

3 РОЗРОБКА АЛГОРИТМУ ПРОГНОЗУВАННЯ ЕЛЕКТРОСПОЖИВАННЯ

Прогнозування електроспоживання є складним завданням, яке має багато невизначеностей, зокрема, залежність від різних факторів, таких як сезонність, погодні умови, промисловість і так далі. Основна ціль даного розділу – розробити алгоритм прогнозування електроспоживання, заснований на вейвлет-аналізі, який би був ефективним і точним.

Вейвлет-аналіз дозволяє фільтрувати шумові сигнали і виявляти тренди та коливання в часових рядах даних про електроспоживання. Вейвлет-розкладання і реконструкція зменшують нестационарність часових рядів даних, тим самим підвищуючи точність прогнозу.

У цьому розділі ми пропонуємо алгоритм, який використовує вейвлет-аналіз для прогнозування часових рядів електроспоживання, використовуючи дискретні вейвлети.

3.1 Визначення вимог до алгоритму

Алгоритм прогнозування електроспоживання, заснований на вейвлет-аналізі, повинен відповідати ряду вимог, щоб забезпечити його ефективність та точність.

1. *Надійність:* Алгоритм повинен відповідати високим стандартам надійності. Це означає, що він повинен стабільно працювати при різних умовах та відповідати очікуванням при використанні різних наборів даних.

2. *Точність прогнозування:* Одним з основних критеріїв ефективності алгоритму є його здатність точно прогнозувати майбутнє електроспоживання. Він повинен мати мінімальну помилку прогнозування.

3. *Швидкість обробки:* Алгоритм повинен бути здатний обробляти великі об'єми даних за короткий час. Це особливо важливо для реального часу прогнозування, де швидкість обробки може бути вирішальною.

4. *Здатність адаптуватися до нових даних:* Алгоритм повинен бути здатний враховувати нові дані та відповідно до них коригувати свої прогнози.

5. *Універсальність:* Алгоритм повинен бути застосовним до різних типів часових рядів даних, не обмежуючись лише даними про електроспоживання.

6. *Економічність:* Розробка та впровадження алгоритму повинні бути економічно виправданими. Це означає, що витрати на його розробку та впровадження повинні бути виправдані отриманими перевагами.

3.2 Проектування алгоритму прогнозування

Розробка алгоритму прогнозування на основі вейвлетів включає кілька ключових кроків:

1. Збираються дані про електроспоживання з різних джерел та проводиться їх передобробка. Це може включати заповнення пропущених значень, нормалізацію та інші методи передобробки даних.

2. Вхідні дані перетворюються за допомогою вейвлет-перетворення. Це дозволяє виділити ключові особливості даних, що значно полегшує наступний етап аналізу. Він ефективно відфільтровує шум і допомагає виявити тенденцію зміни та коливання даних.

3. Аналізуються отримані в результаті вейвлет-перетворення дані. Це може включати в себе виявлення закономірностей, виявлення аномалій, виявлення трендів та інші методи аналізу даних.

4. На основі аналізу даних створюється модель прогнозування. Ця модель використовує виявлені закономірності для прогнозування майбутнього електроспоживання.

5. Після створення моделі прогнозування проводиться її оцінка. Якщо результати прогнозування не відповідають очікуванням, модель вдосконалюється за допомогою додаткового аналізу даних та налаштування параметрів. Або застосувати функцію прогнозування заново, таким чином навчити нейронну мережу.

Згідно із запропонованими кроками розробки алгоритму прогнозування , була наведена блок схема на рис. 3.1.



Рис. 3.1 Блок схема запропонованого алгоритму

3.3 Реалізація алгоритму

Для реалізації цього алгоритму написано код на Python для прогнозування часових рядів: TimeSeriesPredictor та ARIMAPredictor. Перший використовує нейронні мережі, а другий – модель ARIMA. Ці класи містять

методи для підготовки даних, навчання моделі, прогнозування, обчислення метрик та візуалізації результатів.

Клас `WaveletTransform` виконує вейвлет-перетворення сигналу та дозволяє отримати апроксимацію та деталізацію на різних рівнях розкладання. `ScrollableFrame` — це віджет, який реалізує прокрутку у `tkinter`, бібліотеці Python для створення графічного користувацького інтерфейсу.

Програма має графічний інтерфейс, який наведений на рис. 3.2, дозволяє користувачам взаємодіяти з вейвлет-перетвореннями та прогнозувати майбутні дані за допомогою двох моделей: нейронної мережі та моделі ARIMA.

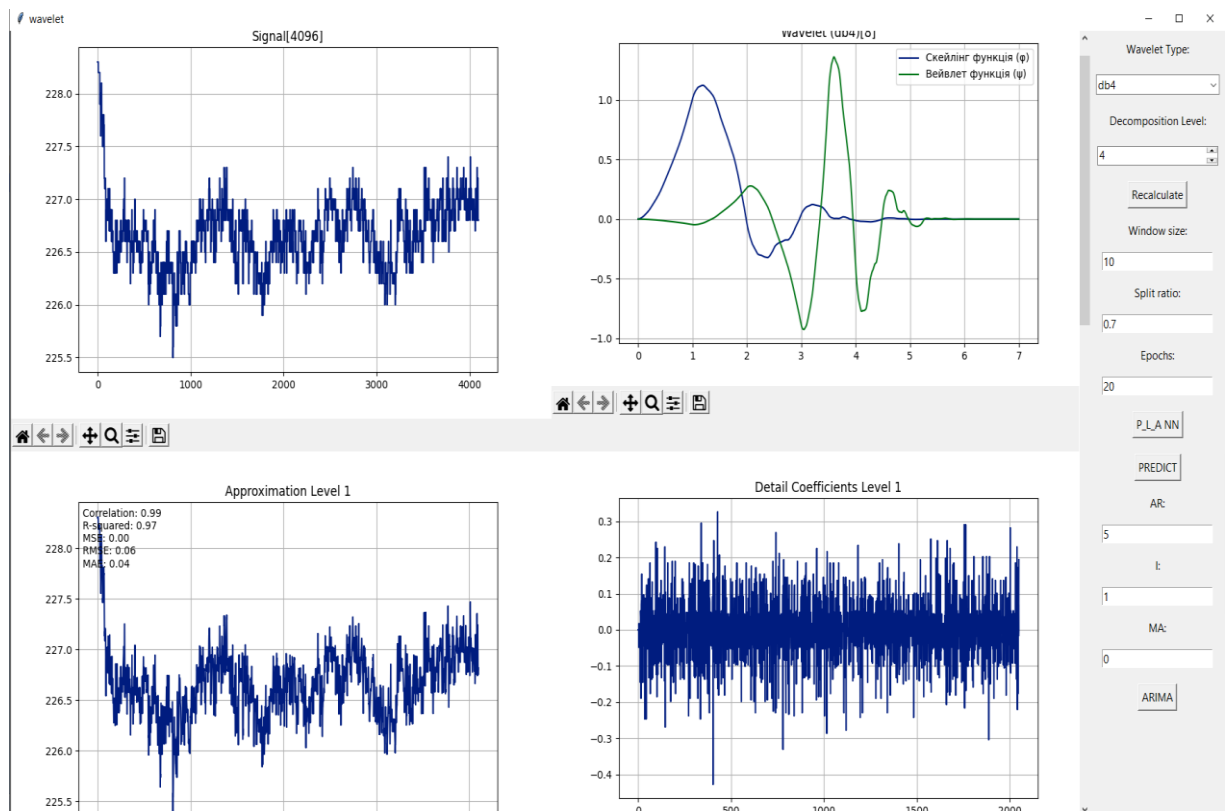


Рис. 3.2 Інструмент для вейвлет-аналізу з графічним інтерфейсом користувача

В програмі використовується найпростіша форма нейронної мережі, яка називається повністю пов'язаною нейронною мережею або нейронною мережею зі зворотним зв'язком FNN. Ось основні характеристики цієї мережі:

1. *Архітектура*. Мережа складається з трьох типів шарів: вхідний шар, один прихований шар і вихідний шар. Кожен шар складається з нейронів, які повністю пов'язані з нейронами наступного шару.

2. *Активация*. Прихований шар використовує функцію активації (Rectified Linear Unit, ReLU). Ця функція виводить вхідний сигнал, якщо він позитивний; в іншому випадку вона виводить 0. Вихідний шар не має функції активації, тому виводиться прямий вихід з нейронів.

3. *Оптимізація та втрати*. Використовується оптимізатор Adam, який є одним з найпопулярніших оптимізаторів для нейронних мереж завдяки своєму адаптивному навчанню та ефективності. Середня квадратична помилка (MSE) використовується як функція втрат, що є стандартним вибором для задач прогнозування.

4. *Кількість epoch*. Кількість епох визначає, скільки разів алгоритм навчання пройде через весь навчальний набір даних. Це контролюється змінною epochs.

Ця мережа є одношаровою, оскільки має лише один прихований шар. Мережі з більшою кількістю прихованих шарів називаються глибокими нейронними мережами.

Вейвлет-перетворення є важливим інструментом для аналізу часових рядів і сигналів, який використовується для виявлення внутрішньої структури сигналу.

Програма виконує дискретне вейвлет-перетворення, використовуючи вейвлети Добеші (db), Симлети(sym), Койвлети(coif) і вейвлет Хаара.

Рівень декомпозиції вказує на глибину розкладання вейвлету. Кожен наступний рівень розкладання ділить сигнал на дві частини: апроксимацію (низькочастотні компоненти) і деталі (високочастотні компоненти). Апроксимація може бути далі розкладена на наступні рівні. Рівень декомпозиції визначається залежно від довжини вхідного сигналу та довжини вейвлет-фільтра.

Метод `get_approximation` повертає апроксимацію сигналу на вказаному рівні. Це низькочастотна частина сигналу.

Метод `get_details` повертає деталі сигналу на вказаному рівні. Це високочастотна частина сигналу, яка містить деталі та шум.

Загалом, використання вейвлет-перетворення дозволяє аналізувати сигнал на різних масштабах і виділяти важливі особливості сигналу.

3.4 Підготовка даних для тестування

Моє основне вікно програми слугує як центральний інтерфейс для обробки даних. Однією з ключових функцій цього вікна є можливість імпортувати даних у форматі csv або xlsx. Цей файл містить важливі дані, такі як величини напруги, струму, активної потужності, реактивної потужності, а також фази.

В програмі передбачено спеціальне вікно налаштувань для імпорту даних, його можемо бачити на рис. 3.3. Дане вікно гарантує безпроблемне завантаження даних з файлу в таблицю.

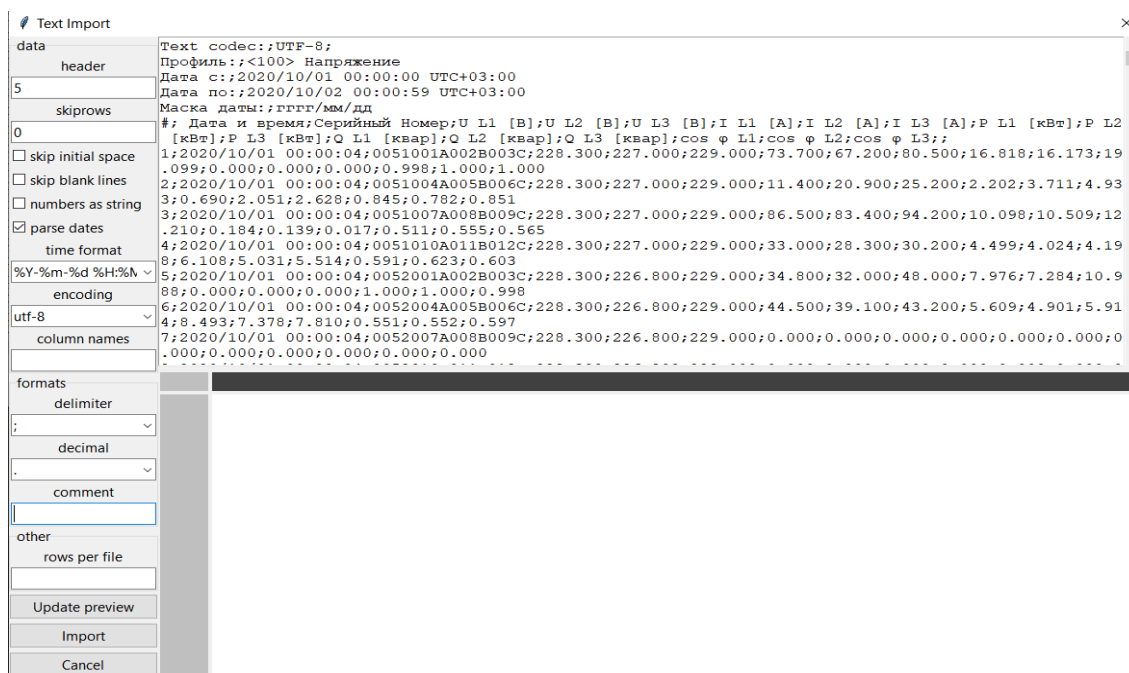


Рис. 3.3 Вікно налаштувань для імпорту даних

Для коректної роботи з даними потрібно внести декілька налаштувань: відмінити включення перших п'яти рядків, оскільки вони не потрібні в таблиці програми; встановити формат дати та часу; визначити роздільник даних, який використовується у файлі, та вимкнути коментування.

Кожен з цих параметрів має відповідну колонку в csv файлі, що дозволяє програмі чітко звертатись до значень величин, і потім аналізувати їх. Для такої взаємодії маємо зручний графічний , який показано на рис. 3.4. Імпортовані дані будуть використанні для подальшої побудови графіків і застосування до них методів прогнозування.

#	Дата и время	Серийный Номер	U L1 [В]	U L2 [В]	U L3 [В]	I L1 [А]	I L2 [А]	I L3 [А]	P L1 [кВт]	P L2 [кВт]	P L3 [кВт]	Q L1 [квар]	Q L2 [квар]	Q L3 [квар]	cos φ L1	cos φ L2	cos φ L3
1	2020/10/01 00:00:04	0051001A002B003C	228.30	227.00	229.00	73.70	67.20	80.50	16.82	16.17	19.10	0	0	0	1	1.00	1.00
2	2020/10/01 00:00:04	0051004A005B006C	228.30	227.00	229.00	11.40	20.90	25.20	2.20	3.71	4.93	0.69	2.05	2.63	0.84	0.78	0.85
3	2020/10/01 00:00:04	0051007A008B009C	228.30	227.00	229.00	86.50	83.40	94.20	10.10	10.51	12.21	0.18	0.14	0.017	0.51	0.56	0.56
4	2020/10/01 00:00:04	0051010A011B012C	228.30	227.00	229.00	33.00	28.30	30.20	4.50	4.02	4.20	6.11	5.03	5.51	0.59	0.62	0.6
5	2020/10/01 00:00:04	0052001A002B003C	228.30	226.80	229.00	34.80	32.00	48.00	7.98	7.28	10.99	0	0	0	1.00	1.00	1
6	2020/10/01 00:00:04	0052004A005B006C	228.30	226.80	229.00	44.50	39.10	43.20	5.61	4.90	5.91	8.49	7.38	7.81	0.55	0.55	0.6
7	2020/10/01 00:00:04	0052007A008B009C	228.30	226.80	229.00	0	0	0	0	0	0	0	0	0	0	0	0
8	2020/10/01 00:00:04	0052010A011B012C	228.30	226.80	229.00	0	0	0	0	0	0	0	0	0	0	0	0
9	2020/10/01 00:00:14	0051001A002B003C	228.20	226.90	228.90	67.20	61.80	76.30	13.97	13.54	16.64	0.33	0.4	0.37	0.91	0.96	0.95
10	2020/10/01 00:00:14	0051004A005B006C	228.20	226.90	228.90	11.60	21.40	27.00	2.51	4.69	5.41	0.6	1.98	2.47	0.95	0.96	0.87
11	2020/10/01 00:00:14	0051007A008B009C	228.20	226.90	228.90	84.30	82.00	94.30	10.10	10.50	12.58	0.19	0.13	0.005	0.52	0.56	0.58
12	2020/10/01 00:00:14	0051010A011B012C	228.20	226.90	228.90	33.00	28.30	30.20	4.47	3.98	4.16	6.11	5.04	5.50	0.59	0.61	0.6
13	2020/10/01 00:00:14	0052001A002B003C	228.20	226.60	228.80	34.80	32.00	48.00	7.97	7.29	10.99	0	0	0	1.00	1.00	1.00
14	2020/10/01 00:00:14	0052004A005B006C	228.20	226.60	228.80	44.50	39.10	43.20	5.65	4.93	6.03	8.48	7.38	7.80	0.56	0.56	0.61
15	2020/10/01 00:00:14	0052007A008B009C	228.20	226.60	228.80	0	0	0	0	0	0	0	0	0	0	0	0
16	2020/10/01 00:00:14	0052010A011B012C	228.20	226.60	228.80	0	0	0	0	0	0	0	0	0	0	0	0
17	2020/10/01 00:00:24	0051001A002B003C	228.20	226.80	228.80	67.00	60.00	74.60	16.90	15.66	19.05	0	0	0	1.00	1.00	1.00
18	2020/10/01 00:00:24	0051004A005B006C	228.20	226.80	228.80	11.60	21.20	26.80	2.19	4.51	5.56	0.53	2.22	2.50	0.82	0.94	0.9
19	2020/10/01 00:00:24	0051007A008B009C	228.20	226.80	228.80	95.30	91.60	103.00	10.08	10.50	12.29	0.19	0.13	0.029	0.46	0.5	0.52
20	2020/10/01 00:00:24	0051010A011B012C	228.20	226.80	228.80	33.00	28.30	30.20	4.45	3.94	4.15	6.10	5.03	5.52	0.59	0.62	0.6
21	2020/10/01 00:00:24	0052001A002B003C	228.20	226.60	228.90	34.80	32.00	48.00	7.97	7.28	11.00	0	0	0	1.00	1.00	1.00
22	2020/10/01 00:00:24	0052004A005B006C	228.20	226.60	228.90	44.50	39.10	43.30	5.64	4.94	6.00	8.48	7.38	7.82	0.55	0.56	0.6
23	2020/10/01 00:00:24	0052007A008B009C	228.20	226.60	228.90	0	0	0	0	0	0	0	0	0	0	0	0
24	2020/10/01 00:00:24	0052010A011B012C	228.20	226.60	228.90	0	0	0	0	0	0	0	0	0	0	0	0
25	2020/10/01 00:00:34	0051001A002B003C	227.90	226.70	228.60	61.90	57.10	70.90	13.40	13.24	16.20	0.36	0	0	0.95	1.00	1
26	2020/10/01 00:00:34	0051004A005B006C	227.90	226.70	228.60	11.10	21.10	26.70	2.21	3.73	5.23	0.57	2.09	2.50	0.87	0.78	0.85
27	2020/10/01 00:00:34	0051007A008B009C	227.90	226.70	228.60	87.70	83.50	97.10	10.05	10.47	12.53	0.2	0.12	0.012	0.5	0.55	0.56
28	2020/10/01 00:00:34	0051010A011B012C	227.90	226.70	228.60	32.90	28.40	30.30	4.47	4.00	4.20	6.07	5.00	5.50	0.59	0.61	0.6
29	2020/10/01 00:00:34	0052001A002B003C	228.00	226.40	228.80	34.80	32.00	47.90	7.96	7.27	10.96	0	0	0	1.00	1.00	1
30	2020/10/01 00:00:34	0052004A005B006C	228.00	226.40	228.80	44.40	39.00	43.20	5.59	4.90	5.99	8.44	7.35	7.80	0.55	0.55	0.6
31	2020/10/01 00:00:34	0052007A008B009C	228.00	226.40	228.80	0	0	0	0	0	0	0	0	0	0	0	0
32	2020/10/01 00:00:34	0052010A011B012C	228.00	226.40	228.80	0	0	0	0	0	0	0	0	0	0	0	0
33	2020/10/01 00:00:44	0051001A002B003C	227.60	226.50	228.60	70.30	63.00	76.20	16.81	15.60	19.00	0	0	0	1.00	1.00	1.00
34	2020/10/01 00:00:44	0051004A005B006C	227.60	226.50	228.60	11.20	15.10	24.40	2.38	2.76	4.90	0.58	1.65	2.46	0.93	0.81	0.87
35	2020/10/01 00:00:44	0051007A008B009C	227.60	226.50	228.60	91.90	85.20	100.20	10.03	10.86	12.53	0.24	0.13	0.007	0.48	0.56	0.55

Рис. 3.4 Завантажені дані в головному вікні програми

Розташована у нижній лівій частині інтерфейсу, кнопка 'Plot Graph' активізує додаткове вікно діалогу. Це вікно дозволяє користувачу точно визначити кількість даних, які будуть застосовуватися при виконанні вейвлет-аналізу та прогнозуванні. Важливо відмітити, що це інструмент, який надає користувачу змогу контролювати глибину аналізу, дозволяючи керувати балансом між детальністю і обчислювальною продуктивністю.

Крім того, це вікно діалогу містить додаткові параметри налаштування, що дають користувачу можливість настроїти візуалізацію графіків відповідно до власних потреб. Можемо бачити на рис. 3.5 наступне вікно програми, яке включає в себе вибір кольорової схеми, типу лінії, масштабу та інших візуальних аспектів. Ці налаштування дають користувачам гнучкість і контроль, дозволяючи їм адаптувати зображення даних до своєї конкретної ситуації або вимог.

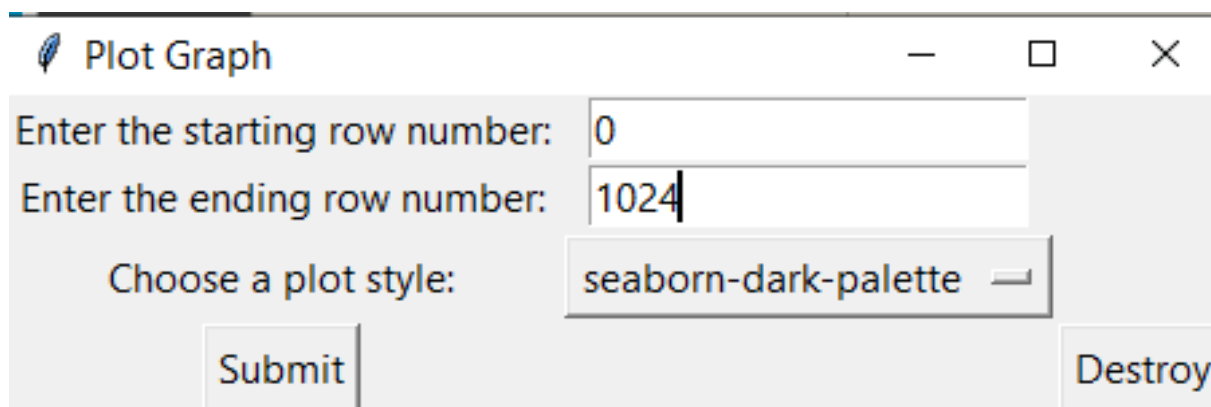


Рис. 3.5 Діалогове вікно

Загальний інтерфейсу на рис. 3.2 можемо описати так, що вікно має розмір 1800x900 пікселів. Воно поділено на дві частини: область для графіків (Scrollable Frame) та панель керування (Function Panel). Область для графіків знаходиться зліва від панелі керування і може містити декілька графіків. Графіки можна прокручувати, щоб переглянути всі доступні дані. Панель керування розташована справа від області для графіків. Вона містить наступні елементи:

1. Вибір типу вейвлету: комбінований список дозволяє вибрати тип вейвлету з передбачених значень.

2. Вибір рівня декомпозиції: «спінбокс» дозволяє вибрати рівень декомпозиції з передбачених значень.

3. Кнопка «Recalculate»: коли цю кнопку натиснуто, програма заново виконує вейвлет-перетворення з вибраними параметрами.

4. Поля вводу для «Window size», «Split ratio», та «Epochs»: ці поля вводу дозволяють користувачу ввести параметри для нейромережі.

5. Кнопка «P_L_A NN»: коли цю кнопку натиснуто, програма виконує прогнозування використовуючи нейронну мережу.

6. Кнопка «PREDICT» теж виконує прогнозування, вже на заздалегіть навченій моделі нейромережі.

7. Поля вводу для «AR», «I», та «MA»: ці поля вводу дозволяють користувачу ввести параметри для моделі ARIMA.

8. Кнопка «ARIMA» запуск моделі.

Вищеописаний інтерфейс дозволяє користувачеві вибирати параметри вейвлет-перетворення та моделі ARIMA, переглядати візуалізації сигналів та результатів моделювання, і робити прогнози на основі введених даних. Також на графіках апроксимованих сигналів, в правому верхньому кутку розміщенні метрики для оцінки схожості апроксимованого сигналу з основним.

Основні параметри нейромережі:

1. signal: це вхідний сигнал, який використовується для тренування та прогнозування моделі.

2. window_size: цей параметр визначає кількість попередніх точок часового ряду, що використовуються для прогнозування наступної точки.

3. split_ratio: це відношення, яке використовується для розподілу вхідних даних на тренувальний та тестовий набори.

4. epochs: це кількість повних проходів через весь набір тренувальних даних, що використовується під час навчання моделі.

5. `model`: це структура нейронної мережі. Вона складається з двох шарів: прихованого шару з 10 нейронами і активаційною функцією 'relu', і вихідного шару з одним нейроном.

6. `optimizer`: це метод оптимізації, який використовується для оновлення ваг моделі під час навчання. В цій моделі використовується оптимізатор 'adam'.

7. `loss`: це функція втрат, яка використовується для обчислення помилки прогнозування моделі. В цій моделі використовується квадратична помилка 'mse' (mean squared error).

8. `metrics`: це метрики, які використовуються для оцінки ефективності моделі. Вони включають кореляцію, коефіцієнт детермінації (R-squared), середньоквадратичну помилку (MSE), кореневу середньоквадратичну помилку (RMSE) та середню абсолютну помилку (MAE).

3.5 Метрики для оцінки похибки

Для об'єктивної оцінки ефективності моделей машинного навчання необхідно використовувати кілька метрик якості. Ці метрики надають нам кількісну інформацію про продуктивність моделі і допомагають виявити, наскільки добре модель передбачає або класифікує дані. В залежності від типу задачі (регресія, класифікація, кластеризація тощо) ми можемо вибрати різні метрики для оцінки нашої моделі. В цьому розділі ми зосередимося на таких ключових метриках для регресії, як кореляція, R-квадрат, середньоквадратична помилка (MSE), коренева середньоквадратична помилка (RMSE), та середня абсолютна помилка (MAE). Детальне розуміння цих метрик необхідне для вироблення правильних висновків про продуктивність та коректність моделей машинного навчання.

3.5.1. Кореляція

Ця метрика вимірює ступінь залежності між двома змінними. Це може допомогти визначити, як одна змінна змінюється при зміні іншої. Коефіцієнт

кореляції Pearsona – це один з найбільш поширених методів для вимірювання лінійної кореляції між змінними і варіює від -1 (ідеальна негативна кореляція) до +1 (ідеальна позитивна кореляція).

$$x^m = (x_1, \dots, x_m), y^m = (y_1, \dots, y_m)$$

$$r_{xy} = \frac{\sum_{i=1}^m (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^m (x_i - \bar{x})^2 \sum_{i=1}^m (y_i - \bar{y})^2}}, \quad (3.1)$$

де \bar{x}, \bar{y} - вибіркові середні x^m і y^m .

3.5.2 R-квадрат (R-squared)

Ця метрика використовується для оцінки якості моделі в контексті її здатності передбачати набір даних. R-квадрат, або коефіцієнт детермінації, показує частку варіативності вихідної змінної, яку можна пояснити введеннями моделі. Значення R^2 зазвичай лежать у діапазоні від 0 до 1. Значення 0 означає, що незалежна змінна не пояснює жодної варіації у залежній змінній, тоді як значення 1 означає, що незалежна змінна повністю пояснює зміни у залежній змінній.

$$R^2 = 1 - \frac{D[y|x]}{D[y]} = 1 - \frac{\sigma_y^2}{\sigma_y^2}, \quad (3.2)$$

де $D[y] = \sigma_y^2$ - дисперсія випадкової величини y , $D[y|x] = \sigma^2$ - умовна дисперсія залежних змінних.

3.5.3 Середньоквадратична помилка (MSE)

Це середнє значення квадратів помилок. Помилка – це різниця між передбаченням моделі і реальним значенням. MSE є популярною метрикою для вимірювання якості моделі прогнозування, особливо для моделей, що використовують неперервні вхідні та вихідні змінні.

Зазвичай, чим менше значення MSE, тим краща точність моделі, оскільки це означає, що помилки прогнозів менші. Однак, необхідно пам'ятати, що

залишки моделі можуть мати різний характер, і наслідки прогнозування можуть бути різними в залежності від конкретного додаткового застосування.

$$MSE = \frac{1}{n} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2, \quad (3.3)$$

де n – кількість спостережень, Y_i – i -тє спостереження, \hat{Y}_i – i -та прогнозована змінна.

3.5.4 Коренева середньоквадратична помилка (RMSE)

Це квадратний корінь від середньоквадратичної помилки. Вона широко використовується, оскільки має ті ж міри, що і вихідна змінна (дані), що дозволяє її легко інтерпретувати. RMSE краще працює, коли великі помилки є більш критичними.

Зазвичай, чим менше значення RMSE, тим краща точність моделі, оскільки це означає, що середнє відхилення прогнозування менша. Однак, так само як і в разі MSE, необхідно аналізувати залишки моделі та їх характер для точнішого розуміння результатів.

$$RMSE = \sqrt{MSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2}, \quad (3.4)$$

де n – кількість спостережень, Y_i – i -тє спостереження, \hat{Y}_i – i -та прогнозована змінна.

3.5.5 Середня абсолютна помилка (MAE)

Це середнє абсолютне значення помилок. Помилка – це різниця між передбаченням моделі і реальним значенням. MAE є метрикою, яка легко інтерпретується, оскільки вона має ті ж міри, що і вихідна змінна. Вона є надійною метрикою для моделей прогнозування, особливо для тих, що використовують неперервні вхідні та вихідні змінні. MAE менш чутлива до великих помилок, ніж MSE або RMSE.

Зазвичай, чим менше значення MAE, тим краща точність моделі, оскільки це означає, що середнє абсолютне відхилення між прогнозованими і спостережуваними значеннями менша. Однак, необхідно пам'ятати, що залишки моделі можуть мати різний характер, і наслідки прогнозування можуть бути різними в залежності від конкретного додаткового застосування

$$MSE = \frac{1}{n} \sum_{i=1}^n |Y_i - \hat{Y}_i|, \quad (3.5)$$

де n – кількість спостережень, Y_i – i -те спостереження, \hat{Y}_i – i -та прогнозована змінна.

4 АНАЛІЗ ТА ПОРІВНЯННЯ МЕТОДІВ ПРОГНОЗУВАННЯ

Аналіз результатів буде виконуватись на основі визначених метрик, що дозволить забезпечити формалізованість та об'єктивність результатів. Хоча на перший погляд результати можуть здаватися подібними, без використання метрик неможливо отримати об'єктивну оцінку отриманих результатів.

Під час практичної роботи було встановлено, що для надійного навчання та прогнозування нейронної мережі важливо забезпечити достатній обсяг вхідних даних для часового ряду. Конкретно, оптимальний обсяг складає 2048 значень.

На рис. 4.1 представлено числовий ряд без застосування вейвлет-аналізу. З його допомогою можна відзначити, що без застосування вейвлет-аналізу числовий ряд не дозволяє отримати бажані значення метрик.

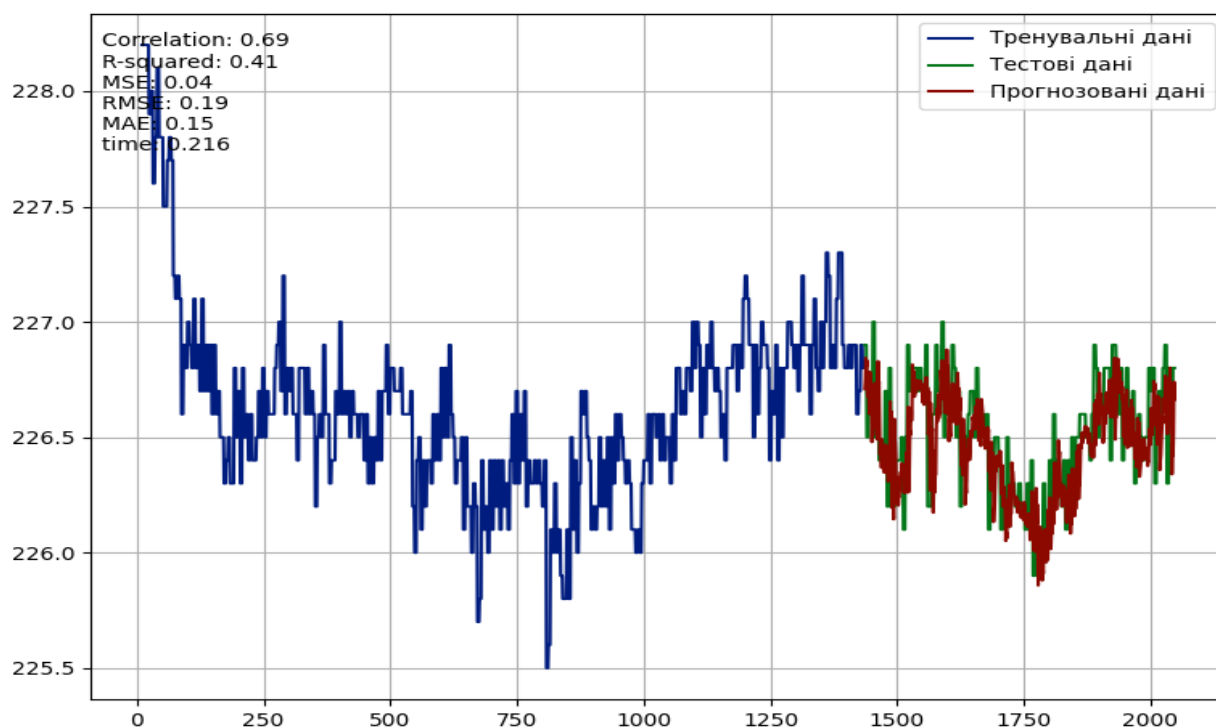


Рис. 4.1 Вхідний числовий ряд з перших 2048 числових значень

Наступні рис. 4.2 , 4.3 представляють розрахунок метрик для всіх вейвлетів Добеші, доступних в програмі. Дані графіки були побудовані з використанням вже навченої нейронної мережі, що дозволило отримати результати за лічені секунди. Були використані базові параметри: 38 вейвлетів Добеші, декомпозиція на четвертий рівень, обчислення всіх метрик, вікно введення для нейронної мережі з 10 значеннями, та 70% від основного числового ряду використано як навчальні дані. При деякій періодичності спостерігались максимальні значення метрик.

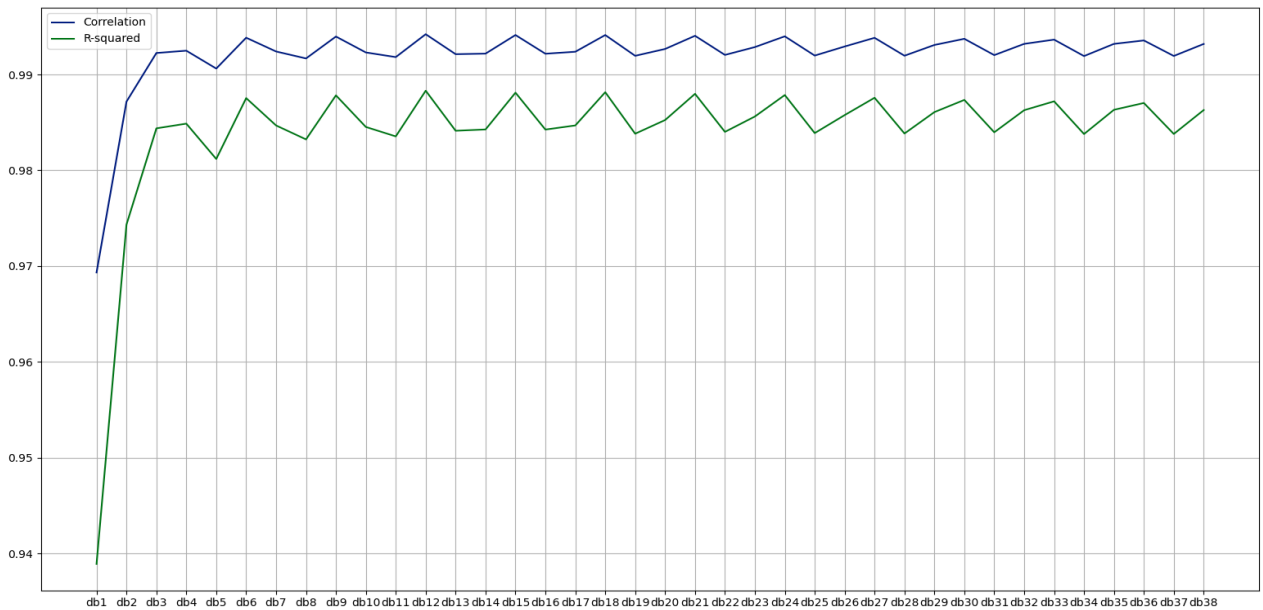


Рис. 4.2 Кореляція і R-квадрат для вейвлетів Добеші

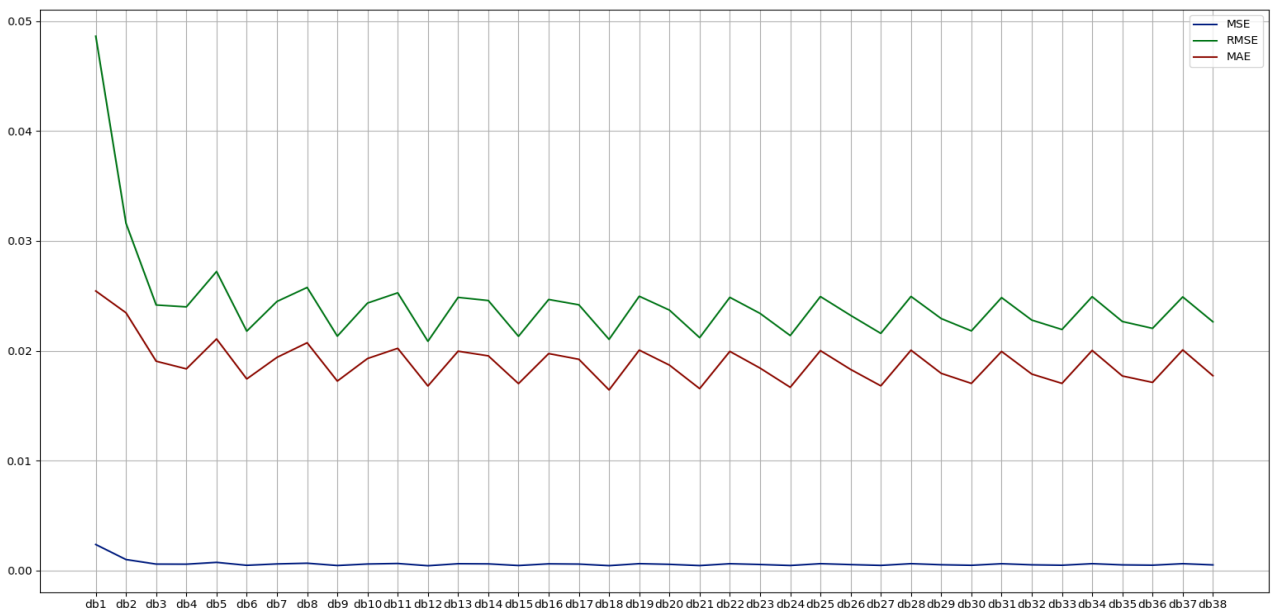


Рис. 4.3 MSE, RMSE, MAE для вейвлетів Добеші

На рис. 4.4 , 4.5 нижче наведені графіки, побудовані з використанням симлетів. Хоча початкові дані для цих графіків є тими ж, що й для Добеші вейвлетів, кількість використаних вейвлетів у цьому випадку становить 20.

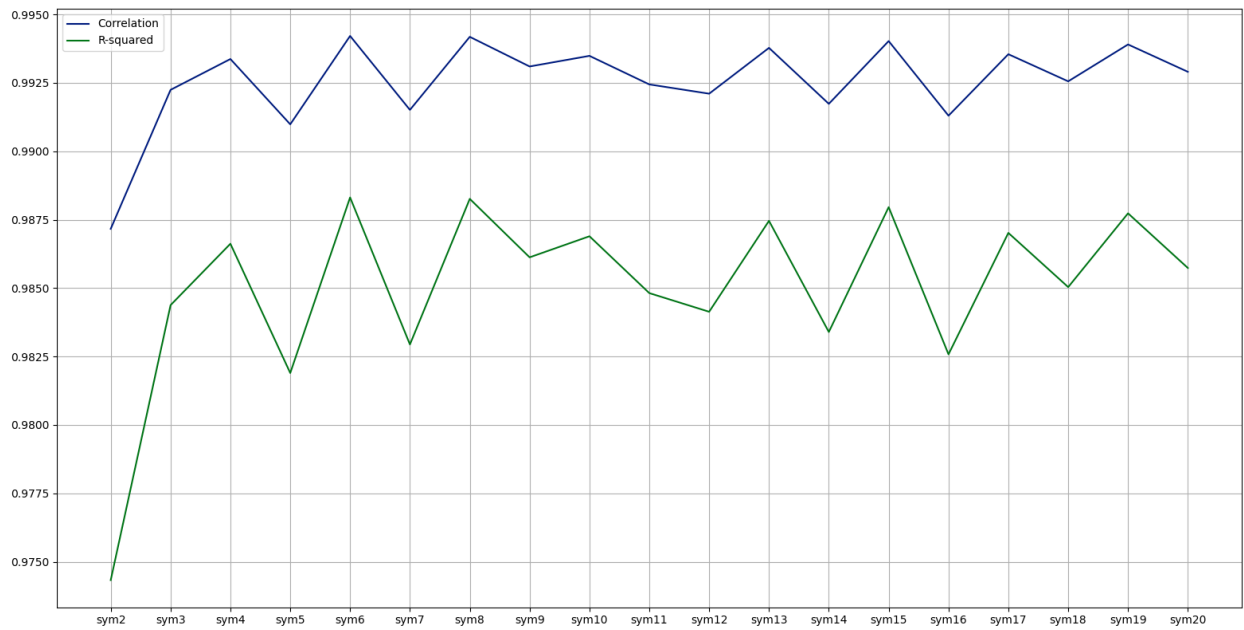


Рис. 4.4 Кореляція і R-квадрат для Симлет вейвлетів

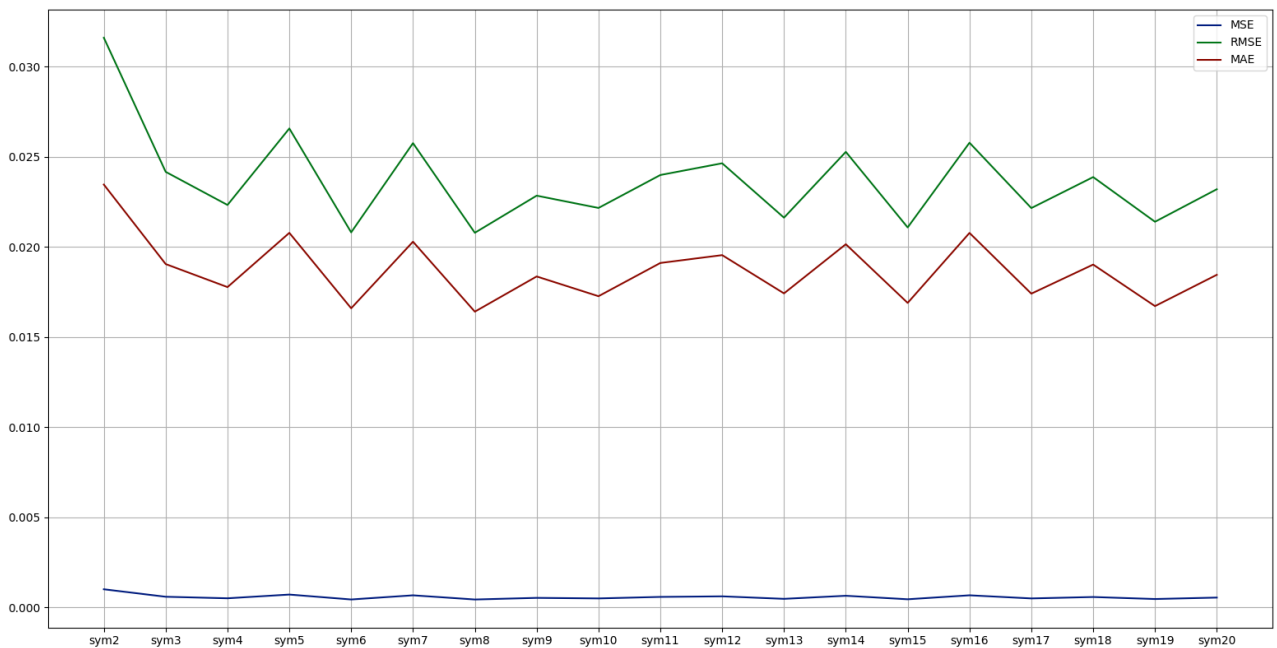


Рис. 4.5 MSE, RMSE, MAE для Симлет вейвлетів

Наступні рис. 4.6 , 4.7 також представлені метрики, але вже з використанням вейвлетів Койфлет. Початкові налаштування залишаються незмінними. Варто зазначити, що на всіх попередньо представлених графіках метрик можна простежити деяку періодичність у найкращих значеннях для окремих вейвлетів.

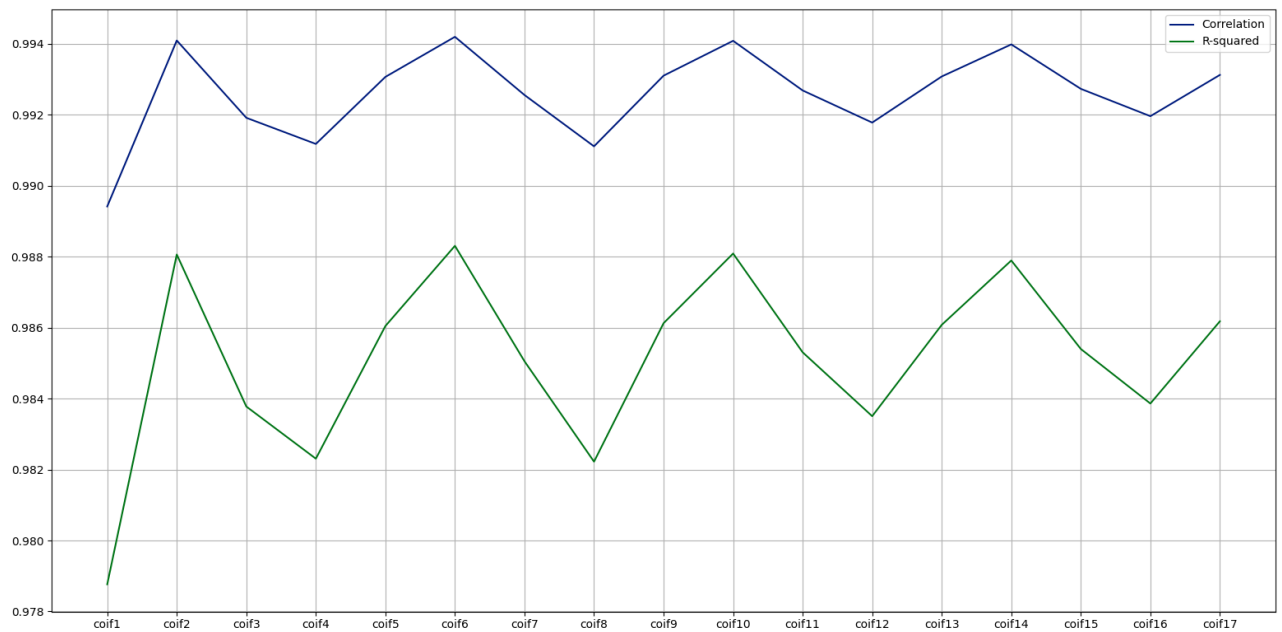


Рис. 4.6 Кореляція і R-квадрат для Койфлет вейвлетів

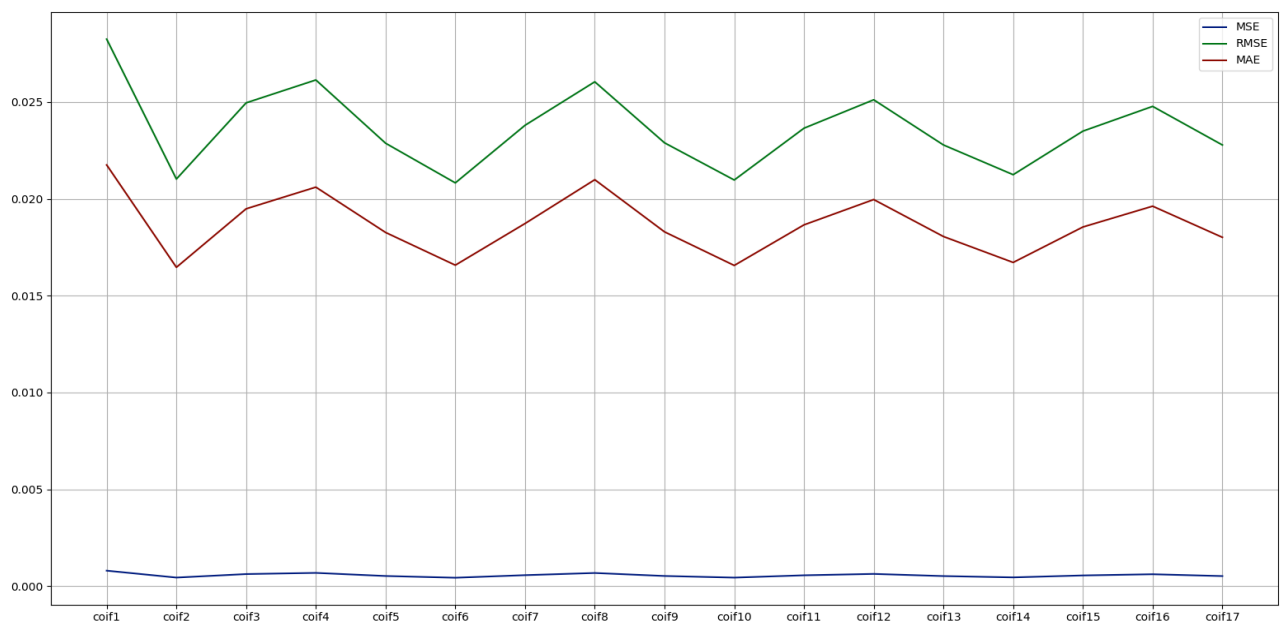


Рис. 4.7 MSE, RMSE, MAE для Койфлет вейвлетів

При розрахунку метрик для числового ряду, що складається з 69168 значень, ми отримуємо кращі результати. В даному випадку, нейронна мережа має більше даних для навчання, що стосуються конкретного числового ряду, відтак, вона забезпечує вищу якість прогнозування. В таких обставинах, підбір

конкретного вейвлету в рамках одного сімейства не має особливого значення. Отримані результати представлені на рис. 4.8 , 4.9.

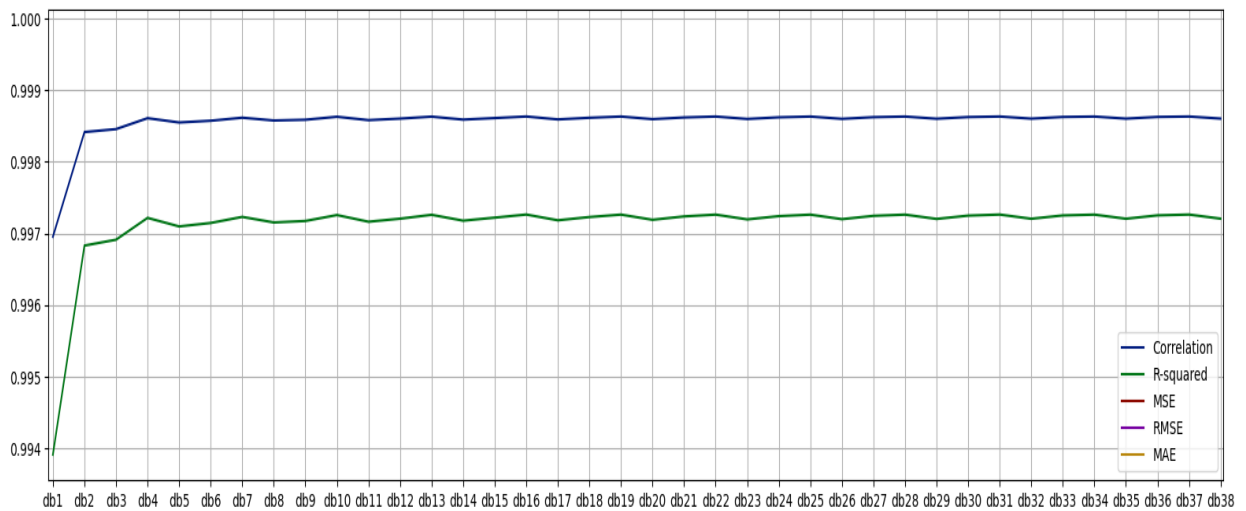


Рис. 4.8 Кореляція і R-квадрат для вейвлетів Добеші(69168 значень)

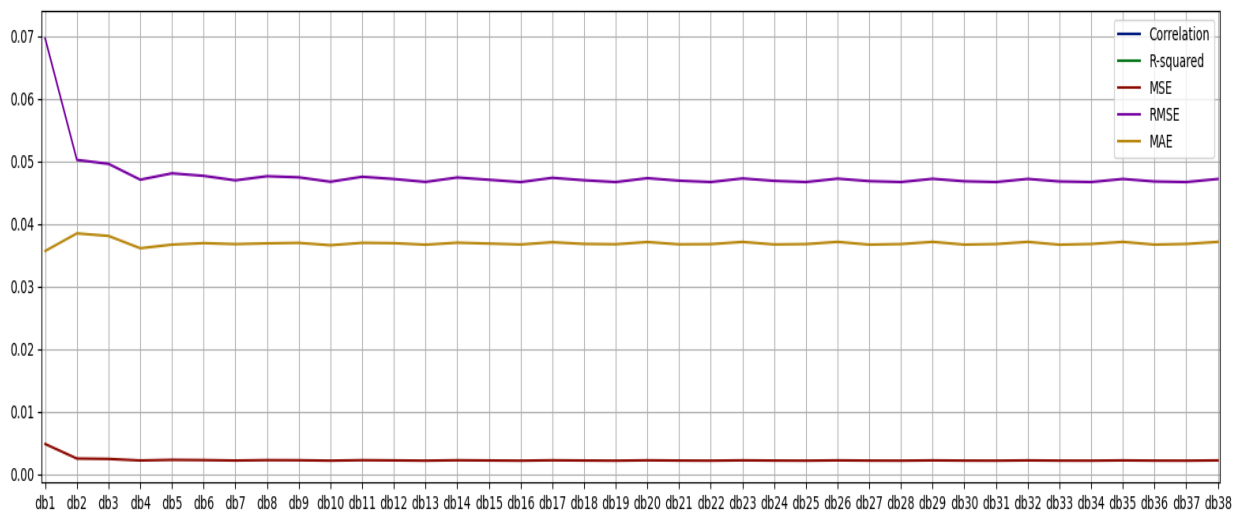


Рис. 4.9 MSE, RMSE, MAE для вейвлетів Добеші(69168 значень)

На наступному рис.4.10 демонстрація застосування моделі ARIMA в двох різних варіантах. У першому варіанті, модель використовує тестові дані як основу для прогнозування, що є аналогічно до методу навчання нейронної мережі. Тобто, кожне наступне значення прогнозується на основі вже відомих тестових даних. На рис. 4.11 представлений другий варіант роботи моделі, де

для прогнозування наступних значень використовується попереднє прогнозоване значення, отримане за допомогою цієї моделі. Як результат, фактично отримуємо медіану числового ряду. Модель ARIMA вимагає значного часу на обчислення. Наприклад, для обробки відносно простого числового ряду знадобилося 234 секунди, як це відображено на графіку. На противагу, нейронна мережа обробляє такі дані набагато швидше. Ще швидшим є процес отримання результатів, якщо використовується попередньо навчена та збережена в окремому файлі нейронна мережа.

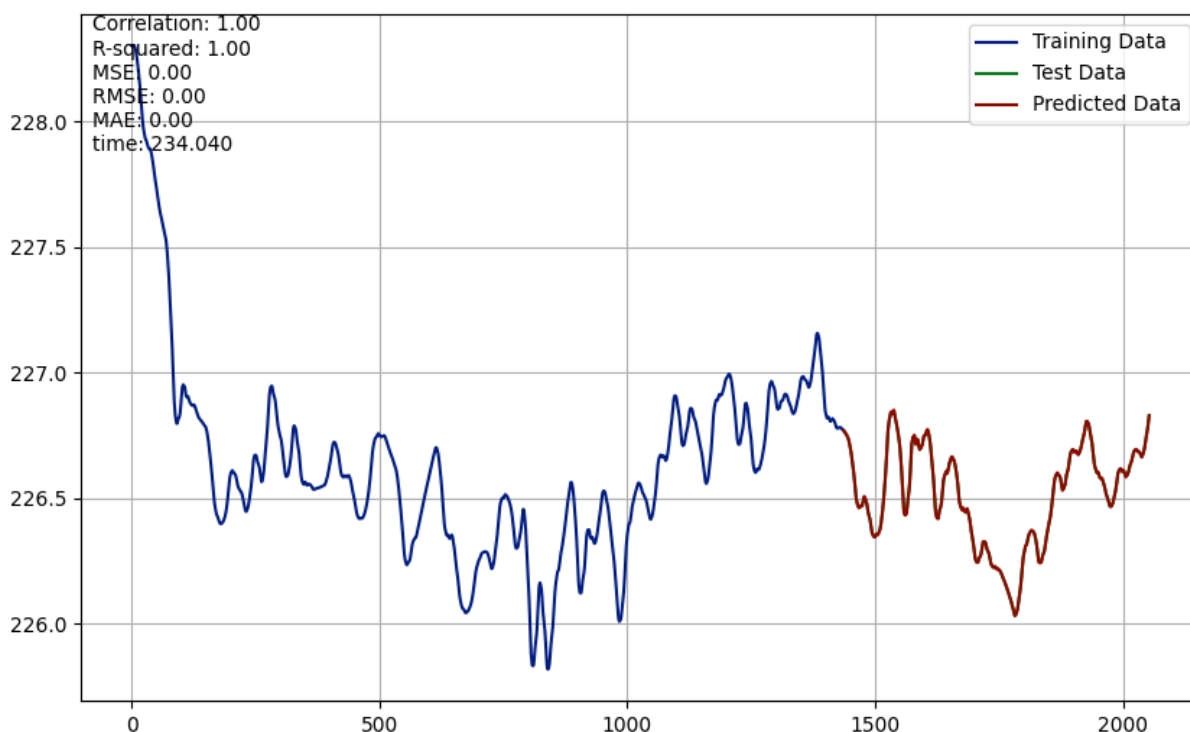


Рис. 4.10 Прогнозування ARIMA перший варіант

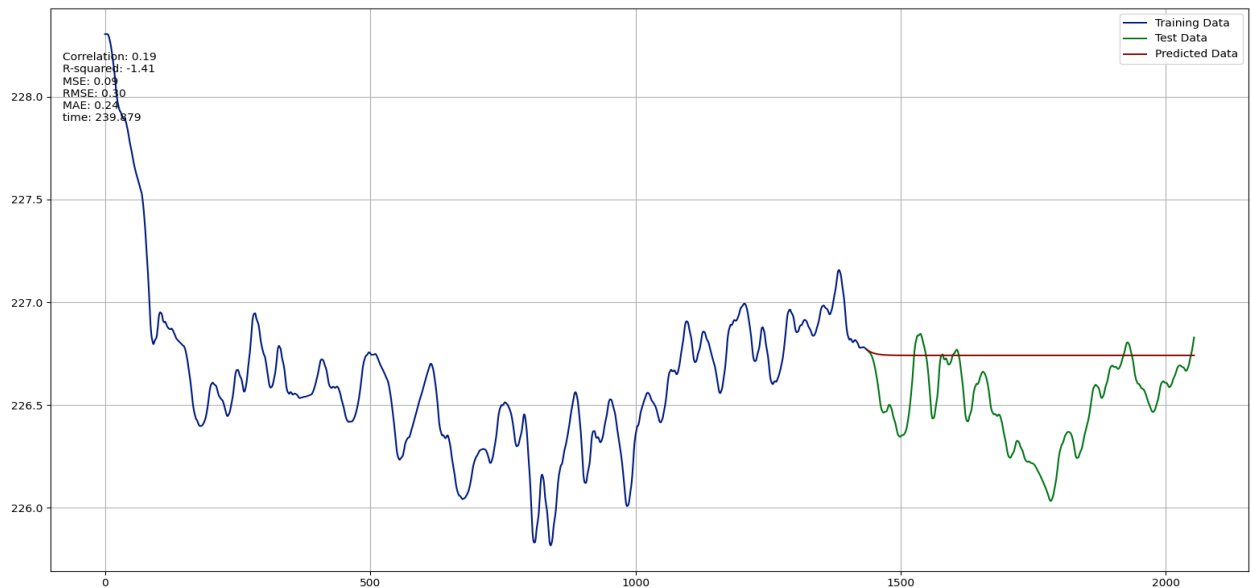


Рис. 4.11 Прогнозування ARIMA другий варіант

Рис. 4.12 акцентує увагу на часі, потрібному для прогнозування числового ряду, що складається з 69168 значень. Розрахунок тривав 54 секунди, що було зумовлено великою кількістю даних та двадцятьма епохами навчання.

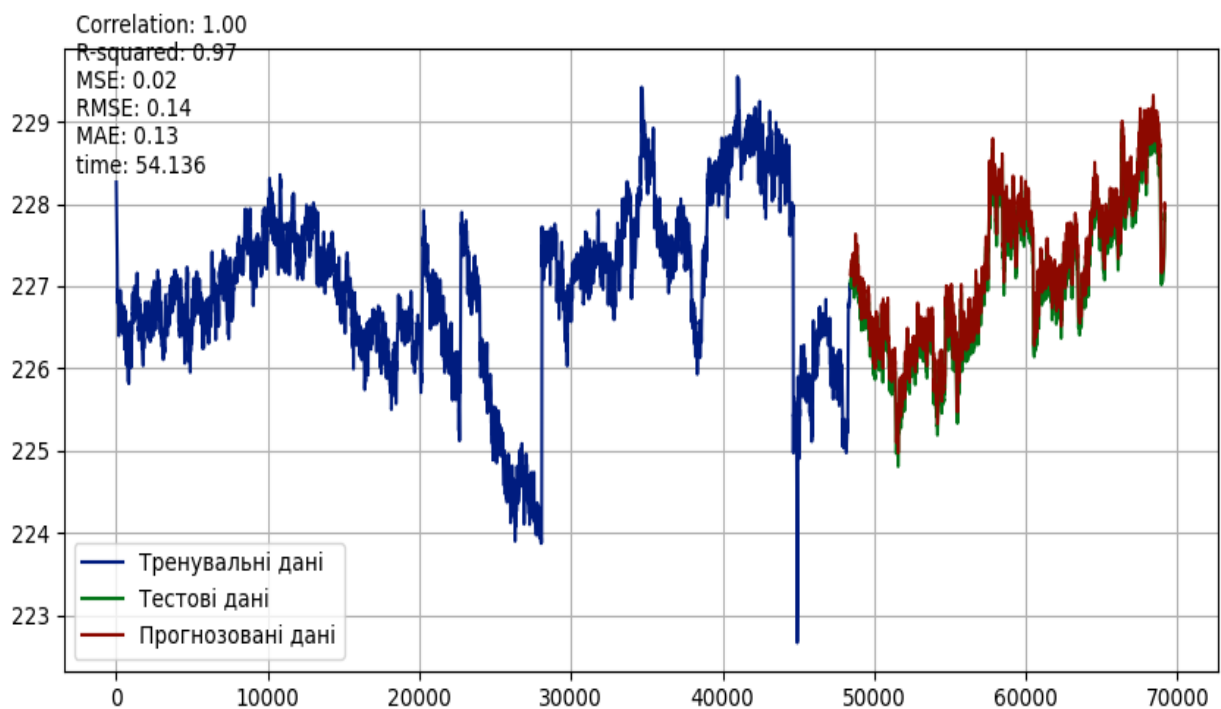


Рис. 4.12 Результат навчання нейронної мережі

Рис. 4.13 демонструє результати вже навченої нейронної мережі, яка здатна обробити таку ж кількість даних лише за одну секунду. Значущих відмінностей в метриках не спостерігається, навіть більше, результати вийшли кращими.

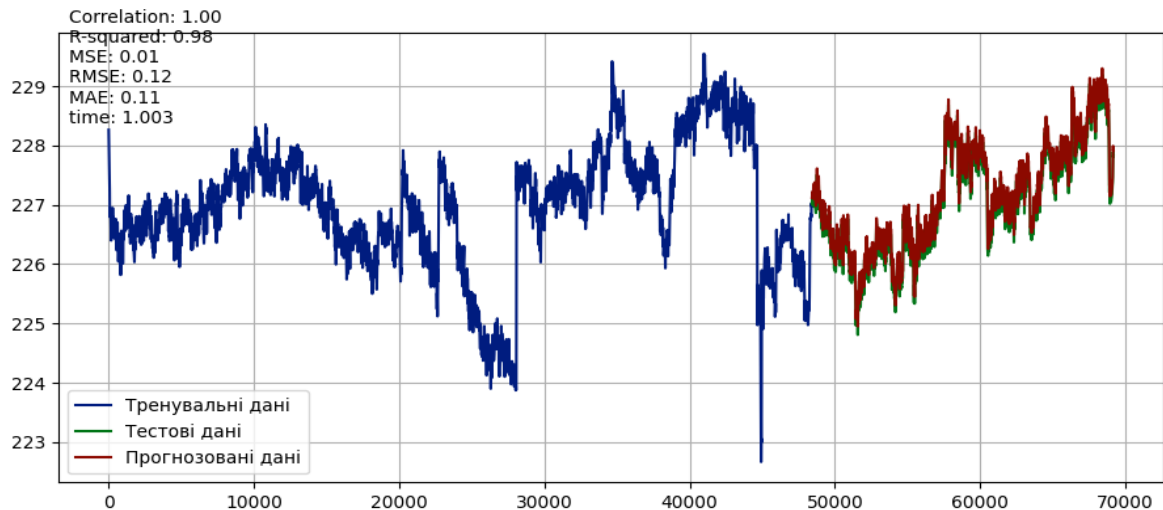


Рис. 4.13 Прогнозування навченої нейронної мережі

Для ілюстрації використано інший набір значень напруги для вже навченої нейронної мережі. Рис. 4.14 чітко демонструє, що нейронна мережа ефективно впоралася з прогнозуванням, що відображено у показниках метриках.

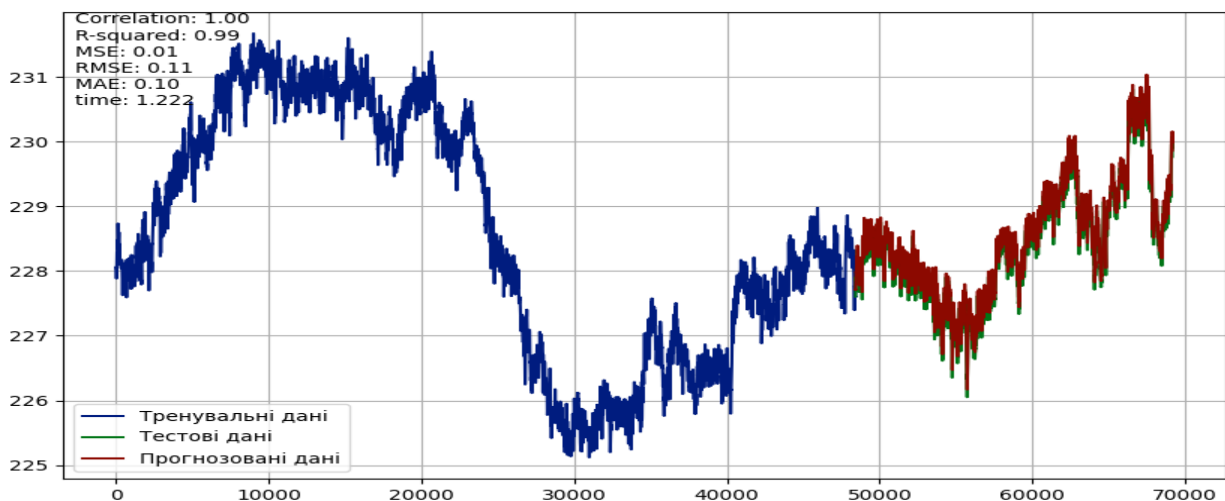


Рис. 4.14 Прогнозування навченої нейронної мережі, для інших значень

Рис. 4.15 ілюструє причину використання кількох метрик. На рисунку чітко видно, що кореляція майже дорівнює одиниці, проте прогнозований числовий ряд лише віддалено нагадує тестові дані, з якими його порівнює нейронна мережа. Середньоквадратична похибка дорівнює 0.13, хоча в ідеальному випадку вона має бути нульовою.



Рис. 4.15 Наведена причина застосування кількох метрик

Висновки до четвертого розділу

У даному розділі було розглянуто використання різних типів вейвлетів для прогнозування електроспоживання, включаючи вейвлети Добеші, Симлети та Койфлети. Було показано, що використання вейвлет-перетворення може допомогти у покращенні якості прогнозування, особливо при великій кількості даних.

Було проведено порівняння різних метрик для оцінки якості прогнозування, включаючи кореляцію, R-квадрат, MSE, RMSE та MAE. Результати показали, що нейронна мережа, навчена на великому наборі даних,

забезпечує вищу якість прогнозування, незалежно від вибору конкретного вейвлету в рамках одного сімейства.

Також було показано, що алгоритм ARIMA може бути використаний для прогнозування електроспоживання, хоча він може бути менш ефективним при великій кількості даних.

В цілому, цей розділ демонструє, як різні методи аналізу часових рядів та машинного навчання можуть бути використані для прогнозування електроспоживання, і підкреслює важливість вибору правильного методу в залежності від обсягу та характеру даних.