verkhovna Rada of Ukraine]*, data zvernennia: 27.10.2023.

- [12] Zakon Ukrainy «Pro rynek elektrychnoi enerhii» vid 13.04.2017 (№ 2019-VIII) Redaktsiia vid 03.09.2023 [Law of Ukraine On the Electricity Market]. *Verkhovna Rada Ukrainy [Verkhovna Rada of Ukraine]*, data zvernennia: 27.10.2023.
- [13] Sobri S., Koohi-Kamali S., Rahim N. A. (2018) Solar photovoltaic generation forecasting methods: A review. *Energy Conversion and Management*, Vol. 156, pp. 459-497. doi:10.1016/j.enconman.2017.11.019.
- [14] Myroshnichenko M. V., Klen K. S. (2022) Prohnozuvannia potuzhnosti na vykhodi soniachnoi paneli [Prediction of the Power of the Solar Panel]. *Microsystems, Electronics and Acoustics*, Vol. 27, Iss. 2, pp. 237737-1-237737-5. doi:10.20535/2523-4455.me.237737.
- [15] Abdel-Nasser M., Mahmoud K. (2019) Accurate photovoltaic power forecasting models using deep LSTM-RNN. *Neural Comput & Applic*, Vol. 31, pp. 2727-2740. doi:10.1007/s00521-017-3225-z.
- [16] Zhang J., Verschae R., Nobuhara S., Lalonde J.-F. (2018) Deep photovoltaic nowcasting. *Solar Energy*, Vol. 176, pp. 267-276. doi:10.1016/j.solener.2018.10.024.
- [17] Li G., Xie S., Wang B., Xin J., Li Y., Du S. (2020) Photovoltaic Power Forecasting With a Hybrid Deep Learning Approach. *IEEE Access*, Vol. 8, pp. 175871-175880. doi:10.1109/ACCESS.2020.3025860.
- [18] Mayer M. J. (2021) Influence of design data availability on the accuracy of physical photovoltaic power forecasts. *Solar Energy*, Vol. 227, pp. 532-540. doi:10.1016/j.solener.2021.09.044.
- [19] Visser L., Lorenz E., Heinemann D., Wilfried G. J. H. M. van Sark (2022) 1.11 – Solar Power Forecasts. *Comprehensive Renewable Energy (Second Editoin)*, Vol. 1, pp. 213-233. doi:10.1016/B978-0-12-819727-1.00135-7.
- [20] Watanabe T., Takenaka H., Nohara D. (2021) Post-processing correction method for surface solar irradiance forecast data from the numerical weather model using geostationary satellite observation data. *Solar Energy*, Vol. 223, pp. 202-216. doi:10.1016/j.solener.2021.05.055.
- [21] Nespoli A., Niccolai A., Ogliari E., Perego G., Collino E., Ronzio D. (2022) Machine Learning techniques for solar irradiation nowcasting: Cloud type classification forecast through satellite data and imagery. *Applied Energy*, Vol. 305, 117834. doi:10.1016/j.apenergy.2021.117834.
- [22] Osypenko K. S., Zhuikov V. Ya. (2017) Otsinka fraktalnoi rozmirnosti ta peredavalnoi funktsii khmar [The evaluation of fractal dimension and transfer function of the clouds]. *Microsystems, Electronics and Acoustics*, Vol. 22, No 5, pp. 13-19. doi:10.20535/2523-4455.2017.22.5.106578.
- [23] Chen Y., Xu J. (2022) Solar and wind power data from the Chinese State Grid Renewable Energy Generation Forecasting Competition. *Scientific Data*, Vol. 9, Article number: 577. doi:10.1038/s41597-022-01696-6.

Algorithm for Increasing Accuracy of Solar Power Forecasting When Applying Nearest Neighbor Method

Sedliarov Ye. V., Klen K. S.

This paper presents the prediction of the output power of a solar power plant based on the nearest neighbor regression method. Data from a 50 MW solar power plant in China with solar radiation, ambient temperature, pressure, and humidity values and corresponding output power values from 01/01/2019 to 12/31/2020 with 15-minute intervals were used to create the output power forecast. Before using the data, they were cleaned of outliers using the standard interquartile range method, data points were divided into test and training groups, and feature scaling was applied using the standardization method to correctly calculate the Euclidean distance between data points. The application of clustering of weather parameters by the k-means method is proposed, which allows for individual selection of the number of neighbors for each cluster and to exclude the influence of points of one cluster on the number of neighbors of another cluster. The number of clusters is selected by determining the silhouette coefficient, the training group of weather parameters with their corresponding output power values is distributed among clusters based on the Euclidean distance to the centroid of the clusters. The test group of weather parameters is divided into clusters, after which forecasting by the method of nearest neighbors takes place within each cluster separately. The output power of the solar power plant is calculated as the weighted arithmetic average of the neighbors of each point of the test group. At the end of the algorithm, the sequence of points of weather parameters of the test group is restored and a time series of the output power forecast is created. The proposed algorithm made it possible to reduce the MSE, RMSE, MAPE, MAE forecast errors for 1 day by 0.5348, 0.2265, 0.38%, 0.1448, respectively, for 7 days, the errors became smaller by 0.1992, 0.0384, 0.1%, 0.0193, respectively. As a result, the relative error in forecasting for 24 hours is 4.22%.

Keywords: nearest neighbor method; power forecasting; clustering; solar energy; solar irradiation; regression