

ДОСЛІДЖЕННЯ ЕФЕКТИВНОСТІ МЕТОДІВ МАШИННОГО НАВЧАННЯ
ДЛЯ ПРОГНОЗУВАННЯ ВИРОБЛЕННЯ ЕЛЕКТРИЧНОЇ ЕНЕРГІЇ СОНЯЧНИМИ
ЕЛЕКТРОСТАНЦІЯМИ

Волобуєв А. С., аспірант, e-mail: andreyvolobuev7777@gmail.com

Савченко О. А., к.т.н., доц., e-mail: savoa@btu.kharkiv.ua

Державний біотехнологічний університет

Актуальність дослідження. У наші дні споживання енергії у світі зростає і стає актуальним вирішення задачі щодо заміни традиційних джерел на альтернативні. Вирішення цього завдання неможливе без попереднього аналізу даних та подальшого прогнозування вироблення енергії альтернативними джерелами. Прогнозування величини виробітку електричної енергії сонячною електростанцією дозволить підвищити обґрунтованість рішень, що приймаються в процесі управління розподілом електроенергії в децентралізованих системах. За наявності інформації про те, скільки електроенергії буде вироблено сонячною електростанцією і передано в мережу щогодини, з'явиться можливість більш обґрунтовано планувати використання виробленої електроенергії та її розподіл. Також наявність достовірного прогнозу дозволить вбудовувати математичну модель у підсистему управління мікромережами, що, своєю чергою, сприятиме інтеграції централізованих електричних мереж та об'єктів розподіленої генерації. В даному дослідженні було проведено оцінювання ефективності застосування алгоритмів машинного навчання для прогнозування виробітку електричної енергії сонячними електростанціями.

Мета досліджень. Підвищення обґрунтованості перспективного планування обсягів виробітку електричної енергії об'єктами альтернативної енергетики (сонячними електростанціями).

Основні матеріали досліджень. Як об'єкт дослідження було розглянуто сонячну електростанцію потужністю 2,6 МВт. Розглядалися фактичні дані про вироблення електроенергії за три роки (2019, 2020, 2021р.). Усі розрахунки було реалізовано мовою програмування python. В якості прогностичних факторів використовувались: спрямована вниз інтенсивність сонячного випромінювання (Вт/м^2), індекс ясності всього неба, альbedo всієї поверхні неба, температура на відстані 2 м, відносна вологість повітря на висоті 2 метрів, відкоригована кількість опадів (мм/год), тиск (кПа), швидкість вітру на висоті 10 метрів (м/с), напрям вітру (градуси). Прогнозована величина - активна потужність СЕС (МВт).

В першу чергу була виконана первинна обробка даних. На цьому етапі дані були попередньо оброблені наступним чином: видалені рядки з нульовим значенням цільової змінної (нічні години), а також знайдені та усунені пропуски даних. Потім набір даних був розділений на навчальну та тестову вибірки у співвідношенні 80/20. Після чого всі ознаки були приведені до єдиного масштабу шляхом застосування мінімаксної нормалізації.

Було побудовано моделі прогнозування для таких методів машинного навчання: k -найближчих сусідів (KNeighborsRegressor), дерево рішень (DecisionTreeRegressor), випадковий ліс (RandomForestRegressor – RFR), багатошаровий перцептрон (MLPRegressor), градієнтний бустинг SkLearn (GradientBoostingRegressor – GBR), екстремальний градієнтний бустинг (XGBRegressor – XGBoost) градієнтний бустинг бібліотеки CatBoost (CatBoostRegressor - CAT).

Модель k -найближчих сусідів прогнозує цільову змінну шляхом локальної інтерполяції значень цільової змінної n найближчих значень. Модель дерев рішень використовує послідовність предикатів для передбачення цільової змінної. Будується так зване дерево рішень, у вузлах якого розташовані предикати, а в листі – передбачення. Дерево рішень можна розглядати як шматково-постійну апроксимацію. Основні параметри моделі: criterion, де задається функція якості поділу, splitter - стратегія при виборі поділу в кожному вузлі, max_depth - максимальна глибина дерева та ін. У порівнянні з моделлю дерев рішень, модель RFR запобігає перенавчанню, що є суттєвою перевагою ансамблевого підходу. При реалізації

моделей градієнтного бустингу відбувається ітеративне додавання моделей в ансамбль в такий спосіб. Створюється модель, яка ініціалізує ансамбль. При цьому модель може мати досить велику похибку прогнозу. Потім починається цикл, в якому при використанні поточного ансамблю генеруються прогнози для кожного спостереження набору даних, які потім додаються в ансамбль, обчислюється функція втрат. Потім шляхом реалізації градієнтного спуску визначаються оптимальні параметри моделі, після чого нова модель додається до ансамблю і всі описані операції повторюються. Модель регресії градієнтного бустингу бібліотеки XGBoost є однією з реалізацій градієнтного бустингу, що використовується для роботи з табличними даними. Для її налаштування необхідний вибір наступних параметрів: параметр `n_estimators` – кількість проходів за циклом моделювання (еквівалентно кількості моделей, які входять у ансамбль), `early_stopping_rounds` – кількість невдало побудованих моделей (тобто тих, послідовне додавання в ансамбль яких не збільшило продуктивність моделі і викликало ранню зупинку виконання циклу, тобто без необхідності прогону алгоритму). Модель регресії CatBoost реалізує алгоритм градієнтного бустингу, є ансамблевим методом, що базується на деревах рішень. Алгоритм градієнтного бустингу CatBoostRegressor підтримує можливість використання категоріальних змінних, L2-регуляризацію та інші можливості. Основними параметрами моделі є: `iterations` – кількість ітерацій, `loss_function` – функція втрат, `learning_rate` – швидкість навчання, `depth` – глибина дерев. Відмінністю даного алгоритму від алгоритму випадкового лісу є послідовне вдосконалення моделей дерев рішень. Тобто, кожна наступна в ансамблі модель дерева рішень будується з урахуванням результатів попередньої моделі.

Для кожної моделі була проведена процедура підбору параметрів за допомогою інструмента перехресної перевірки GridSearchCV. Для оцінки якості прогнозних моделей були використані середня абсолютна помилка прогнозу (mean absolute error - MAE), коефіцієнт детермінації (R^2). Найкращі результати за величиною середньої абсолютної помилки отримані моделлю на основі алгоритму градієнтного бустингу бібліотеки CatBoost (MAE складає близько 85 Вт/м²). Також у цієї моделі найменше перенавчання. Інші алгоритми градієнтного бустингу дали схожі результати, особливо це стосується моделі випадкового лісу. Тому в подальших дослідженнях можна рекомендувати поєднання їх за технологією стекінгу.

Висновок. В результаті проведеного дослідження було створено моделі машинного навчання, що базуються на сучасних алгоритмах інтелектуального аналізу даних. У ході виконання роботи було проаналізовано фактори тимчасового та метеорологічного характеру та оцінено їх ступінь впливу на цільовий результат – вироблення електроенергії сонячною електростанцією. Розрахунки показують, що найбільшу точність прогнозування мають моделі градієнтного бустингу. Досить високу точність прогнозування мають також моделі випадкового лісу. Однією з перспектив для подальших досліджень є проведення непрямого прогнозування величини виробітку електроенергії та його порівняльний аналіз з отриманими в цьому дослідженні результатами.

ПЕРЕЛІК ПОСИЛАНЬ

1. Gensler, A., Henze, J., Sick, B., Raabe, N. Deep Learning for solar power forecasting—An approach using AutoEncoder and LSTM Neural Networks. In Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics (SMC), Budapest, Hungary, 9–12 October 2016.
2. Hendouzi, A., Bourouhou, A. Solar Photovoltaic Power Forecasting. J. Electr. Comput. Eng. 2020
3. Mellit, A., Pavan, A., Oglari, E., Leva, S., Lughi, V. Advanced Methods for Photovoltaic Output Power Forecasting: A Review. Appl. Sci. 2020, 10, 487.
4. Li, P., Zhou, K., Yang, S. Photovoltaic Power Forecasting: Models and Methods. In Proceedings of the 2nd IEEE Conference Energy Internet Energy System Integration, Beijing, China, 20–22 October 2018.