

ІМОВІРНІСНЕ ПРОГНОЗУВАННЯ ПРОМИСЛОВОГО НАВАНТАЖЕННЯ: LIGHTGBM ТА TFT З УРАХУВАННЯМ РИНКОВИХ ОБМЕЖЕНЬ

В роботі представлено орієнтований на графік роботи обладнання підхід до середньострокового прогнозування промислового навантаження, який передбачає декомпозицію споживання на детермінований базовий профіль, сформований із графіків роботи обладнання, та стохастичного залишку, що моделюється за допомогою LightGBM і Temporal Fusion Transformer (TFT). Режимні зсуви визначаються за допомогою сегментації PELT, тоді як багатогоризонтні квантильні прогнози забезпечуються неперетинними та узгодженими між часовими рівнями завдяки методам MinT/THieF. Ринкові обмеження інтегруються через вибір релевантних горизонтів прогнозування, що відповідають правилам закриття «воріт» для двосторонніх договорів і ринку «на добу наперед», а також через застосування вартісно-зваженої *pinball*-втрати, яка відображає асиметричні витрати небалансів у критичні години. На відкладеній тестовій вибірці TFT перевершує LGBM за точністю точкових прогнозів. Використання конформної квантильної регресії (CQR) для LGBM і γ -звуження для TFT покращує емпіричне покриття до номінального рівня та зменшує ширину інтервалів. Запропонований підхід є безпосередньо придатним для планування закупівель відповідно до правил українського ринку електроенергії та узгоджується з «*decision-centric*» оцінюванням економічної вартості прогнозів.

Ключові слова: прогнозування електричного навантаження, промислове підприємство, квантильні прогнози, темпоральне узгодження, ринкові фактори електроенергетики.

Вступ

Прогнозування навантаження та попиту на електричну енергію промислового підприємства є важливим інструментом для експлуатації та планування його діяльності. Особливо враховуючи, що підприємство, як непобутовий споживач, згідно Закону України «Про ринок електричної енергії» є повноцінним учасником ринку електричної енергії. Такий споживач має право купувати електричну енергію для власного споживання за двосторонніми договорами та на організованих сегментах ринку [1]. Це відкриває для промислового підприємства, як нові можливості, так і відповідальність за власні небаланси електричної енергії, що фактично стимулює таких споживачів покращувати точність власного прогнозування і забезпечувати мінімальну можливу його помилку, з метою зменшення витрат на власне споживання з врахуванням ринкових умов. Промисловий споживач має потенціал для активної участі у впровадженні сучасних ринкових механізмів, таких як гнучке споживання, розподілена генерація, а також оптимізація процесів імпорту та експорту електроенергії.

Електричне навантаження промислового підприємства формується під впливом множини взаємопов'язаних чинників: клімату та погоди, календарно-сезонних шаблонів, макроекономічної активності/обсягів виробництва, а також цінових сигналів на сегментах ринку електричної енергії. Систематичні огляди прогнозування навантаження послідовно підкреслюють роль цих факторів прогнозування та їхню значущість на різних горизонтах прогнозу, зокрема для промислових споживачів [2, 3].

Перспективи розвитку нових приладів обліку, що можуть передавати дані в режимі реального часу, створюють як нові можливості (детальніша ідентифікація режимів), так і нові виклики для моделей (виражена нелінійність, характеризується стохастичністю) [4, 5].

На практиці прогнозування традиційно розрізняють завдання за горизонтом: від дуже короткого/короткого (хвилини-дні) через середньострокове (тижні-місяці) до довгострокового (роки) [6]. Водночас для задачі формування оптимальної структури закупівлі електроенергії за двосторонніми договорами ключовим є інтервал, у межах якого промислові споживачі приймають рішення щодо фіксації базового навантаження: практично доцільним є діапазон від 3 місяців до 2 тижнів до початку постачання. Саме цей проміжок забезпечує достатню стабільність виробничих планів і водночас дозволяє використовувати актуалізовані зовнішні чинники (календарно-сезонні шаблони, погодне нормування, очікування щодо ринкових цін), знижуючи ризик не узгодити план випуску продукції та контрактованими обсягами.

Для промислового споживача довгострокова задача задоволення попиту може бути поділена на: задоволення потреби базового навантаження та реалізація залишкових відхилень. Укладання двосторонніх

договорів дає можливість завчасно зафіксувати гарантовані обсяги споживання електричної енергії, що забезпечує покриття базового навантаження підприємства та знижує залежність від цінових коливань на короткострокових сегментах ринку. У свою чергу залишкові відхилення оптимально коригувати на «ринку на добу наперед», внутрішньодобовому та балансуєчному сегментах ринку електричної енергії. У даному дослідженні увага зосереджена на методах прогнозування електричного навантаження, які є доцільними для задоволення попиту підприємства на ринку двосторонніх договорів, а також на розробленні методики подальшого уточнення структури закупівлі електроенергії через ринок «на добу наперед» та внутрішньодобовий ринок без порушення виробничого графіка.

У роботі аналізуються моделі середньострокового (від кількох тижнів до кількох місяців) і довгострокового (від кількох місяців до років) прогнозування електроспоживання на рівні промислового підприємства із залученням історичних даних, метеорологічних чинників та виробничих графіків роботи обладнання.

Метою дослідження є розробка та оцінка орієнтований на графік роботи обладнання підходу середньострокового прогнозування навантаження промислового підприємства, що поєднує детерміновану базу від графіка роботи обладнання, режимну сегментацію, ймовірнісне моделювання залишку і темпоральне узгодження з ринково-орієнтованою функцією втрат, релевантною до «закриття воріт» та вартості небалансів.

Визначення горизонту прогнозування

Аналіз поточної структури ринку електричної енергії України дає змогу виокремити його основні сегменти та ідентифікувати специфічні обмеження щодо участі суб'єкта на кожному з них. Ключовим елементом функціонування є часові рамки торгівлі, які визначають доступність укладання угод. Ці часові обмеження відображено на відповідній часовій шкалі (рис. 1), що ілюструє моменти відкриття та закриття «воріт» на різних сегментах ринку. Такий підхід дозволяє систематизувати особливості роботи двосторонніх договорів, ринку «на добу наперед» та внутрішньодобового ринку, а також врахувати регуляторні вимоги та технічні умови, що накладаються на учасників ринку в кожному часовому проміжку.



Рисунок 1 – Сегменти ринку електричної енергії України (відкриття та закриття «воріт» на відповідних сегментах ринку електричної енергії) [1, 7, 8, 9]

Ключову роль у забезпеченні стабільності постачання для промислових споживачів відіграє сегмент двосторонніх договорів. Саме через нього формується базове навантаження, що дозволяє мінімізувати ризики волатильності короткострокових сегментів та забезпечувати прогнозованість виробничих процесів. Важливо підкреслити, що укладання двосторонніх договорів здійснюється заздалегідь, нерідко на горизонтах від кількох тижнів до одного року і більше, а також супроводжується регламентованими часовими обмеженнями – зокрема закриттям «воріт» о 10:00 за один день до доби постачання [7]. Це зумовлює необхідність наявності достовірного прогнозу споживання з достатнім випередженням.

Для ефективної участі у сегменті двосторонніх договорів потрібне поєднання декількох підходів: довгострокового та середньострокового прогнозування – з метою формування загального профілю навантаження та визначення обсягів, які будуть зафіксовані в договорах заздалегідь, а також короткострокового прогнозування, що дозволяє уточнювати споживання напередодні постачання та узгоджувати його з виробничим графіком. Таким чином, завдання прогнозування у контексті двосторонніх договорів набуває комплексного характеру, поєднуючи стратегічне та оперативне планування.

Аналіз актуальних досліджень

Прогнозування на основі комплексних та численних показників раннього попередження

Автори [10] пропонують середньо- та довгостроковий метод прогнозування навантаження, що об'єднує кілька «ранніх індикаторів», переводячи їх у спільну дискретну шкалу й інтегруючи за принципом мінімальних інформаційних втрат (MIB). Результатом є ймовірнісне передбачення інтервалів зміни сезонної компоненти попиту, з подальшим переходом до точкового прогнозу (середина інтервалу).

На першому етапі працюють з попередньою сезонною корекцією (X-12-ARIMA), далі із циклічним приростом зміни несезонної компоненти. Наступним кроком виконується вибір індикаторів із фізичним/економічним сенсом, дискретизація індикаторів і цільової змінної та оцінка їхніх розподілів. Індикатори об'єднуються за принципом MIB. І в кінці повернення від зміни сезонної компоненти до реального попиту.

Перевагами такого підходу можна зазначити єдину шкалу для різномірних індикаторів, стійкість до шуму (через дискретизацію) та краща поведінка в «точках зламу» (коли структура попиту змінюється). Але наявні недоліки, а саме використання місячних рядів, що у контексті ринку електричної енергії не є доцільним, адже ціни змінюються погодинно. Також потрібен осмислений вибір індикаторів. Дискретизація спрощує (може втрачатися нюанс неперервних зв'язків).

Детерміновано-агреговані моделі на основі графіка роботи обладнання

Характеризуються побудовою планового профілю навантаження як суми одиничних графіків роботи технологічних агрегатів і/або цехів з урахуванням їхніх режимів, пускових струмів та процесів розігріву, обмежень на зміну станів (гістерезису) та календаря виробництва. Такий підхід природно спирається на відомі наперед регламенти функціонування й добре масштабується на складні виробничі системи [11, 12].

Практично це формує «базову траєкторію», яка надалі коригується статистичною чи ML-моделлю для врахування залишкових ефектів (похибки плану, впливу погоди на непроцесних споживачів, мікроподій). Такий двошаровий підхід є ключовим для досягнення високої точності на середніх горизонтах прогнозу, коли розклад операцій уже набуває достатньої стабільності.

Регресійні моделі із зовнішніми регресорами

Регресія з ARIMA-похибкою (ARIMAX). Такий підхід дає змогу включати відомі наперед коваріати (наприклад, календарні фактори, інформацію про зміни або планові операції) та одночасно моделювати автокореляційну структуру ряду через ARIMA-компоненту похибки. Метод доцільний для середньострокових горизонтів прогнозування за наявності стабільних зовнішніх предикторів [13].

Динамічна гармонічна регресія (DHR). Сезонні коливання у цьому підході описуються за допомогою системи гармонічних (Фур'є) термів, тоді як короткострокова залежність моделюється ARMA-похибкою. Така специфікація ефективна у випадках, коли в часових рядах поєднується кілька циклів (наприклад, добові, тижневі або змінні) [13].

TBATS. Це універсальна модельна рамка, яка поєднує трансформацію Бокса-Кокса, ARMA-похибку, компоненту тренду та тригонометричне подання сезонності, що дозволяє адекватно відтворювати складні сезонні структури й багаточастотні дані [15].

Ключовими перевагами наведених підходів є прозорість, інтерпретованість та відносна простота інтеграції відомих у майбутньому змінних, таких як службові індикатори робочих змін, простоїв або інших виробничих регламентів.

Гладкі адитивні моделі (GAM/SCAM) із фізично орієнтованими обмеженнями

Гладкі адитивні моделі добре описують нелінійні залежності попиту від температури, календарних факторів та їхніх взаємодій із режимом роботи. Для відображення циклічних закономірностей застосовуються циклічні сплайни, наприклад, для змін «година-тиждень» чи «місяць-рік». У випадках, коли потрібні фізично обґрунтовані обмеження, доцільно використовувати SCAM. Такий підхід дозволяє накладати апіорні умови, зокрема монотонне зростання реакції навантаження або U-подібну залежність від температури [15, 16].

Основними сильними сторонами цих підходів є інтерпретованість отриманих ефектів, можливість фізично коректного екстраполювання на помірних горизонтах прогнозування за умови стабільної структури предикторів та гнучкість у моделюванні нелінійних та циклічних зв'язків.

Ансамблі дерев рішень для точкових і квантильних прогнозів (GBM/XGBoost/LightGBM/CatBoost; QRF)

Градiєнтне підсилення дерев (GBM). Цей підхід є стандартом для роботи з табличними даними, що містять велику кількість ознак (лагові змінні, ковні статистики, календарні та режимні індикатори). Реалізації, зокрема XGBoost, LightGBM і CatBoost, демонструють високу точність, забезпечують механізми регуляризації та, у випадку LightGBM, підтримують накладання монотонних обмежень. Це особливо важливо при моделюванні температурних ефектів, де очікується фізично обґрунтована монотонність реакції навантаження [17, 18, 19].

Для побудови інтервальних прогнозів застосовуються: квантильне навчання бустингових моделей із використанням функції втрат pinball-loss, що дозволяє безпосередньо оцінювати τ -квантілі розподілу умовного попиту; Quantile Regression Forests (QRF) як непараметрична альтернатива, яка формує емпіричний розподіл на основі ансамблю дерев і дозволяє отримувати будь-які квантілі постфактум [20]. У практичних застосуваннях необхідно забезпечувати узгодженість квантильних прогнозів, а також перевіряти покриття прогнозних інтервалів на валідаційних вибірках для досягнення цільового рівня довіри.

Нейромережеві архітектури для багатогоризонтного прогнозування.

LSTM/GRU та Temporal Convolutional Networks (TCN). Рекурентні мережі типу LSTM та GRU, а також каузальні згорткові мережі TCN, здатні відображати довгострокові залежності у багатовимірних часових рядах і добре інтегруються з відомими наперед коваріатами [21].

N-BEATS та N-HiTS. Архітектури на основі багатосарових персептронів, які показали високу точність на міжнародних змаганнях із прогнозування та продемонстрували здатність до узагальнення. Вони однаково ефективні як для одиничних, так і для групових часових рядів [22, 23].

Temporal Fusion Transformer (TFT). Спеціалізована трансформерна архітектура для багатогоризонтного прогнозування, яка здійснює розподіл ознак на статичні, відомі у майбутньому та спостережувані. Завдяки інтерпретованим механізмам уваги TFT особливо доречний для випадків, коли графік роботи відомий наперед і може бути поданий як known-future inputs [24].

PatchTST та інші трансформери для довгих послідовностей. Сучасні архітектури, що базуються на патч-поданні, демонструють високу ефективність у роботі з довгими часовими рядами (годинні або півгодинні дані) та уникають «вибуху» обчислювальних витрат пам'яті, властивого класичним трансформерам [25].

Режимно-орієнтований шар: виявлення та моделювання точок зламу

На середньострокових горизонтах прогнозування часто спостерігаються зміни виробничих режимів, пов'язані з перезапуском технологічних ліній, проведенням ремонтів або сезонними переходами. Для підвищення стабільності моделей доцільно застосовувати такі підходи.

Один з таких є алгоритм PELT, що має лінійну складність і дозволяє сегментувати часовий ряд та закодувати режими за допомогою one-hot представлення [26, 27]. Також за допомогою методу BOCPD (Bayesian Online Change Point Detection), який забезпечує оперативне оновлення прогнозів у режимі реального часу [28].

Альтернативним підходом є використання регресійних моделей зі «switch»-режимами або прихованих марковських моделей (Hidden Markov Models, HMM), які дозволяють моделювати зміни станів на основі календарних та режимних індикаторів.

Ієрархічне та темпоральне узгодження прогнозів

Для забезпечення когерентності прогнозів як між часовими рівнями (година-доба-тиждень-місяць), так і між виробничими підсистемами (лінія-цех-завод), доцільно застосовувати методи *forecast reconciliation*. Найбільш поширеними підходами є MinT (minimum-trace reconciliation) та темпоральні ієрархії (THieF). Їхнє використання підвищує точність оцінок на всіх рівнях агрегації та забезпечує узгодженість планування [29, 30, 32].

Постановка задачі

Нехай задано часовий ряд активної потужності (або, еквівалентно, ряду споживання електроенергії, перерахованого у потужність) з годинною дискретністю y_t для $t = 1, \dots, T$. Відомі в майбутньому коваріати x_t , а саме детермінований графік роботи обладнання s_t , календарні коваріати c_t , погода (температура, CDD/HDD тощо) w_t , лагові та ковзані значення l_t :

$$x_t = (s_t, c_t, w_t, l_t) \quad (1)$$

Пропонується декомпозиція навантаження у вигляді:

$$y_t = b_t + r_t, \quad b_t := \sum_{k=1}^K s_{k,t} P_k \quad (2)$$

де b_t – детермінований базовий профіль, отриманий із графіка роботи обладнання (індикатори $s_{k,y} \in \{0,1\}$ та його номінальна потужність P_k), а r_t – стохастичний залишок, який відображає відхилення від планового профілю. Така декомпозиція є природною для промислових споживачів, де значна частка споживання визначається виробничими режимами [32].

Задача дослідження – оцінити багатогоризонтні квантильні прогнози $\hat{y}_{T+h}^{(q)}$, для $h = 1:H, q \in \{0.1, 0.5, 0.9\}$, із забезпеченням когерентності між годинними, добовими та тижневими рівнями агрегації.

Режимно-орієнтовне виявлення зламів у залишках

На першому кроці оцінюємо залишок $r_t = y_t - b_t$ та виконуємо **сегментацію** на однорідні режими методом PELT (Pruned Exact Linear Time) з квадратичною вартістю [26]:

$$\min_{\tau_1:M} \sum_{m=0}^M C(\tau_m + 1: \tau_{m+1}) + \beta M, \quad C(\cdot) = \sum (r_t - \bar{r})^2, \quad (3)$$

Де τ_m – межі сегментів, β – штраф. Реалізаційно використовуємо алгоритми бібліотеки *ruptures* [27].

Стохастична модель залишку

Для кожного $q \in \{0.1, 0.5, 0.9\}$ навчаємо апроксиматор $f_q: \mathbf{x}_t \rightarrow \hat{y}_t^{(q)}$. Навчальна ціль – вартісно-зважена “pinball”-втрата [33]:

$$\mathcal{L}_q = \sum_{t=1}^T \omega_t p_q(r_t - f_q(\mathbf{x}_t)), \quad p_q(u) = u(q - 1\{u < 0\}) \quad (4)$$

де ваги ω_t відображають економіку небалансів (наприклад, підсилення робочих годин та «закриття воріт»), як вартісний пріоритет у функції втрат, коли помилка прогнозу найбільш дорога з точки зору контрактів/небалансів:

$$\omega_t = 1 + \alpha_1 1\{t \in \text{peak}\} + \alpha_2 1\{t = h^*\} \quad (5)$$

Як базову реалізацію беремо LightGBM з GPU-прискоренням [18]. Фізичні shape-обмеження (монотонний вплив CDD/HDD) задаємо через монотонні обмеження ознак у медіанній моделі $f_{0.5}$ з L2-втратою:

$$\min_f \sum_t \omega_t (r_t - f(\mathbf{x}_t))^2 \quad \frac{\partial f}{\partial \text{CDD}} \geq 0, \frac{\partial f}{\partial \text{HDD}} \geq 0 \quad (6)$$

Таке рішення узгоджується з ідеологією SCAM (Shape-Constrained Additive Models) [15], а також із підтримкою монотонних обмежень у дерев’яних ансамблях (зокрема, у LightGBM для регресійних задач). Водночас у LightGBM монотонні обмеження несумісні з квантильною ціллю. Тому на практиці застосовується гібридний підхід: медіану $\hat{r}^{(0.5)}$ оцінюють за допомогою L2-моделі з монотонними обмеженнями, тоді як крайні квантілі $q = (0.1, 0.9)$, прогнозують квантильними моделями без обмежень, додаючи прогноз медіани $\hat{r}^{(0.5)}$ як додаткову ознаку. Такий підхід забезпечує емпіричну стабілізацію крайових квантилів і зменшує ризик їхнього перетину.

Альтернативою або компонентом ансамблю виступає Temporal Fusion Transformer (TFT) - спеціалізована трансформерна архітектура для інтервальних багатогоризонтних прогнозів, яка використовує відомі наперед коваріати та поєднує механізми уваги керування потоками [24]. Для кожного рівня квантілі q навчається модель g_q , що мінімізує відповідну квантильну функцію втрат $\mathcal{L}_q$ на залишках:

$$\hat{r}_t^{(q)} = g_q(\{\mathbf{x}_{t-L:t}\}, \{\mathbf{x}_{t+1:t+H}^{\text{відомі}}\}) \quad (7)$$

TFT є придатним для моделювання складних нелінійних взаємодій типу «режим $\times$ погода $\times$ календар» і природно узгоджується з постановкою «*known-in-advance*» у задачах промислового планування, коли графік роботи обладнання є детермінованим і відомий заздалегідь [24].

Рекурсивне багатогоризонтне прогнозування та узгодженість квантилів.

Нехай $t = T + k, k = 1: H$. На кроці k будемо $\mathbf{x}_{T+k}$, який включає відомі календарні та планові фактори, а також оновлені лагові й ковзні статистики, побудовані на основі вже прогнозованої медіани. Далі прогноз для квантілі. Далі прогноз для квантілі q визначається як:

$$\hat{y}_{T+k}^{(q)} = b_{T+k} + \hat{r}_{T+k}^{(q)}, \quad \hat{r}_{T+k}^{(q)} \in \{f_q(\mathbf{x}_{T+k}), g_q(\cdot)\} \quad (8)$$

Для забезпечення неперетину квантильних прогнозів застосовується впорядкування $\{\hat{y}^{(0.1)} \leq \hat{y}^{(0.5)} \leq \hat{y}^{(0.9)}\}$, що відповідає підходам квантильної реконструкції [34].

Узгодження між часовими рівнями

Позначимо $\mathbf{y}^{\text{баз}} \in \mathbb{R}^n$ вектор базових прогнозів усіх рівнів часової ієрархії (година/доба/тиждень). Тоді узгоджений прогноз у методі MinT визначається як

$$\tilde{\mathbf{y}} = S(S^T W^{-1} S)^{-1} S^T W^{-1} \mathbf{y}^{\text{баз}} \quad (9)$$

де S – матриця агрегації, яка відображає зв’язок між базовими та агрегованими рівнями; W – (структурно) узагальнена коваріація похибок (*wls_struct*). Такий підхід гарантує **когерентність прогнозів** між рівнями ієрархії та зменшує середньоквадратичну похибку узгоджених оцінок [29]. Побудову темпоральних ієрархій і практичні аспекти застосування описано в [31].

Остаточні квантильні прогнози набувають вигляду:

$$\hat{y}_{T+1:T+H}^{(q)} = \mathcal{R}(b_{T+1:T+H} + \hat{r}_{T+1:T+H}^{(q)}), \quad q \in \{0.1, 0.5, 0.9\} \quad (10)$$

де $\mathcal{R}(\cdot)$ - оператор узгодження $\tilde{\mathbf{y}}$ на рівні $q = 0.5$ за схемою MinT, а також процедура впорядкування (non-crossing) для множини квантилів (0.1, 0.5, 0.9).

Обґрунтування вибору моделей. Запропонована архітектура спирається на поєднання сучасних методів прогнозування, кожен з яких має власні сильні сторони. По-перше, використання орієнтованого на графік роботи обладнання підходу декомпозиції дозволяє відокремити детерміновану компоненту (виробничий план, номінали обладнання) від стохастичної частини, що знижує дисперсію залишків і підвищує здатність моделі узагальнювати між різними режимами [32]. По-друге, сегментація за допомогою алгоритму PELT забезпечує асимптотично лінійну обчислювальну складність та стійкість до довгих часових рядів, а бібліотека *ruptures* пропонує перевірену реалізацію цього підходу [26, 27]. По-третє, ансамблеві моделі градієнтного бустингу, зокрема LightGBM, дозволяють інтегрувати фізично обґрунтовані залежності, наприклад вплив температури [18]. Додатково застосовується Temporal Fusion Transformer (TFT), спеціалізований для багато-горизонтних задач із відомими наперед коваріатами, здатний відображати довготривалі залежності та взаємодії між режимними факторами завдяки механізмам уваги, що забезпечує конкурентну точність та інтерпретованість результатів [24]. Використання вартісно-звážеної функції втрат «pinball» дозволяє узгодити процес навчання із метриками, які враховують небаланси та договірні зобов'язання [33]. Крім того, застосування методів узгодження прогнозів, зокрема MinT та темпоральних ієрархій (THieF), гарантує когерентність між годинними, добовими та тижневими рішеннями й зменшує помилки на агрегованих рівнях [29, 31]. Нарешті, використання процедури *non-crossing* забезпечує достовірність інтервальних прогнозів [34].

Метрики та валідація

Оцінювання здійснювалося на відкладеному тестовому інтервалі з протоколом «розширюваного вікна» та окремою калібрувальною вибіркою. Точкову якість вимірювали за MAE, RMSE, sMAPE; інтервальну за Pinball-втратою для $\tau \in \{0.1, 0.5, 0.9\}$, покриттям ICP_{80} , шириною AIW_{80} і критерієм $Winkler_{80}$. Відомі у майбутньому ознаки (графік/календар/номінали) включалися без витоку; лагові та ковзні статистики обчислювалися лише з минулих значень.

Точкові метрики:

$$MAE = \frac{1}{N} \sum_{t \in \mathcal{T}} |y_t - \hat{y}_t|, \quad RMSE = \sqrt{\frac{1}{N} \sum_{t \in \mathcal{T}} (y_t - \hat{y}_t)^2}, \quad sMAPE = \frac{2}{N} \sum_{t \in \mathcal{T}} \frac{|y_t - \hat{y}_t|}{|y_t| + |\hat{y}_t| + \varepsilon} \quad (11)$$

Pinball-втрата для $\tau \in (0,1)$:

$$Pinball_{\tau} = \frac{1}{N} \sum_{t \in \mathcal{T}} \rho_{\tau}(t_t - \hat{y}_t^{(\tau)}), \quad \rho_{\tau}(u) = \begin{cases} \tau u, & u \geq 0 \\ (\tau - 1)u, & u < 0 \end{cases} \quad (12)$$

Покриття інтервалу (ICP) для рівня $1 - \alpha = 0.8$ та середня ширина інтервалу (AIW):

$$ICP_{80} = \frac{1}{N} \sum_{t \in \mathcal{T}} 1\{L_t \leq y_t \leq U_t\} \quad (13)$$

$$AIW_{80} = \frac{1}{N} \sum_{t \in \mathcal{T}} (U_t - L_t) \quad (14)$$

Winkler-бал (інтервальний критерій) для рівня $1 - \alpha$:

$$Winkler_{1-\alpha} = \frac{1}{N} \sum_{t \in \mathcal{T}} \left[(U_t - L_t) + \frac{2}{\alpha} (L_t - y_t) 1\{y_t < L_t\} + \frac{2}{\alpha} (y_t - U_t) 1\{y_t > U_t\} \right] \quad (15)$$

Результати експерименту

Дослідження виконано на погодинному графіку навантаження існуючого промислового підприємства хімічної промисловості з відомими наперед коваріатами (виробничий розклад/номінали, календар, погода). Модельний підхід є орієнтованим на виробничий графік: споживання декомпонується на детерміновану базу b_t , розраховану з графіка роботи обладнання, та стохастичний залишок r_t , для якого навчаються точкові моделі та моделі квантильного прогнозування (LightGBM і TFT) з подальшим темпоральним узгодженням MinT. Опис підходу і мотивацію використання MinT/THieF у контексті планування промислового навантаження наведено вище.

Точкові метрики на тесті (Табл. 1) свідчать про суттєву перевагу TFT над LGBM: MAE зменшено на 52.1 %, RMSE на 40.9 %, sMAPE на 52.9 %. Отже, поєднання орієнтованої на виробничий графік бази з багатогоризонтною неймережею краще відтворює нелінійні взаємодії «графік роботи обладнання-календар-погода» та довгі часові залежності, які важко повністю закодувати в ансамблях із фіксованими лагами/ковзними ознаками.

За ймовірнісними метриками (Табл. 2) до калібрування спостерігаємо діаметрально протилежні зсуви. LGBM має різке недопокриття 80 %-інтервалів ($ICP_{80} \approx 0.045$) та надто вузькі інтервали ($AIW_{80} \approx 240$), що підтверджують діаграми (Рис. 2): «тонкі» інтервали системно промахуються під час

півків/запусків. Натомість TFT дає помірне недопокриття ($ICP_{80} \approx 0.731$) із відносно широкими інтервалами ($AIW_{80} \approx 3964$), тобто поводить себе обережніше, але ще не номінально калібровано. Після калібрування (CQR – для LGBM, γ -звуження – для TFT) покриття для обох моделей у наведеному прикладі стало наближено $ICP_{80} \approx 0.8$. Отже, наявне достатнє покриття фактичного графіка та помірні інтервали.

Таблиця 1 – Точкові метрики прогнозу навантаження

	MAE	RMSE	sMAPE
LGBM	1950.93	2241.72	0.04
TFT	933.904	1323.92	0.019
$\Delta\%$	-52.13	-40.941	-52.922

Таблиця 2 – Ймовірнісні метрики прогнозу навантаження до та після калібрування

	$Pinball_{10}$		$Pinball_{50}$		$Pinball_{90}$		$Winkler_{80}$		ICP_{80}		AIW_{80}	
	до	після	до	після	до	після	до	після	до	після	до	після
LGBM	842.752	329.795	566.289	566.289	231.204	245.542	10739.6	5753.36	0.045	0.746	240.059	3234.37
TFT	353.286	330.864	609.722	609.722	332.758	260.652	6860.44	5915.16	0.731	0.746	3963.57	3351.08

Вхідний набір містить виміри споживання лише для найпотужніших споживачів і не охоплює дрібні/допоміжні навантаження (вентилятори, приводи, HVAC тощо), які сумарно створюють істотний фон. Через це «детермінована база» b_t є заниженою, а її відсутній компонент фактично переноситься у залишок r_t . Неповна специфікація бази підвищує дисперсію r_t і спотворює інтервальні оцінки: дерева, треновані на «вужчому» наборі предикторів, схильні до недопокриття (як у LGBM-до), тоді як TFT частково компенсує латентні ефекти за рахунок уваги/керування потоками, але потребує окремого калібрування ширини. Це узгоджується з графіками (Рис. 2) та числовими показниками (Табл. 2). Практичний висновок: щоб підвищити як точність, так і каліброваність інтервалів, варто донаситити b_t (включити більше обладнання/операційні режими) або додати проксі-ознаки для «неінструментованих» споживачів; альтернативно застосувати ієрархічну/латентну регуляризацию залишку.

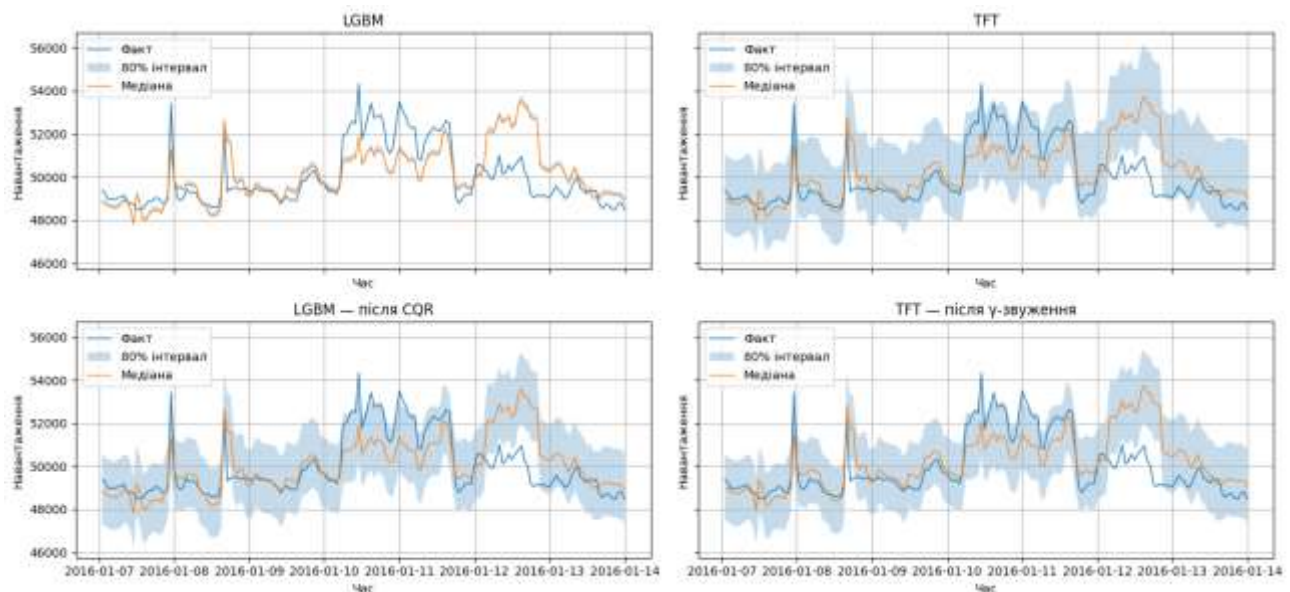


Рисунок 2 – Діаграми прогнозів навантаження до та після калібрування

З огляду на $Pinball$ -втрати, після калібрування спостерігаються очікувані зрушення на краях розподілу, а саме для LGBM $Pinball$ -втрата для квантилю 0.1 знижується з 842.8 до 329.8, а $Pinball$ -втрата для квантилю 0.9 з 231.2 до 245.5 (перерозподіл маси знизу/зверху, консистентний із розширенням інтервалів). У свою чергу для TFT $Pinball$ -втрата для квантилю 0.9 зменшується з 332.8 до 260.7 при майже сталому $Pinball$ -втрата для квантилю 0.5 (609.7), що типово для γ -звуження, а саме звуження інтервалів із помірним впливом на медіану. Сукупно це підтверджує коректність обраної процедури калібрування та напрям подальшої оптимізації.

Висновки.

У роботі запропоновано орієнтовану на графік роботи обладнання архітектуру середньострокового прогнозування навантаження промислового підприємства, яка поєднує детерміновану базу від графіка обладнання з режимною сегментацією PELT, стохастичним моделюванням залишку і темпоральним узгодженням MinT/THieF. Ринковий чинник інтегровано на рівні вибору горизонту та оперативного календаря прогнозу узгоджено з регламентом торгівлі та «закриттям воріт» та на рівні навчання ведеться за вартісно-зваженою функцією втрат, яка відображає асиметрію бізнес-ризиків небалансів у пікові та «критичні» години. До калібрування спостерігались крайові зсуви покриття, які усунено методами CQR та γ -звуження, що дало наближене до номіналу покриття 80 % і суттєве зменшення ширини інтервалів. Подальші дослідження буде зосереджено на «*decision-centric*» оцінюванні економічної вартості прогнозів: моделюванні процесів закупівлі (РДД, РДН/ВДР та врахування небалансів) під різними моделями та схемами калібрування, а також на аналізі чутливості до тарифів небалансів та різницею між ціною продажу та купівлі на різних сегментах. Додаткову увагу варто приділити розширенню набору відомих наперед коваріатів і забезпеченню крос-секційної когерентності.

Список використаної літератури

1. Про ринок електричної енергії: Закон України від 13.04.2017 № 2019-VIII : станом на 09.04.2025 URL: <https://zakon.rada.gov.ua/laws/show/2019-19#Text> (дата звернення: 10.08.2025).
2. Comprehensive review of load forecasting with emphasis on intelligent computing approaches / H. Wang та ін. *Energy Reports*. 2022. Т. 8. С. 13189–13198. URL: <https://doi.org/10.1016/j.egy.2022.10.016> (дата звернення: 10.08.2025).
3. Pinheiro M. G., Madeira S. C., Francisco A. P. Short-term electricity load forecasting—A systematic approach from system level to secondary substations. *Applied Energy*. 2023. Т. 332. С. 120493. URL: <https://doi.org/10.1016/j.apenergy.2022.120493> (дата звернення: 10.08.2025).
4. Forecasting Uncertainty in Electricity Smart Meter Data by Boosting Additive Quantile Regression / S. Ben Taieb та ін. *IEEE Transactions on Smart Grid*. 2016. Т. 7, № 5. С. 2448–2455. URL: <https://doi.org/10.1109/tsg.2016.2527820> (дата звернення: 10.08.2025).
5. A review of distribution network applications based on smart meter data analytics / C. L. Athanasiadis та ін. *Renewable and Sustainable Energy Reviews*. 2024. Т. 191. С. 114151. URL: <https://doi.org/10.1016/j.rser.2023.114151> (дата звернення: 10.08.2025).
6. Electricity load forecasting: a systematic review / I. K. Nti та ін. *Journal of Electrical Systems and Information Technology*. 2020. Т. 7, № 1. URL: <https://doi.org/10.1186/s43067-020-00021-8> (дата звернення: 10.08.2025).
7. Про затвердження Правил ринку: Постанова від 14.03.2018 № 307 : станом на 30.07.2025 URL: <https://zakon.rada.gov.ua/laws/show/v0307874-18#Text> (дата звернення: 10.08.2025).
8. Про затвердження Правил ринку «на добу наперед» та внутрішньодобового ринку: Постанова від 14.03.2018 № 308 : станом на 23.02.2024 URL: <https://zakon.rada.gov.ua/laws/show/v0308874-18#Text> (дата звернення: 10.08.2025).
9. Про затвердження Правил управління обмеженнями та Порядку розподілу пропускної спроможності міждержавних перетинів: Постанова від 03.04.2020 № 763 : станом на 01.02.2022 URL: <https://zakon.rada.gov.ua/laws/show/v0763874-20#Text> (дата звернення: 10.08.2025).
10. Medium and Long-term Load Intelligent Forecasting Method Based on the Comprehensive and Multiple Early Warning Indicators / Y. Wang та ін. *Journal of Physics: Conference Series*. 2020. Т. 1449. С. 012124. URL: <https://doi.org/10.1088/1742-6596/1449/1/012124> (дата звернення: 10.08.2025).
11. Energy Load Profile Analysis on Machine Level / H. Teiwes та ін. *Procedia CIRP*. 2018. Т. 69. С. 271–276. URL: <https://doi.org/10.1016/j.procir.2017.11.073> (дата звернення: 10.08.2025).
12. Alkadi N. E., Starke M. R., Ma O. Assessment of Industrial Load for Demand Response across Western Interconnect. Office of Scientific and Technical Information (OSTI), 2013. URL: <https://doi.org/10.2172/1111445> (дата звернення: 10.08.2025).
13. Forecast reconciliation: A geometric view with new insights on bias correction / A. Panagiotelis та ін. *International Journal of Forecasting*. 2021. Т. 37, № 1. С. 343–359. URL: <https://doi.org/10.1016/j.ijforecast.2020.06.004> (дата звернення: 10.08.2025).
14. De Livera A. M., Hyndman R. J., Snyder R. D. Forecasting Time Series With Complex Seasonal Patterns Using Exponential Smoothing. *Journal of the American Statistical Association*. 2011. Т. 106, № 496. С. 1513–1527. URL: <https://doi.org/10.1198/jasa.2011.tm09771> (дата звернення: 10.08.2025).
15. Pya N., Wood S. N. Shape constrained additive models. *Statistics and Computing*. 2014. Т. 25, № 3. С. 543–559. URL: <https://doi.org/10.1007/s11222-013-9448-7> (дата звернення: 10.08.2025).
16. Zimmermann M., Ziel F. Efficient mid-term forecasting of hourly electricity load using generalized additive models. *Applied Energy*. 2025. Т. 388. С. 125444. URL: <https://doi.org/10.1016/j.apenergy.2025.125444> (дата звернення: 10.08.2025).
17. Chen T., Guestrin C. XGBoost. *KDD '16: The 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, м. San Francisco California USA. New York, NY, USA, 2016. URL: <https://doi.org/10.1145/2939672.2939785> (дата звернення: 10.08.2025).

18. G. Ke et al. LightGBM: A Highly Efficient Gradient Boosting Decision Tree, NeurIPS, 2017
19. L. Prokhorenkova, G. Gusev, A. Vorobev, A. V. Dorogush, i A. Gulin, «CatBoost: unbiased boosting with categorical features», 2017, arXiv. doi: 10.48550/ARXIV.1706.09516.
20. Meinshausen, N. Quantile Regression Forests. *Journal of Machine Learning Research* 7 (2006) 983–999.
21. S. Bai, J. Z. Kolter, i V. Koltun, «An Empirical Evaluation of Generic Convolutional and Recurrent Networks for Sequence Modeling», 2018, arXiv. doi: 10.48550/ARXIV.1803.01271.
22. B. N. Oreshkin, D. Carpo, N. Chapados, i Y. Bengio, «N-BEATS: Neural basis expansion analysis for interpretable time series forecasting», 2019, arXiv. doi: 10.48550/ARXIV.1905.10437.
23. C. Challu, K. G. Olivares, B. N. Oreshkin, F. Garza, M. Mergenthaler-Canseco, i A. Dubrawski, «N-HiTS: Neural Hierarchical Interpolation for Time Series Forecasting», 2022, arXiv. doi: 10.48550/ARXIV.2201.12886.
24. B. Lim, S. O. Arik, N. Loeff, i T. Pfister, «Temporal Fusion Transformers for Interpretable Multi-horizon Time Series Forecasting», 2019, arXiv. doi: 10.48550/ARXIV.1912.09363.
25. Y. Nie, N. H. Nguyen, P. Sinthong, i J. Kalagnanam, «A Time Series is Worth 64 Words: Long-term Forecasting with Transformers», 2022, arXiv. doi: 10.48550/ARXIV.2211.14730.
26. R. Killick, P. Fearnhead, i I. A. Eckley, «Optimal detection of changepoints with a linear computational cost», arXiv, 2011, doi: 10.48550/ARXIV.1101.1438.
27. C. Truong, L. Oudre, N. Vayatis. “ruptures: change point detection in Python”, JMLR, 2020.
28. Adams & MacKay, 2007
29. S. Wickramasuriya, G. Athanasopoulos, R. Hyndman. “Optimal forecast reconciliation for hierarchical and grouped time series through trace minimization (MinT)”, JASA, 2019.
30. Wickramasuriya S. L., Athanasopoulos G., Hyndman R. J. Optimal Forecast Reconciliation for Hierarchical and Grouped Time Series Through Trace Minimization. *Journal of the American Statistical Association*. 2018. T. 114, № 526. C. 804–819. URL: <https://doi.org/10.1080/01621459.2018.1448825> (дата звернення: 10.08.2025).
31. Forecasting with temporal hierarchies / G. Athanasopoulos та ін. *European Journal of Operational Research*. 2017. T. 262, № 1. C. 60–74. URL: <https://doi.org/10.1016/j.ejor.2017.02.046> (дата звернення: 10.08.2025).
32. Hong T., Fan S. Probabilistic electric load forecasting: A tutorial review. *International Journal of Forecasting*. 2016. T. 32, № 3. C. 914–938. URL: <https://doi.org/10.1016/j.ijforecast.2015.11.011> (дата звернення: 10.08.2025).
33. Koenker R., Bassett G. Regression Quantiles. *Econometrica*. 1978. T. 46, № 1. C. 33. URL: <https://doi.org/10.2307/1913643>.
34. Quantile and Probability Curves Without Crossing. *Econometrica*. 2010. T. 78, № 3. C. 1093–1125. URL: <https://doi.org/10.3982/ecta7880> (дата звернення: 10.08.2025).

V. Kalinchyk¹, Cand. Sc. (Eng.), Assoc. Prof, ORCID 0000-0003-4028-0185

O. Kopchykov¹, Ph. D. student, ORCID 0000-0001-7885-4055

¹National Technical University of Ukraine «Igor Sikorsky Kyiv Polytechnic Institute»

PROBABILISTIC FORECASTING OF INDUSTRIAL LOAD: LIGHTGBM AND TFT WITH MARKET CONSTRAINTS

In this work, we present an equipment-schedule-oriented approach to medium-term industrial load forecasting that decomposes consumption into a deterministic baseline profile constructed from equipment operating schedules and a stochastic residual modeled using LightGBM (LGBM) and the Temporal Fusion Transformer (TFT). Regime shifts are detected via PELT (Pruned Exact Linear Time) change-point segmentation, while multi-horizon quantile forecasts are enforced to be non-crossing and coherent across temporal aggregation levels through MinT/THieF reconciliation. Market constraints are integrated by selecting forecasting horizons aligned with the gate-closure rules for bilateral contracts and the day-ahead market, and by employing a cost-weighted pinball loss that reflects asymmetric imbalance costs during critical hours. On a hold-out test set, TFT outperforms LGBM in point-forecast accuracy. Applying Conformal Quantile Regression (CQR) to LGBM and γ -shrinkage (gamma-shrinkage) to TFT improves empirical coverage to the nominal level and reduces interval widths. The proposed approach is directly applicable to procurement planning under the rules of the Ukrainian electricity market and is consistent with decision-centric evaluation of the economic value of forecasts.

Keywords: electric load forecasting, industrial enterprise, quantile forecasting, temporal reconciliation, electricity market factors.

Надійшла: 02.09.2025

Received: 02.09.2025