

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
СУМСЬКИЙ ДЕРЖАВНИЙ УНІВЕРСИТЕТ

Кафедра прикладної математики та моделювання складних систем

Допущено до захисту
Завідувач кафедри ПМ та МСС
_____ Коплик І.В.

«__» _____ 20__ р.

КВАЛІФІКАЦІЙНА РОБОТА

на здобуття освітнього ступеня «магістр»

спеціальність 113 «Прикладна математика»

освітньо-професійна програма «Прикладна математика»

тема роботи:

**«ПРОГНОЗУВАННЯ СПОЖИВАННЯ ЕЛЕКТРОЕНЕРГІЇ:
ПОРІВНЯЛЬНИЙ АНАЛІЗ МОДЕЛЕЙ ЧАСОВИХ РЯДІВ»**

Виконавець

студент факультету ЕЛІТ

Холявка Євген Петрович _____

Науковий керівник

доцент, к.е.н.

Маринич Тетяна Олександрівна _____

Суми – 2020 р.

СУМСЬКИЙ ДЕРЖАВНИЙ УНІВЕРСИТЕТ

Факультет **електроніки та інформаційних технологій**
Кафедра **прикладної математики та моделювання складних систем**

Рівень вищої освіти другий
(перший (бакалавр) або другий (магістр))

Галузь знань **11 Математика та статистика**

Спеціальність **113 Прикладна математика**

Освітня програма **освітньо-професійна «Прикладна математика»**

ЗАТВЕРДЖУЮ
Завідувач кафедри ПМ та МСС
Коплик І.В. _____
«___» _____ 20__ р.

ЗАВДАННЯ

НА КВАЛІФІКАЦІЙНУ РОБОТУ ЗДОБУВАЧУ ВИЩОЇ ОСВІТИ

Холявка Євген Петрович

(прізвище, ім'я, по батькові)

1. Тема роботи «Прогнозування споживання електроенергії: порівняльний аналіз моделей часових рядів»

Керівник роботи Маринич Тетяна Олександрівна доцент, к. е. н.
(прізвище, ім'я, по батькові, науковий ступінь, вчене звання, посада)

затверджено наказом по факультету ЕЛІТ від «__» __ 20__ р. № _____

2. Термін подання роботи студентом «__» _____ 20__ р.

3. Вихідні дані до роботи провести аналіз даних часових рядів споживання електроенергії. За основу взято дані споживання електроенергії головним корпусом Сумського державного університету. Провести моделювання часових рядів. Оцінити адекватність створених моделей прогнозування. Зробити порівняльний аналіз моделей часових рядів

4. Зміст розрахунково-пояснювальної записки (перелік питань, що їй належить розробити) Провести аналітичний огляд літератури споживання і прогнозування споживання електричної енергії. Огляд методів і підходів прогнозування часових рядів. Побудувати математичних і комп'ютерних моделей прогнозування часових рядів. Порівняльний аналіз моделей прогнозування. Висновки.

5. Перелік графічного матеріалу:

Рис. 3.2 – Діаграма розмаху залежності споживання електроенергії від дня тижня.

Рис. 3.4 – Декомпозиція(розкладання) часового ряду.

Рис. 3.6 – Діаграма розсіювання залежності часових рядів.

Рис. 3.7 – Кореляційна матриця часового ряду.

Рис. 3.10 – Діагностичні тести залишків моделей.

Рис. 3.21 – Графік залежності прогнозованого значення від реального.

6. Консультанти до проекту (роботи), із значенням розділів проекту, що стосується їх

Розділ	Прізвище, ініціали та посада консультанта	Підпис, дата	
		Завдання видав	Завдання прийняв
1.	Маринич Т. О., доцент, к.е.н.		
2.	Маринич Т. О., доцент, к.е.н.		
3.	Маринич Т. О., доцент, к.е.н.		

7. Дата видачі завдання «___»_____20__р.

КАЛЕНДАРНИЙ ПЛАН

№ п/п	Назва етапів кваліфікаційної роботи	Термін виконання роботи	Примітка
1.	Аналіз поточного стану світової електроенергетики	02.11-8.11	
2.	Аналіз існуючих моделей прогнозування часових рядів.	09.11-15.11	
3.	Побудова моделей прогнозування часових рядів	16.11-06.12	
4.	Порівняльний аналіз моделей прогнозування	07.12-13.12	

Здобувач вищої освіти

(підпис)

Холявка Є. П.
(прізвище та ініціали)

Керівник роботи

(підпис)

Маринич Т. О.
(прізвище та ініціали)

РЕФЕРАТ

Магістерська робота складається з 82 сторінок, містить 24 рисунка та 33 посилання на літературу.

Об'єкт дослідження – щоденні показники споживання електроенергії Сумським державним університетом.

Предмет дослідження – визначення факторів, які впливають на обсяги споживання електричної енергії та пошук оптимальних високоточних моделей для прогнозування.

Мета роботи – є розробка алгоритму на основі часових рядів для прогнозування споживання електричної енергії на рівні підприємства, установи, організації.

Планування і прогнозування енергоспоживання є важливим завданням як для держави в цілому, електрогенеруючих та розподільчих компаній, так і для підприємств та організацій, які є споживачами електроенергії. Своєчасне отримання інформації щодо майбутнього навантаження системи дозволяє оптимізувати електроспоживання та забезпечити його безперебійність. Якщо раніше прогнозування обсягів споживання електроенергії здійснювалося звичайною екстраполяцією попередніх показників, то зараз активно впроваджуються технології машинної обробки та моделювання на підставі даних лічильників. При цьому актуальним стає вимірювання можливого впливу зовнішніх факторів на обсяги електроспоживання, таких як погодні умови, календарні ефекти, політика енергоефективності тощо.

В роботі використовуються статистичні моделі прогнозування часових рядів, запропоновані алгоритми провідних компаній і моделі машинного навчання.

Оцінено прогнозну якість кожної моделі досліджено фактори впливу на споживання, обрано моделі, які забезпечують мінімальну похибку прогнозування.

ЗМІСТ

ВСТУП	6
1. СПОЖИВАННЯ ЕЛЕКТРОЕНЕРГІЇ ЯК ОДНА З СКЛАДОВИХ СТАЛОГО РОЗВИТКУ	7
1.1 Сучасний стан світової електроенергетики	7
1.2 Прогнозування споживаної електроенергії: виклики та можливості	9
1.3 Постановка завдання дослідження.....	13
2. ОГЛЯД МЕТОДІВ І ПІДХОДІВ АНАЛІЗУ ЧАСОВИХ РЯДІВ	15
2.1 Статистичний аналіз часових рядів та їх компонентів	15
2.2 Математичні моделі часових рядів	19
2.2.1 Авторегресійні моделі прогнозування	19
2.2.2 Моделі експоненціального згладжування	25
2.2.3 Моделі машинного навчання	27
2.3 Перевірка адекватності моделей	32
2.4 Перевірка прогнозної якості моделей.....	35
3. ПРАКТИЧНА РЕАЛІЗАЦІЯ.....	38
3.1 Аналіз вхідних даних.....	39
3.2 Модель на основі регресійного аналізу	44
3.3 Моделі експоненціального згладжування	48
3.4 Моделі на основі машинного навчання	52
3.4.1 Модель Prophet	52
3.4.2 Рекурентні нейронні мережі.....	55
3.5 Порівняльний аналіз моделей прогнозування	57
ВИСНОВКИ.....	61
ПЕРЕЛІК ДЖЕРЕЛ ІНФОРМАЦІЇ	62
ДОДАТКИ.....	67
Додаток А.....	67
Додаток Б	70
Додаток В.....	73
Додаток Г	76

ВСТУП

Актуальність. Планування і прогнозування енергоспоживання є важливим завданням як для держави в цілому, електрогенеруючих та розподільчих компаній, так і для підприємств та організацій, які є споживачами електроенергії. Своєчасне отримання інформації щодо майбутнього навантаження системи дозволяє оптимізувати електроспоживання та забезпечити його безперебійність. Якщо раніше прогнозування обсягів споживання електроенергії здійснювалося звичайною екстраполяцією попередніх показників, то зараз активно впроваджуються технології машинної обробки та моделювання на підставі даних лічильників. При цьому актуальним стає вимірювання можливого впливу зовнішніх факторів на обсяги електроспоживання, таких як погодні умови, календарні ефекти, політика енергоефективності тощо.

Метою роботи є розробка алгоритму прогнозування часових рядів для споживання електричної енергії на рівні підприємства, установи, організації.

Об'єкт дослідження. Споживання електричної енергії підприємством, установою, організацією.

Предмет дослідження. Визначення факторів, які впливають на обсяги споживання електричної енергії та пошук оптимальних високоточних моделей для прогнозування.

Практична цінність. Розроблена методика прогнозування споживання електричної енергії дає можливість знизити фінансові витрати підприємства, установи, організації за рахунок визначення впливових факторів та підвищення прогнозної якості моделей.

1. СПОЖИВАННЯ ЕЛЕКТРОЕНЕРГІЇ ЯК ОДНА З СКЛАДОВИХ СТАЛОГО РОЗВИТКУ

1.1 Сучасний стан світової електроенергетики

Доступ до енергії є ключовою опорою для добробуту людей, економічного розвитку та подолання бідності. Забезпечення достатнього доступу для всіх є постійним та нагальним викликом для глобального розвитку.

Загалом, щонайменше 1,6 мільярда людей – одна четверта населення світу – в даний час живуть без електроенергії, і ця кількість майже не змінилася в абсолютних показниках з 1970 року. Економіка, країн що розвиваються, стикаються з подвійною енергетичною проблемою у 21 столітті: задоволення потреб мільярдів людей, які все ще не мають доступу до основних сучасних енергетичних послуг, одночасно беручи участь у глобальному переході до чистих низьковуглецевих енергетичних систем. І історичні темпи прогресу у напрямку підвищення ефективності, декарбонізації, більшої різноманітності палива та зниження викидів забруднюючих речовин потребують значного прискорення для цього [1].

Існуючі енергетичні системи також мають значний вплив на навколишнє середовище. В історичних та сучасних енергетичних системах переважають викопні види палива (вугілля, нафта та газ), які виробляють вуглекислий газ (CO₂) та інші парникові гази – основний двигун глобальних кліматичних змін. Для того щоб досягти наших глобальних кліматичних цілей та уникнути небезпечних кліматичних змін, світ потребує значного та узгодженого переходу до екологічно чистих джерел енергії [2].

Балансуючи між розвитком та навколишнім середовищем забезпечує мету: забезпечити кожному доступ до достатньої кількості стійкої енергії для підтримання високого рівня життя.

Останні роки спостерігається глобальний перехід до відновлюваних джерел енергії, який стає дедалі дешевшим. За прогнозами, сонячна енергія буде домінуючою технологією у відновлюваному ринковому просторі [2]. У звичайному виробництві електроенергії газ, як прогнозується, сприятиме майбутньому зростанню потужностей, оскільки він здатний забезпечити швидке та чисте джерело базової потужності та швидко зростає через екологічні проблеми вугілля та ядерних технологій. Незважаючи на це зростання в газовому секторі, вугілля буде продовжувати відігравати певну роль в країнах Азії та Африки завдяки менш жорстким екологічним директивам та нормам, а також через велику кількість внутрішніх недорогих ресурсів. [1]

Поява та розвиток рішень для накопичення енергії створює умови для революції базового навантаження відновлюваної енергії, завдяки поєднанню енергії, що переривається, з швидкою та легкодоступною резервною потужністю. Відповідно, швидкозростаюча динаміка розподіленого виробництва електроенергії та поява позамеревих енергосистем чинять тиск на традиційну односпрямовану модель електроенергетичних систем та ведуть до нових ринкових можливостей [3].

У 2018 та 2019 роках Європейський Союз розпочав реалізацію пакету "Чиста енергія для всіх європейців", який складається з восьми законодавчих актів, включаючи зміни на ринку електроенергії для кращого сприяння інтеграції періодично відновлюваних джерел енергії, транскордонну співпрацю, продовжує орієнтувати електроенергетичну систему регулювання щодо ринкової конкуренції та надають споживачам більший вибір на ринку. Пакет чистої енергії встановлює правову базу для сприяння досягненню енергетичних та кліматичних цілей на 2030 рік. [4]

Пакет включає оновлення Директиви про електроенергію та Регламенту електроенергії, покликані поетапно припинити плату за потужності електростанціям, що викидають більше 550 грамів CO₂ за кіловат-годину

(ефективно видаляючи вугільні установки без технології уловлювання вуглецю з ринку потужності), і вимагатиме від країн забезпечити, щоб ринки потужності використовувались як крайній захід і щоб національне законодавство не перешкоджало транскордонному потоку електроенергії. Оновлення також гарантують споживачам нові права, спрямовані на збільшення вибору споживачів, включаючи обмеження часу, необхідного для заміни постачальника, право вимагати розумного лічильника та динамічну структуру цін, а також право на безкоштовний доступ хоча б до одного інструменту для порівняння роздрібні постачальники електроенергії [4].

1.2 Прогнозування споживаної електроенергії: виклики та можливості

Оскільки електроенергія є важливим ресурсом і важелем економічного розвитку, передбачення попиту на нього є актуальним завданням як для держави у цілому, так і для економічних суб'єктів – генеруючих та розподільчих компаній, а також бізнес-структур, комунальних установ та організацій. Перевагами прогнозування обсягів електроспоживання є забезпечення безперебійного надійного постачання, підвищення енергоефективності та оптимізація фінансових витрат.

На сьогоднішній день ринок електроенергії пред'являє все більш жорсткі вимоги до обсягів закупуваної і продаваної електроенергії, у відповідність з цим питання прогнозування є фундаментальним для фінансового планування.

Проблема прогнозування електроспоживання полягає в тому, що необхідно врахувати величезну кількість чинників, що має вплив на зміни обсягів енергоспоживання, серед яких:

– Кліматичні та погодні фактори – температура повітря, швидкість вітру, опади, тиск, видимість, рівень освітленості і вологість. Зазначені фактори

можуть впливати на короточасні та сезонні зміни споживання електроенергії. Багато з них моделюються систематичним чином, але іноді для обробки таких ситуацій, як раптові морські бризи в прибережних районах або тривала спекотна погода, передбачаються спеціальні коригувальні процедури.

– Телебачення. Популярні телевізійні програми можуть робити істотний вплив на конфігурацію вечірньої частини добових графіків споживання електричної енергії.

– Вихідні та свята. Національні свята викликають перерви в роботі автоматичних систем прогнозування електроспоживання. У деяких роботах [23] згадується про існування методики придушення помилок прогнозу, обумовлених святами, що дозволяють обійтися без звичайного переоцінювання параметрів прогнозуючої моделі. Навіть шкільні канікули можуть викликати певні зміни конфігурації добових графіків споживання електроенергії (особливо в ранкові години). Метод прогнозування електроспоживання святкових та інших нерегулярних днів відсутня. Вирішення цього завдання вимагає великих теоретичних уявлень про імовірнісний механізм утворення, протікання і розвитку режиму електроспоживання, і відповідних статистичних досліджень.

– Наступ темряви. Прогнозування часу настання темної частини доби і тривалості використання електричного освітлення.

– Переведення часу. Переведення годинника з літнього часу на зимовий викликає ефекти запізнювання, які необхідно моделювати спеціальним чином.

– Регулювання електроспоживання. Широке застосування регулювання електроспоживання енергосистемами вимагає спеціальних методів прогнозування як "нерегульованого", так і "регульованого" споживання.

– Попередня перевірка достовірності даних телевимірювань - необхідна умова запобігання появи помилок в інформаційній базі і отримання поганих прогнозів.

– Промислові споживачі. Споживання великих промислових підприємств в разі аварій, страйків або незвичайних режимів роботи слід розглядати окремо. [5]

Існує багато досліджень, які вивчають підходи, методи та алгоритми прогнозування часових рядів споживання електроенергії та алгоритми зниження похибок моделей [8, 23-27]. Їх актуальність не втрачає силу через необхідність прогнозування даних у реальному часі, врахування різних джерел постачання електроенергії, кліматичних змін, а також змін у регулюванні ринку електроенергії.

Актуальним стало питання створення моделі прогнозування обсягів електроспоживання для постачальників електроенергії з метою збільшення ефективного розподілу електроенергії. У зв'язку з необхідністю створення запасу потужності, транспортування електроенергії і її перерозподілу важливим питанням є прогнозування обсягів електроенергії необхідних кінцевому споживачеві.

Перед розподільчими та генеруючими компаніями постає завдання розрахувати необхідне навантаження мереж і енергоблоків, на яку впливають різні фактори: клімат, географічне положення, місяць року, час доби і економічні чинники [24]. Прогнозування обсягів споживання електроенергії дозволяє зменшити ризики при прийнятті рішень.

Підприємства, установи, організації представлені значеннями витрат електричної енергії в вихідній вибірці, здійснюють різні види діяльності: промислове виробництво; сфера послуг; охорону здоров'я; освіта; житлово-комунальне обслуговування. Дослідивши розподіл величини витрати електроенергії між групами споживачів, освіченими шляхом класифікації споживачів за ознакою здійснення однакових видів діяльності, і виявивши існуючі закономірності, виконується прогноз електроспоживання на середньострокову перспективу.

Залежно від терміну, на який планується величина енергоспоживання, прогнози діляться на оперативні (on-line) і неоперативні (off-line). Відповідно до сучасної термінологією в літературі [19, 26] по прогнозуванню і управлінню в енергетиці, використовується наступна класифікація інтервалів прогнозування:

- оперативний прогноз – від декількох хвилин до декількох годин в межах поточної доби;
- короткостроковий – від однієї доби до десяти;
- середньостроковий – від місяця до кількох місяців;
- довгостроковий – від одного року до п'яти років;
- перспективний – на кілька років і більше.

В даний час умови ринку електроенергії пред'являють все більш жорсткі вимоги до обсягів закупуваної і продаваної електроенергії, у відповідність з цим питання прогнозування є фундаментальним для фінансового планування суб'єктів.

Проблема прогнозування електроспоживання полягає в тому, що необхідно врахувати величезну кількість чинників, що має вплив на зміну енергоспоживання. На сьогоднішній день проведено безліч досліджень по цій темі і створена велика кількість моделей прогнозування часових рядів [23, 24].

Тривають розробки для зниження похибок моделей прогнозування часових рядів. У зв'язку з жорсткістю вимог до прогнозування дана задача не тільки вдосконалюється, а й ускладнюється. Враховується все більшу кількість факторів, використовуються нові методи, ускладнюються алгоритми розрахунків [8].

Прогнозування споживання енергії є важливим інструментом, який може допомогти у вимірювання та управління гнучкістю попиту. У цьому аспекті мають бути певні виклики подолати в майбутньому. [3]

Майбутні напрямки досліджень повинні заохочувати використання сучасних методів прогнозування, дозволяючи їм бути актуальними для

представлення енергії будівель в різних масштабах і в різних умовах навколишнього середовища [26].

У деяких методах прогнозування, заснованих на машинному навчанні, враховуються дані моделювання шуму. Таким чином, різноманітність енергетичних показників є більш передбачуваним, ніж енергетичне моделювання результатів, які призводять до перенавчання або навіть до неправильної підгонки [27].

Враховуючи вищезазначені виклики, не можна було б використовувати єдиний метод прогнозування який можна застосувати до будь-якого завдання, щоб задовольнити всі вимоги різних сценаріїв застосування. Тому слід максимізувати потенціал кожного з методів за допомогою критеріїв вибору найкращого для конкретного завдань.

Побудова моделей прогнозування дає економічну та екологічну вигоду за рахунок створення стратегій гнучкості постачання електроенергії, такі як стратегічне збереження, гнучка форма навантаження, обрізка піків та перенесення навантаження.

1.3 Постановка завдання дослідження

Проведений вище аналіз дозволив сформулювати цілі і завдання дослідження. Метою проведеного дослідження є розробка методики прогнозу споживання електричної енергії підприємством, організацією, установою, забезпечивши зменшення похибки прогнозу без залучення спеціалізованих пакетів програмного забезпечення.

Об'єктом дослідження є показники щоденного споживання електроенергії одним із корпусів Сумського державного університету, на основі емпіричних даних лічильників.

Загальне енергоспоживання залежить від внутрішніх змін в будівлі корпусу університету, а також від змін у секторі споживачів, які впливають в цілому на енергоспоживання всього університету.

Для прогнозування енергоспоживання необхідно пройти ряд етапів:

- здійснити графічний або описовий аналіз наявної вихідної інформації;
- дослідити отримані часові ряди;
- вибрати методи прогнозування і скласти моделі прогнозу;
- зробити порівняльний аналіз моделей прогнозування;
- оцінити отримані прогнозні значення.

2. ОГЛЯД МЕТОДІВ І ПІДХОДІВ АНАЛІЗУ ЧАСОВИХ РЯДІВ

2.1 Статистичний аналіз часових рядів та їх компонентів

Під часовим рядом (time series) розуміється послідовність спостережуваних значень деякої змінної, вироблених через рівні проміжки часу. Якщо прийняти довжину такого проміжку за одиницю часу (рік, квартал, день), то можна вважати, що послідовні спостереження x_1, \dots, x_n зроблені в моменти $t = 1, \dots, n$ [7].

Введення тимчасової шкали в часовий ряд істотно відрізняє його від простої (випадкової) вибірки статистичних даних. Ключова особливість часового ряду – прив'язка значень (вимірювань) до відповідних моментів часу.

Аналіз часового ряду переслідує дві головні цілі:

- визначення його структури (природи);
- прогноз майбутніх значень ряду на основі поточних і минулих вимірів. [8]

Обидві тісно взаємопов'язані. Вирішення першого завдання необхідно для побудови математичної моделі часового ряду, її коректної ідентифікації та формалізації. Математична модель стане, свого роду, лабораторією для дослідження часового ряду і фундаментом для щодо точних (з допустимою нормою похибки) передбачень по ряду.

Часовому ряду притаманні чотири основні компоненти: тренд, сезонність, циклічність і нерегулярні компоненти. Ці ознаки відрізняють часовий ряд від звичайних спостережуваних даних. [6]

Оцінка лінійного тренду – це статистичний прийом, який допомагає інтерпретувати дані (рис 2.1). Коли серія вимірювань процесу розглядається як часовий ряд, оцінка тренду може бути використана для складання та обґрунтування тверджень про тренд в даних, шляхом співвіднесення

вимірювань із часом, коли вони відбулися. Потім створену модель можна використовувати для опису поведінки спостережуваних даних, не пояснюючи цього. У цьому випадку оцінка лінійного тренду виражає дані як лінійну функцію часу, а також може бути використана для визначення значущості відмінностей у наборі даних, пов'язаних категоріальним фактором. [8]

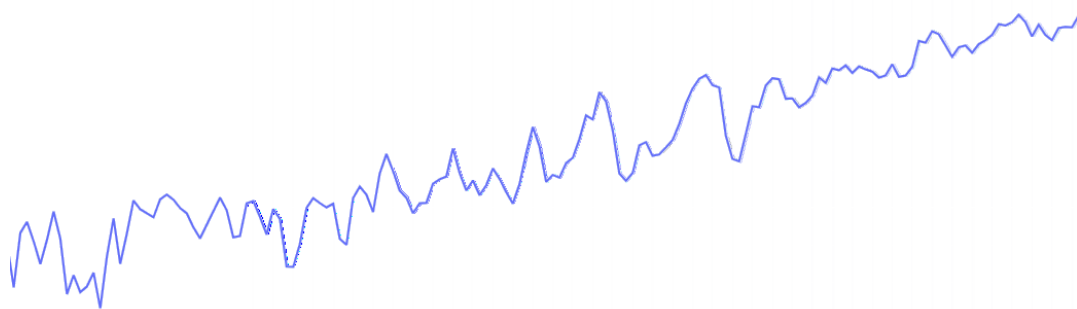


Рис. 2.1 – Приклад зростаючого тренду

Враховуючи набір даних і бажання створити якусь модель цих даних, існує безліч функцій, які можна вибрати для відповідності. Якщо немає попереднього розуміння даних, то найпростішою функцією, яка підходить, є пряма лінія зі значеннями даних на осі y та часом ($t = 1, 2, 3, \dots$) на осі x .

Даний набір моментів часу t та значення даних y_t спостерігаються для тих моментів часу значення a і b вибрані так, що:

$$\sum_t [y_t - (\hat{a}t + \hat{b})]^2, \quad (2.1)$$

зведено до мінімуму. Тут \hat{a} і \hat{b} – лінія тренду, тому сума квадратів відхилень від лінії тренду – це те, що мінімізується. Це завжди можна зробити у закритій формі, оскільки це випадок простої лінійної регресії. [6]

У даних часових рядів сезонність – це наявність змін, які відбуваються через певні регулярні проміжки часу менше року, наприклад, щотижня, щомісяця чи кварталу. Сезонність може бути спричинена різними факторами, такими як погода, відпустка та свята, і вона складається з періодичних,

повторюваних та загалом регулярних та передбачуваних закономірностей на рівнях часового ряду.

Сезонні коливання в часових рядах можна протиставити циклічним закономірностям. Останні трапляються, коли дані експозицій зростають і падають, які не мають фіксованого періоду. Такі несезонні коливання, як правило, зумовлені економічними умовами і часто пов'язані з «діловим циклом»; їхній період, як правило, перевищує один рік, а коливання, як правило, становлять щонайменше два роки.

У адитивній моделі часових рядів сезонний компонент оцінюється як:

$$S = Y - (T + C + I), \quad (2.2)$$

де S – сезонні значення, Y – для дійсних значень даних часових рядів, T – для значень тренду, C – для циклічних значень, I – для нерегулярних значень.

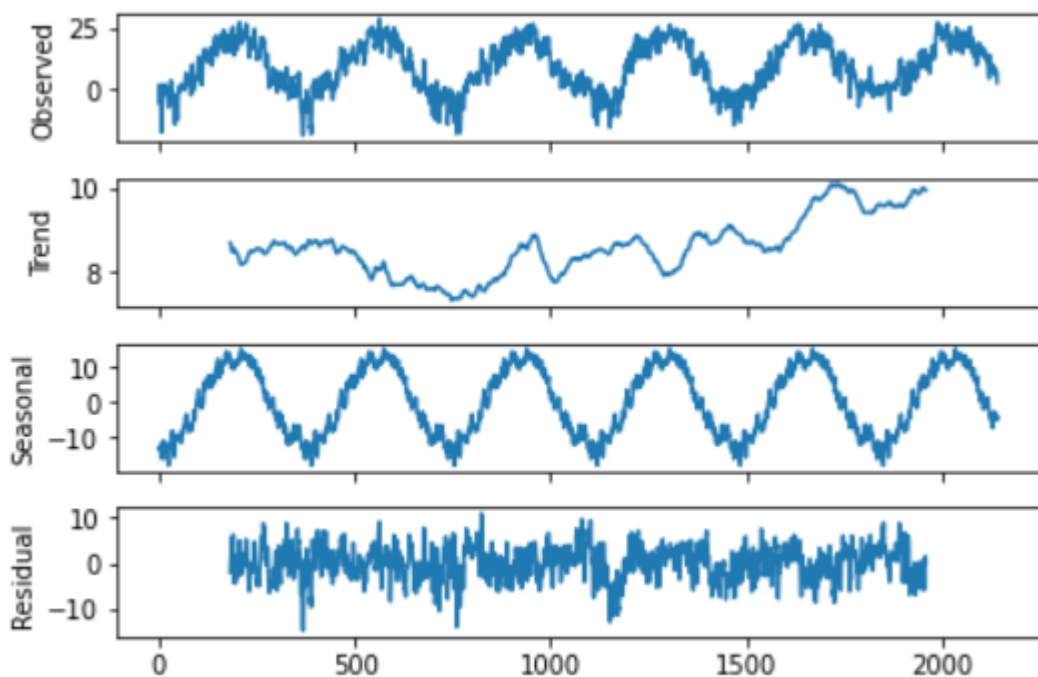


Рис. 2.2 – Приклад декомпозиції часового ряду

Циклічна складова відображається в даних які мають несезонні коливання. Тривалість цих коливань зазвичай становить не менше 2 років.

Прикладом є ділові цикли, які зазвичай тривають кілька років, але де тривалість поточного циклу заздалегідь невідома.

Багато людей плутають циклічну поведінку із сезонною, але насправді вони зовсім інші. Якщо коливання не мають фіксованого періоду, то вони циклічні; якщо період незмінний і пов'язаний з якимось аспектом календаря, то модель є сезонною. Взагалі, середня тривалість циклів перевищує довжину сезонної схеми, і величина циклів, як правило, є більш мінливою, ніж величина сезонних моделей.

Нерегулярний часовий ряд зберігає дані для послідовності довільних часових точок. Нерегулярні часові ряди доречні, коли дані надходять непередбачувано, наприклад, коли програма фіксує кожен торгівлю акціями або коли лічильники електроенергії фіксують випадкові події, такі як попередження про низький рівень заряду акумулятора або показники низької напруги. Нестандартні часові ряди також потрібні для упакованих даних, які включають герцові дані та стислі цифрові дані.

Нерегулярні часові ряди зберігають мітки часу для кожного елемента, а не зберігають зсуви, оскільки інтервал між кожним елементом може бути різною довжиною. Неправильні елементи зберігаються до наступного елемента за замовчуванням і не можуть бути нульовими.

Незважаючи на те, що нелінійні системи можуть значно відрізнятися в конкретних проявах і деталях, існують глибокі аналогії їх організації та функціонуванні. Це зумовило інтерес до методів, які розвиваються в рамках теорії динамічних систем, як універсального інструменту дослідження об'єктів самої різної природи. Найбільш явно такий об'єднуючий підхід проявляється при аналізі часових рядів.

В даний час існує два якісно різних підходи до дослідженню часових рядів:

- статистичні;
- динамічні.

До статистичних підходів відносяться ймовірнісні моделі. До динамічних теорія Такенса (Такенса-Мане). Сучасно уявлення про можливість опису спостережуваних дає ембедологія (від англ. embedding - вкладення), яка об'єднує елементи теорії розмірності, теорії інформації, топології, диференціальної динаміки і теорії динамічних систем.

Основа стохастичних методів обробки даних є ряд y_n і шум – послідовність некорелюючих однаково розподілених випадкових величин ξ_i з нульовим середнім. Тоді можна записати, що:

$$y_n = F(y_{n-1}, \dots, y_{n-m}, \xi_n, \dots, \xi_{n-k}), \quad (2.3)$$

де k і m деякі кінцеві числа.

2.2 Математичні моделі часових рядів

2.2.1 Авторегресійні моделі прогнозування

В основу авторегресійних моделей закладено припущення, яке означає, що значення процесу $Z(t)$ залежить лінійно від попередніх значень процесу $Z(t-1), \dots, Z(t-p)$.

В області аналізу часових рядів модель авторегресії (autoregressive, AR) і модель ковзного середнього (moving average, MA) є однією з найбільш використовуваних. [9]

Згідно роботі, модель авторегресії є виключно корисною для опису деяких зустрічаються на практиці часових рядів. У цій моделі поточне значення процесу відображається у вигляді кінцевої лінійної сукупності попередніх значень процесу і помилок, який називається «білим шумом». Розглянемо наступну формулу:

$$Z(t) = C + \sum_{i=1}^p \varphi_i X_{t-i} + \varepsilon_t, \quad (2.4)$$

де $Z(t)$ - процес авторегресії порядку p , C - константа, $\varphi_1, \dots, \varphi_p$ - коефіцієнти моделі, ε_t - помилка моделі. [10]

Функція автокореляції процесу AR може бути виражена як:

$$\rho(\tau) = \sum_{k=1}^p a_k y_k^{-|\tau|}, \quad (2.5)$$

де y_k – корені многочлена:

$$\phi(B) = 1 - \sum_{k=1}^p \varphi_k B^k, \quad (2.6)$$

де B - оператор лагу, $\phi()$ - це функція, що визначає авторегресію, і де φ_k - коефіцієнти в авторегресії. [7]

Модель ковзного середнього (MA), також відома як процес ковзної середньої, є загальним підходом для моделювання одновимірних часових рядів. Модель ковзного середнього визначає, що вихідна змінна лінійно залежить від поточного та різних минулих значень стохастичного (недосконало прогнозування) терміну.

Разом з моделлю авторегресії (AR) модель ковзного середнього є особливим випадком та ключовим компонентом більш загальних моделей ARMA та ARIMA часових рядів, які мають більш складну стохастичну структуру. [9]

Модель ковзного середнього не слід плутати з ковзною середньою, що є окремим поняттям, незважаючи на деякі подібності.

На відміну від моделі AR, кінцева модель MA завжди нерухома.

Позначення MA відноситься до моделі ковзного середнього порядку q :

$$X_t = \mu + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q}, \quad (2.7)$$

де μ – середнє значення ряду, $\theta_1, \dots, \theta_q$ – параметри моделі, ε_t – білий шум.

Значення q називається порядком моделі МА. Це можна еквівалентно записати в термінах оператора лагу B як:

$$X_t = \mu + (1 + \theta_1 B + \dots + \theta_q B^q) \varepsilon_t, \quad (2.8)$$

де μ – середнє значення ряду, $\theta_1, \dots, \theta_q$ – параметри моделі, ε_t – білий шум, B – оператор лагу.

Таким чином, модель ковзного середнього є концептуально лінійною регресією поточного значення серії попередніх (спостережених) умов похибки білого шуму або випадкових поштовхів. Випадкові поштовхи в кожній точці вважаються взаємно незалежними і походять від одного і того ж розподілу, як правило, нормального розподілу, з розташуванням у нульовому та постійному масштабі.

У класичному статистичному аналізі часових рядів, модель (ARMA) забезпечують представлення стаціонарного випадкового процесу в термінах двох поліномів, один для авторегресії (AR), а другий для ковзного середнього (МА). Загальна модель ARMA була описана в 1951 р. дисертації Пітера Уітла тестування гіпотез для аналізу часових рядів [22].

Враховуючи часовий ряд даних X_t , модель ARMA є інструментом для описання та прогнозування майбутніх значень часового ряду. Частина AR передбачає регресію змінної за власними попередніми (минулими) значеннями часового ряду. Частина МА передбачає моделювання терміну помилки як лінійної комбінації термінів помилок, що трапляються одночасно і в різні періоди минулого. Модель зазвичай називають моделлю ARMA (p, q), де p – позначає AR, а q – позначає МА.[9]

Позначення ARMA(p, q) відноситься до моделі з p – авторегресорними компонентами та q – компонентами ковзного середнього. Модель містить дві моделі: AR (p) та МА (q),

$$X_t = C + \varepsilon_t + \sum_{i=0}^p \varphi_i X_{t-i} + \sum_{i=0}^q \theta_i \varepsilon_{t-i}. \quad (2.9)$$

Умови помилки ε_t зазвичай прийнято вважати незалежними одна від одної розподіленими випадковими величинами, відібраними з нормального розподілу з нульовим середнім: $\varepsilon_t \sim N(0, \sigma^2)$, де σ^2 - дисперсія. Ці припущення можуть бути послаблені, але послаблюючи їх – змінюються властивості моделі. Зміна значення припущення про випадкову величину призведе до досить принципової різниці.

Залежність X_t від попередніх значень та термінів помилок ε_t вважається лінійною, якщо не вказано інше. Якщо залежність нелінійна, модель спеціально називається нелінійною ковзною середньою (NMA), нелінійною авторегресивною (NAR) або нелінійною авторегресією ковзною середньою моделлю (NARMA).[7]

Модель авторегресії ковзного середнього можуть бути узагальнені іншими способами. Таким чином утворені моделі авторегресивної умовної гетероскедастичності (ARCH) та моделі авторегресивної інтегрованої ковзної середньої (ARIMA). Якщо потрібно встановити кілька часових рядів, тоді може бути встановлена векторна модель ARIMA (або VARIMA). Якщо часові ряди в питанні мають довготривалу пам'ять, то дробовий ARIMA (FARIMA, іноді називають ARFIMA) моделювання може бути доречним: авторегресії частинної інтегрованої ковзної середньої. Якщо припускається, що дані мають сезонні ефекти, вони можуть моделюватися за допомогою SARIMA (сезонна ARIMA) або періодичної моделі ARMA.

Іншим типом є багатомасштабна модель авторегресії (MAR). Модель MAR індексується вузлами дерева, а стандартна (дискретний час) авторегресивна модель індексується цілими числами.

Модель ARMA є одновимірною моделлю. Для дослідження багатовимірного часового ряду є векторна авторегресія (VAR) і векторна авторегресія ковзного середнього (VARMA).

Модель авторегресійної інтегрованої ковзної середньої (ARIMA) є узагальненням моделі авторегресивної ковзної середньої (ARMA). Обидві ці

моделі пристосовані до даних часових рядів або для кращого розуміння даних, або для прогнозування майбутніх моментів у серії (прогнозування). Моделі ARIMA застосовуються в деяких випадках, коли дані свідчать про нестационарність, де початковий етап диференціювання (що відповідає «інтегрований» частині моделі) може застосовуватися один або кілька разів для усунення нестационарності.

Несезонні моделі ARIMA зазвичай позначають ARIMA (p, d, q), де параметри p, d і q є невід'ємними цілими числами, p - порядок (кількість часових лагів) авторегресивної моделі, d - ступінь диференціювання (скільки разів дані віднімали минулі значення), а q - порядок моделі ковзного середнього. Сезонні моделі ARIMA зазвичай позначають ARIMA (p, d, q) (P, D, Q) m, де m - кількість періодів у кожному сезоні, а великі регістри P, D, Q - авторегресивні, диференційовані та ковзні середні умови для сезонної частини моделі ARIMA.

Коли два з трьох термінів є нулями, на модель можна посилатись на основі ненульового параметра, виключаючи " AR ", " I " або " MA " зі скорочення, що описує модель.

$$\left(1 - \sum_{k=1}^p \alpha_k L^k\right) \Delta^d y_t = \left(1 - \sum_{k=1}^q b_k L^k\right) \varepsilon_t. \quad (2.10)$$

Оператор різниці Δ можна перенести з тимчасового ряду y_t на послідовність коефіцієнтів, причому кількість ненульових доданків збільшиться на одиницю.

Моделі авторегресії та ковзного середнього працюють зі стаціонарними та лінійними даними. Однак у багатьох випадках дані нестационарні. Авторегресійна інтегрована ковзного середнього (ARIMA) моделі використовуються для роботи з нестационарними даними. Узагальненою формою моделі ARIMA, відомою як сезонна ARIMA (SARIMA), використовується для обробки сезонності даних. Цей клас моделей ARIMA

явно стосується сезонності в даних із використанням сезонних AR, MA та диференційованих термінів у моделі.

Зовнішні змінні також можуть бути додані до моделі за допомогою екзогенного регресора. Сезонна ARIMA з екзогенними регресорами (SARIMAX) дозволяє користувачеві додавати ефекти зовнішні змінні до моделі.

Моделі SARIMAX записується як: SARIMAX (p, d, q) (P, D, Q) s де:

- p – порядок AR;
- d – порядок диференціювання;
- q – порядок MA;
- P – порядок сезонного AR;
- D – порядок сезонного диференціювання;
- Q – порядок MA;
- S – кількість періодів у сезоні.

SARIMAX (p, d, q) (P, D, Q) s математично представлений як [9]:

$$y_t = \beta_0 + \beta_1 X_{1,t} + \beta_2 X_{2,t} + \dots + \beta_k X_{k,t} + \frac{(1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q)(1 - \Theta_1 B^s - \Theta_2 B^{2s} - \dots - \Theta_Q B^{Qs})}{(1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p)(1 - \Phi_1 B^s - \Phi_2 B^{2s} - \dots - \Phi_P B^{Ps})} Z_t, \quad (2.11)$$

де:

- y_t – значення часового ряду в момент часу t;
- $X_{1,t}, X_{2,t}, \dots, X_{k,t}$ – екзогенні змінні;
- $\beta_0, \beta_1, \dots, \beta_k$ – параметри регресії;
- $\phi_1, \phi_2, \dots, \phi_p$ – несезонна авторегресія;
- $\Phi_1, \Phi_2, \dots, \Phi_P$ – сезонна авторегресія;
- $\theta_1, \theta_2, \dots, \theta_q$ – несезонні умови ковзного середнього;
- $\Theta_1, \Theta_2, \dots, \Theta_Q$ – сезонні умови ковзного середнього;
- B^s – оператор лагу;
- Z_t – шум.

2.2.2 Моделі експоненціального згладжування

Альтернативний варіант побудови прогнозу розглянуті – адаптивні методи прогнозування, зокрема, різні типи експоненціального згладжування.

Експоненціальне згладжування включає постійне оновлення моделі за допомогою найбільш актуальних даних. Цей метод ґрунтується на усередненні часових рядів попередніх спостережень в експоненціальному напрямку. Тобто пізнішим подій присвоюється більшу вагу. [10]

Здійснення вирівнювання часового ряду на основі методу експоненціального згладжування, ідея якого полягає в тому, що рівні вихідного часового ряду зважуються ковзної середньої, з експоненціальним характером зміни ваг. Загальна формула (рекурентна формула) розрахунку експоненційної середньої має вигляд:

$$S_t(y) = \alpha \cdot y_1 + (1 - \alpha) \cdot S_{t-1}(y), \alpha \in (0; 1), \quad (2.12)$$

де S_t – значення експоненційної середньої, α – параметр згладжування,

У згладженому вигляді прогнозування буде визначатися як зважене середнє значення останнього спостереження момент часу t і її попереднього прогнозу на той самий період:

$$\hat{Y}_{t+1} = \alpha Y_t + (1 - \alpha) \hat{Y}_t, \quad (2.13)$$

де:

\hat{Y}_{t+1} - прогнозоване значення;

α - стала згладжування;

Y_t - спостережуване величини за поточний період t ;

\hat{Y}_t - попередній згладжуваний прогноз за період t .

Модель функціонує тільки при невеликому інтервалі прогнозування. Тренд і сезонні параметри не враховуються. Для врахування впливу лінійного тренду, використовується модель Хольта (Holt) [10]:

$$\hat{y}_{t+d} = a_t + d \cdot b_t, \quad (2.14)$$

де a_t - прогноз, очищений від тренду (експоненціальне згладжування), b_t - параметр лінійного тренду.

$$\begin{aligned} a_t &= \alpha_1 y_t + (1 - \alpha_1)(a_{t-1} - b_{t-1}), \\ b_t &= \alpha_2(a_t - a_{t-1}) + (1 - \alpha_2)b_{t-1}. \end{aligned} \quad (2.15)$$

Важливою проблемою є вибір коефіцієнтів $\alpha_1, \alpha_2 \in (0,1)$, які визначають чутливість моделі. Чутлива модель швидко реагує на реальні зміни, а нечутлива не реагує на шум і випадкові відхилення.

Модель Хольта-Уінтерс є розвитком моделі Хольта, в ній з'являється сезонна складова, в результаті чого виходить система рівнянь з трьома постійними згладжування:

$$\begin{aligned} \hat{y}_{t+d} &= (a_t + d \cdot b_t)c_{t-s+d}, \\ a_t &= \alpha_1 \frac{y_t}{c_{t-s}} + (1 - \alpha_1)(a_{t-1} - b_{t-1}), \\ b_t &= \alpha_2(a_t - a_{t-1}) + (1 - \alpha_2)b_{t-1}, \\ c_t &= \alpha_3 \frac{y_t}{a_t} + (1 - \alpha_3)c_{t-s}, \end{aligned} \quad (2.16)$$

де:

- $\alpha_1, \alpha_2, \alpha_3$ приймають значення з діапазону $[0; 1]$,
- y - спостереження,
- a - згладжене значення спостереження,
- b - коефіцієнт тренду,
- c - індекс сезонності,
- t - індекс поточного спостереження. [10]

Алгоритм побудови моделі полягає в тому, що спочатку розраховуються коефіцієнти a_0 і b_0 лінійного тренду, далі розраховуються сезонні складові c_t за формулою:

$$c_t = \frac{y_t}{a_0 + b_0 t}. \quad (2.17)$$

Адитивні та мультиплікативні моделі Холт-Вінтерса запропонували вдосконалити модель [11] з урахуванням тренду та сезонних моделей [12].

Досконаліші моделі експоненціального згладжування простору станів з адитивними чи мультиплікативними помилками містять рівняння вимірювання, яке описує спостережувані дані, та деякі рівняння стану, що описують, як змінюються з часом неспостережувані компоненти чи стани (рівень, тренд, сезонність) [12]. Одною із успішних нещодавніх у моделей експоненціального згладжування простору є модель TBATS із перетворенням Бокса-Кокса, помилкою ARMA, тренду та поданням сезонних компонентів [13]. Цей підхід дає прогнози високої точності, що обробляють багаторазові вкладені та не вкладені сезонні показники. Хоча це вимагає додаткового часу обчислення, особливо для великих даних часових рядів.

Загальне представлення моделі TBATS (2.18) включає рівняння рівня (2.19), тренду (2.20), сезонних (2.21) та рівнянь помилки ARMA (2.22):

$$y_t^{(\omega)} = l_{t-1} + \phi b_{t-1} + \sum_{i=1}^T s_t^{(i)} + d_t. \quad (2.18)$$

$$l_t = l_{t-1} + \phi b_{t-1} + \alpha d_t. \quad (2.19)$$

$$b_t = (1 - \phi)b + \phi b_{t-1} + \beta d_t. \quad (2.20)$$

$$s_t^{(i)} = s_{j,t-1}^{(i)} \cos \lambda_j^{(i)} + s_{j,t-1}^{(i)} \sin \lambda_j^{(i)} + \gamma_i d_t. \quad (2.21)$$

$$d_t = \sum_{i=1}^p \varphi_i d_{t-i} + \sum_{i=1}^q \theta_i \varepsilon_{t-1} + \varepsilon_t. \quad (2.22)$$

2.2.3 Моделі машинного навчання

Штучні нейронні мережі - це методи прогнозування, які базуються на простих математичних моделях нейрона. Вони допускають складні нелінійні взаємозв'язки між змінною відповіді та її предикторами. [7]

Нейронну мережу (NN) можна розглядати як мережу «нейронів», які організовані в шари. Прогнози (або вхідні дані) утворюють нижній шар, а

прогнози (або результати) - верхній шар. Також можуть бути проміжні шари, що містять «приховані нейрони». [12]

Найпростіші мережі не містять прихованих шарів і еквівалентні лінійним регресіям. Коефіцієнти, прикріплені до цих предикторів, називаються "вагами". Прогнози отримують за допомогою лінійної комбінації вхідних даних. Ваги вибираються в рамках нейронної мережі за допомогою «алгоритму навчання», який мінімізує «функцію витрат», таку як MSE. Звичайно, у цьому простому прикладі можна використати лінійну регресію, яка є набагато ефективнішим методом навчання моделі.

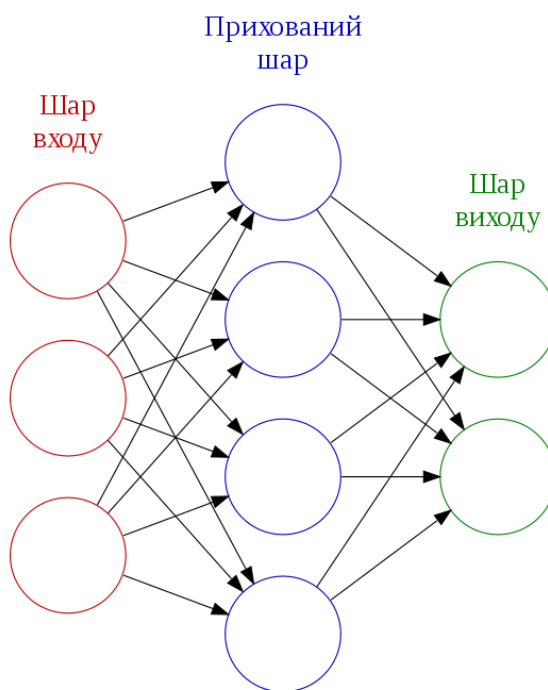


Рис. 2.3 – Загальний вигляд нейронної мережі

Як тільки додається проміжний шар із прихованими нейронами, нейронна мережа стає нелінійною (рис 2.3). [12]

Це відомо як багатошарова мережа прямого пересилання, де кожен рівень вузлів отримує вхідні дані від попередніх шарів. Виходи вузлів в одному шарі є входами для наступного шару. Входи кожного вузла об'єднуються за допомогою зваженої лінійної комбінації. Потім результат

модифікується нелінійною функцією перед виведенням. Наприклад, вхідні дані у прихований нейрон j (рис 2.3) об'єднані лінійно, щоб отримати:

$$z_j = b_j + \sum_{i=1}^n w_{i,j} x_i. \quad (2.23)$$

У прихованому шарі модифікується за допомогою нелінійної функції, такої як сигмоїда:

$$s(z) = \frac{1}{1 + e^{-z}}, \quad (2.24)$$

щоб дати вхід для наступного шару. Це має тенденцію до зменшення ефекту екстремальних вхідних значень, роблячи таким чином мережу надійнішою до вищого рівня. [12]

Параметри b_j, w_i залежать від даних. Значення ваг w_i часто обмежують, щоб запобігти їх надмірному зростанню. Параметр, який обмежує ваги, відомий як "параметр затухання", і часто встановлюється рівним 0,1.

Для початку ваги приймають випадкові значення, які потім оновлюються з використанням спостережуваних даних. У прогнозуваннях, створених нейронною мережею, є елемент випадковості. Тому мережу навчають кілька разів, використовуючи різні випадкові вихідні точки, і результати усереднюють.

Кількість прихованих шарів та кількість вузлів у кожному прихованому шарі повинні бути вказані заздалегідь.

Найбільш важливою здатністю нейронних мереж порівняно з іншими нелінійними моделями є їх гнучкість у моделюванні будь-якого типу нелінійних шаблонів без попереднього припущення про основний процес генерації даних [16]. Найпопулярнішою моделлю нейронної мережі для прогнозування часових рядів є тришарова модель прямої мережі, яку можна записати як:

$$y_t = \alpha_0 + \sum_{j=1}^n \alpha_j f\left(\sum_{i=1}^m \beta_{ij} y_{t-i} + \beta_{0j}\right) + \varepsilon_t, \quad (2.25)$$

де m - кількість вхідних вузлів, n - кількість прихованих вузлів, f - сигмоїдна передавальна функція, $\{\alpha_j, j=0, 1, \dots, jn = \dots\}$ - це вектор ваг від прихованих до вихідних вузлів, α_0 та β_{0j} - це ваги дуг, що ведуть із зміщених членів.

Для проблеми прогнозування часових рядів побудова моделі NN еквівалентна визначенню як кількості вхідних вузлів, так і кількості прихованих вузлів [17]. Вхідні вузли – це минулі спостереження, за допомогою яких може бути отримана основна структура автокореляції даних. Хоча теорія універсального наближення NN вказує на хороше наближення може знадобитися велика кількість прихованих вузлів, лише невелика кількість часто потрібна в реальних додатках [16].

У 2017 року фахівці компанії Facebook оголосили про розроблений ними новий алгоритм для прогнозування часових рядів – Prophet. Prophet вирішує недоліки вище перелічених моделей і дозволяє в автоматичному режимі створювати моделі, що володіють високою точністю прогнозувань. [15]

Prophet базується на розкладній адитивній моделі, де нелінійні тренди відповідають сезонності, а також враховує наслідки відпусток, свят, вихідних [14].

Загальна ідея моделі подібна до узагальненої адитивної моделі. Prophet відповідає, як уже згадувалося вище, тренду, сезонності та святам:

$$y(t) = g(t) + s(t) + h(t) + e(t), \quad (2.26)$$

де:

- $g(t)$ – функція тренду;
- $s(t)$ – функція сезонності;
- $h(t)$ – функція відпусток;
- $e(t)$ – помилка;
- $y(t)$ – прогноз. [15]

Prophet пропонує нам дві моделі, одна з них – це модель логістичного зростання, а інша – частинно-лінійна модель. За замовчуванням Prophet використовує частинно-лінійну модель, але її можна змінити, вказавши іншу

модель. Вибір моделі є делікатним, оскільки він залежить від різноманітних факторів, таких як розмір компанії, темпи зростання, бізнес-модель тощо. Якщо дані, що підлягають прогнозуванню, мають нелінійні дані (зростають нелінійно і після досягнення точки насичення, майже не демонструє зростання або зменшується і демонструє лише деякі сезонні зміни), тоді логістична модель зростання – найкращий варіант. Тим не менше, якщо дані показують лінійні властивості і мали тренд зростаючий чи спадаючий в минулому, то частинно-лінійна модель є кращим вибором. [14, 15]

Модель логістичного зростання підходить за допомогою наступного статистичного рівняння:

$$g(t) = \frac{C}{1 + e^{-k(t-m)}}, \quad (2.27)$$

де:

- C – точка зміни тренду (порогове значення);
- k – темп зростання;
- m – параметр зміщення. [15]

Частинно-лінійна модель підходить за допомогою наступних статистичних рівнянь:

$$y = \begin{cases} \beta_0 + \beta_1 x & x \leq c \\ \beta_0 - \beta_2 c + (\beta_1 + \beta_2)x & x > c \end{cases} \quad (2.28)$$

де c - точка зміни тренду (визначає зміну тренду).

Часові ряди часто мають багатоперіодну сезонність в результаті людської поведінки, яку вони представляють. Наприклад, 5-денний робочий тиждень може давати вплив на часовий ряд, який повторюється щотижня, тоді як графіки відпусток та перерв у школі можуть створювати ефекти, що повторюються щороку [15]. Для прогнозування цих ефектів вказується сезонна функція часового ряду:

$$s(t) = \sum_{n=1}^N \left(a_n \cos\left(\frac{2\pi n t}{P}\right) + b_n \sin\left(\frac{2\pi n t}{P}\right) \right), \quad (2.29)$$

де P – період.

Свята та вихідні значно впливають на прогнозування часових рядів і часто не мають періодичного зразка, тому їх наслідки не добре моделюються плавним циклом. У багатьох країнах світу є великі свята, які дотримуються місячного календаря. Вплив конкретного свята на часові ряди часто подібний рік у рік, тому важливо включити його в прогноз.

Prophet дозволяє надавати перелік минулих та майбутніх подій. Для вирішення проблеми прогнозування використовується об'єднання загальної сукупності свят специфічних для кожної країни.

2.3 Перевірка адекватності моделей

Одним з підходів в оцінці адекватності полягає в порівнянні виходів моделі і реальної системи при однакових значеннях входів. Для оцінки його застосовують методи статистичної теорії оцінювання і перевірки гіпотез.

Адекватність досліджуваної моделі можна перевірити по дисперсія відхилень значень моделі від середнього значення системи.

Порівняння дисперсії проводять за допомогою F - критерію (критерію Фішера):

$$F_{f_{\text{теор}}f_{\text{експ}}} = \frac{S_{\text{теор}}^2}{S_{\text{експ}}^2}, \quad (2.30)$$

де:

- $S_{\text{теор}}^2$ – дисперсія, яка характеризує помилку моделі,
- $S_{\text{експ}}^2$ – дисперсія, що характеризує помилку експерименту,
- $f_{\text{теор}}f_{\text{експ}}$ – число ступенів свободи моделі і експерименту відповідно.

[18]

Дисперсія, яка характеризує помилку моделі, розраховується за формулою:

$$S_{\text{теор}}^2 = \frac{\sum_{i=1}^N (y_i^{\text{теор}} - y_i^{\text{експ}})^2}{f_{\text{теор}}}, \quad (2.31)$$

де:

- $\sum_{i=1}^N (y_i^{\text{теор}} - y_i^{\text{експ}})^2$ – залишкова сума квадратів, що характеризує відхилення від регресії,
- $f_{\text{теор}}$ – число ступенів свободи моделі,
- n – число експериментальних точок;
- m – кількість значущих коефіцієнтів моделі в рівнянні регресії.

Числом ступенів свободи в статистиці називається різниця між числом експериментів і числом коефіцієнтів рівняння регресії.

Значимість коефіцієнтів лінійної регресії визначається за допомогою t -критерію Стюдента. Для цього необхідно обчислити розрахункові (фактичні) значення t – критерію для параметра a_0 (2.32), для параметра a_1 (2.33).

$$t_{a_0} = |a_0| \frac{\sqrt{n-2}}{S_{\text{зал}}}, \quad (2.32)$$

$$t_{a_1} = |a_1| \frac{\sqrt{n-2}}{S_{\text{зал}}} S_x, \quad (2.33)$$

де $S_{\text{зал}}$ – середньоквадратичне відхилення результативної ознаки від середнього арифметичного значення, S_x – середньоквадратичне відхилення факторної ознаки від середнього арифметичного значення. [18]

Розраховані за допомогою формул (2.32, 2.33) значення порівнюють з критичними t , значення якого отримується з таблиці Стюдента беручи до уваги рівня значущості α і кількість ступенів свободи $\nu=n-2$. Якщо $t_{\text{знайд}} > t_{\text{табл}}$, гіпотеза про значущість коефіцієнта a_1 приймається, в іншому випадку коефіцієнт вважається незначним і прирівнюється до нуля.

Для порівняння моделей з різним числом параметрів використовують Інформаційні критерії.

Інформаційний критерій Акаїке (AIC) заснований на інформації Кульбака-Лайблера, запропонованої в роботі [19]. Розглянемо кілька

популярних прикладів, в яких часто використовуються інформаційні критерії та їх модифікації. Оскільки часто метою є прогнозування економічних показників, як в даній роботі, ми обмежимо розгляд авторегресійну моделями. АІС визначається як:

$$AIC = \ln(\hat{\sigma}^2) + \frac{2(p+q+1)}{n}, \quad (2.34)$$

де $\hat{\sigma}^2$ – очікувана вибіркова дисперсія.

Байєсівський інформаційний критерій (BIC) – критерій вибору моделі з класу параметризованих моделей, що залежать від різного числа параметрів. Для оцінювання моделі зазвичай використовується метод знаходження максимуму функції правдоподібності, значення якого можна збільшити додаванням додаткових параметрів. Байєсівський критерій усуває проблему перенавчання, штрафуючи збільшення числа параметрів моделі. Він пов'язаний з критерієм Акаїке, але в разі байєсівського критерію збільшення параметрів штрафується суворіше. BIC визначається як:

$$BIC = \ln(\hat{\sigma}^2) + \frac{(p + q + 1) \ln(n)}{n}, \quad (2.35)$$

де $\hat{\sigma}^2$ – очікувана вибіркова дисперсія.

Після підгонки лінійної моделі до даних перевіряють адекватність відповідної моделі, для чого обчислюються різні залишки.

Для визначення автокореляції залишків першого порядку в регресійній моделі використовують статистичний критерій Дарбіна-Уотсона (або DW-критерій) [20]. Критерій названий на честь Джеймса Дарбина і Джеффри Уотсона. Значення критерію Дарбіна-Уотсона обчислюється за формулою:

$$DW = \sum_{t=2}^n \frac{(y_t - y_{t-1})^2}{\sum_{t=1}^n y_t^2}, \quad (2.36)$$

де y_{t-1} і y_t – послідовні значення часового ряду.

За відсутності автокореляції помилок $DW = 2$, при позитивній автокореляції DW прагне до нуля, а при негативній прагне до 4.

Тест Дарбіна-Уотсона можна застосовувати тільки якщо виконуються наступні умови:

- в регресійному рівнянні присутній вільний член;
- регресорів є нестохастична;
- в регресійному рівнянні немає лагових значень залежної змінної.

Для надійності виведення перевіряється статистична гіпотеза про нормальність розподілу за допомогою критерію згоди Пірсона [21]. Було висунуто гіпотеза про нормальний закон розподілу залишків. Для перевірки цієї гіпотези використовують:

$$\chi^2 = \sum_{t=2}^n \frac{(f_i - f_e)^2}{f_e}, \quad (2.37)$$

де: f_i - дані, f_e – гіпотеза.

2.4 Перевірка прогнозної якості моделей

У процесі прогнозування і вибору початкових значень або параметрів моделі виникає задача оцінки точності прогнозування. Оцінка точності важлива і при порівнянні між собою різних моделей або при визначенні спроможності отриманого прогнозу. Відома велика кількість оцінок, що визначають точність прогнозування, але для обчислення будь-якої з них необхідно знати на кожному кроці помилку прогнозування.

Помилка прогнозування на один крок вперед в момент часу t дорівнює:

$$e_t = X_t - \hat{X}_t, \quad (2.38)$$

де X_t – значення вхідної послідовності в момент часу t , \hat{X}_t – прогноз на момент часу t зроблений на попередньому кроці.

Напевно, найбільш поширеною оцінкою точності прогнозування є середнє значення квадратів помилок (Mean Squared Error):

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2, \quad (2.39)$$

де n – кількість елементів послідовності.

Іноді, як недолік оцінки MSE вказують на її надмірну чутливість до одиночних помилок великої величини. Це пояснюється тим, що значення помилки при обчисленні MSE зводиться в квадрат. У цьому випадку в якості альтернативи пропонується використовувати середнє значення абсолютної помилки (Mean Absolute Error):

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|. \quad (2.40)$$

В MAE операція піднесення значення помилки в квадрат замінена на використання абсолютного її значення. Передбачається, що при використанні MAE можна отримати більш стійкі оцінки.

Обидві ці оцінки добре підходять, наприклад, для визначення точності прогнозу однієї і тієї ж послідовності при різних параметрах моделі або при використанні різних моделей, але вони виявляються малопродатними для порівняння між собою результатів прогнозування, отриманих на різних послідовностях.

Крім того, безпосередньо по величині цих оцінок можна визначити, наскільки хороший результат прогнозування.

Для того щоб мати можливість порівнювати між собою точність прогнозу різних послідовностей можна використовувати відносні оцінки RelMSE і RelMAE:

$$RelMSE = \frac{MSE}{MSE_b}, \quad (2.41)$$

$$RelMAE = \frac{MAE}{MAE_b}. \quad (2.42)$$

Середня квадратична помилка (Root Mean Square Error, RMSE) розраховується відповідно до формули:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}, \quad (2.43)$$

де:

- n - кількість спостережень,
- y_i - значення вихідного ряду,
- \hat{y}_i - прогнознi значення.

В якості помилки прогнозування часового ряду можна скористатися середньої абсолютною помилкою в процентах (Mean Absolute Percentage Error):

$$MAPE = \frac{100}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right|, \% \quad (2.44)$$

де n - кількість спостережень, y_i - значення вихідного ряду, \hat{y}_i - прогнознi значення.

3. ПРАКТИЧНА РЕАЛІЗАЦІЯ

Емпіричний набір даних представлений собою показниками лічильників щоденного споживання електроенергії головним корпусом Сумського державного університету. Дані розміщено в 6-ти файлах, окремо за кожний рік, починаючи з 2015 року закінчуючи поточним. Дані мають наступні змінні:

- «Точка обліку» – місце знаходження лічильника, номер лічильника;
- «Напрямок» – напрямок постачання, одиниці виміру споживання електроенергії;
- «Дата» – дата спостереження;
- «Споживання» – число спожитої енергії.

Для побудови точнішої моделі було обрано додаткові погодні чинники, котрі впливають на споживання електроенергії. Набір даних містить погодинні показники з метеостанції, що знаходиться поблизу Сумського державного університету. Приклад даних представлено в таблиці 3.1.

Таблиця 3.1

Характеристика змінних часового ряду

Змінна	Значення
Date	Дата, місцевий час
T	Температура повітря (°C) на висоті 2 м над поверхнею землі
P ₀	Атмосферний тиск на рівні станції (mm Hg)
P	Атмосферний тиск, приведений до середнього рівня моря (mm Hg)
P _a	Барична тенденція: зміна атмосферного тиска за останні 3 години (mm Hg)

U	Відносна вологість повітря (%) на висоті 2 м над поверхнею землі
DD	Напрямок вітру на висоті 10-12 м над поверхнею землі
Ff	Швидкість вітру на висоті 10-12 м над поверхнею землі
N	Загальна хмарність

3.1 Аналіз вхідних даних

Об'єднавши дані було створено новий набір даних. Новий набір містить щоденні показники споживання електроенергії і середні погодні показники. Для статистичного аналізу наведено часовий ряд змінної «Споживання» (рис 3.1).

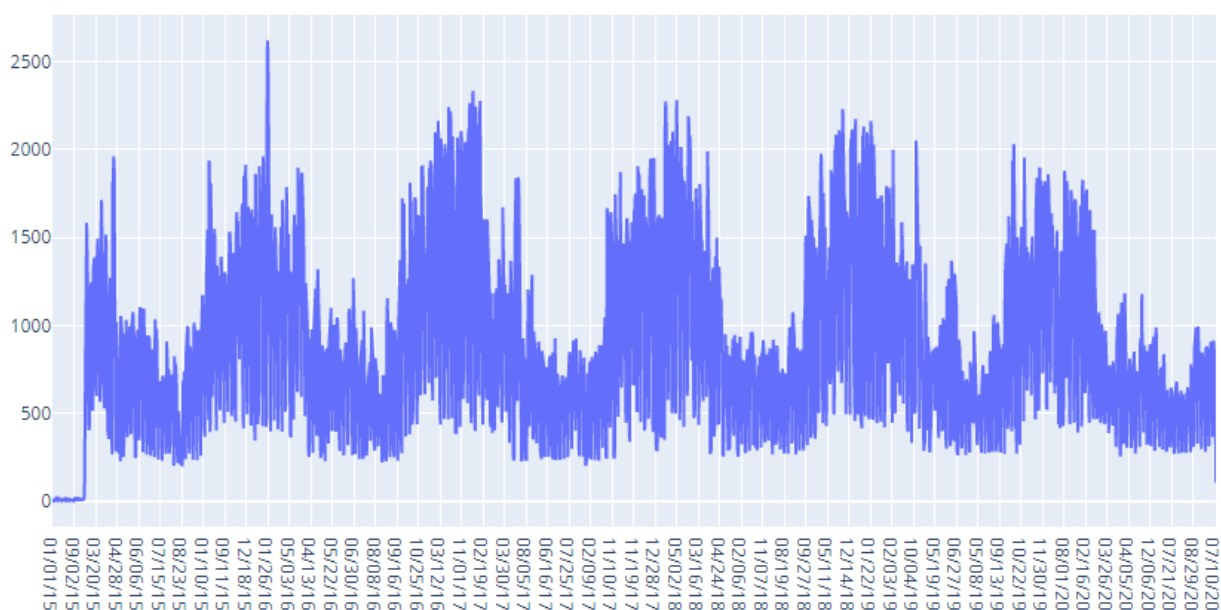


Рис. 3.1 – Графік часового ряду споживання електроенергії

Аналізуючи дані стало помітно аномально низькі показники споживання електроенергії на початку 2015 року, далі такі значення не спостерігались, отже ними можна знехтувати.

Досліджуючи дані стало очевидним наявність впливання на споживання електроенергії вихідних та святкових днів, вплив відпусток та канікул (рис 3.2, 3.3).

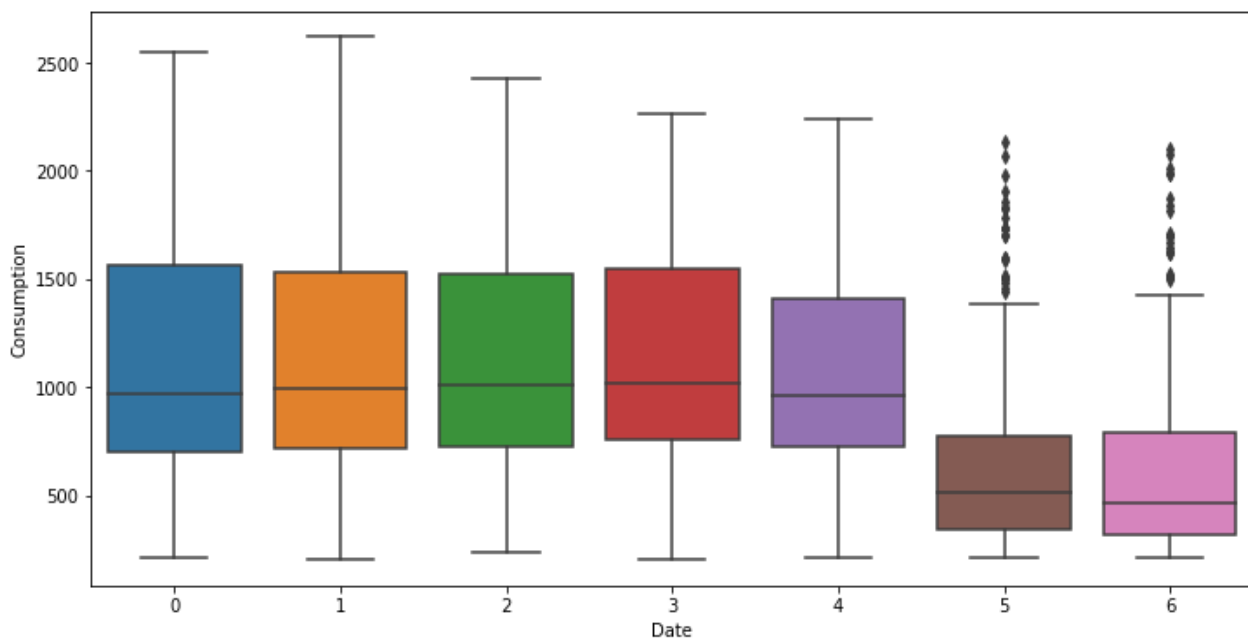


Рис. 3.2 – Споживання електроенергії по днях тижня

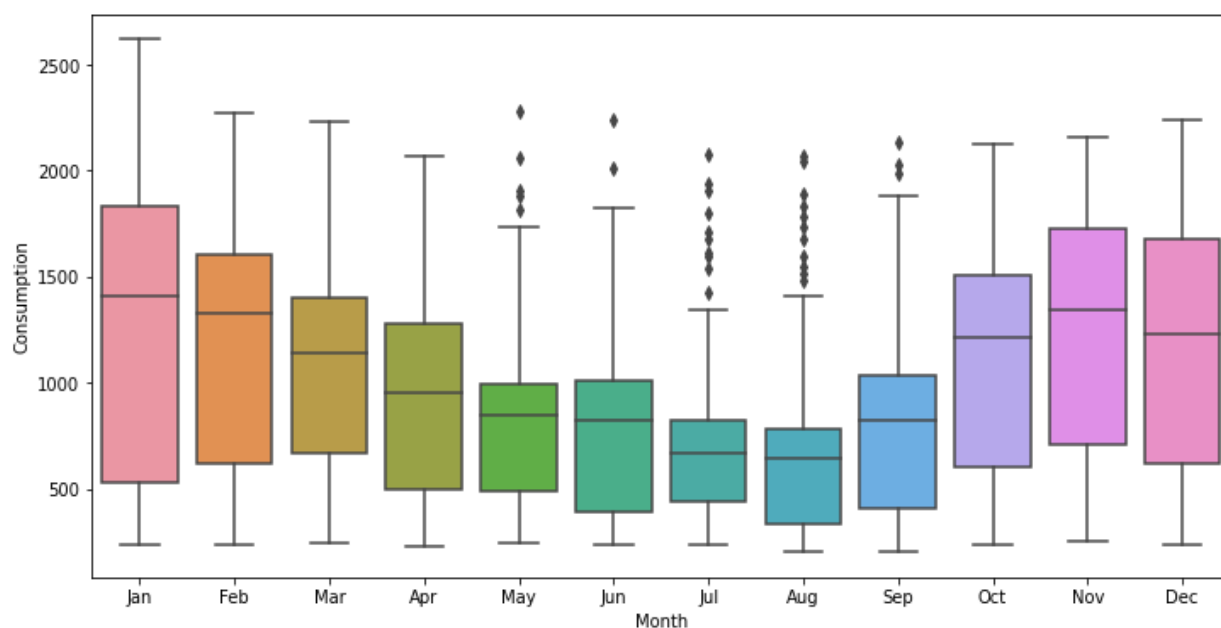


Рис. 3.3 – Споживання електроенергії по місяцях

Побудувавши декомпозицію часового ряду стало очевидним відсутності тренду, наявності сезонності і великої кількості залишків.

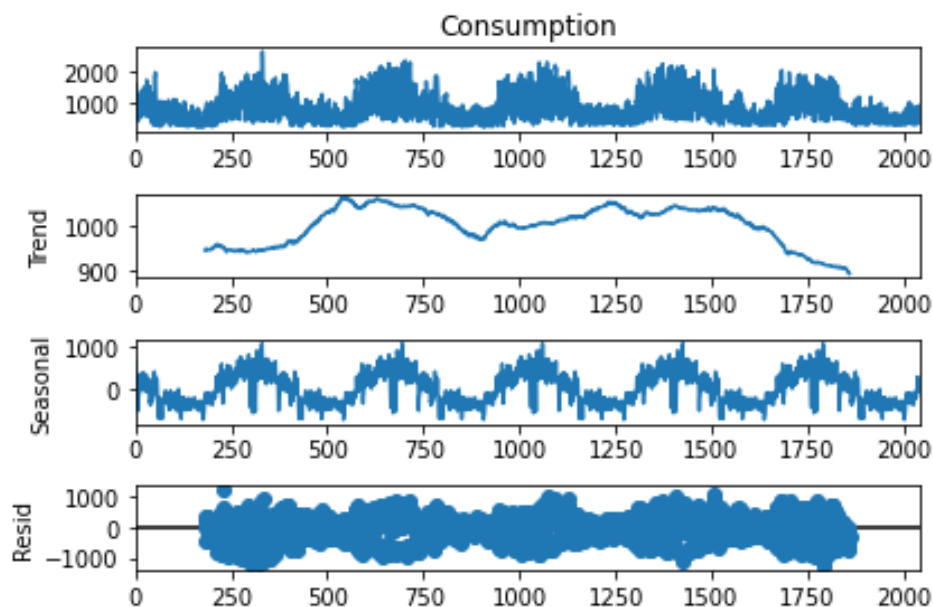


Рис. 3.4 – Декомпозиція споживання електроенергії

Для перевірки щільності розподілу використовується графік «Q-Q» (рис 3.5), він показує відхилення емпіричних даних від нормального розподілу.

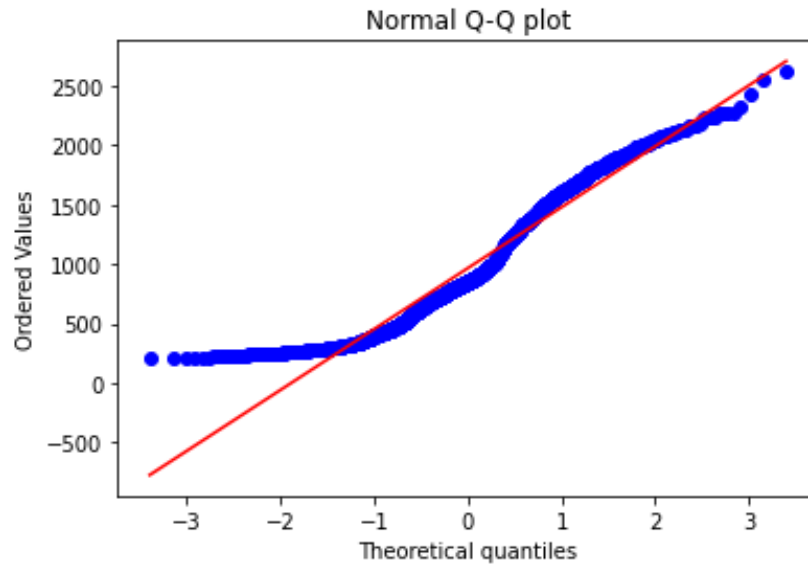


Рис. 3.5 – Графік розподілу часового ряду за допомогою квантильного розподілу

Для перевірки залежності змінних набору даних використовується кореляційна матриця (рис. 3.7).

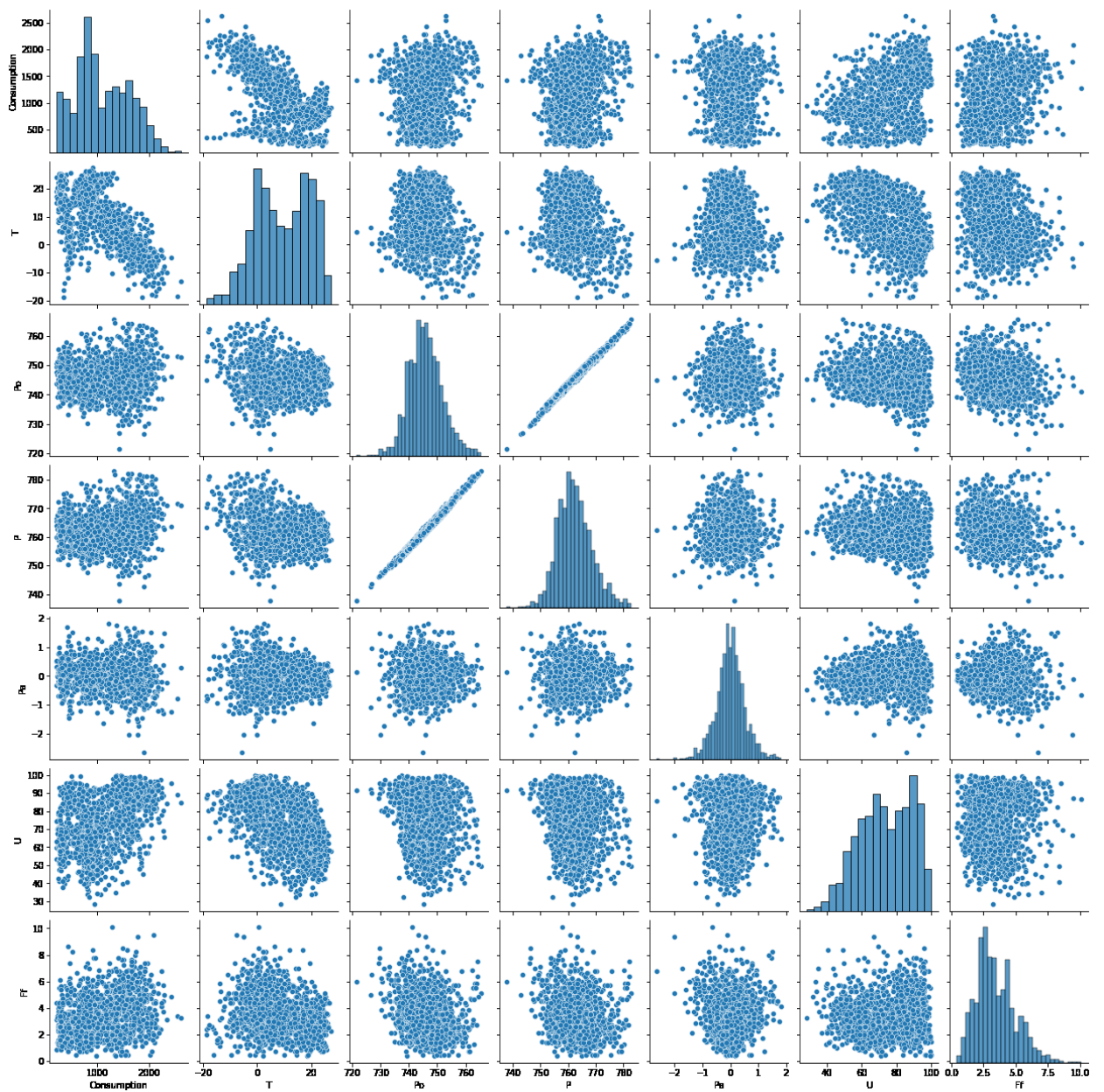


Рис. 3.6 – Діаграма розсіювання змінних набору даних

	Consumption	T	Po	P	Pa	U	Ff
Consumption	1.000000	-0.659832	0.139537	0.201545	-0.036218	0.414159	0.259364
T	-0.659832	1.000000	-0.208573	-0.302464	-0.047818	-0.527609	-0.288880
Po	0.139537	-0.208573	1.000000	0.995270	0.019948	-0.183598	-0.219937
P	0.201545	-0.302464	0.995270	1.000000	0.023784	-0.128207	-0.186047
Pa	-0.036218	-0.047818	0.019948	0.023784	1.000000	0.026969	-0.068214
U	0.414159	-0.527609	-0.183598	-0.128207	0.026969	1.000000	0.118262
Ff	0.259364	-0.288880	-0.219937	-0.186047	-0.068214	0.118262	1.000000

Рис. 3.7 – Кореляційна залежність між змінними набору даних

Згідно кореляційної матриці (рис. 3.7), найбільш корелюємі змінні це «Consumption» і «Т» (споживання електроенергії і температура).

Побувавши діаграму розсіювання (рис. 3.6) дало змогу побачити як вихідні і святкові дні впливають на споживання електроенергії. (рис. 3.8)

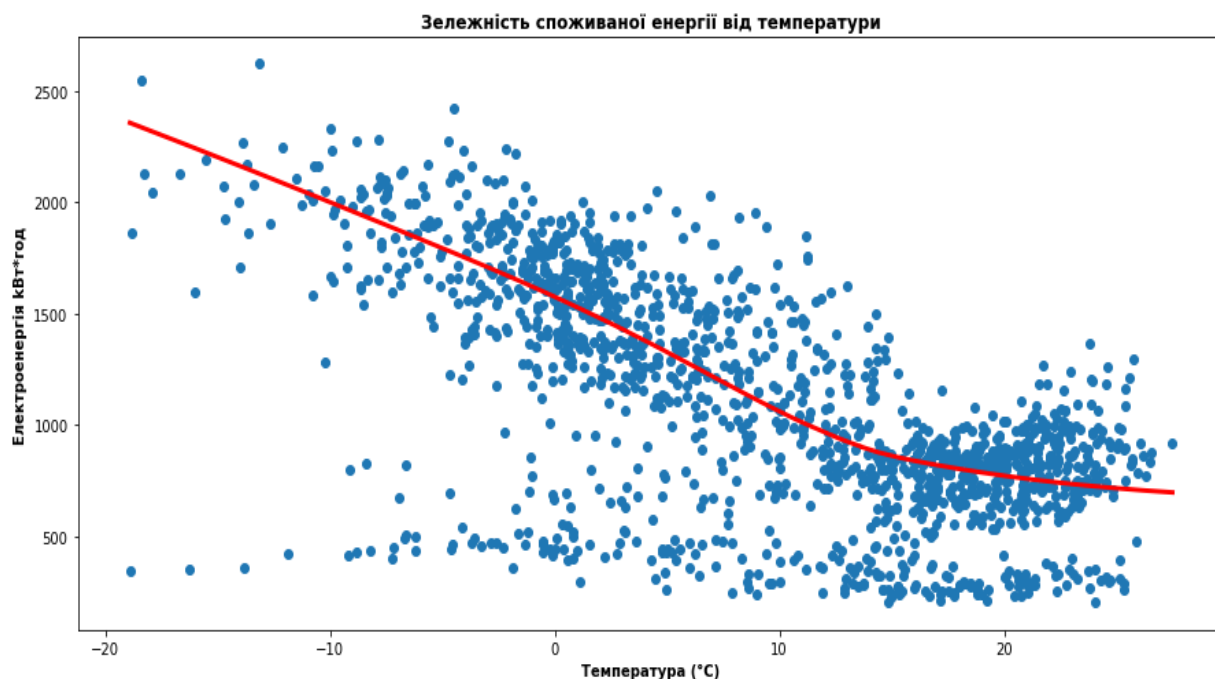


Рис. 3.8 – Діаграма розсіювання змінних «Consumption» і «Т»

3.2 Модель на основі регресійного аналізу

Провівши аналіз даних стало очевидним використати модель SARIMAX, ця модель враховує змінні моделі ARIMA з додавання сезонних показників.

Для вибору кращих параметрів моделі SARIMAX, використовується інформаційний критерій AIC. При значення $p=7$, $d=1$, $q=2$, $P=5$, $D=1$, $Q=1$, $S=12$ показник Акаїке показав найменші значення (рис. 3.9).

SARIMAX Results

Dep. Variable:	Consumption	No. Observations:	1767
Model:	SARIMAX(7, 1, 2)x(5, 1, [1], 12)	Log Likelihood	-12045.339
Date:	Tue, 08 Dec 2020	AIC	24124.678
Time:	17:03:53	BIC	24217.000
Sample:	0	HQIC	24158.869
	- 1767		

Covariance Type: opg

	coef	std err	z	P> z	[0.025	0.975]
T	-14.3467	3.440	-4.170	0.000	-21.089	-7.604
ar.L1	-0.3046	0.207	-1.470	0.142	-0.711	0.102
ar.L2	-0.5495	0.184	-2.983	0.003	-0.911	-0.189
ar.L3	-0.3962	0.176	-2.249	0.024	-0.741	-0.051
ar.L4	-0.4356	0.175	-2.485	0.013	-0.779	-0.092
ar.L5	-0.4764	0.167	-2.853	0.004	-0.804	-0.149
ar.L6	-0.2889	0.178	-1.623	0.105	-0.638	0.060
ar.L7	0.3057	0.151	2.031	0.042	0.011	0.601
ma.L1	-0.1737	0.204	-0.851	0.395	-0.574	0.227
ma.L2	0.0864	0.105	0.820	0.412	-0.120	0.293
ar.S.L12	-0.0052	0.052	-0.101	0.920	-0.107	0.096
ar.S.L24	-0.0184	0.055	-0.334	0.739	-0.126	0.090
ar.S.L36	-0.0692	0.055	-1.262	0.207	-0.177	0.038
ar.S.L48	-0.0692	0.051	-1.356	0.175	-0.169	0.031
ar.S.L60	0.0007	0.048	0.014	0.989	-0.094	0.095
ma.S.L12	-0.9709	0.022	-43.931	0.000	-1.014	-0.928
sigma2	1.498e+05	6887.903	21.745	0.000	1.36e+05	1.63e+05

Ljung-Box (L1) (Q): 0.00 Jarque-Bera (JB): 895.99
 Prob(Q): 0.99 Prob(JB): 0.00
 Heteroskedasticity (H): 1.19 Skew: -0.02
 Prob(H) (two-sided): 0.04 Kurtosis: 6.57

Рис. 3.9 – Результати моделі SARIMAX

Результати діагностичних тестів на SARIMAX (7,1,2) (5,1,1,12) наведені на рисунку 3.10. Стандартизовані залишки та корелограми показують, що залишки - це білий шум. Лінія KDE на гістограмі та графіку Q-Q показують, що залишки приблизно дотримуються нормального розподілу.

Результати діагностичних тестів задовольняють припущення щодо моделювання SARIMAX.

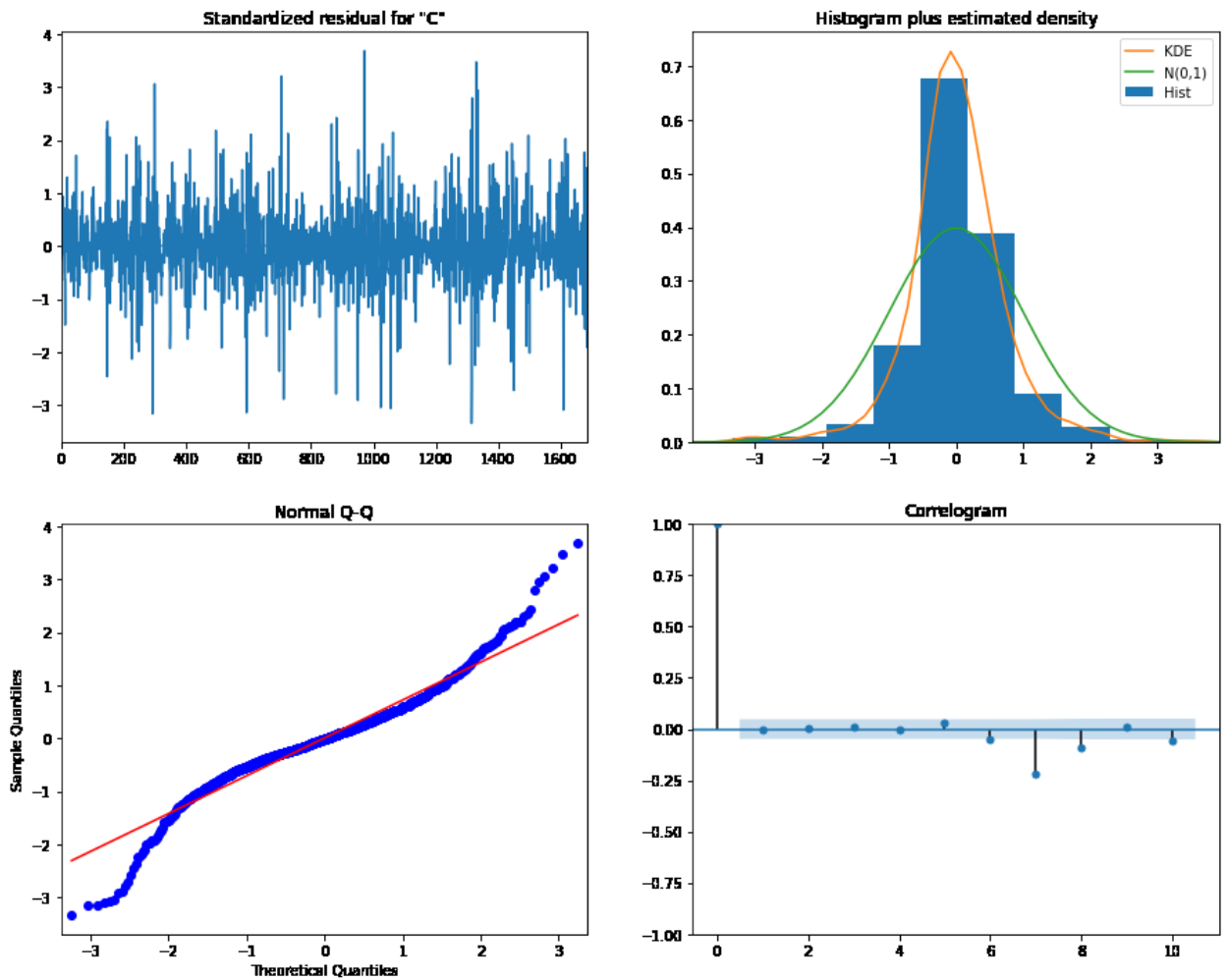


Рис. 3.10 – Діагностичні тести на залишки моделі SARIMAX

Середньодобове споживання електроенергії з 2015 по 2019 рік було прогнозовано з використанням моделі SARIMAX на тренувальних даних. Дані за 2020 рік використовувались як набір тестувань і не були включені в етапи моделювання. На рисунку 3.11 показано прогнозовані значення накладені на фактичні значення. Синя лінія показує фактичні значення, а помаранчева - прогнозовані значення в кВт-год.

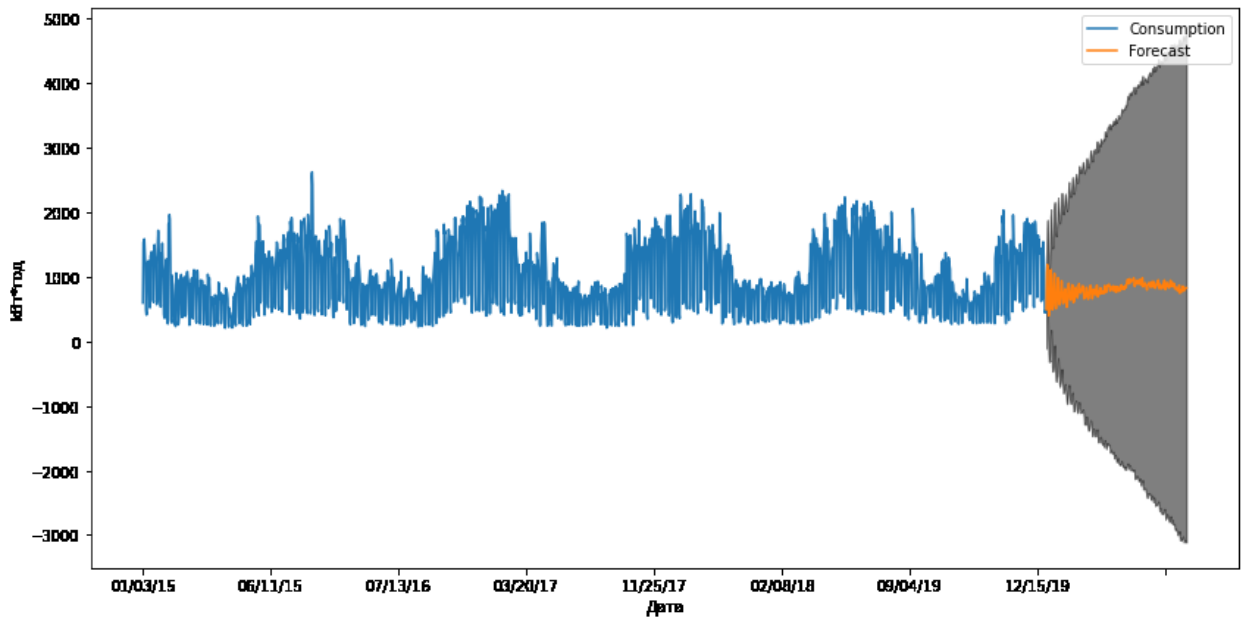


Рис. 3.11 – Прогнозовані та фактичні значення середньодобового споживання електроенергії

Перевірка прогнозової якості наведена в таблиці 3.2.

Таблиця 3.2

Результати перевірки прогнозової якості моделі
SARIMAX (7,1,2) (5,1,1,12)

	Навчальні дані	Тестові дані
MAE	206.13	344.74
RMSE	298.32	298.93
MAPE (%)	28.67	59.19
AIC	24127.67	
BIC	27217.00	

В результаті перевірки виявлено що модель SARIMAX (7,1,2) (5,1,1,12) не задовольняє вимогам, оцінка точності моделі низька.

3.3 Моделі експоненціального згладжування

Одним із методів експоненціального згладжування є метод Хольта-Уінтерс. Цей метод враховує сезонні характеристики часового ряду. Для вибору кращих параметрів методу Хольта-Уінтерс використовується інформаційний критерій Акаїке, зі значенням періоду $P=365$ показники інформаційного критерію найкращі (рис. 3.12).

ExponentialSmoothing Model Results			
Dep. Variable:	Consumption	No. Observations:	1767
Model:	ExponentialSmoothing	SSE	422064281.491
Optimized:	True	AIC	22615.874
Trend:	None	BIC	24625.947
Seasonal:	Multiplicative	AICC	22811.335
Seasonal Periods:	365	Date:	Wed, 09 Dec 2020
Box-Cox:	False	Time:	10:53:21
Box-Cox Coeff.:	None		
	coeff	code	optimized
smoothing_level	0.0050000	alpha	True
smoothing_seasonal	0.6396429	gamma	True
initial_level	1501.1280	l.0	True
initial_seasons.0	0.3952414	s.0	True
initial_seasons.1	1.0196945	s.1	True
initial_seasons.2	1.0208270	s.2	True
initial_seasons.3	1.0541539	s.3	True
initial_seasons.4	0.9389046	s.4	True
initial_seasons.5	0.8010509	s.5	True
initial_seasons.6	0.3187763	s.6	True
initial_seasons.7	0.2709456	s.7	True
initial_seasons.8	0.2849511	s.8	True
initial_seasons.9	0.8152323	s.9	True

Рис. 3.12 – Результати моделі Хольта-Уінтерс

Середньодобове споживання електроенергії з 2015 по 2019 рік було прогнозовано з використанням адитивної (рис. 3.13) і мультиплікативної (рис. 3.14) моделі Хольта-Уінтерса на тренувальних даних. Дані за 2020 рік використовувались як набір тестувань і не були включені в етапи моделювання. На графіках (рис. 3.13, 3.14) показано прогнозовані значення накладені на фактичні значення. Зелена лінія показує фактичні значення, а червона - прогнозовані значення в кВт-год.

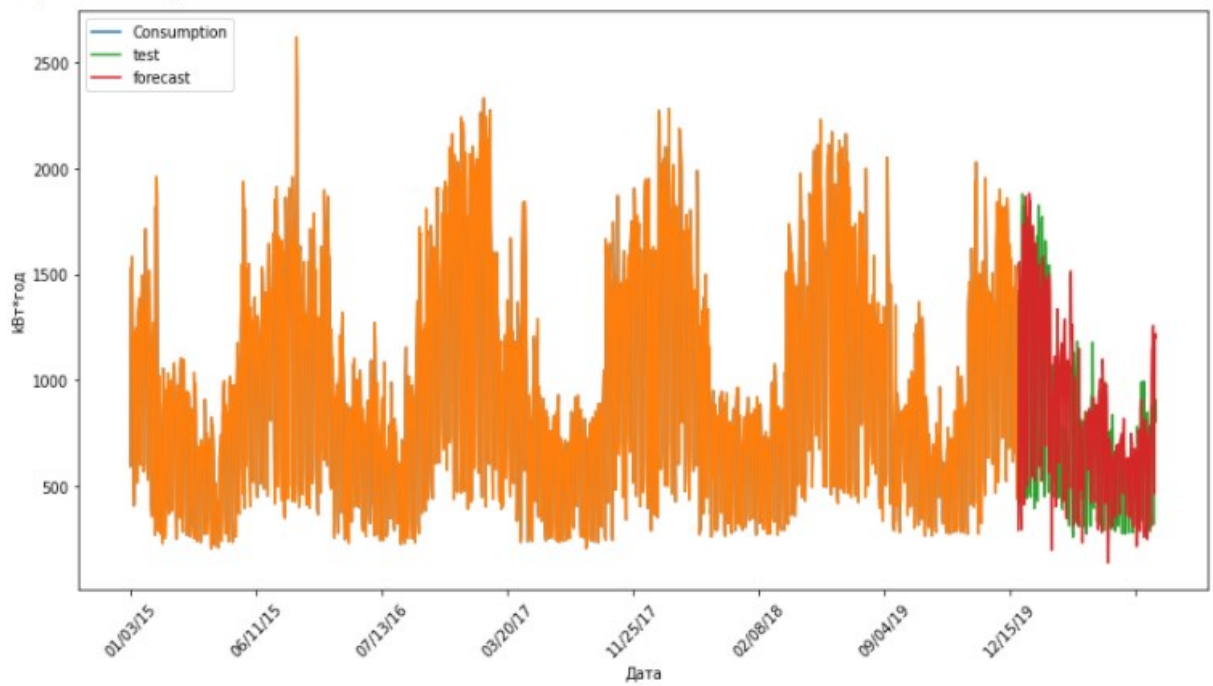


Рис. 3.13 – Прогнозовані та фактичні значення з використанням адитивної сезонності моделі Хольта-Уінтерса

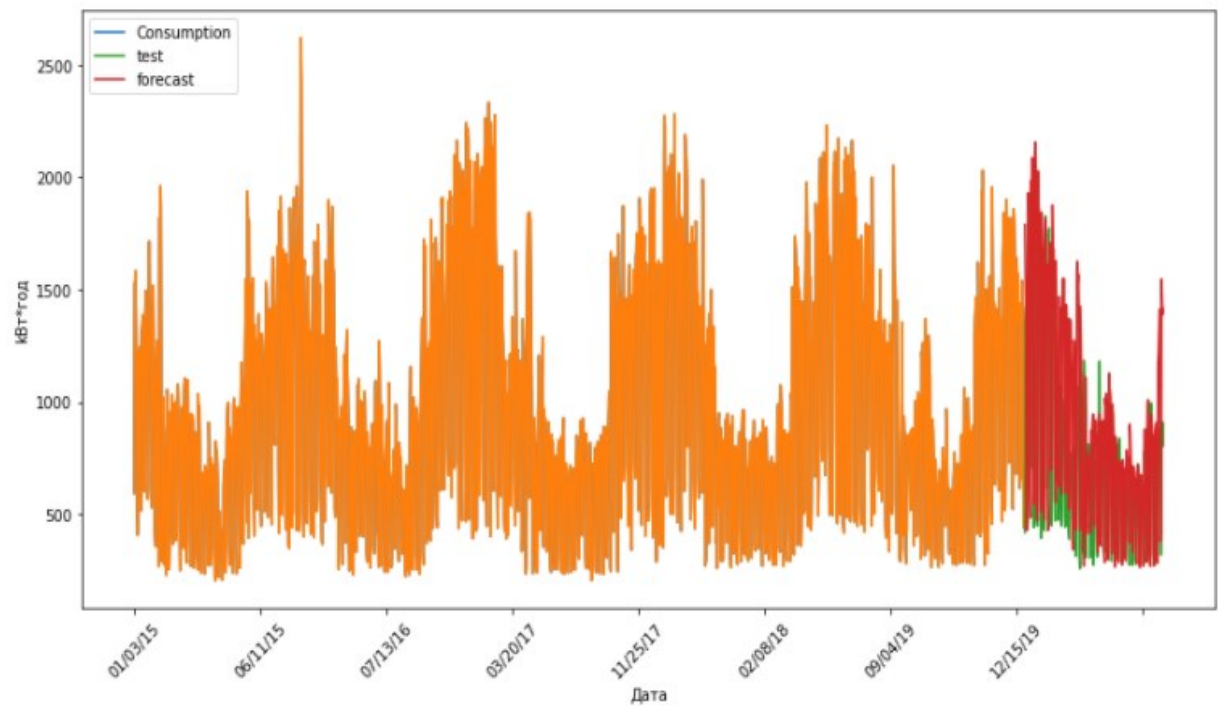


Рис. 3.14 – Прогнозовані та фактичні значення з використанням мультиплікативної сезонності моделі Хольта-Уінтерса

Результати діагностичних тестів на залишки моделі Хольта-Уінтерса на 3.15. Стандартизовані залишки та корелограми показують, що залишки - це білий шум. Лінія KDE на гістограмі та графіку Q-Q показують, що залишки приблизно дотримуються нормального розподілу.

Результати діагностичних тестів задовольняють припущення щодо моделювання Хольта-Уінтерса.

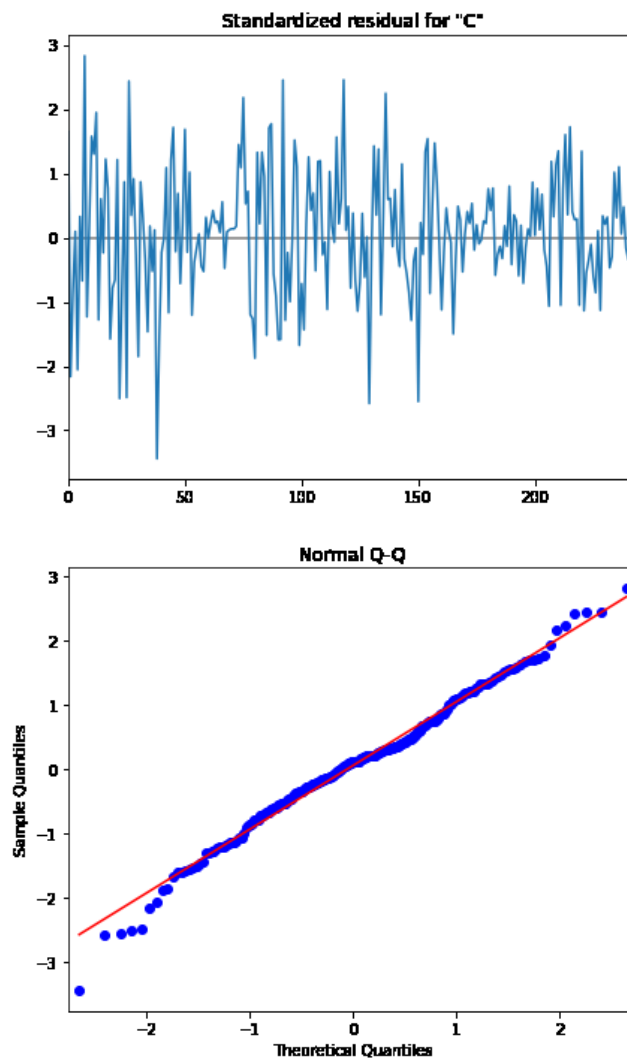


Рис. 3.15 – Залишки моделі Хольта-Уінтерса

Гістограма залишків наведено на рисунку 3.16, по ній можна зробити висновки що залишки відповідають нормальному закону розподілу.

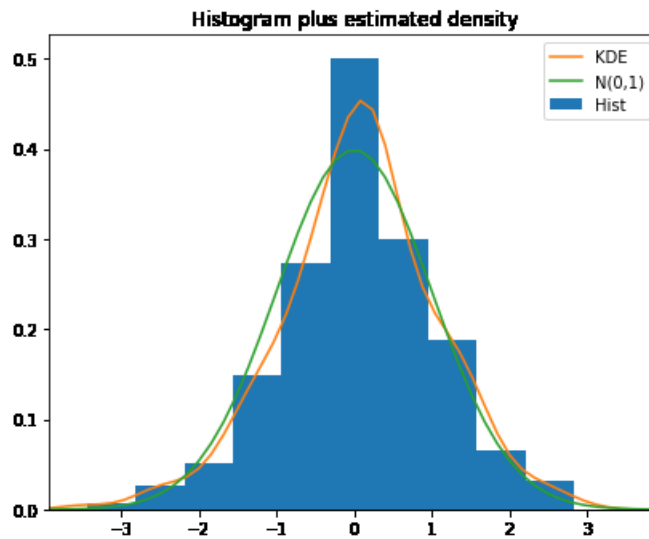


Рис. 3.16 – Гістограми залишків моделі Хольта-Уінтерса

Перевірка прогнозної якості наведена в таблиці 3.3.

Таблиця 3.3

Результати перевірки прогнозної якості моделей Хольта-Уінтерса

	Навчальні дані	Тестові дані	Навчальні дані	Тестові дані
	Адитивна сезонність		Мультиплікативна сезонність	
MAE	214.47	285.00	181.84	212.64
RMSE	300.02	231.56	193.99	80.03
MAPE (%)	31.20	35.47	27.49	31.77
AIC	22615.874		19401.263	
BIC	24265.947		21405.652	

3.4 Моделі на основі машинного навчання

3.4.1 Модель Prophet

Модель Prophet враховує сезонну складову часового ряду і впливання вихідних і свят. Вхідними даними моделі є показники електроспоживання, температури, значення дня тижня і період сезонності ($P=365$).

Вказавши вхідні дані, одразу будується модель, прогнозовані значення наведено на рисунку 3.17.

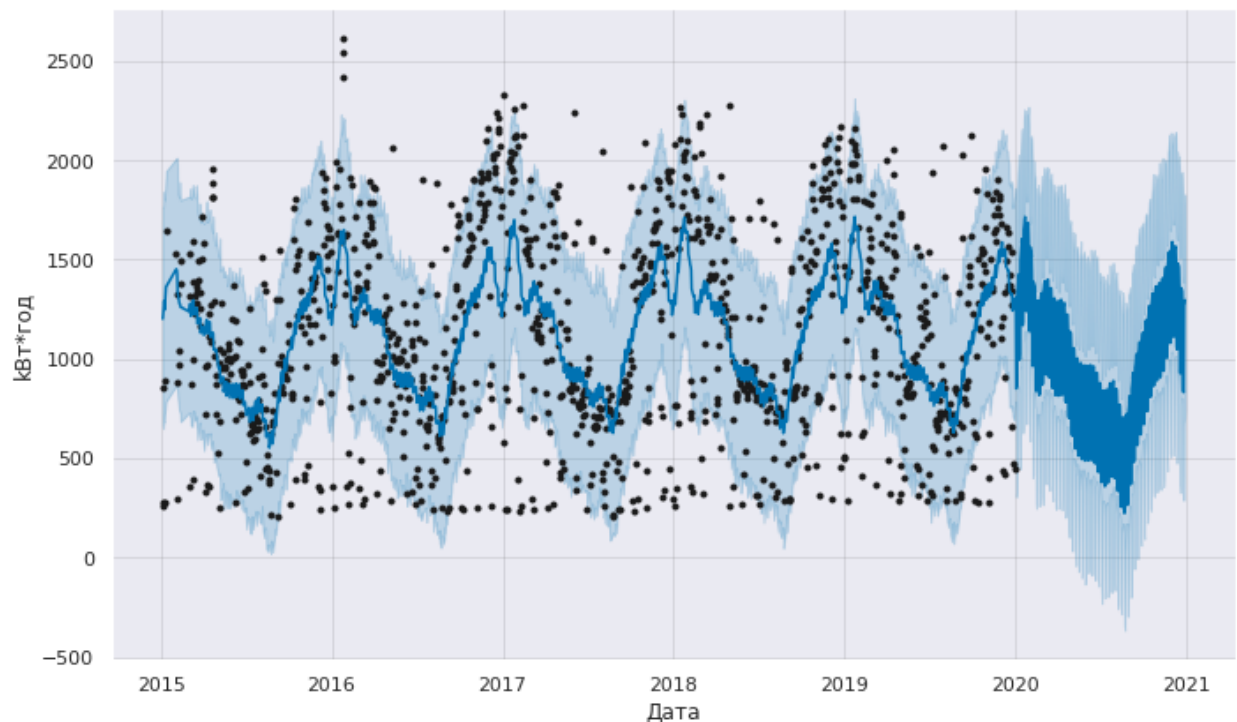


Рис. 3.17 – Прогнозовані та фактичні значення з використанням моделі Prophet

Як було сказано раніше, Prophet враховує такі компоненти часового ряду як тренд, сезонність і вплив вихідних днів (рис 3.18). Модель помилково

початок даних розуміє як тренд, так як на початку року спостерігається зростання споживання електроенергії.

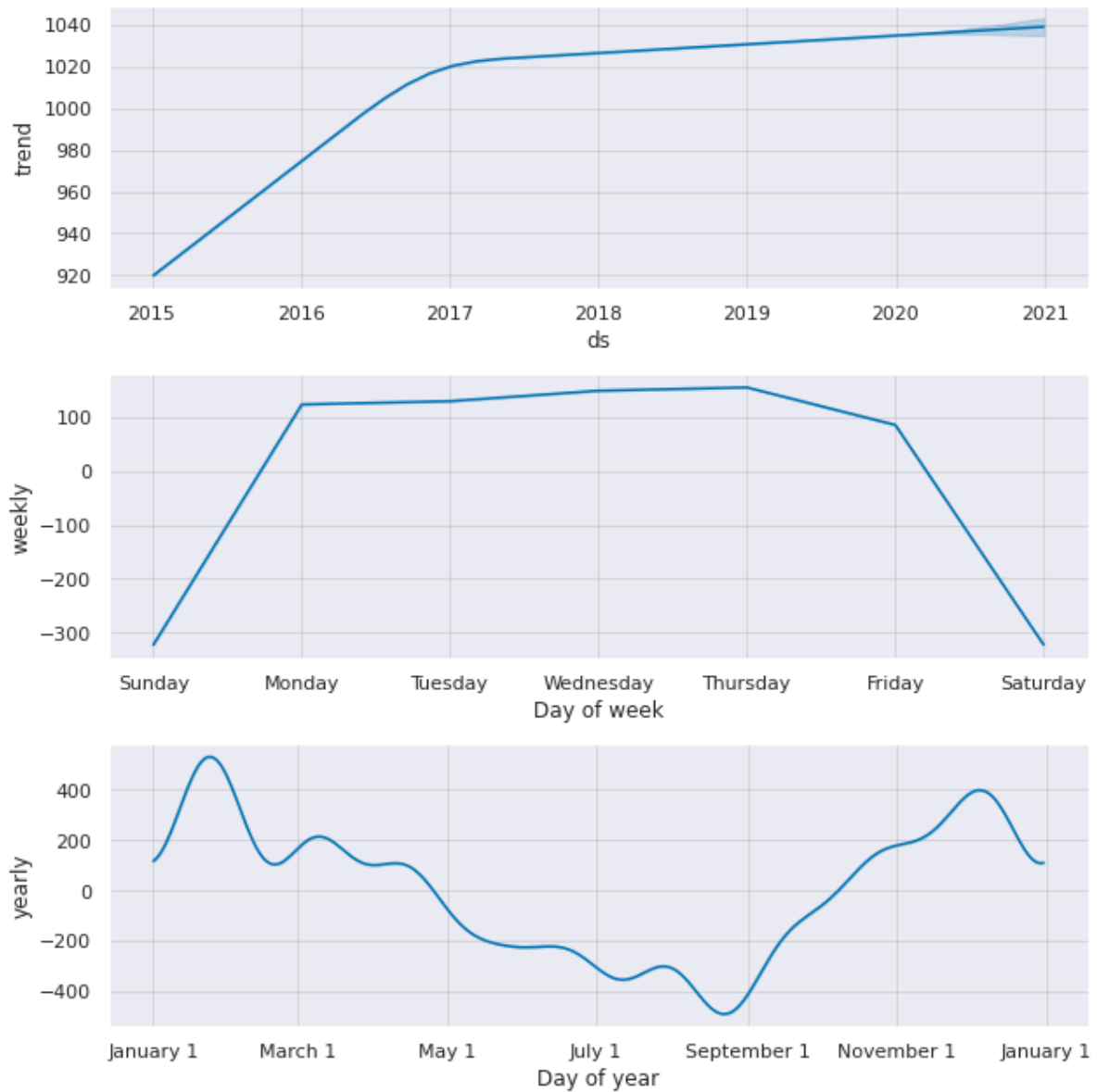


Рис. 3.18 – Компоненти часового ряду, котрі виділяє модель Prophet

На рисунку 3.19 зображено прогнозоване споживання електроенергії на місяць вперед.

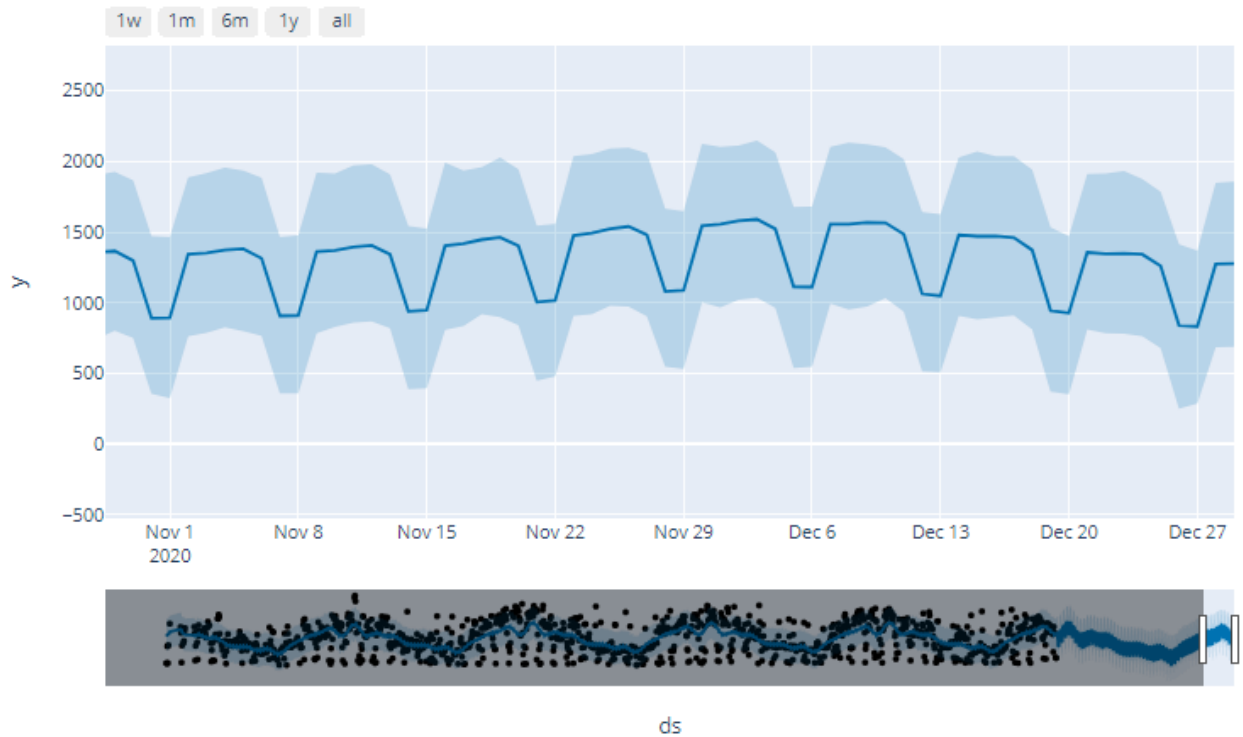


Рис. 3.19 – Прогнозування споживання електроенергії на місяць з використанням моделі Prophet

Перевірка прогностної якості моделі наведено в таблиці 3.3.

Таблиця 3.3

Результати перевірки прогностної якості моделі Prophet

	Навчальні дані	Тестові дані
MAE	189.30	276.33
RMSE	324.87	416.43
MAPE (%)	33.04	43.19

3.4.2 Рекурентні нейронні мережі

Наступним кроком є розробка моделі нейронної мережі для аналізу часових рядів. За основу було взято рекурентну нейронну мережу (RNN). RNN обробляє послідовність даних у часі.

Одним із видів рекурентних нейронних мереж є мережі довгої короткострокової пам'яті (LSTM). LSTM-мережа добре пристосована до навчання на задачах класифікації, обробки і прогнозування часових рядів у випадках, коли важливі події розділені часовими лагами з невизначеною тривалістю і кордонами.

Вхідними даними моделі є показники електроспоживання, результати моделі наведено на рисунку 3.20.

```
Model: "sequential_11"
```

Layer (type)	Output Shape	Param #
lstm_10 (LSTM)	(None, 100)	52400
dropout_3 (Dropout)	(None, 100)	0
dense_10 (Dense)	(None, 1)	101

```
Total params: 52,501  
Trainable params: 52,501  
Non-trainable params: 0
```

Рис. 3.20 – Результати моделі LSTM

Функція втрат знаходиться в центрі нейронної мережі. Вона використовується для розрахунку помилки між реальними і прогнозованими значеннями. Глобальна мета – мінімізувати цю помилку. Таким чином, функція втрат ефективно наближає навчання нейронної мережі до цієї мети.

Функція втрат вимірює «наскільки хороша» нейронна мережа щодо даної навчальної вибірки і прогнозованих значень. Вона також може залежати від таких змінних, як ваги і зміщення.

Функція втрат одномірна і не є вектором, оскільки вона оцінює, наскільки добре нейронна мережа працює в цілому.

Епохи (epochs) – кількість раз проходження даних через нейронну мережу в прямому і зворотному напрямку. Так як одна epoch занадто велика для комп'ютера, набір даних діляться на маленькі партії (batches).

Залежність функції витрат від епохи наведено на рисунку 3.21.

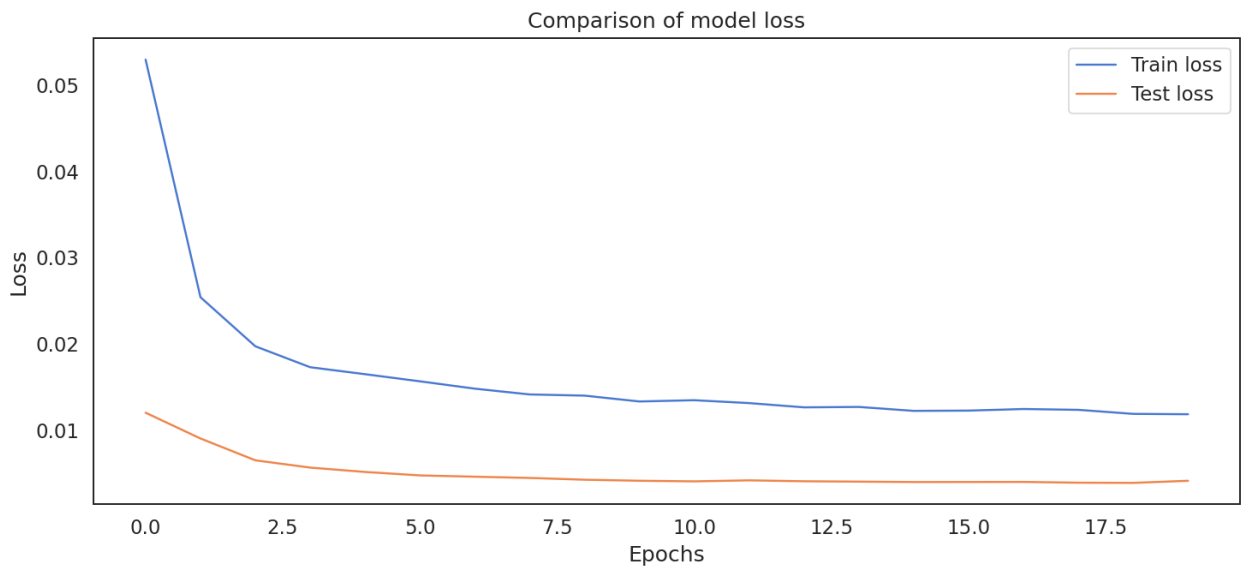


Рис. 3.21 – Залежність втрат моделі від кількості епох навчання

Навчання нейронної мережі проводилось з характеристиками: epochs = 20, batches = 100 на навчальних даних. Прогнозоване споживання електроенергії та тестове значення наведено на рисунку 3.22.

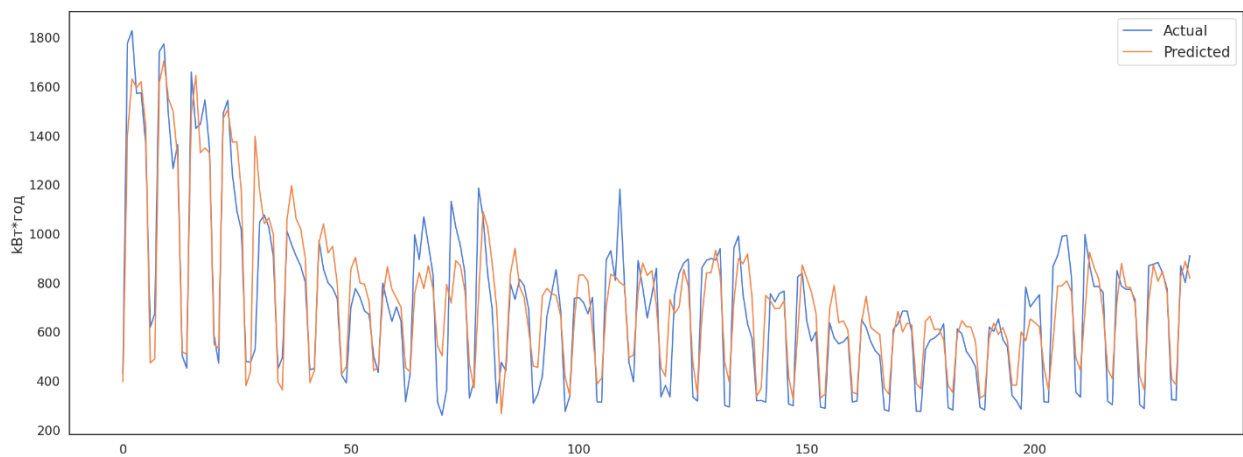


Рис. 3.22 – Прогнозовані та фактичні значення споживання електроенергії використовуючи LSTM-мережу

Перевірка прогнозної якості LSTM-мережі наведено на рисунку 3.4.

Таблиця 3.4

Результати перевірки прогнозної якості моделі LSTM

	Навчальні дані	Тестові дані
MAE	179.91	110.95
RMSE	262.11	150.29
MAPE (%)	25.60	19.53

3.5 Порівняльний аналіз моделей прогнозування

Побудувавши моделі прогнозування часових рядів виникає нагальна потреба проведення порівняльного аналізу отриманих результатів. Порівнюються моделі: сезонна авторегресійна інтегрована ковзного середнього з використанням екзогенних змінних (SARIMAX), експоненціального згладжування (Хольта-Уінтерса), Prophet і рекурентна нейронна мережа (LSTM).

Оцінюються похибки навчальних та тестових наборів даних. Для всіх моделей брався навчальний набір даних, який включав показники споживання електроенергії за період 2015 – 2019 рр. Похибка моделей прогнозування на навчальних даних наведено в таблиці 3.5.

Таблиця 3.5

Похибка моделей прогнозування на навчальних даних

	MAE	RMSE	MAPE (%)
SARIMAX	206.13	298.32	28.67
Хольта-Уінтерса	181.84	193.99	27.49
Prophet	189.30	224.87	33.04
LSTM	179.91	262.11	25.60

Моделі тестувались на наборі даних який містить показники споживання електроенергії за 2020 рік (табл. 3.6).

Таблиця 3.6

Похибка моделей прогнозування на тестових даних

	MAE	RMSE	MAPE (%)
SARIMAX	344.74	298.93	59.19
Хольта-Уінтерса	212.64	80.03	31.77
Prophet	316.43	416.43	43.19
LSTM	110.95	150.29	19.53

Порівнюючи моделі прогнозування часових рядів, можна побачити, що моделі машинного навчання забезпечують кращі показники прогнозової якості такі як MAE, RMSE, MAPE.

Великі значення похибки показують вплив на споживання електроенергії в 2020 зумовлене карантинном (рис. 3.23).

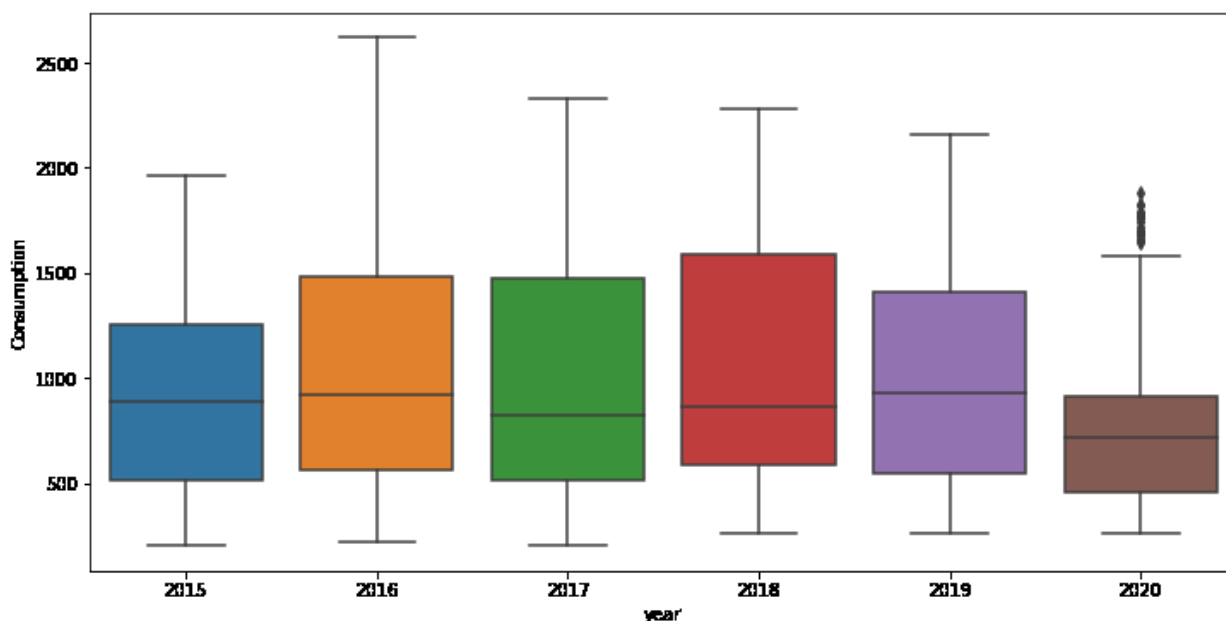


Рис. 3.23 – Показники споживання електроенергії по роках

Розглянуті моделі прогнозування мають свої переваги і недоліки у прогнозуванні споживання електроенергії. Для короткострокового прогнозу можна використовувати авторегресійні моделі прогнозування. Для середньострокового прогнозу можна використовувати моделі експоненціального згладжування (Хольта-Вінтерса) і Prophet. Рекурентні нейронні мережі є універсальним методом прогнозування споживання електроенергії, цей метод показує однаково точні результати прогнозування як для короткострокових та і для довгострокових інтервалів прогнозування.

В залежності від погодних умов, дня тижня, поточного місяця значення точність прогнозування покращується. Використовуючи поєднання цих

факторів можна зробити якісний прогноз споживання електроенергії на майбутній рік (рис. 3.24).

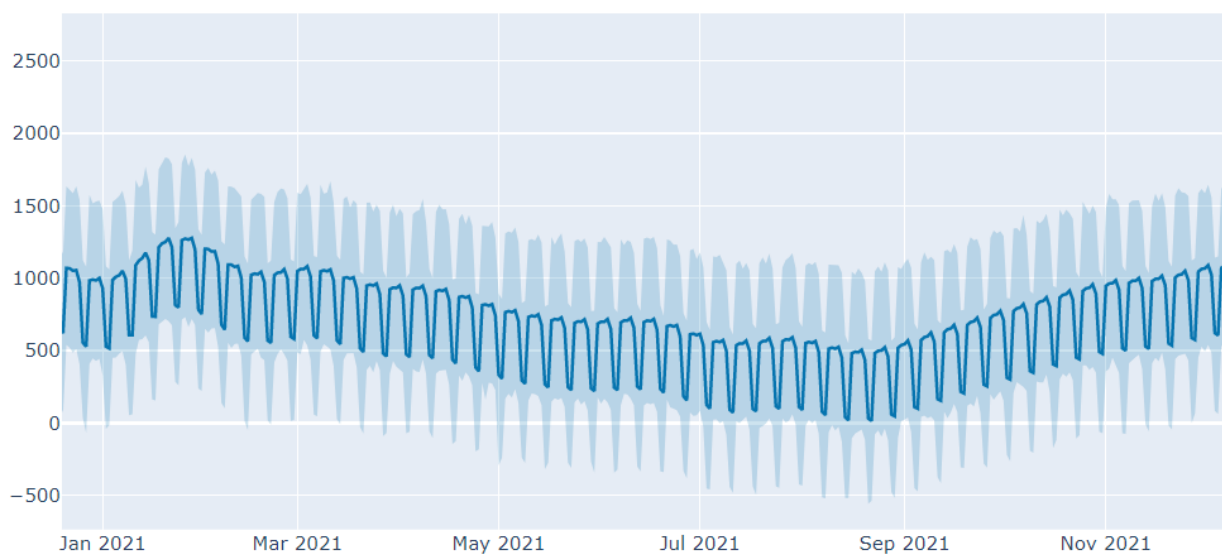


Рис. 3.24 – Прогноз споживання електроенергії головним корпусом університету на 2021 рік

ВИСНОВКИ

Зроблено аналіз вихідної інформації по споживанню електричної енергії головним корпусом Сумського державного університету. Проведено аналіз елементів що впливають на споживання електроенергії.

Визначено стійку статистичну залежність між обсягами споживання електроенергії в певні місяці та дні тижня.

Проведено перевірку прогнозних якостей побудованих моделей, обрано моделі котрі показують найкращу прогнозну якість серед різних типів прогнозних моделей.

Досліджено моделі прогнозування часових рядів і визначено, що застосування класичних методів експоненціального згладжування і регресійного аналізу дає велику похибку прогнозу.

Розраховано очікуваного обсягу споживання електричної енергії до кінця 2020 року, який показує, що максимальна похибка прогнозу становить 19.53%

Пропонована методика прогнозування споживання електричної енергії на основі даних університету, котра може застосовуватись до інших підприємств, установ, організацій.

ПЕРЕЛІК ДЖЕРЕЛ ІНФОРМАЦІЇ

1. Sustainable energy for developing countries. journals. URL: <https://journals.openedition.org/sapiens/823>. (date of access: 02.12.2020).
2. Energy Country Profiles. Our World in Data. URL: <https://ourworldindata.org/energy> (date of access: 02.12.2020).
3. Global Power Sector Overview. Nebras Power. URL: https://nebras-power.com/about/global-power-sector-overview?language_content_entity=en (date of access: 02.12.2020).
4. Ladislav S. Transforming the U.S. and EU Electric Power Sectors. csis. URL: <https://www.csis.org/analysis/transforming-us-and-eu-electric-power-sectors> (date of access: 02.12.2020).
5. Короткострокове прогнозування споживання електричної енергії на основі інформації АСКОЕ для АСПР (автоматизованої системи планових розрахунків) регіонального диспетчерського управління енергосистеми. "Мала енергетика-2005". 2005. Т. 4, № 7. С. 285—297. URL: <http://www.combienergy.ru/stat/990-Kratkosrochnoe-prognozirovanie-potrebleniya-elektricheskoy> (дата звернення: 02.12.2020).
6. Gourieroux C., Monfort A. Time Series and Dynamic Models (Themes in Modern Econometrics) : підручник. 467th ed. Лондон, 1996. 688 p.
7. Spall J. Bayesian Analysis of Time Series and Dynamic Models : підручник. 978th ed. Лондон : CRC Press, 1988. 576 p.
8. Kalimoldayev M., Marinich T. Analysis of modern approaches for the prediction of electric energy consumption. Open Engineering. 2020. Vol. 1, no. 10. P. 350—361. URL: <https://doi.org/10.1515/eng-2020-0028> (date of access: 02.12.2020).
9. Cools M., Moons E. Investigating the Variability in Daily Traffic Counts through use of ARIMAX and SARIMAX Models: Assessing the Effect of Holidays on Two Site Locations. Transportation Research Record: Journal of

- the Transportation Research Board Journal Indexing and Metrics. 2009. Vol. 4, no. 6. P. 57—66. URL: <https://doi.org/10.3141/2136-07> (date of access: 02.12.2020).
10. Alon I. Time Series Model Forecasting of Boot Using Holt, Winter and Decomposition Method. *Journal of Industrial Safety Engineering*. 2015. Vol. 2, no. 2. P. 23—31. URL: https://www.researchgate.net/publication/327847645_Time_Series_Model_Forecasting_of_Boot_Using_HoltWinter_and_Decomposition_Method/link/5ba9273d92851ca9ed2255a1/download (date of access: 03.12.2020).
 11. Zhang L., Wu J., Liu H. Policies to enhance the drivers of green housing development in China, *Energy Policy*, 2018, Vol 4 no 121 P, 225 – 235. URL: <https://doi.org/10.1016/j.enpol.2018.06.029>. (date of access: 03.12.2020).
 12. Hyndman R., Athanasopoulos G. *Forecasting: principles and practice : підручник*. 2nd ed. Australia : Melbourne, 2018. 292 p. URL: <https://otexts.com/fpp2/> (date of access: 03.12.2020).
 13. De Livera A., Hyndman R., and Snyder R. Forecasting time series with complex seasonal patterns using exponential smoothing. *Journal of the American Statistical Association*, 2011. Vol. 106, no. 46. P. 1513 –1527. URL: <https://doi.org/10.1198/jasa.2011.tm09771>.
 14. Sean J. Taylor and Benjamin Letham. Forecasting at scale. *The American Statistician*, (just-accepted), 2017. Vol 10 no 3. P. 37-45. URL: <https://doi.org/10.1080/00031305.2017.1380080>.
 15. Sean J., Taylor T. Forecasting at Scale. *PeerJ Preprints*. 2017. Vol. 12, no. 4. P. 4—41. URL: <https://peerj.com/preprints/3190/> (date of access: 03.12.2020).
 16. Ching C., Zhang P. A comparative study of linear and nonlinear models for aggregate retail sales forecasting. *International Journal of Production Economics*. 2003. Vol. 86, no. 3. P. 217—231.

17. Sena D., Nagwani N. Application of time series based prediction model to forecast per capita disposable income. International Conference on Advance Computing Conference. 2015. Vol. 15, no. 10. P. 454—457.
18. Fisher's F-Test in R Programming. Geeksforgeeks. URL: <https://www.geeksforgeeks.org/fishers-f-test-in-r-programming/> (date of access: 02.12.2020).
19. Kullback S., Leibler R. On information and sufficiency // The annals of mathematical statistics. — 1951. — Vol. 22, no. 1. — P. 79–86.
20. Вербік М. Тести на наявність автокореляції залишків першого порядку. studme. URL: https://studme.org/72661/ekonomika/testy_nalichie_avtokorrelyatsii_ostatkov_pervogo_poryadka (дата звернення: 02.12.2020).
21. Агашина Ю. Перевірка адекватності і засоби верифікації прогнозних моделей. Современные наукоемкие технологии. 2007. Т. 2. С. 75—79. URL: <http://www.top-technologies.ru/ru/article/view?id=24252> (дата звернення: 02.12.2020).
22. Han P., Wang K., Zhand S. Drought forecasting based on the sensing data using ARIMA model. Math Comput Model. 2010.- Vol. 51(11). – P. 1398-1403
23. Su Y. Electricity demand in industrial and services sectors in Taiwan. Energy Efficiency. 2018, Vol 11 no 6. 1541–1557. URL: <https://doi.org/10.1007/s12053-018-9615-y>.
24. Carvalho A. Energy efficiency in transition economies: A stochastic frontier approach. Economics of Transition. 2018. Vol 26 no 3. P. 553– 578. URL: <https://doi.org/10.1111/ecot.12152>.
25. Bianco V. Linear Regression Models to Forecast Electricity Consumption in Italy. Energy Sources. Part B: Economics, Planning, and Policy. 2013. Vol. 8 no 1. P. 86-93.

26. Gassar, A., Cha, S. Energy prediction techniques for large-scale buildings towards a sustainable built environment: A review. *Energy Build.* 2020. Vol. 8 no 12. P. 224-238.
27. Zhang L., Alahmad M., Wen J. Comparison of time-frequency-analysis techniques applied in building energy data noise cancellation for building load forecasting: A real-building case study. *Energy Build.* 2020. Vol 43 no 9. P. 377-389.
28. Graves J., Schmidhuber K. Framewise phoneme classification with bidirectional LSTM and other neural network architectures. *Neural Networks*, 2005. Vol. 12 no 18. P. 602–610.
29. Graves, A., Fernández, S., Gomez, F. Connectionist temporal classification: Labelling unsegmented sequence data with recurrent neural networks. In *Proceedings of the International Conference on Machine Learning. ICML.* 2006. Vol. 45 no 6. P. 369—376.
30. Socher R., Perelygin, A., Chuang, J. Recursive Deep Models for Semantic Compositionality Over a Sentiment Treebank. *Emnlp.* 2013. Vol. 34 no 6. P. 145-156.
31. Han P. Drought forecasting based on the sensing data using ARIMA model. *Math Comput Model.* 2010. Vol. 51 no 11. P. 1398-1403.
32. Bianco V., Manca O., Nardini S. Linear Regression Models to Forecast Electricity Consumption in Italy. *Energy Sources. Part B: Economics, Planning, and Policy.* 2013. Vol. 8 no 1. P. 86-93.
33. Рудної Г.І. Вибір функцій активації при прогнозуванні нейронних мереж. *Машинне навчання та аналіз Даних.* 2011. - Т.1, №1. С.16-39.
34. Pandas is a fiscally sponsored project of NumFOCUS. Pandas. URL: <https://pandas.pydata.org/> (date of access: 11.12.2020).
35. Installation Prophet. Facebook.github.io. URL: <https://facebook.github.io/prophet/docs/installation.html#python> (date of access: 10.12.2020).

36. Python & NumPy utilities. Keras. URL:
<https://keras.io/search.html?query=python> (date of access: 11.12.2020).
37. Seaborn: statistical data visualization. Seaborn. URL:
<https://seaborn.pydata.org/#:~:text=Seaborn is a Python data,can read the introductory notes.> (date of access: 11.12.2020).
38. Getting started. Scipy.org. URL: <https://www.scipy.org/getting-started.html>
(date of access: 10.12.2020).
39. SARIMAX: Introduction. Statsmodels. URL:
https://www.statsmodels.org/dev/examples/notebooks/generated/statespace_sarimax_stata.html (date of access: 10.12.2020).
40. Holt Winter's Exponential Smoothing. Statsmodels. URL:
<https://www.statsmodels.org/dev/generated/statsmodels.tsa.holtwinters.ExponentialSmoothing.html> (date of access: 11.12.2020).

ДОДАТКИ

Додаток А

Приклад даних для прогнозування

Точка учета	Направл.	Дата	Потребление
Корпус Г, Ввод 1 (4541562)	A+ (кВт*час)	01.01.2015	2,27940
Корпус Г, Ввод 1 (4541562)	R+ (кВар*час)	01.01.2015	1,47750
Корпус Г, Ввод 2 (4541560)	A+ (кВт*час)	01.01.2015	2,42790
Корпус Г, Ввод 2 (4541560)	R+ (кВар*час)	01.01.2015	0,11840
Корпус Г, Ввод 1 (4541562)	A+ (кВт*час)	02.01.2015	2,33810
Корпус Г, Ввод 1 (4541562)	R+ (кВар*час)	02.01.2015	1,39540
Корпус Г, Ввод 2 (4541560)	A+ (кВт*час)	02.01.2015	2,17900
Корпус Г, Ввод 2 (4541560)	R+ (кВар*час)	02.01.2015	0,05810
Корпус Г, Ввод 1 (4541562)	A+ (кВт*час)	03.01.2015	2,77500
Корпус Г, Ввод 1 (4541562)	R+ (кВар*час)	03.01.2015	1,36830
Корпус Г, Ввод 2 (4541560)	A+ (кВт*час)	03.01.2015	2,33480
Корпус Г, Ввод 2 (4541560)	R+ (кВар*час)	03.01.2015	0,15090
Корпус Г, Ввод 1 (4541562)	A+ (кВт*час)	04.01.2015	2,32520
Корпус Г, Ввод 1 (4541562)	R+ (кВар*час)	04.01.2015	1,31240
Корпус Г, Ввод 2 (4541560)	A+ (кВт*час)	04.01.2015	2,10010
Корпус Г, Ввод 2 (4541560)	R+ (кВар*час)	04.01.2015	0,13560
Корпус Г, Ввод 1 (4541562)	A+ (кВт*час)	05.01.2015	9,55130
Корпус Г, Ввод 1 (4541562)	R+ (кВар*час)	05.01.2015	2,48750
Корпус Г, Ввод 2 (4541560)	A+ (кВт*час)	05.01.2015	9,87190
Корпус Г, Ввод 2 (4541560)	R+ (кВар*час)	05.01.2015	1,73840
Корпус Г, Ввод 1 (4541562)	A+ (кВт*час)	06.01.2015	8,77350
Корпус Г, Ввод 1 (4541562)	R+ (кВар*час)	06.01.2015	2,04990
Корпус Г, Ввод 2 (4541560)	A+ (кВт*час)	06.01.2015	7,25450
Корпус Г, Ввод 2 (4541560)	R+ (кВар*час)	06.01.2015	0,94970
Корпус Г, Ввод 1 (4541562)	A+ (кВт*час)	07.01.2015	2,35240
Корпус Г, Ввод 1 (4541562)	R+ (кВар*час)	07.01.2015	1,41800
Корпус Г, Ввод 2 (4541560)	A+ (кВт*час)	07.01.2015	2,26940
Корпус Г, Ввод 2 (4541560)	R+ (кВар*час)	07.01.2015	0,14200
Корпус Г, Ввод 1 (4541562)	A+ (кВт*час)	08.01.2015	11,67710
Корпус Г, Ввод 1 (4541562)	R+ (кВар*час)	08.01.2015	2,06470
Корпус Г, Ввод 2 (4541560)	A+ (кВт*час)	08.01.2015	9,71650
Корпус Г, Ввод 2 (4541560)	R+ (кВар*час)	08.01.2015	1,18600
Корпус Г, Ввод 1 (4541562)	A+ (кВт*час)	09.01.2015	11,87620
Корпус Г, Ввод 1 (4541562)	R+ (кВар*час)	09.01.2015	2,06170
Корпус Г, Ввод 2 (4541560)	A+ (кВт*час)	09.01.2015	10,22780
Корпус Г, Ввод 2 (4541560)	R+ (кВар*час)	09.01.2015	1,32510
Корпус Г, Ввод 1 (4541562)	A+ (кВт*час)	10.01.2015	2,86400
Корпус Г, Ввод 1 (4541562)	R+ (кВар*час)	10.01.2015	1,35760
Корпус Г, Ввод 2 (4541560)	A+ (кВт*час)	10.01.2015	3,03510
Корпус Г, Ввод 2 (4541560)	R+ (кВар*час)	10.01.2015	0,19860
Корпус Г, Ввод 1 (4541562)	A+ (кВт*час)	11.01.2015	2,37340

Корпус Г, Ввод 1 (4541562)	R+ (кВар*час)	11.01.2015	1,29190
Корпус Г, Ввод 2 (4541560)	A+ (кВт*час)	11.01.2015	2,52970
Корпус Г, Ввод 2 (4541560)	R+ (кВар*час)	11.01.2015	0,10770
Корпус Г, Ввод 1 (4541562)	A+ (кВт*час)	12.01.2015	9,66800
Корпус Г, Ввод 1 (4541562)	R+ (кВар*час)	12.01.2015	2,46320
Корпус Г, Ввод 2 (4541560)	A+ (кВт*час)	12.01.2015	9,53750
Корпус Г, Ввод 2 (4541560)	R+ (кВар*час)	12.01.2015	1,57950
Корпус Г, Ввод 1 (4541562)	A+ (кВт*час)	13.01.2015	9,25820
Корпус Г, Ввод 1 (4541562)	R+ (кВар*час)	13.01.2015	2,46610
Корпус Г, Ввод 2 (4541560)	A+ (кВт*час)	13.01.2015	9,14310
Корпус Г, Ввод 2 (4541560)	R+ (кВар*час)	13.01.2015	1,81510
Корпус Г, Ввод 1 (4541562)	A+ (кВт*час)	14.01.2015	8,09220
Корпус Г, Ввод 1 (4541562)	R+ (кВар*час)	14.01.2015	2,31560
Корпус Г, Ввод 2 (4541560)	A+ (кВт*час)	14.01.2015	6,48390
Корпус Г, Ввод 2 (4541560)	R+ (кВар*час)	14.01.2015	1,07490
Корпус Г, Ввод 1 (4541562)	A+ (кВт*час)	15.01.2015	8,63870
Корпус Г, Ввод 1 (4541562)	R+ (кВар*час)	15.01.2015	2,59470
Корпус Г, Ввод 2 (4541560)	A+ (кВт*час)	15.01.2015	7,97110
Корпус Г, Ввод 2 (4541560)	R+ (кВар*час)	15.01.2015	1,95250
Корпус Г, Ввод 1 (4541562)	A+ (кВт*час)	16.01.2015	8,13100
Корпус Г, Ввод 1 (4541562)	R+ (кВар*час)	16.01.2015	2,44330
Корпус Г, Ввод 2 (4541560)	A+ (кВт*час)	16.01.2015	6,71670
Корпус Г, Ввод 2 (4541560)	R+ (кВар*час)	16.01.2015	1,48540
Корпус Г, Ввод 1 (4541562)	A+ (кВт*час)	17.01.2015	6,62010
Корпус Г, Ввод 1 (4541562)	R+ (кВар*час)	17.01.2015	2,01510
Корпус Г, Ввод 2 (4541560)	A+ (кВт*час)	17.01.2015	5,12770
Корпус Г, Ввод 2 (4541560)	R+ (кВар*час)	17.01.2015	1,09730
Корпус Г, Ввод 1 (4541562)	A+ (кВт*час)	18.01.2015	3,23750
Корпус Г, Ввод 1 (4541562)	R+ (кВар*час)	18.01.2015	1,38830
Корпус Г, Ввод 2 (4541560)	A+ (кВт*час)	18.01.2015	1,92830
Корпус Г, Ввод 2 (4541560)	R+ (кВар*час)	18.01.2015	0,16460
Корпус Г, Ввод 1 (4541562)	A+ (кВт*час)	19.01.2015	7,91830
Корпус Г, Ввод 1 (4541562)	R+ (кВар*час)	19.01.2015	2,57340
Корпус Г, Ввод 2 (4541560)	A+ (кВт*час)	19.01.2015	7,85780
Корпус Г, Ввод 2 (4541560)	R+ (кВар*час)	19.01.2015	1,90640
Корпус Г, Ввод 1 (4541562)	A+ (кВт*час)	20.01.2015	7,95370
Корпус Г, Ввод 1 (4541562)	R+ (кВар*час)	20.01.2015	2,48870
Корпус Г, Ввод 2 (4541560)	A+ (кВт*час)	20.01.2015	6,84990
Корпус Г, Ввод 2 (4541560)	R+ (кВар*час)	20.01.2015	1,72410
Корпус Г, Ввод 1 (4541562)	A+ (кВт*час)	21.01.2015	8,65550
Корпус Г, Ввод 1 (4541562)	R+ (кВар*час)	21.01.2015	2,42010
Корпус Г, Ввод 2 (4541560)	A+ (кВт*час)	21.01.2015	6,69760
Корпус Г, Ввод 2 (4541560)	R+ (кВар*час)	21.01.2015	1,28090
Корпус Г, Ввод 1 (4541562)	A+ (кВт*час)	22.01.2015	8,20910
Корпус Г, Ввод 1 (4541562)	R+ (кВар*час)	22.01.2015	2,37300
Корпус Г, Ввод 2 (4541560)	A+ (кВт*час)	22.01.2015	7,14360
Корпус Г, Ввод 2 (4541560)	R+ (кВар*час)	22.01.2015	1,34200
Корпус Г, Ввод 1 (4541562)	A+ (кВт*час)	23.01.2015	7,60880
Корпус Г, Ввод 1 (4541562)	R+ (кВар*час)	23.01.2015	2,25500
Корпус Г, Ввод 2 (4541560)	A+ (кВт*час)	23.01.2015	7,12510
Корпус Г, Ввод 2 (4541560)	R+ (кВар*час)	23.01.2015	1,39130
Корпус Г, Ввод 1 (4541562)	A+ (кВт*час)	24.01.2015	2,81940
Корпус Г, Ввод 1 (4541562)	R+ (кВар*час)	24.01.2015	1,43910

Корпус Г, Ввод 2 (4541560)	A+ (кВт*час)	24.01.2015	2,53360
Корпус Г, Ввод 2 (4541560)	R+ (кВар*час)	24.01.2015	0,25940
Корпус Г, Ввод 1 (4541562)	A+ (кВт*час)	25.01.2015	2,36560
Корпус Г, Ввод 1 (4541562)	R+ (кВар*час)	25.01.2015	1,27990
Корпус Г, Ввод 2 (4541560)	A+ (кВт*час)	25.01.2015	1,92510
Корпус Г, Ввод 2 (4541560)	R+ (кВар*час)	25.01.2015	0,13910
Корпус Г, Ввод 1 (4541562)	A+ (кВт*час)	26.01.2015	9,79990
Корпус Г, Ввод 1 (4541562)	R+ (кВар*час)	26.01.2015	2,46170
Корпус Г, Ввод 2 (4541560)	A+ (кВт*час)	26.01.2015	8,63040
Корпус Г, Ввод 2 (4541560)	R+ (кВар*час)	26.01.2015	1,39130
Корпус Г, Ввод 1 (4541562)	A+ (кВт*час)	27.01.2015	9,15110
Корпус Г, Ввод 1 (4541562)	R+ (кВар*час)	27.01.2015	2,45860
Корпус Г, Ввод 2 (4541560)	A+ (кВт*час)	27.01.2015	8,36810
Корпус Г, Ввод 2 (4541560)	R+ (кВар*час)	27.01.2015	1,38610
Корпус Г, Ввод 1 (4541562)	A+ (кВт*час)	28.01.2015	9,21630
Корпус Г, Ввод 1 (4541562)	R+ (кВар*час)	28.01.2015	2,63160
Корпус Г, Ввод 2 (4541560)	A+ (кВт*час)	28.01.2015	7,81010
Корпус Г, Ввод 2 (4541560)	R+ (кВар*час)	28.01.2015	1,12360
Корпус Г, Ввод 1 (4541562)	A+ (кВт*час)	29.01.2015	8,46090
Корпус Г, Ввод 1 (4541562)	R+ (кВар*час)	29.01.2015	2,63490
Корпус Г, Ввод 2 (4541560)	A+ (кВт*час)	29.01.2015	7,16800
Корпус Г, Ввод 2 (4541560)	R+ (кВар*час)	29.01.2015	1,25590
Корпус Г, Ввод 1 (4541562)	A+ (кВт*час)	30.01.2015	8,05020
Корпус Г, Ввод 1 (4541562)	R+ (кВар*час)	30.01.2015	2,48400
Корпус Г, Ввод 2 (4541560)	A+ (кВт*час)	30.01.2015	5,76410
Корпус Г, Ввод 2 (4541560)	R+ (кВар*час)	30.01.2015	0,87720
Корпус Г, Ввод 1 (4541562)	A+ (кВт*час)	31.01.2015	2,90710
Корпус Г, Ввод 1 (4541562)	R+ (кВар*час)	31.01.2015	1,63980
Корпус Г, Ввод 2 (4541560)	A+ (кВт*час)	31.01.2015	2,13140
Корпус Г, Ввод 2 (4541560)	R+ (кВар*час)	31.01.2015	0,14400
Корпус Г, Ввод 1 (4541562)	A+ (кВт*час)	01.02.2015	2,50000
Корпус Г, Ввод 1 (4541562)	R+ (кВар*час)	01.02.2015	1,46590
Корпус Г, Ввод 2 (4541560)	A+ (кВт*час)	01.02.2015	2,19640
Корпус Г, Ввод 2 (4541560)	R+ (кВар*час)	01.02.2015	0,26660
Корпус Г, Ввод 1 (4541562)	A+ (кВт*час)	02.02.2015	7,96520
Корпус Г, Ввод 1 (4541562)	R+ (кВар*час)	02.02.2015	2,60550
Корпус Г, Ввод 2 (4541560)	A+ (кВт*час)	02.02.2015	5,24950
Корпус Г, Ввод 2 (4541560)	R+ (кВар*час)	02.02.2015	0,91430
Корпус Г, Ввод 1 (4541562)	A+ (кВт*час)	03.02.2015	8,08440
Корпус Г, Ввод 1 (4541562)	R+ (кВар*час)	03.02.2015	2,41140
Корпус Г, Ввод 2 (4541560)	A+ (кВт*час)	03.02.2015	5,31330
Корпус Г, Ввод 2 (4541560)	R+ (кВар*час)	03.02.2015	0,98000
Корпус Г, Ввод 1 (4541562)	A+ (кВт*час)	04.02.2015	8,05040
Корпус Г, Ввод 1 (4541562)	R+ (кВар*час)	04.02.2015	2,61800
Корпус Г, Ввод 2 (4541560)	A+ (кВт*час)	04.02.2015	5,27410
Корпус Г, Ввод 2 (4541560)	R+ (кВар*час)	04.02.2015	0,92690
Корпус Г, Ввод 1 (4541562)	A+ (кВт*час)	05.02.2015	8,10090
Корпус Г, Ввод 1 (4541562)	R+ (кВар*час)	05.02.2015	2,51050
Корпус Г, Ввод 2 (4541560)	A+ (кВт*час)	05.02.2015	5,92970
Корпус Г, Ввод 2 (4541560)	R+ (кВар*час)	05.02.2015	1,10690
Корпус Г, Ввод 1 (4541562)	A+ (кВт*час)	06.02.2015	8,51330
Корпус Г, Ввод 1 (4541562)	R+ (кВар*час)	06.02.2015	2,58840
Корпус Г, Ввод 2 (4541560)	A+ (кВт*час)	06.02.2015	5,31140

Додаток Б

Приклад додаткових даних

Date	T	Po	P	Pa	U	DD
10.11.2020 5:00	0.8	754.5	771.8	0.2	89	Ветер, дующий с западо-северо-запада
10.11.2020 2:00	04.фев	754.3	771.3	0.3	88	Ветер, дующий с северо-запада
09.11.2020 23:00	06.июн	754.0	770.9	0.3	91	Ветер, дующий с северо-северо-запада
09.11.2020 20:00	07.фев	753.7	770.6	0.2	92	Ветер, дующий с западо-северо-запада
09.11.2020 17:00	07.мар	753.5	770.4	0.1	88	Ветер, дующий с севера
09.11.2020 14:00	08.апр	753.4	770.2	-0.3	90	Ветер, дующий с западо-северо-запада
09.11.2020 11:00	07.июн	753.7	770.5	0.7	92	Ветер, дующий с севера
09.11.2020 8:00	06.мар	753.0	769.9	0.8	95	Ветер, дующий с западо-северо-запада
09.11.2020 5:00	06.янв	752.2	769.0	0.9	92	Ветер, дующий с западо-северо-запада
09.11.2020 2:00	6.0	751.3	768.2	0.7	91	Ветер, дующий с западо-северо-запада
08.11.2020 23:00	06.мар	750.6	767.5	0.5	93	Ветер, дующий с западо-северо-запада
08.11.2020 20:00	06.май	750.1	766.9	0.3	95	Ветер, дующий с западо-северо-запада
08.11.2020 17:00	06.сен	749.8	766.6	-0.2	95	Ветер, дующий с запада
08.11.2020 14:00	06.авг	750.0	766.8	-1.2	95	Ветер, дующий с юго-запада
08.11.2020 11:00	07.янв	751.2	768.0	-0.7	97	Ветер, дующий с юго-запада
08.11.2020 8:00	06.июл	751.9	768.7	-0.2	100	Ветер, дующий с западо-юго-запада
08.11.2020 5:00	06.апр	752.1	769.0	-0.4	97	Ветер, дующий с западо-юго-запада
08.11.2020 2:00	07.фев	752.5	769.3	-0.3	95	Ветер, дующий с западо-юго-запада
07.11.2020 23:00	05.мар	752.8	769.7	0.5	89	Ветер, дующий с запада
07.11.2020 20:00	07.янв	752.3	769.1	0.3	92	Ветер, дующий с запада
07.11.2020 17:00	07.июн	752.0	768.8	1.0	90	Ветер, дующий с запада
07.11.2020 14:00	8.0	751.0	767.8	0.5	92	Ветер, дующий с западо-северо-запада
07.11.2020 11:00	07.янв	750.5	767.3	0.3	94	Ветер, дующий с западо-северо-запада
07.11.2020 8:00	05.апр	750.2	767.1	0.3	95	Ветер, дующий с запада
07.11.2020 5:00	05.фев	749.9	766.8	0.4	95	Ветер, дующий с западо-северо-запада

07.11.2020 2:00	05.май	749.5	766.4	01.фев	95	Ветер, дующий с западо-северо-запада
06.11.2020 23:00	06.апр	748.3	765.1	0.0	87	Ветер, дующий с западо-северо-запада
06.11.2020 20:00	06.апр	748.3	765.1	0.7	87	Ветер, дующий с западо-северо-запада
06.11.2020 17:00	06.июл	747.6	764.3	-0.4	88	Ветер, дующий с запада
06.11.2020 14:00	06.фев	748.0	764.8	-0.6	88	Ветер, дующий с запада
06.11.2020 11:00	05.мар	748.6	765.4	-0.3	95	Ветер, дующий с западо-северо-запада
06.11.2020 8:00	04.мар	748.9	765.8	-0.6	93	Ветер, дующий с запада
06.11.2020 5:00	03.авг	749.5	766.5	-1.5	90	Ветер, дующий с западо-юго-запада
06.11.2020 2:00	04.фев	751.0	767.9	-1.1	89	Ветер, дующий с запада
05.11.2020 23:00	06.мар	752.1	769.0	-0.5	82	Ветер, дующий с запада
05.11.2020 20:00	07.апр	752.6	769.4	0.1	84	Ветер, дующий с западо-северо-запада
05.11.2020 17:00	08.апр	752.5	769.2	-0.5	87	Ветер, дующий с запада
05.11.2020 14:00	09.фев	753.0	769.7	-1.2	89	Ветер, дующий с запада
05.11.2020 11:00	06.авг	754.2	771.1	-0.6	100	Штиль, безветрие
05.11.2020 8:00	06.фев	754.8	771.7	-0.2	100	Штиль, безветрие
05.11.2020 5:00	06.апр	755.0	772.0	-0.6	95	Штиль, безветрие
05.11.2020 2:00	06.апр	755.6	772.6	0.0	100	Штиль, безветрие
04.11.2020 23:00	06.июл	755.6	772.6	-0.3	100	Штиль, безветрие
04.11.2020 20:00	06.мар	755.9	772.9	0.1	100	Штиль, безветрие
04.11.2020 17:00	7.0	755.8	772.6	0.6	94	Ветер, дующий с востока
04.11.2020 14:00	11.мар	755.2	771.9	-0.3	86	Ветер, дующий с востоко-юго-востока
04.11.2020 11:00	10.апр	755.5	772.3	0.4	92	Штиль, безветрие
04.11.2020 8:00	06.июл	755.1	772.0	0.7	95	Ветер, дующий с северо-северо-востока
04.11.2020 5:00	07.авг	754.4	771.2	0.4	95	Штиль, безветрие
04.11.2020 2:00	07.июл	754.0	770.8	0.7	95	Штиль, безветрие
03.11.2020 23:00	07.июл	753.3	770.1	0.9	95	Штиль, безветрие
03.11.2020 20:00	07.апр	752.4	769.2	01.фев	95	Ветер, дующий с востока
03.11.2020 17:00	08.авг	751.2	767.8	01.янв	96	Ветер, дующий с востоко-юго-востока
03.11.2020 14:00	09.июл	750.1	766.7	0.8	88	Ветер, дующий с юго-востока
03.11.2020 11:00	09.июн	749.3	765.9	1.0	95	Ветер, дующий с востоко-юго-востока
03.11.2020 8:00	09.янв	748.3	764.8	01.фев	96	Ветер, дующий с северо-северо-востока
03.11.2020 5:00	08.июн	747.1	763.7	0.7	95	Ветер, дующий с севера
03.11.2020 2:00	08.апр	746.4	763.0	0.7	94	Ветер, дующий с севера
02.11.2020 23:00	08.май	745.7	762.3	0.5	93	Ветер, дующий с запада
02.11.2020 20:00	08.авг	745.2	761.8	0.8	95	Ветер, дующий с северо-запада

02.11.2020 17:00	10.0	744.4	760.8	0.9	95	Ветер, дующий с северо-запада
02.11.2020 14:00	10.июн	743.5	760.0	-0.3	94	Ветер, дующий с северо-запада
02.11.2020 11:00	09.июл	743.8	760.3	0.3	99	Ветер, дующий с севера
02.11.2020 8:00	08.июл	743.5	760.0	0.0	97	Ветер, дующий с северо-северо-востока
02.11.2020 5:00	08.май	743.5	760.0	-0.7	97	Ветер, дующий с северо-северо-востока
02.11.2020 2:00	08.янв	744.2	760.8	-0.1	96	Ветер, дующий с северо-северо-востока
01.11.2020 23:00	07.сен	744.3	760.9	-0.1	97	Ветер, дующий с востоко-северо-востока
01.11.2020 20:00	08.фев	744.4	760.9	0.9	97	Ветер, дующий с юго-востока
01.11.2020 17:00	08.июн	743.5	760.0	01.янв	96	Ветер, дующий с юга
01.11.2020 14:00	09.фев	742.4	758.9	0.6	96	Ветер, дующий с юго-юго-запада
01.11.2020 11:00	08.янв	741.8	758.3	0.3	97	Ветер, дующий с западо-северо-запада
01.11.2020 8:00	07.июл	741.5	758.1	0.1	93	Ветер, дующий с севера
01.11.2020 5:00	7.0	741.4	757.9	-0.7	95	Ветер, дующий с севера
01.11.2020 2:00	06.июн	742.1	758.8	-0.7	95	Ветер, дующий с севера
31.10.2020 23:00	06.авг	742.8	759.4	-0.7	95	Ветер, дующий с севера
31.10.2020 20:00	7.0	743.5	760.1	-0.1	95	Ветер, дующий с северо-северо-востока
31.10.2020 17:00	07.фев	743.6	760.3	-0.2	95	Ветер, дующий с северо-северо-востока
31.10.2020 14:00	07.апр	743.8	760.4	-0.4	94	Ветер, дующий с северо-северо-востока
31.10.2020 11:00	07.сен	744.2	760.8	-0.3	93	Ветер, дующий с северо-северо-востока
31.10.2020 8:00	7.0	744.5	761.2	0.1	97	Ветер, дующий с северо-северо-запада
31.10.2020 5:00	06.сен	744.4	761.1	-0.4	93	Ветер, дующий с севера
31.10.2020 2:00	06.авг	744.8	761.5	-0.1	88	Ветер, дующий с северо-запада

Додаток В

Відформатований набір даних

Date	Consumption	T	U	Ff	Month	Day	year	quarter	month	day
2015-01-03	593,308	0,8375	96,125	0,375	Jan	5	2015	1	1	3
2015-02-03	1530,692	0,1375	93	3,5	Feb	1	2015	1	2	3
2015-03-03	1532,392	1,0375	95,5	5	Mar	1	2015	1	3	3
2015-04-03	1582,42	0,55	97,25	1,875	Apr	4	2015	2	4	3
2015-05-03	1409,416	0,0625	90,5	2,125	May	6	2015	2	5	3
2015-06-03	1202,48	0,3	81	2,375	Jun	2	2015	2	6	3
2015-07-03	478,524	-0,5	83,25	2	Jul	4	2015	3	7	3
2015-08-03	406,724	2,2	70,625	3,25	Aug	0	2015	3	8	3
2015-09-03	427,748	5,1875	69	3,75	Sep	3	2015	3	9	3
2015-10-03	1223,768	5,725	67,25	2,375	Oct	5	2015	4	10	3
2015-11-03	1129,768	6,4875	70,5	2,625	Nov	1	2015	4	11	3
2015-12-03	1244,36	5,5625	76,5	3,875	Dec	3	2015	4	12	3
2015-03-13	1225,776	3,8375	73,25	7,375	Mar	4	2015	1	3	13
2015-03-14	631,964	2,7875	84,5	6,5	Mar	5	2015	1	3	14
2015-03-15	515,776	3,925	64,75	4,625	Mar	6	2015	1	3	15
2015-03-16	1318,64	3,85	49,375	5,125	Mar	0	2015	1	3	16
2015-03-17	1337,06	2,1375	50,5	4,875	Mar	1	2015	1	3	17
2015-03-18	1385,836	1,2375	62,375	4,25	Mar	2	2015	1	3	18
2015-03-19	1357,804	2,6125	61,125	2,75	Mar	3	2015	1	3	19
2015-03-20	1280,6	3,475	62,25	1,5	Mar	4	2015	1	3	20
2015-03-21	642,464	5,1	58,75	2,875	Mar	5	2015	1	3	21
2015-03-22	601,068	2,5125	68,125	4,875	Mar	6	2015	1	3	22

2015-03-23	1400,772	-2,575	51,375	2,75	Mar	0	2015	1	3	23
2015-03-24	1494,156	2,2875	39,75	5,5	Mar	1	2015	1	3	24
2015-03-25	1332,976	5,6	42,125	5,125	Mar	2	2015	1	3	25
2015-03-26	1255,22	5,6625	54,125	3	Mar	3	2015	1	3	26
2015-03-27	1162,184	9,875	48	5,125	Mar	4	2015	1	3	27
2015-03-28	567,74	6,125	48,25	11	Mar	5	2015	1	3	28
2015-03-29	623,6	-0,8625	83,5	10,5	Mar	6	2015	1	3	29
2015-03-30	1507,296	0,0625	93,625	7,625	Mar	0	2015	1	3	30
2015-03-31	1714,104	0,95	96,125	7,5	Mar	1	2015	1	3	31
2015-01-04	1507,532	3,6875	94	4,875	Jan	6	2015	1	1	4
2015-02-04	1323,884	3,65	77,875	2,5	Feb	2	2015	1	2	4
2015-03-04	1313,396	3,2	73,75	2,5	Mar	2	2015	1	3	4
2015-04-04	667,568	3,05	86,75	3	Apr	5	2015	2	4	4
2015-05-04	529,904	3,0875	79,5	4,375	May	0	2015	2	5	4
2015-06-04	1366,628	3,2375	69	2,125	Jun	3	2015	2	6	4
2015-07-04	1313,308	4,8875	63,5	5	Jul	5	2015	3	7	4
2015-08-04	1516,46	3,8	87,25	5,25	Aug	1	2015	3	8	4
2015-09-04	1301,1	4,5125	61,75	6,875	Sep	4	2015	3	9	4
2015-10-04	973,28	8,3125	48,5	4,75	Oct	6	2015	4	10	4
2015-11-04	417,232	12,3625	42	3,25	Nov	2	2015	4	11	4
2015-12-04	359,412	13,1625	39,875	2,25	Dec	4	2015	4	12	4
2015-04-13	356,652	11,175	42,625	3,125	Apr	0	2015	2	4	13
2015-04-14	1109,888	5,4375	68,75	4,5	Apr	1	2015	2	4	14
2015-04-15	1232,22	3,5875	84,5	4,75	Apr	2	2015	2	4	15
2015-04-16	1270,848	10,525	71,25	6,75	Apr	3	2015	2	4	16

2015-04-17	1087,292	9,6	50,125	3,25	Apr	4	2015	2	4	17
2015-04-18	408,88	8,55	59	2,25	Apr	5	2015	2	4	18
2015-04-19	268,544	4,0375	68,875	4	Apr	6	2015	2	4	19
2015-04-20	1817,448	1,575	83,875	3,25	Apr	0	2015	2	4	20
2015-04-21	1813,368	2,65	64,375	3,625	Apr	1	2015	2	4	21
2015-04-22	1959,512	5,4125	56,75	3,625	Apr	2	2015	2	4	22
2015-04-23	1886,004	9,4	44,875	6,5	Apr	3	2015	2	4	23
2015-04-24	1289,124	14,375	43,125	5,5	Apr	4	2015	2	4	24
2015-04-25	331,108	13,6625	43,5	4	Apr	5	2015	2	4	25
2015-04-26	286,816	15,05	41,875	4,75	Apr	6	2015	2	4	26
2015-04-27	1019,692	16,05	54,75	4	Apr	0	2015	2	4	27
2015-04-28	917,172	16,2125	46,875	4,25	Apr	1	2015	2	4	28
2015-04-29	870,752	14,9625	59,125	3,375	Apr	2	2015	2	4	29
2015-04-30	719,204	14,3375	76,125	3,375	Apr	3	2015	2	4	30
2015-01-05	273,384	13,525	76,375	2,25	Jan	0	2015	1	1	5
2015-02-05	292,288	13,775	82,5	3,625	Feb	3	2015	1	2	5
2015-03-05	361,84	11,8875	93,375	3,25	Mar	3	2015	1	3	5
2015-04-05	229,128	11,125	68	3	Apr	6	2015	2	4	5
2015-05-05	1055,168	11,3625	52,125	2,625	May	1	2015	2	5	5
2015-06-05	1002,952	12,0625	48,5	3,625	Jun	4	2015	2	6	5
2015-07-05	944,536	13	47,125	1,75	Jul	6	2015	3	7	5
2015-08-05	917,88	10,9125	89,5	2,75	Aug	2	2015	3	8	5
2015-09-05	254,772	13,5	62,375	2,375	Sep	5	2015	3	9	5
2015-10-05	268,968	15,9875	41,25	2,25	Oct	0	2015	4	10	5

Додаток Г

Лістинг програми

Перелік використаних бібліотек

```
# Basic packages
import numpy as np # linear algebra
import pandas as pd # data processing, CSV file I/O (e.g. pd.read_csv)
import random as rd # generating random numbers
import datetime # manipulating date formats
# Viz
import matplotlib.pyplot as plt # basic plotting
import seaborn as sns # for prettier plots
import plotly.graph_objs as go
import plotly.offline as py
from plotly.offline import init_notebook_mode, iplot
# TIME SERIES
#from pyramid.arima import auto_arima
from statsmodels.tsa.arima_model import ARIMA
from statsmodels.tsa.statespace.sarimax import SARIMAX
from pandas.plotting import autocorrelation_plot
from statsmodels.tsa.stattools import adfuller, acf, pacf, arma_order_select_ic
import statsmodels.formula.api as smf
import statsmodels.tsa.api as smt
import statsmodels.api as sm
import scipy.stats as scs
import datetime as dt
from keras.models import Sequential
from keras.layers import Dense
from keras.layers import LSTM
#import pmdarima as pm
from google.colab import drive
# settings
import warnings
warnings.filterwarnings("ignore")
```

Підключення ресурсів

```
%time
weather=pd.read_csv("/content/drive/MyDrive/Colab Notebooks/weather1.csv",
    sep=';', parse_dates=['Date'])

G2015=pd.read_excel("/content/drive/MyDrive/Colab Notebooks/2015_day_G.xlsx",parse_dates=['Дата'])
```

```

G2016=pd.read_excel("/content/drive/MyDrive/Colab Notebooks/2016_day_G.xls
x",parse_dates=['Дата'])
G2017=pd.read_excel("/content/drive/MyDrive/Colab Notebooks/2017_day_G.xls
x",parse_dates=['Дата'])
G2018=pd.read_excel("/content/drive/MyDrive/Colab Notebooks/2018_day_G.xls
x",parse_dates=['Дата'])
G2019=pd.read_excel("/content/drive/MyDrive/Colab Notebooks/2019_day_G.xls
x",parse_dates=['Дата'])
G2020=pd.read_excel("/content/drive/MyDrive/Colab Notebooks/2020_day_G.xls
x",parse_dates=['Дата'])
#Удаление пропусков
G2015=G2015.drop(G2015.index[[0,1,1462,1463]])
G2016=G2016.drop(G2016.index[[0,1,1466,1467]])
G2017=G2017.drop(G2017.index[[0,1,1462,1463]])
G2018=G2018.drop(G2018.index[[0,1,1462,1463]])
G2019=G2019.drop(G2019.index[[0,1,1462,1463]])
G2020=G2020.drop(G2020.index[[0,1,1134,1135]])
#Изменения форматов даты
weather['Date'] = weather['Date'].apply(lambda x: dt.datetime.strptime(x,
'%D'))

```

Обработка данных

```

#Обработка данных о погоде
weather = weather.iloc[::-1]
weather1=weather[['Date', 'T', 'Po', 'P', 'Pa', 'U', 'DD', 'Ff']]
weather1 = weather1.set_index([pd.Index(weather1['Date'])])
#weather1 = weather1.reset_index(drop=True)
weather1 = weather1.groupby(pd.Grouper(key='Date')).mean()
weather1

#Обработка данных о расходе электроэнергии
frames=[G2015,G2016,G2017,G2018,G2019,G2020]
#frames=[G2019,G2020]
df3=pd.concat(frames)

#Удаление реактивной энергии
df3.drop(df3.index[df3['Направл.'] == 'R+ (кВар*час)'], inplace = True)

df3['Consumption'] = df3['Потребление']
df3['Date'] = df3['Дата']
df3=df3[['Date', 'Consumption']]
df2=df3[['Date', 'Consumption']]
df3['Date'] = df3['Date'].apply(lambda x: dt.datetime.strptime(x, '%D'))
#df3 = df3.reset_index(drop=True)
#df3 = df3.set_index([pd.Index(df3['Date'])])
df3 = df3.groupby(['Date']).sum()

```

```

#Соединение данных о погоде и потребление электроэнергии
#Final = pd.merge(weather2,fg,on=['Date','Date'])
Final = pd.merge(weather1,df3,on=['Date'])
#Final = Final.set_index([pd.Index(Final['Date'])])
Final=Final[['Consumption','T','Po','P','Pa','U','Ff']]
Final = Final.apply (pd.to_numeric, errors='coerce')
#Final = Final.dropna()

Final=Final.loc['01/03/15':'01/10/20']

Final.reset_index(inplace=True)
res = sm.tsa.seasonal_decompose(Final['Consumption'],
                               freq=365,
                               model='additive')

resplot = res.plot()

df=Final
df.index = pd.to_datetime(df.index)
#df=df[Final.index.dayofweek < 5]
df['Date'] =pd.to_datetime(df.index)
df['year'] = df['Date'].apply(lambda x: x.year)
df['Month'] = df['Date'].dt.strftime('%b')
df['Day'] = df['Date'].apply(lambda time: time.dayofweek)
df['Date'] = df['Date'].apply(lambda x: dt.datetime.strftime(x, '%Y-%m-%d'))
df=df.set_index('Date')
df

# Scatterplot of "KWh" ~ "Temp"
plt.figure(figsize=(15,7))
plt.scatter(Final['T'], Final['Consumption'])
ys = sm.nonparametric.lowess(Final['Consumption'].values, Final['T'].values)
plt.plot(ys[:,0], ys[:,1], 'r-', linewidth=3)
plt.title('Залежність споживаної енергії від температури',fontweight='bold')
plt.xlabel('Температура (\u00b0C)', fontweight='bold')
plt.ylabel('Електроенергія Вт', fontweight='bold')
plt.show()

corr = df.corr()
corr.style.background_gradient(cmap='coolwarm')

fig, ax = plt.subplots()

```

```
fig.set_size_inches((12,6))
sns.boxplot(x='year', y='Consumption', data=df, ax=ax)
plt.show()
```

```
df.index = pd.to_datetime(df.index)
ts = pd.Series(df['Consumption'], index=df.index)
import seaborn
```

```
fig, ax = plt.subplots(figsize=(12,6))
seaborn.boxplot(ts.index.dayofweek, ts, ax=ax)
```

SARIMAX

```
mod = sm.tsa.statespace.SARIMAX(test['Consumption'],
                                order=(5, 1, 2),
                                seasonal_order=(1, 1, 1, 12),
                                enforce_stationarity=False,
                                enforce_invertibility=False)
```

```
results = mod.fit()
results.summary()
```

Залишки моделі

```
results.plot_diagnostics(figsize=(15, 12))
plt.show()
```

Прогнозна якість моделі

```
test['residual'] = abs(test['Consumption']-test['pred'])
MAE = test['residual'].sum()/len(test)
RMSE = (((test['Consumption']**2)-
(test['pred']**2)**0.5).sum()/len(test)
MAPE = (abs(test['residual'])/test['Consumption']).sum()*100/len(test)
print("MAE:", MAE)
print("RMSE:", RMSE)
print("MAPE:", MAPE)
test.head()
```

Модель машинного навчання

```
def create_dataset(X, y, time_steps=1):
    Xs, ys = [], []
    for i in range(len(X) - time_steps):
        v = X.iloc[i:(i + time_steps)].values
        Xs.append(v)
        ys.append(y.iloc[i + time_steps])
```

```

    return np.array(Xs), np.array(ys)

train_size = int(len(data) * 0.87)
train, test = data[:train_size,], data[train_size:,]

def Create_Dataset(df, lookback=1):
    X, Y = [], []
    for i in range(len(df) - lookback - 1):
        X.append(df[i:(i+lookback), 0])
        Y.append(df[i + lookback, 0])
    return np.array(X), np.array(Y)

lookback = 30
X_train, Y_train = Create_Dataset(train, lookback)
X_test, Y_test = Create_Dataset(test, lookback)

X_train = np.reshape(X_train, (X_train.shape[0], 1, X_train.shape[1]))
X_test = np.reshape(X_test, (X_test.shape[0], 1, X_test.shape[1]))

from tensorflow.keras.models import Sequential
from tensorflow.keras.layers import Dense, Dropout, LSTM, Bidirectional
from tensorflow.keras.callbacks import EarlyStopping

model = Sequential()
model.add(LSTM(100, input_shape=(X_train.shape[1], X_train.shape[2])))
model.add(Dropout(0.2))
model.add(Dense(1))

model.compile(optimizer='adam', loss='mean_squared_error')

model.summary()

train_predict = model.predict(X_train)
test_predict = model.predict(X_test)

# Invert the predictions to original scale
train_predict = scaler.inverse_transform(train_predict)
Y_train = scaler.inverse_transform([Y_train])

test_predict = scaler.inverse_transform(test_predict)
Y_test = scaler.inverse_transform([Y_test])

import numpy as np

def mean_absolute_percentage_error(y_true, y_pred):
    y_true, y_pred = np.array(y_true), np.array(y_pred)
    return np.mean(np.abs((y_true - y_pred) / y_true)) * 100

```



```
def mean_absolute_error(y_true, y_pred):
    y_true, y_pred = np.array(y_true), np.array(y_pred)
    return np.mean(np.abs((y_true - y_pred)))
```

Перевірка прогнозної якості моделі

```
from sklearn.metrics import mean_squared_error

print("Train RMSE : ", np.sqrt(mean_squared_error(Y_train[0], train_predict[:,0])))
print("Test RMSE : ", np.sqrt(mean_squared_error(Y_test[0], test_predict[:,0])))
print("Train MAPE :", mean_absolute_percentage_error(Y_train[0], train_predict[:,0]))
print("Test MAPE :", mean_absolute_percentage_error(Y_test[0], test_predict[:,0]))
print("Train MAE :", mean_absolute_error(Y_train[0], train_predict[:,0]))
print("Test MAE :", mean_absolute_error(Y_test[0], test_predict[:,0]))
print("Train MSE : ", (mean_squared_error(Y_train[0], train_predict[:,0]))
)
print("Test MSE : ", (mean_squared_error(Y_test[0], test_predict[:,0])))

plt.figure(figsize=(12,5))
plt.title("Порівняння втрати моделі")
plt.plot(history.history['loss'], label='Train loss')
plt.plot(history.history['val_loss'], label='Test loss')
plt.xlabel('Epochs')
plt.ylabel('Loss')
plt.legend(loc='upper right')
plt.show()
```

Побудова прогнозованих значень

```
test = pd.DataFrame(Y_test[0][:3000])
predicted = pd.DataFrame(test_predict[:,0][:3000])

plt.figure(figsize=(16,6))
plt.plot(test, label='Actual')
plt.plot(predicted, label='Predicted')
plt.ylabel('кВт*год', size=13)
plt.tight_layout()
plt.legend(fontsize=13)
plt.show()
```

Перевірка на залишки моделі

```
plt.figure(figsize=(14,6))
```

```
plt.subplot(1,2,1)
df['Consumption'].hist(bins=50,edgecolor='black', grid=False)
plt.title('Consumption Distribution')

plt.subplot(1,2,2)
stats.probplot(dfp['Consumption'], plot=plt);
#print("Skewness: %f" % df['Consumption'].skew())
#print("Kurtosis: %f" % df['Consumption'].kurt())
```