Feature Engineering for Sequences

Raw Sequence Data

↓

Windowing

Creating fixed-size segments from sequences

Lag Features

Using previous timesteps as predictors

Rolling Statistics

Moving averages, standard deviations

Differencing

Removing trends to achieve stationarity

Fourier Transforms

Extracting frequency domain features

Embeddings

Representations for categorical sequential data

↓

Engineered Features

Ready for sequence modeling and prediction

Real-World Application Examples

Windowing - Stock Price Prediction

Original Data
Day 1: $100
Day 2: $102
Day 3: $101
Day 4: $105
Day 5: $107

3-Day Window Applied
Window 1: [100, 102, 101] → 105
Window 2: [102, 101, 105] → 107
Predict next day from past 3 days

Lag Features - Website Traffic

Hourly Visitors
10:00 - 150 visitors
11:00 - 180 visitors
12:00 - 220 visitors
13:00 - 200 visitors

Lag Features Created
Current: 220, 1hr ago: 180
Current: 200, 1hr ago: 220
Utilize previous hour information

Rolling Statistics - Temperature Data

Daily High Temperature (°C)
Mon: 22°C

3-Day Moving Average
Wed: 23.0°C (22+24+23)/3

Tue: 24°C	Thu: 24.0°C (24+23+25)/3
Wed: 23°C	Fri: 24.7°C (23+25+26)/3
Thu: 25°C	Reduce volatility, identify trends
Fri: 26°C	

📉 Differencing - Sales Data

Monthly Cumulative Sales ($M)	Monthly Growth (Difference)
Jan: $100M	Feb: +$30M
Feb: $130M	Mar: +$15M
Mar: $145M	Apr: +$25M
Apr: $170M	Remove trend, analyze changes

🌊 Fourier Transforms - ECG Signal

Time Domain Signal	Frequency Domain Features
t=0ms: 0.1V	Heart Rate: 75 BPM (1.25Hz)
t=10ms: 0.8V	High Frequency: 0.02
t=20ms: 0.2V	Low Frequency: 0.15
Voltage changes over time	Extract periodic patterns

Embeddings - User Behavior Sequence

Click Event Sequence

Homepage → Search

Product Page → Cart

Checkout → Complete

Categorical sequential data

Embedding Vectors

Homepage: [0.2, 0.8, 0.1]

Search: [0.3, 0.6, 0.4]

Product: [0.7, 0.2, 0.5]

Transform to semantic representation