



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

28 października 2025

Spis treści

1	Abstrakt	2
2	Wstęp	3
3	Podstawy teoretyczne i przegląd literatury	4
3.1	Przewidywanie szeregów czasowych na rynkach finansowych	4
4	Metodyka badań	5
5	Wyniki i dyskusja	5
6	Podsumowanie i wnioski	5
7	Bibliografia	5
8	Spis rysunków i tabel	5
9	Załączniki	5

1. Abstrakt

2. 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.

Mimo że sieci LSTM wykazują dużą skuteczność w modelowaniu szeregów czasowych, ich praktyczne zastosowanie w prognozowaniu cen giełdowych napotyka na istotne wyzwania. Głównym problemem jest wysoka wrażliwość tych sieci na szum oraz losowe wahania, które są nieodłącznym elementem danych finansowych. Te zniekształcenia wynikają z nieprzewidywalnych zdarzeń rynkowych. Model, zamiast uczyć się rzeczywistych trendów, może zacząć modelować wspomniane zakłócenia, co obniża jego zdolność do generalizacji na nowych danych. W rezultacie, predykcje mogą być obciążone znacznym błędem, co podważa ich użyteczność w podejmowaniu decyzji inwestycyjnych.

Aby zapobiec negatywnemu wpływowi szumu na modele LSTM jest zastosowanie technik redukcji szumu w danych wejściowych. Jednym z klasycznych narzędzi tego typu jest filtr Kalmana – metoda estymacji stanu w systemach dynamicznych skażonych szumem – który może pełnić rolę modułu wygładzania i korekty obserwacji, podnosząc stabilność i dokładność predykcji. Taki filtr może działać jako warstwa przed przetwarzaniem przez LSTM, oczyszczając dane wejściowe z fluktuacji. Podanie na wejście modelu LSTM danych przefiltrowanych przez filtr Kalmana może zredukować wpływ szumu, pozwalając sieci skupić się na istotnych wzorcach i trendach.

Wybór tematu pracy wyniknął z chęci zdobycia wiedzy na temat działania sieci 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.

Celem jest zbadanie skuteczności integracji filtrów Kalmana z modelami LSTM w kontekście prognozowania cen akcji. Zakres obejmuje:

1. Stworzenie bazowego modelu LSTM do prognozowania cen akcji.
2. Stworzenie modelu LSTM na bazie danych wzbogaconych o dodatkowe parametry techniczne
3. Implementację filtra Kalmana do przetwarzania danych wejściowych.
4. Stworzenie modelu LSTM bazującego na danych przefiltrowanych przez filtr Kalmana.
5. Porównanie wydajności trzech podejść na rzeczywistych danych giełdowych.

W dalszej części pracy przedstawiono przegląd literatury dotyczącej prognozowania cen akcji i metod redukcji szumu. Następnie opisano zastosowaną metodykę badawczą, w tym szczegóły implementacji modeli i filtra Kalmana oraz informacje na temat wykorzystanych technologii. Kolejna sekcja prezentuje wyniki eksperymentów oraz ich analizę. Pracę kończy podsumowanie z wnioskami i propozycjami dalszych badań.

3. Podstawy teoretyczne i przegląd literatury

3.1. Przewidywanie szeregów czasowych na rynkach finansowych

Dane finansowe, takie jak np. ceny akcji, zawierają obserwacje tworzące szeregi czasowe. Innymi słowy, są to dane, zebrane w określonych odstępach czasu (np. godzinowych, dziennych, tygodniowych), reprezentujące zmiany wartości finansowych.

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

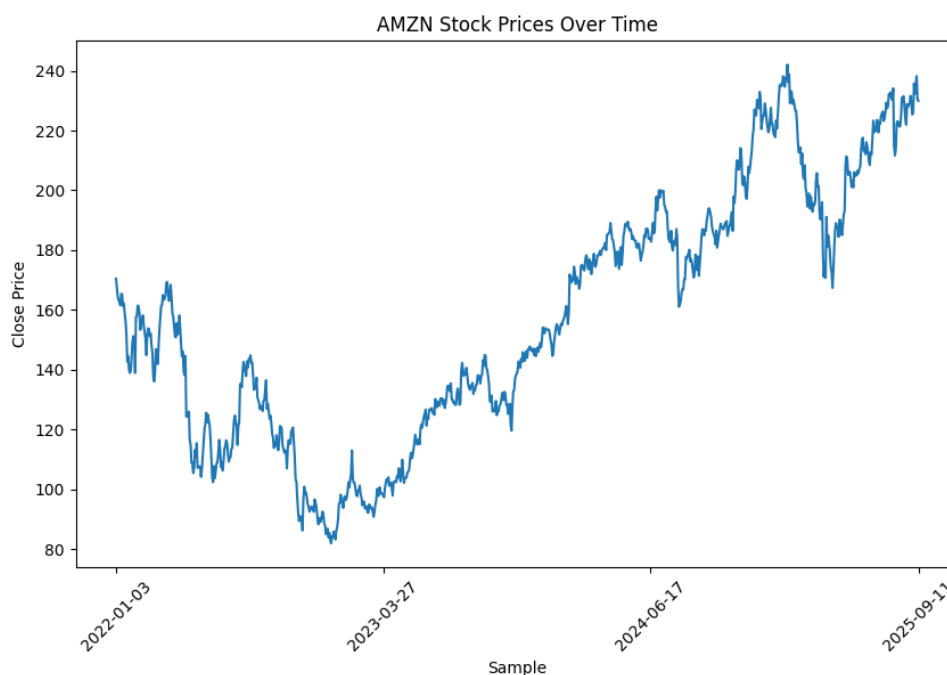
- **Energetyka:** Godzinowe zużycie energii elektrycznej
- **Meteorologia:** Codzienny pomiar temperatury
- **Sport:** Wyniki meczów drużyny na przestrzeni sezonu

Najczęściej spotykanym formatem danych finansowych są tzw. dane OHLC. Przedstawione w takiej postaci informacje, są szeregiem czasowym, w którym każdy odstęp czasowy ma przypisane cztery składniki:

- **Open (O):** Cena otwarcia - cena, po której dany instrument finansowy rozpoczął notowania w określonym 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 określonym okresie.

Na potrzeby tej pracy, wykorzystano wartość *Close*, ze względu na jej powszechne wykorzystanie w analizie i prognozie cen akcji.

Rysunek 1 przedstawia przykładowy szereg czasowy wartości *Close* spółki Amazon.com Inc (AMZN) na przestrzeni kilku lat w odstępach dziennych. Dane przedstawione na wykresie wyglądają na 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: Szereg czasowy wartości *Close* Amazon.com Inc (AMZN) w latach 2022-2025

Aby analizować takie dane, można wykorzystać modele uczenia maszynowego. Do najpopularniejszych podejść należą sieci rekurencyjne (RNN), w tym ich zaawansowane warianty, takie jak *LSTM* (Long Short-Term Memory) czy *GRU* (Gated Recurrent Units). Sieci te są zdolne do uchwycenia złożonych wzorców i zależności w szeregach czasowych, jednak wymagają odpowiedniego przygotowania danych. Oznacza to odpowiednie przygotowanie danych w taki sposób, aby model mógł je efektywnie przetworzyć i wyciągnąć z nich istotne informacje.

4. Metodyka badań

5. Wyniki i dyskusja

6. Podsumowanie i wnioski

7. Bibliografia

8. Spis rysunków i tabel

9. Załączniki