

## ПІДВИЩЕННЯ ТОЧНОСТІ ПРОГНОЗУ ВИРОБНИЦТВА ЕЛЕКТРОЕНЕРГІЇ ФОТОЕЛЕКТРИЧНИМИ СТАНЦІЯМИ НА ОСНОВІ МЕТОДУ RANDOM FOREST

<sup>1</sup>Вінницький національний технічний університет

*В роботі було розглянуто метод Random Forest як потужний інструмент для прогнозування генерування фотоелектричних станцій, що є важливим елементом для підвищення ефективності функціонування енергосистеми. Прогнозування генерації електроенергії від фотоелектричних станцій є складним завданням через їхню високу залежність від метеорологічних умов, таких як інтенсивність сонячного випромінювання, температура повітря, хмарність, опади тощо. Ці фактори створюють значні коливання у виробництві енергії, що ускладнює точний прогноз на короткі та довгострокові періоди. Метод Random Forest дозволяє ефективно моделювати залежності між численними змінними та враховувати складні взаємозв'язки між ними, що підвищує якість прогнозування в умовах невизначеності. У ході дослідження було запропоновано ефективний алгоритм використання цього методу для підвищення точності прогнозу, що зменшує похибки й дозволяє забезпечити більш надійне планування електричних резервів в енергосистемі. Особлива увага приділялася адаптивності моделі, тобто її здатності підлаштовуватися до змін у погодних умовах, сезонних коливань і характерних кліматичних особливостей регіону. Цей підхід забезпечує стабільність роботи енергосистеми, оскільки дозволяє операторам більш точно планувати обсяги необхідних резервів потужності, враховуючи можливі перепади генерації відновлюваних джерел. Застосування Random Forest також є корисним для інтеграції прогнозів генерації з іншими елементами системи управління енергоспоживанням, що полегшує процес прийняття рішень. Результати дослідження свідчать про те, що метод Random Forest є перспективним інструментом для прогнозування генерації електроенергії фотоелектричними станціями, забезпечуючи прийнятний рівень точності для планування енергетичних ресурсів. Це особливо актуально в умовах збільшення частки відновлюваних джерел у структурі енергетичного балансу, що вимагає більш гнучкого підходу до управління енергетичними ресурсами.*

**Ключові слова:** відновлювані джерела енергії, генерування, прогнозування, баланс.

### Вступ

Сучасні енергосистеми стикаються зі значними викликами, пов'язаними зі зростанням використання відновлюваних джерел енергії (ВДЕ), таких як сонячна енергія. Хоча ці джерела є досить ефективними та екологічно чистими, вони мають одну важливу особливість – мінливий характер генерування. Прогнозування [1] генерування відновлюваних джерел енергії, таких як фотоелектричні станції (ФЕС), або вітрові електростанції (ВЕС), має ключове значення для ефективного формування структури балансу потужності в енергосистемі. Формування балансу потужності з відповідними резервами має вирішальне значення для забезпечення стійкості та надійності роботи енергосистеми та забезпечення її операційної безпеки. Прогнозування генерування ФЕС поряд з засобами внутрішньодобового перерозподілу енергії є ефективним інструментом для вирішення зазначених завдань. Прогнозування генерування дозволяє передбачити потенційні коливання у виробництві електроенергії від ФЕС і вжити відповідних заходів для формування резервів потужності в енергосистемі у певні періоди часу. На даний час найпоширенішим підходом до прогнозування є застосування методів машинного навчання та штучних нейронних мереж. Ці алгоритми базуються на аналізі великих обсягів даних та виявленні складних залежностей між вхідними даними [2] (наприклад, погодними умовами) та генеруванням електроенергії. Процес оцінки графіків генерування ФЕС має враховувати низку характеристик самої станції, таких як встановлена потужність, схемні рішення, стан та орієнтування сонячних панелей, наявність трекерів, а також засобів регулювання генерування залежно від потреб системи. Важливо відзначити, що прогнозування генерування від ФЕС та ВЕС є складним завданням. Воно пов'язане з необхідністю прогнозування погодних умов, параметри яких мають складно-передбачувану (випадкову) складову. Погодні умови можуть раптово змінюватися, що призводить до різких коливань у генеруванні енергії. Виходячи з цього точність прогнозів істотно залежить від доступності актуальних даних про впливові параметри навколишнього середовища та від ефективності обраних алгоритмів прогнозування. Прогрес у галузі прогнозування генерування ФЕС дає змогу енергетичним компаніям та операторам систем передачі (ОСП) підвищити ефективність формування оперативних резервів енергії для балансування. Це зменшує негативний вплив негарантованого

генерування від ВДЕ на операційну безпеку енергосистеми. Таким чином, якісне прогнозування графіків видачі енергії ФЕС та ВЕС відіграє важливу роль у досягненні сталого розвитку та забезпечення надійності енергосистеми. Це дає змогу збільшити використання ВДЕ для оптимізації процесів в енергосистемі, забезпечити перехід до чистої та стійкої енергетики, зменшуючи залежність від традиційних джерел та мінімізуючи негативний вплив на довкілля.

### Результати дослідження

Розглянемо Random Forest [2], як один з можливих методів прогнозування генерування ФЕС. Random Forest – це ансамблевий метод машинного навчання, який використовується як для класифікації, так і для регресії. Основна ідея полягає у побудові численних дерев прийняття рішень під час тренування моделі й продукування моди для класів (класифікацій) або усереднення прогнозу (регресії) побудованих дерев. Перевагами даного методу є [3]:

- здатність ефективно обробляти дані з великою кількістю ознак і класів;
- нечутливість до масштабування значень ознак;
- можливість обробляти як неперервні, так і дискретні ознаки;
- внутрішня система оцінювання здатності побудованої моделі до узагальнення (тест out-of-bag);
- адаптивність до паралельних обчислень з багатьма потоками.

Random Forest використовує багато різних дерев прийняття рішень для досягнення більш точних та стійких прогнозів.

Основні етапи використання Random Forest у прогнозуванні включають:

1. Формування ансамблю. Алгоритм створює випадковий ансамбль дерев прийняття рішень. Кожне дерево формується на випадковому підмножині тренувальних даних та випадковому виборі факторів для розгляду на кожному рішенні.

2. Прогнозування кожним деревом. Кожне дерево в ансамблі прогнозує результат на основі свого власного піднабору даних. Прогнози цих дерев об'єднуються для отримання узагальненого прогнозу.

3. Оцінка і ваговий внесок. Кожне дерево отримує вагу на основі його точності на тренувальних даних. Це дозволяє Random Forest автоматично враховувати внесок кожного дерева при узагальненні прогнозу.

4. Відбір важливих факторів. Random Forest автоматично вимірює важливість кожного фактора у прогнозуванні. Це може служити як важливий етап відбору впливових факторів.

5. Виявлення взаємозв'язків. Алгоритм може виявляти взаємозв'язки між факторами, дозволяючи отримати більш глибоке розуміння взаємодії змінних у моделі.

Враховуючи здатність Random Forest працювати з великою кількістю факторів, а також виявляти складні взаємозв'язки, цей метод є ефективним інструментом для прогнозування та аналізу даних, а також для відбору важливих факторів та виявлення взаємозв'язків між ними. В якості даних (табл. 1) було обрано датасет «ФЕС-1». Датасет містить ретроспективу даних за 12 місяців, зокрема мітку дати та часу «time», згенеровану електроенергію «fact\_gen» (кВт\*год), а також метеопараметри: температуру повітря «temperature\_c» (°C), атмосферний тиск «pressure\_in» (inHg), кількість опадів «precip\_mm» (мм), вологість «humidity» (в.о.), хмарність «cloud» (в.о) та точку роси «dewpoint\_c» (°C).

Вибір саме цих метеопараметрів для прогнозування генерування ФЕС не є випадковим. Аналіз показав, що саме їх зміни найбільше корелюють із добовими графіками згенерованої електроенергії. Висока міра кореляції дає підстави для використання цих даних у подальшому прогнозуванні. На адекватність вимірів фізичних величин можуть чинити негативний вплив збої в роботі датчиків, відсутність з'єднання або живлення, а також інші зовнішні збурення. Тому перед аналізом даних і подальшим навчанням моделей машинного навчання необхідно виконувати двоступеневу обробку вхідних даних. На першому етапі необхідно виключити з навчальної вибірки «ФЕС-1» виміри електрогенерування та дані про погодні умови, які відповідають годинам обмежень, годинам з аварійними відключеннями, годинам з ремонтними роботами тощо.

На другому етапі, щоб позбутись помилкових відхилень (викидів), необхідно проаналізувати результати вимірювання метеопараметрів і відфільтрувати години, коли моніторинг працював некоректно, наприклад, у випадках, коли протягом кількох годин поспіль значення параметрів не змінюються або ж відсутні. На цьому ж етапі використовуючи взаємозв'язок між окремими параметрами виявляються їх нетипові зміни. Такі дані маркуються як дані з низькою точністю.

Табл. 1 – Частина даних з датасету по ФЕС-1.

time	fact_gen	temperature_c	pressure_in	precip_mm	humidity	cloud	dewpoint_c
2023-06-01 00:00:00	0	10,8	30,08	0	74	3	6,4
2023-06-01 01:00:00	0	10,4	30,08	0	74	0	5,9
2023-06-01 02:00:00	0	9,9	30,08	0	73	1	5,2
2023-06-01 03:00:00	0	9,3	30,08	0	72	0	4,5
2023-06-01 04:00:00	0	8,9	30,09	0	72	0	4,2
2023-06-01 05:00:00	112,4739	9,8	30,09	0	72	0	5
2023-06-01 06:00:00	686,0913	12,7	30,1	0	64	0	6,1
2023-06-01 07:00:00	2334,6	14,7	30,1	0	56	0	6,1
2023-06-01 08:00:00	4260,6	16,5	30,09	0	51	0	6,2
2023-06-01 09:00:00	5816,7	17,8	30,09	0	45	0	5,8
2023-06-01 10:00:00	6857,1	18,8	30,09	0	42	0	5,7
2023-06-01 11:00:00	6994,8	21,6	30,08	0	41	0	7,6
2023-06-01 12:00:00	6990,3	22,5	30,07	0	40	0	8,1
2023-06-01 13:00:00	6987,6	22	30,05	0	39	4	7,6
2023-06-01 14:00:00	6957,9	21,8	30,03	0	40	12	7,5
2023-06-01 15:00:00	6205,5	21,8	30,02	0	40	24	7,7
2023-06-01 16:00:00	4855,5	21,7	30,02	0	41	26	8
2023-06-01 17:00:00	3194,1	21,7	30,02	0	44	14	8,8
2023-06-01 18:00:00	1301,0826	21,1	30,02	0	54	9	11,4
2023-06-01 19:00:00	263,0652	18,2	30,03	0	62	9	10,8
2023-06-01 20:00:00	0	14,6	30,03	0	63	6	7,7
2023-06-01 21:00:00	0	13,8	30,03	0	65	4	7,2
2023-06-01 22:00:00	0	13,2	30,03	0	66	4	7,1
2023-06-01 23:00:00	0	12,8	30,02	0	67	3	6,9

В якості першого тестового періоду для моделі розрахуємо червень 2023 року. Важливо зазначити, що задля об'єктивної оцінки точності, з навчальної вибірки було виключено період, на який розраховувався прогноз. Отримані результати (рис. 1) показали, що поточні налаштування моделі не забезпечили високої точності прогнозу, адже, відхилення склало 55%.

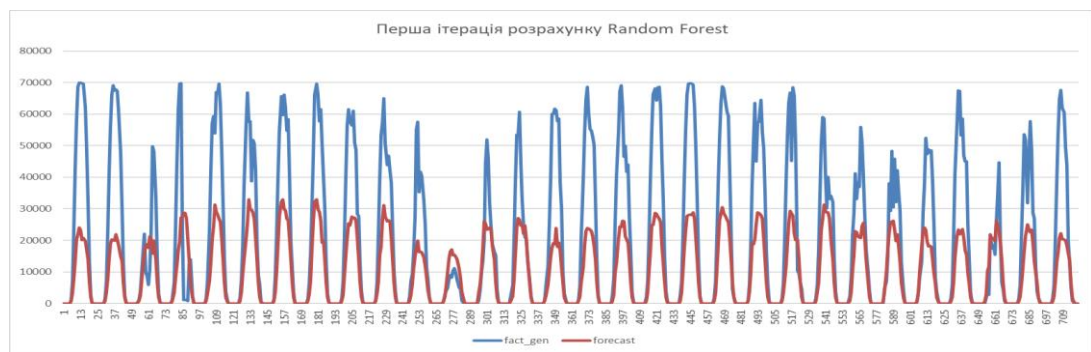


Рис. 1 – Перша ітерація розрахунку прогнозу з використанням Random Forest

З урахуванням того, що прогноз характеризується негативними відхиленнями (рис. 1), запропоновано уточнити його, врахувавши сезонність генерування прогнозованої ФЕС. Для цього розрахуємо середнє значення генерування для кожної години доби подібного за характером генерування періоду, наприклад, червня 2022 року.

Значення усередненого генерування для кожної  $i$ -ї години доби розраховуємо як середнє значення генерування усіх  $i$ -тих годин за період червня 2022 року:

$$Paverage\_gen(i) = AVERAGE(Pgen\_i-1 + Pgen\_i-2 + \dots + Pgen\_i-n),$$

де  $Paverage(i)$  - середнє значення генерування для  $i$ -ї години,

$Pgen\_i-1 + Pgen\_i-2 + \dots + Pgen\_i-n$  - значення генерування для  $i$ -ї години за червень 2022 року.

Відповідно кінцевий прогноз розраховуємо як середнє між значенням середнього генерування та значенням розрахованого прогнозу моделі Random Forest:

$$Pforecast-2(i) = AVERAGE(Pforecast-1(i) + Paverage\_gen(i)),$$

де  $Pforecast-2(i)$  – прогноз генерування  $i$ -ї години 2-ї ітерації розрахунку,

$Pforecast-1(i)$  – прогноз генерування  $i$ -ї години 1-ї ітерації розрахунку.

Цей підхід дозволяє більш детально врахувати вплив сезонності, забезпечуючи гнучкість в регулюванні прогнозу. Провівши дану ітерацію уточнення прогнозу, отримуємо відхилення в 42%, що є кращим результатом за попередній.



Рис. 2 - Друга ітерація розрахунку прогнозу з використанням Random Forest та врахуванням сезонності генерування «ФЕС-1»

Уточнивши прогноз було зменшено обсяг від'ємних відхилень та покращено точність, проте з'явився значний обсяг позитивних відхилень. Для того, щоб зменшити обсяг небалансів додамо до навчальної вибірки дані генерування періоду, на який прогнозуємо, тобто, червня 2023 року. Важливо зауважити, модель розраховує погодинні значення для кожної прогнозованої доби, тому, задля об'єктивності, виключатимемо при розрахунку прогнозовану годину із навчальної вибірки.



Рис. 3 - Третя ітерація розрахунку прогнозу з використанням Random Forest, з врахуванням генерування у червні 2023 року

Враховавши у навчальній вибірці генерування червня 2023 року, було отримано відхилення в 28%.

Для підвищення точності прогнозу, розширимо навчальну вибірку, включивши в неї дані за періоди червня та липня 2021 року. Цей крок дозволить моделі урізноманітнити вивчення залежностей та трендів, що сприятиме покращенню її здатності до прогнозування. Включення цих періодів ключе більшу базу для навчання, що в свою чергу сприятиме ефективнішому функціонуванню моделі в реальних умовах.



Рис. 4 - Четверта ітерація розрахунку прогнозу з використанням Random Forest, з розширенням навчальної вибірки «ФЕС-1»

В результаті розширення навчальної вибірки, було отримано відхилення в 24%.

### Висновки

Застосування даного методу виявило певний потенціал для прогнозування, проте аналіз результатів вказує на значну похибку в прогнозуванні протягом обраного періоду. Це свідчить про необхідність подальшого вдосконалення методу. На етапі роботи вже було застосовано декілька підходів для підвищення точності прогнозу, зокрема було проведено попереднє фільтрування даних, щоб уникнути впливу аномальних показників на результати. Крім того, у вибірці для навчання моделі було враховано основні метеорологічні показники, що могли найбільше впливати на генерацію енергії. Проте важливо продовжити роботу над покращенням результатів. Зокрема, для підвищення точності прогнозу необхідно глибше аналізувати сезонні зміни та їхній вплив на результати, а також залучати додаткові метеорологічні параметри. Одним із перспективних напрямків є використання даних від різних постачальників метеопараметрів, що дозволить знизити ризик похибки через неточність прогнозу. Також слід розширювати період навчання моделі, враховуючи довгострокові тренди та закономірності. Таким чином, подальша робота буде зосереджена на удосконаленні вибірки для навчання та більш точному урахуванні впливу зовнішніх факторів.

### СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

- 1) A. Loureno et al. "Comparison of forecasting models for photovoltaic power generation". Energy Conversion and Management, 118, 2016, 404-418.
- 2) "Introduction to Machine Learning with Python: A Guide for Data Scientists" by Andreas C. Müller and Sarah Guido (2016)
- 3) A. C. Cadena et al. "Weather Forecasting for Photovoltaic Power Prediction Using Machine Learning Techniques". Energies, 11(6), 2018, 1362.

**Кулик Володимир Володимирович** - д. т. н., доц. каф. ЕСС, Вінницький національний технічний університет, Вінниця, e-mail: volodymyrvkulyk@gmail.com  
Вінницький національний технічний університет

**Затхей Максим Вікторович** - аспірант, факультет електроенергетики та електромеханіки, Вінницький національний технічний університет, Вінниця, e-mail: maxzatkey@gmail.com  
Вінницький національний технічний університет

**Kulyk V.V.<sup>1</sup>**  
**Zatkhei M.V.<sup>1</sup>**

# **IMPROVING THE ACCURACY OF THE FORECAST OF ELECTRICITY PRODUCTION BY PHOTOELECTRIC PLANTS BASED ON THE RANDOM FOREST METHOD**

<sup>1</sup> Vinnytsia National Technical University

*The paper considered the Random Forest method as a tool for forecasting the generation of photovoltaic plants, which is a key element for optimizing the structure of the electricity balance in the power system. Photoelectric stations are characterized by significant dependence on meteorological conditions, which creates difficulties in accurately forecasting their generation. The use of the Random Forest method allows taking into account numerous variables and complex interrelationships between them, which increases the accuracy of the forecast in conditions of significant uncertainty inherent in renewable energy. On the basis of the conducted research, an effective Random Forest application algorithm was proposed, which allows to reduce forecast errors. Particular attention is paid to how the model can adapt to changes in weather conditions and seasonal variations, which are important factors in maintaining the stable operation of the power system. The results of the study demonstrate that the application of the Random Forest method for forecasting the generation of photovoltaic plants provides an acceptable level of accuracy for planning power reserves in the power system. In particular, it allows power system operators to more effectively manage the balance of electricity, planning in advance the necessary power reserves depending on the forecasted volumes of renewable energy generation. This is especially important to ensure the stability of the power system during peak loads and in conditions of a growing share of renewable energy sources in the overall generation structure.*