

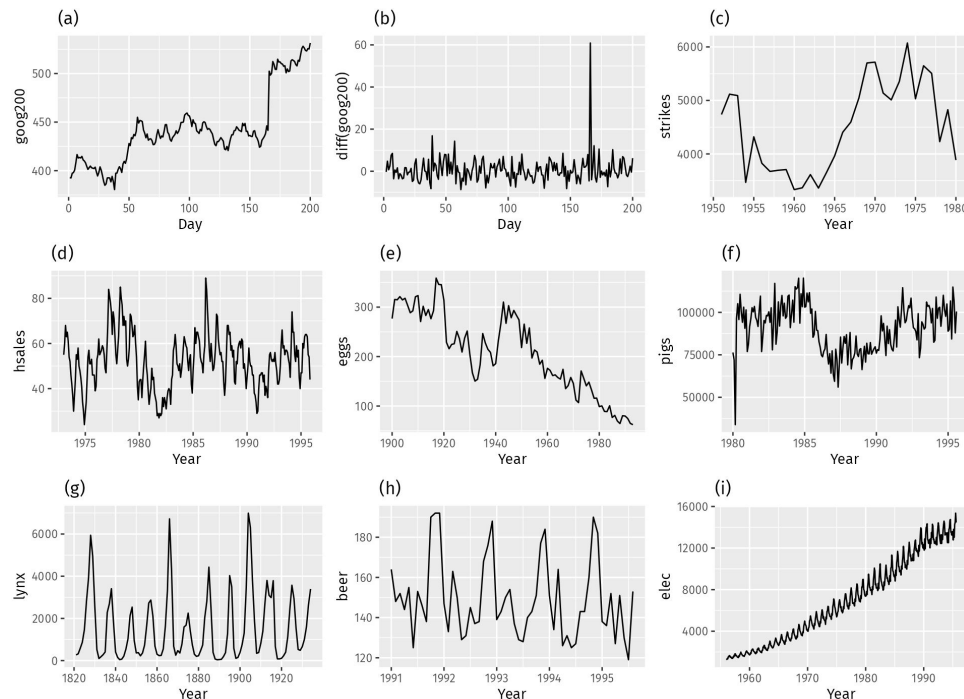
Часові ряди



Часові ряди. Вступ

Часовий ряд (time series) — це ряд точок даних, проіндексованих (або перелічених, або відкладених на графіку) в хронологічному порядку. Найчастіше часовий ряд є послідовністю, взятою на рівновіддалених точках в часі, які йдуть одна за одною.

Одно/багатокритеріальний часовий ряд



Часові ряди. Основні питання

Скільки даних у вас є і чи можете ви зібрати їх усі? Більше даних часто є більш корисним, пропонуючи більше можливостей для дослідницького аналізу даних, тестування та налаштування моделі та точності моделі.

Який необхідний часовий горизонт прогнозів? Короткостроковий, середньостроковий чи довгостроковий? Коротші часові горизонти часто легше прогнозувати з більшою впевненістю.

Чи можна прогнози часто оновлювати з часом або їх потрібно робити один раз і залишатися статичними? Оновлення прогнозів у міру появи нової інформації часто призводить до більш точних прогнозів.

З якою часовою частотою потрібні прогнози? Часто прогнози можна робити на нижчих або вищих частотах, що дає змогу використовувати низьку та збільшену вибірку даних, що, у свою чергу, може запропонувати переваги під час моделювання (година, тиждень, місяць, рік, ...).

Часто проблеми часових рядів виникають у реальному часі, що постійно надає нові можливості для прогнозування.



Часові ряди. Дані

Finance and Economics:

- Stock market forecasting.
- Economic indicators and forecasting.
- Risk management and portfolio optimization.
- Anomaly detection for fraud detection in financial transactions.

Forecasting:

- Sales forecasting for businesses.
- Demand forecasting for supply chain management.
- Weather forecasting.
- Energy consumption forecasting.
- Traffic and transportation demand prediction.

Healthcare and Medicine:

- Patient monitoring and predicting disease outbreaks.
- Electrocardiogram (ECG) analysis for heart disease detection.
- Longitudinal medical data analysis.
- Health resource planning and optimization.

Environmental Science:

- Climate data analysis and modeling.
- Air quality monitoring and prediction.
- Natural disaster prediction and early warning systems.

Manufacturing and Quality Control:

- Process control and quality assurance in manufacturing.
- Equipment maintenance and failure prediction.
- Supply chain optimization.

Energy:

- Energy consumption analysis and optimization.
- Energy price forecasting.
- Anomaly detection in energy grids.

Marketing and Customer Behavior:

- Customer churn prediction.
- Marketing campaign effectiveness analysis.
- Web traffic and user behavior analysis.
- Social media sentiment analysis.

Telecommunications:

- Network traffic analysis and optimization.
- Anomaly detection for network security.
- Call volume and data usage forecasting.

Government and Public Policy:

- Economic policy evaluation.
- Crime rate prediction.
- Social and demographic data analysis.

Sports Analytics:

- Player performance analysis.
- Game outcome prediction.
- Injury prediction and prevention.

Astronomy and Astrophysics:

- Time series data from telescopes for astronomical object tracking.
- Signal analysis for gravitational wave detection.

Energy Finance:

- Analysis of energy prices and commodities.
- Renewable energy generation prediction.

Human Resources:

- Workforce demand forecasting.
- Employee retention and turnover prediction.

Web Analytics:

- Website traffic analysis and prediction.
- User behavior modeling.

Education:

- Student performance analysis.
- Educational resource planning.

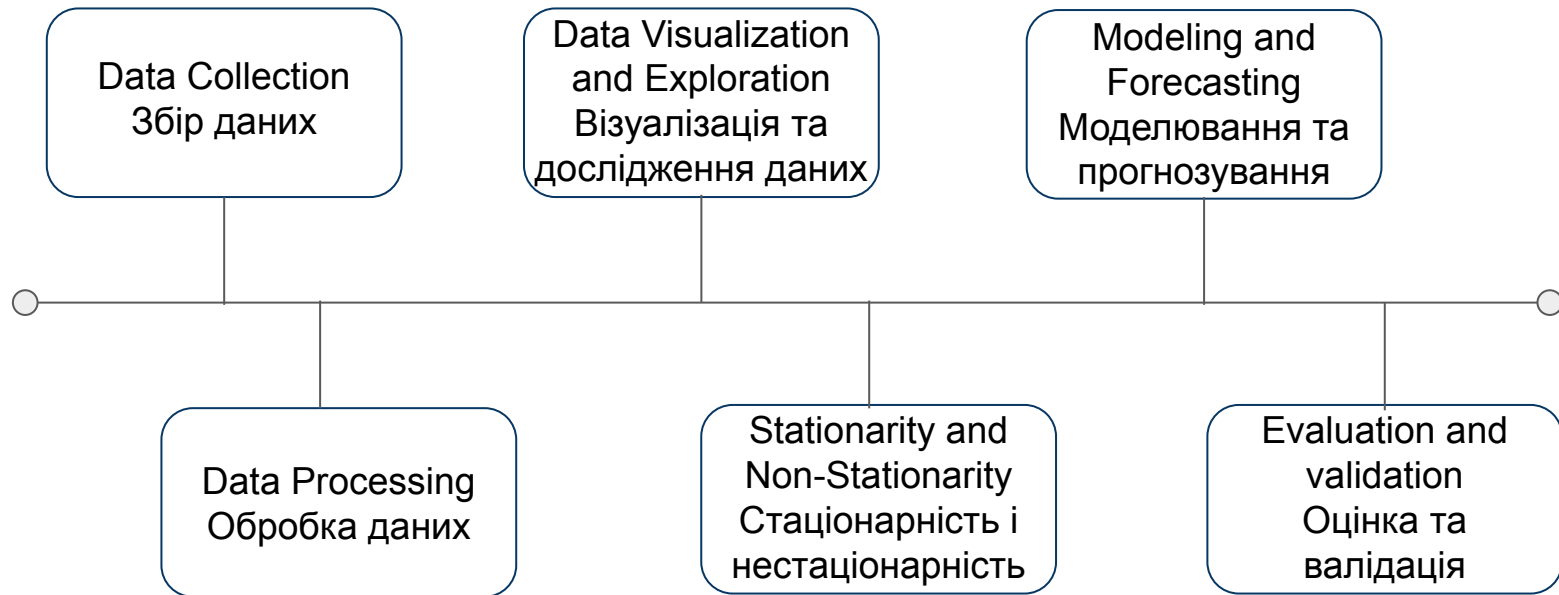
Sensors and IoT:

- Analysis of sensor data from IoT devices.
- Predictive maintenance for machinery and equipment.

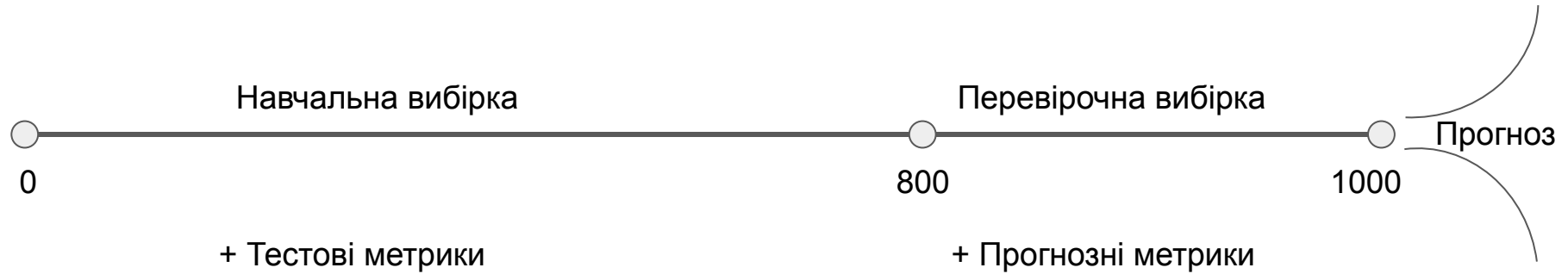
Psychology and Psychiatry:

- Mood and emotion tracking.
- Mental health monitoring.

Етапи аналізу часових рядів

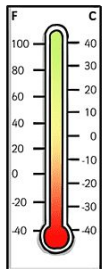


Часові ряди. Прогнозування



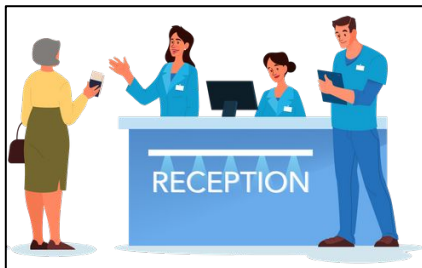
Часові ряди. Вступ

Передбачається, що часові ряди генеруються через певні часові відмітки. Коли дані в часовому ряду є регулярними, вони називаються **regular time series**, а коли вони не є регулярними або синхронізованими, їх називають **irregular time series**.



Regular time series - це такий, у якому часові інтервали між точками даних мають однакову довжину та послідовно розподілені.

Наприклад, добова температура є звичайним часовим рядом, оскільки точки даних можна збирати в один і той же час кожного дня.



Irregular time series - це такий, у якому часові інтервали між точками даних не однакові або непостійні.

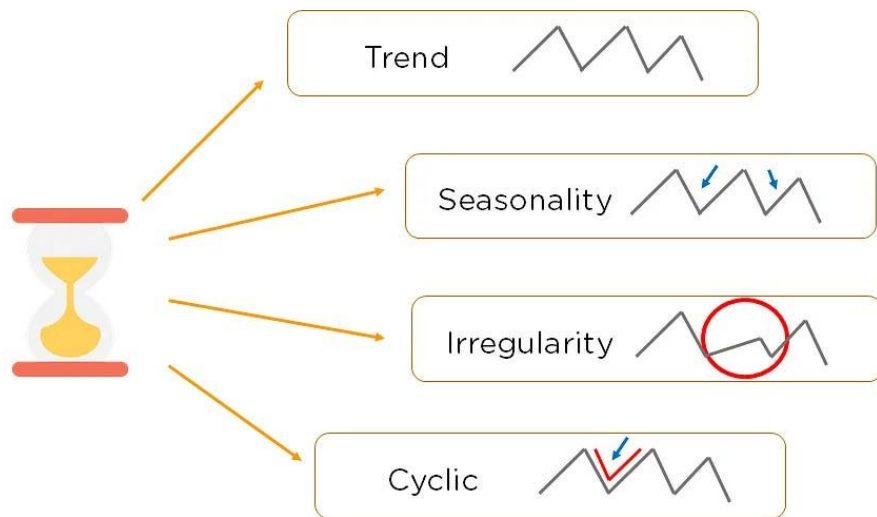
Наприклад, кількість відвідувань відділення невідкладної допомоги в лікарні щодня є нерегулярним часовим рядом, оскільки кількість відвідувань може сильно відрізнятися від дня до дня, а точки даних не збираються через регулярні послідовні проміжки часу.

Буде розглянуто далі

Часові ряди. Вимоги до даних

- Часові марки (timestamp): Кожне спостереження в часовому ряді повинно мати відповідний часовий маркер, що вказує на момент часу, коли було проведено вимірювання або отримано дані.
- Регулярність: Дані в часовому ряді повинні бути зібрані на регулярних інтервалах часу. Наприклад, щоденні, щомісячні або щорічні вимірювання.
- Відсутність пропущених значень: Важливо, щоб не було пропущених даних в часовому ряді. Якщо дані відсутні, можуть виникнути проблеми при аналізі та прогнозуванні.
- Стабільність: Часовий ряд повинен мати стабільну структуру в часі. Це означає, що статистичні властивості ряду (такі як середнє значення та дисперсія) не повинні суттєво змінюватися з часом.
- Наявність великої кількості спостережень: Зазвичай, чим більше даних в часовому ряді, тим краще. Це дозволяє отримати більш точні статистичні оцінки та забезпечує більш надійне прогнозування.
- Якість даних: Перевірте якість даних на предмет помилок, неточностей або аномалій. Якість даних може значно впливати на результати аналізу та прогнозування.
- Джерела даних: Перевірте надійність джерел звідки ви отримуєте дані для роботи.

Часові ряди. Складові



1. Рівень (Level): це базове значення ряду, якби це була пряма лінія. Це центральне, базове значення, навколо якого коливається часовий ряд.

2. Тенденція (Trend): це необов'язкова і часто лінійна (зростаюча або спадна) поведінка ряду з часом. Він фіксує довгостроковий напрямок, у якому рухається часовий ряд.

3. Сезонність: також необов'язкова, сезонність враховує повторювані шаблони чи цикли поведінки з часом. Наприклад, у роздрібній торгівлі може бути сезонна модель, пов'язана зі святами.

4. Шум: цей компонент представляє мінливість у спостереженнях, яку не можна пояснити за допомогою моделі. Він охоплює випадкові коливання та нерівності в даних.

$$y = \text{level} + \text{trend} + \text{seasonality} + \text{noise}$$

Часові ряди. Методи прогнозування часових рядів

Autoregressive Integrated Moving Average (ARIMA)

Цей метод поєднує в собі модель авторегресії та ковзного середнього. У моделях ARIMA кроки є послідовними як лінійна функція. Різниця в спостереженні та залишкові помилки знаходяться на попередніх кроках. Цей метод найкраще працює для одновимірних часових рядів із тенденціями та без сезонних компонентів.

Seasonal Autoregressive Integrated Moving-Average (SARIMA)

У моделі SARIMA наступні кроки виконуються у стилі лінійної функції на основі різних спостережень, помилок, відмінностей у сезонних спостереженнях і сезонних помилок на попередніх часових кроках. SARIMA пристосований для одновимірного часового ряду з тенденцією або сезонними компонентами.

Vector Autoregression (VAR)

Цей метод моделює наступний крок у кожному часовому ряді за допомогою методів AR. Він використовує узагальнення AR для множення паралельних часових рядів, наприклад багатовимірних часових рядів.

Simple Exponential Smoothing (SES)

У цьому методі модель використовує наступний крок як експоненціально зважену лінійну функцію спостереження на попередніх часових кроках. Цей метод найкраще працює для однофакторних часових рядів без трендів і сезонних компонентів.

Модель HGARCH

Це статистична модель, яка використовується для аналізу фінансових даних і прогнозування коливань вартості фінансових активів, таких як акції, валютні курси, товари і інші фінансові інструменти. Ця модель є розширенням класичної моделі GARCH і враховує гетероскедастичність.

Буде розглянуто далі

Що почитати/подивитись:

- Warren M. Persons, 1919. General Considerations and Assumptions. The Review of Economics and Statistics. 1, 5–107.
<https://doi.org/10.2307/1928754>
- Practical Time Series Analysis
<https://www.classcentral.com/course/practical-time-series-analysis-10151>
- Sequences, Time Series and Prediction
<https://www.classcentral.com/course/tensorflow-sequences-time-series-and-pre-13802>
- Applied Time-Series Analysis
<https://www.classcentral.com/course/swayam-applied-time-series-analysis-7889>
- Stanford Seminar - TSAR - Anirudh Todi of Twitter
<https://www.classcentral.com/course/youtube-stanford-seminar-tsar-the-timeseries-aggregator-anirudh-todi-of-twitter-110051>
- <https://www.youtube.com/playlist?list=PLsZ8kVwX52ZEFZsVVYs60lf7idJuKKUO>
- Time Series Analysis in Python
<https://www.datacamp.com/courses/time-series-analysis-in-python>

