



Polsko-Japońska Akademia Technik Komputerowych

Wydział Informatyki

# Zastosowanie filtrów Kalmana do poprawy predykcji cen giełdowych przy użyciu sieci LSTM

Praca Dyplomowa

**Autor:** Mikołaj Warda (s28034)

**Kierunek studiów:** Informatyka

**Specjalizacja:** Data Science

**Promotor:** dr Sinh Hoa Nguyen Thi

31 października 2025

# **Spis treści**

<b>1</b>	<b>Wstęp</b>	<b>3</b>
<b>2</b>	<b>Podstawy teoretyczne</b>	<b>4</b>
2.1	Przewidywanie szeregów czasowych na rynkach finansowych . . . . .	4
2.1.1	Wskaźnik siły względnej (Relative Strength Index, RSI) . . . . .	5
2.1.2	Wstęgi Bollingera (Bollinger Bands) . . . . .	5

## **Streszczenie**

# 1. Wstęp

Prognozowanie cen akcji odgrywa kluczową rolę w finansach, wspierając inwestorów w podejmowaniu świadomych decyzji zarządzania swoim portfolio. Chaotyczny charakter rynków sprawia jednak, że trafne przewidywanie notowań pozostaje trudnym zadaniem. Tradycyjna analiza techniczna bywa niewystarczająca - jest wrażliwa na szum, a wnioski często zawierają element subiektywności.

W ostatnich latach dynamiczny rozwój uczenia maszynowego, zwłaszcza sieci neuronowych, znacząco zmienił podejście do modelowania danych czasowych (w tym giełdowych). Architektury takie jak LSTM (Long Short-Term Memory) potrafią uchwycić złożone zależności i długookresowe relacje w danych, co czyni je obiecującymi narzędziami do prognozowania cen, jednak ich skuteczność nadal zależy od jakości danych wejściowych. Głównym problemem, jak w przypadku podejść tradycyjnych, jest wysoka wrażliwość na losowe wahania, które są nieodłącznym elementem danych finansowych. Te zniekształcenia wynikają z nieprzewidywalnych zdarzeń rynkowych. Model, zamiast uczyć się rzeczywistych trendów, modeluje wspomniane zakłócenia, obniżając swoją zdolność do generalizacji na nowych danych. W rezultacie, predykcje mogą być obarczone znacznym błędem, co podważa ich użyteczność w podejmowaniu decyzji inwestycyjnych.

Można temu zapobiegać stosując techniki filtracji na danych wejściowych. Jednym z klasycznych narzędzi tego typu jest filtr Kalmana, który może pełnić rolę modułu wygładzania i korekty obserwacji, podnosząc stabilność i dokładność predykcji. Trenowanie sieci neuronowej na przefiltrowanych danych przy użyciu filtra Kalmana może zredukować wpływ szumu, pozwalając na lepsze uchwycenie istotnych wzorców i trendów w danych.

W ramach niniejszej pracy dyplomowej, podjęto się zbadania skuteczności zastosowania filtra Kalmana jako etapu przetwarzania danych dla modeli LSTM w kontekście prognozowania cen akcji spółki Amazon.com Inc (AMZN). Wybór tematu pracy wyniknął z chęci zdobycia wiedzy na temat działania architektury LSTM oraz zbadania wpływu filtracji danych na jakość prognoz. Dodatkową motywacją była możliwość wszechstronnego rozwoju ze względu na złożony charakter problemu, łączącego zagadnienia z dziedziny uczenia maszynowego, analizy szeregów czasowych oraz teorii filtrów.

W dalszej części pracy, w sekcji "*Podstawy teoretyczne*", przedstawiono wszelkie niezbędne pojęcia związane z tematem pracy i przebiegiem badań ze szczególnym naciskiem na architekturę sieci LSTM oraz filtr Kalmana. Następnie, w sekcji "*Metodyka badań*", opisano podejście badawcze, w tym przygotowanie danych, implementację modelu, metryki oceny oraz wykorzystane narzędzia. Kolejna sekcja "*Wyniki i dyskusja*" prezentuje uzyskane wyniki eksperymentów wraz z ich analizą i interpretacją. Na zakończenie, w sekcji "*Podsumowanie i wnioski*", podsumowano przeprowadzone badania, przedstawiono kluczowe wnioski oraz zasugerowano kierunki dalszych badań w tym obszarze.

## 2. Podstawy teoretyczne

### 2.1. Przewidywanie szeregów czasowych na rynkach finansowych

Dane finansowe, takie jak np. ceny akcji, zawierają obserwacje tworzące szeregi czasowe. Innymi słowy, są to dane, które reprezentują zmiany wartości w określonych odstępach czasu. Formalnie szeregiem czasowym określa się zbiór uporządkowanych obserwacji w czasie, gdzie każda obserwacja jest powiązana z określona chwilą czasową:

$$X = \{x_1, x_2, x_3, \dots, x_n\} \quad (1)$$

gdzie  $x_i$  reprezentuje wartość obserwacji w czasie  $t_i$ , a  $n$  to liczba obserwacji w szeregu czasowym.

Abstrahując od rynków finansowych, przedstawię kilka przykładów szeregów czasowych z innych dziedzin, celem lepszego zobrazowania tej koncepcji:

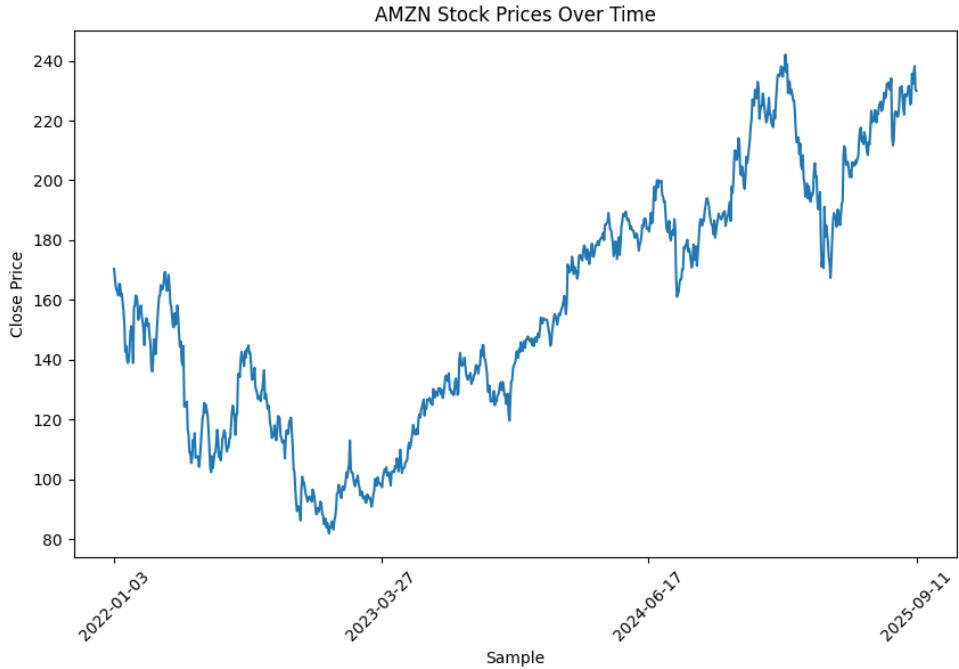
- **Energetyka:** Godzinowe zużycie energii elektrycznej
- **Meteorologia:** Codzienny pomiar temperatury
- **Sport:** Wyniki meczów drużyny na przestrzeni sezonu

Najczęściej spotykany formatem danych finansowych są tzw. dane OHLC. Przedstawione w takiej postaci informacje tworzą szereg czasowy, w którym każda obserwacja składa się z czterech części:

- **Open (O):** Cena otwarcia - cena, po której dany instrument finansowy rozpoczął notowania w danym okresie.
- **High (H):** Cena najwyższa - najwyższa cena osiągnięta przez instrument finansowy w danym okresie.
- **Low (L):** Cena najniższa - najniższa cena osiągnięta przez instrument finansowy w danym okresie.
- **Close (C):** Cena zamknięcia - cena, po której instrument finansowy zakończył notowania w danym okresie.

Na potrzeby tej pracy, wykorzystano składnik *Close*, ze względu na jego powszechnie wykorzystanie w analizie i prognozie cen akcji.

Rysunek 1 przedstawia przykładowy wykres cenowy względem składnika *Close* spółki Amazon.com Inc (AMZN) na przestrzeni kilku lat w odstępach dziennych. Dane przedstawione na wykresie są nieregularne i zawierają liczne fluktuacje, co jest charakterystyczne dla danych finansowych. Ta wysoka zmienność stanowi główne wyzwanie dla modeli predykcyjnych, co motywuje do poszukiwania skutecznych metod filtrujących.



Rysunek 1: Wykres cenowy względem składnika *Close* Amazon.com Inc (AMZN) w latach 2022-2025

Tradycyjnie analiza danych rynkowych opierała się o wskaźniki techniczne. Można je określić mianem narzędzi statystycznych reprezentujących różne aspekty zachowań cen. W niniejszej pracy wykorzystano trzy wskaźniki techniczne, które opiszę w następnych podsekcjach.

### 2.1.1 Wskaźnik siły względnej (Relative Strength Index, RSI)

Wskaźnik siły względnej (RSI) to popularny wskaźnik techniczny używany do oceny siły i prędkości zmian cen aktywów finansowych. Wyraża się go wzorem:

$$RSI = 100 - \frac{100}{1 + RS} \quad (2)$$

gdzie  $RS$  (Relative Strength) to stosunek średnich wzrostów do średnich spadków cen w określonym czasie. Według badań, optymalny okres do obliczania RSI wynosi 14 dni.

Produktorem końcowym RSI jest liczba z zakresu od 0 do 100, która interpretowana jest następująco:

- Wartości powyżej 70 sugerują, że akcja jest wykupiona i może nastąpić spadek cen.
- Wartości poniżej 30 sugerują, że akcja jest wyprzedana i może nastąpić wzrost cen.
- Wartości pomiędzy 30 a 70 wskazują na neutralny stan rynku.

### 2.1.2 Wstęgi Bollingera (Bollinger Bands)

Wstęgi Bollingera to narzędzie analizy technicznej, dostarczające informacji o zmienności rynku. Składają się z trzech linii na wykresie cenowym:

- **Środkowa linia:** Prosta średnia krocząca (SMA) obliczona na podstawie cen zamknięcia w określonym czasie. Najczęściej używanym okresem jest 20 dni.
- **Górna wstęga:** Obliczana jako suma środkowa linia i dwukrotności odchylenia standardowego cen w tym okresie.
- **Dolna wstęga:** Obliczana jako różnica środkowa linia i dwukrotności odchylenia standardowego cen w tym okresie.

Na podstawie ww. składowych można zinterpretować dwie nowe miary, które zostały wykorzystane w niniejszej pracy:

- **Bollinger Bandwidth (BBW):** Miara szerokości wstęg Bollingera, obliczana jako stosunek różnicy między górną a dolną wstęgą do środkowej linii:

$$BBW = \frac{UpperBand - LowerBand}{MiddleLine} \quad (3)$$

- **Bollinger %B (BB%):** Miara położenia ceny względem wstęg Bollingera, obliczana jako stosunek różnicy między ceną zamknięcia a dolną wstęgą do różnicy między górną a dolną wstęgą:

$$BB\% = \frac{Close - LowerBand}{UpperBand - LowerBand} \quad (4)$$

Opisane powyżej wskaźniki techniczne - RSI, BBW oraz BB% - dostarczają cennych informacji o dynamice rynku. W tradycyjnej analizie mogą posłużyć do subiektywnej interpretacji przez analityków. W nowoczesnych podejściach, opartych na modelach uczenia maszynowego, można je wykorzystać do wzbogacenia danych o dodatkowe cechy, co potencjalnie może poprawić jakość prognoz.