

Prognozowanie cen akcji przy użyciu klasycznych modeli predykcyjnych

Adrian Dobkowski

Czerwiec 2025

Temat pracy

Celem niniejszej pracy jest zbadanie, jak skuteczne mogą być wybrane klasyczne modele uczenia maszynowego w przewidywaniu cen akcji spółek notowanych na giełdzie papierów wartościowych. W ramach projektu porównano działanie trzech popularnych algorytmów: regresji liniowej, lasu losowego (Random Forest) oraz Gradient Boosting.

Dodatkowo, na potrzeby analizy stworzono interaktywną aplikację internetową opartą na bibliotece Streamlit, która automatycznie pobiera dane giełdowe, przetwarza je wstępnie i umożliwia generowanie prognoz w czasie rzeczywistym.

1 Wstęp

Przewidywanie cen akcji odgrywa ważną rolę w inwestowaniu, zarządzaniu ryzykiem i podejmowaniu decyzji na rynku giełdowym. Uczenie maszynowe w znacznym stopniu ułatwia nam analizę trendów i zachowań akcji różnych spółek i po dokładnym wyuczeniu potrafi lepiej i szybciej przewidywać zachowania giełdy niż człowiek.

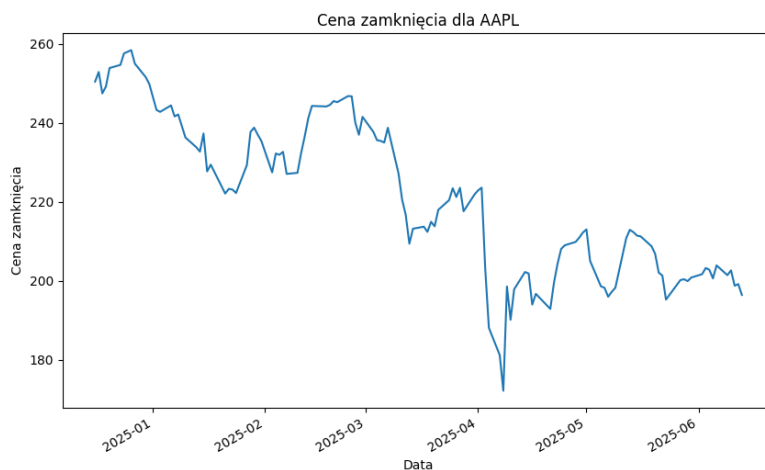
W projekcie skupiono się na wykorzystaniu klasycznych modeli regresyjnych do prognozowania przyszłych cen akcji na podstawie danych historycznych. Choć modele te są prostsze niż zaawansowane algorytmy, takie jak głębokie sieci neuronowe, w wielu przypadkach mogą zapewniać dużą dokładność, szczególnie przy prognozach krótkoterminowych. Co więcej, opracowana aplikacja webowa umożliwia użytkownikowi samodzielne analizowanie danych oraz testowanie modeli na wybranych spółkach notowanych na giełdzie, co czyni całe rozwiązanie praktycznym i łatwym w obsłudze.

2 Dane i metodyka

2.1 Opis danych

Dane historyczne dotyczące notowań akcji zostały pobrane z serwisu Yahoo Finance za pomocą dedykowanej funkcji `download_stocks()`. W analizie skupiono się na spółce Apple Inc. (ticker: AAPL). Otrzymany zbiór danych zawierał następujące atrybuty:

- Data – dzień sesji giełdowej,
- Open – cena otwarcia,
- High – najwyższa cena w danym dniu,
- Low – najniższa cena,
- Close – cena zamknięcia,
- Volume – wolumen obrotu.



Rysunek 1: Ceny akcji AAPL w przeciągu sześciu miesięcy

Na podstawie danych pierwotnych obliczono dodatkowe wskaźniki techniczne, takie jak średnie kroczące (Moving Averages) oraz wskaźnik siły względnej (RSI). Do zbioru danych wprowadzono również zmienną docelową **Target**, obliczona na podstawie przesunięcia czasowego ceny zamknięcia, służąca jako podstawa do nauki modeli.

2.2 Modele i metody

Do realizacji zadania predykcji cen wykorzystano trzy klasyczne modele regresyjne:

- **Linear Regression** – liniowy model regresji służący do modelowania zależności pomiędzy zmiennymi niezależnymi a zmienna zależna.
- **Random Forest Regressor** – zespołowy model bazujący na wielu drzewach decyzyjnych, stosujący mechanizm losowego wyboru prób i cech.
- **Gradient Boosting Regressor** – model oparty na iteracyjnym ulepszaniu słabych klasyfikatorów za pomocą metody gradientowego wzmacniania błędu.

Każdy z modeli oceniono za pomocą standardowych metryk regresyjnych:

- **MAE (Mean Absolute Error)** – średni błąd bezwzględny,
- **MSE (Mean Squared Error)** – średni błąd kwadratowy,
- **R² Score** – współczynnik determinacji,
- **Accuracy** – dokładność prognoz w ramach tolerancji 10 jednostek cenowych.

3 Wyniki

3.1 Metryki modeli

Model	MAE	MSE	R ²
Linear Regression	2.515	12.8376	0.5721
Random Forest	5.7668	48.9519	-0.6318
Gradient Boosting	6.2163	58.2887	-0.9431

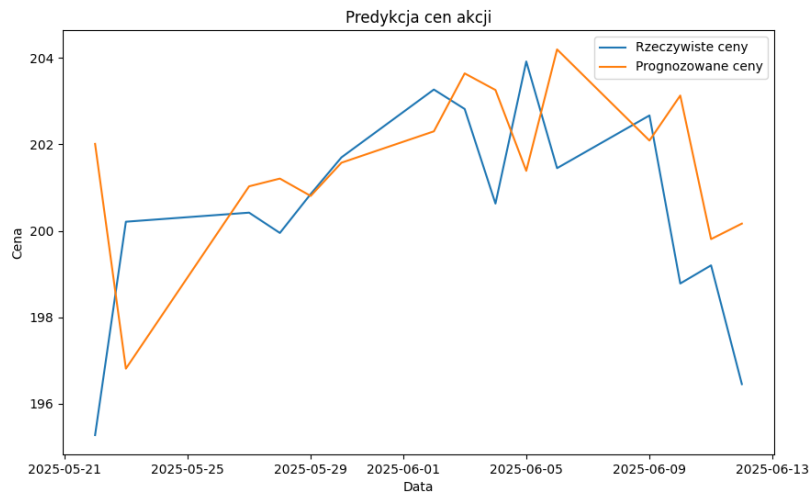
Tabela 1: Porównanie metryk błędów i współczynnika determinacji

3.2 Dokładność modeli

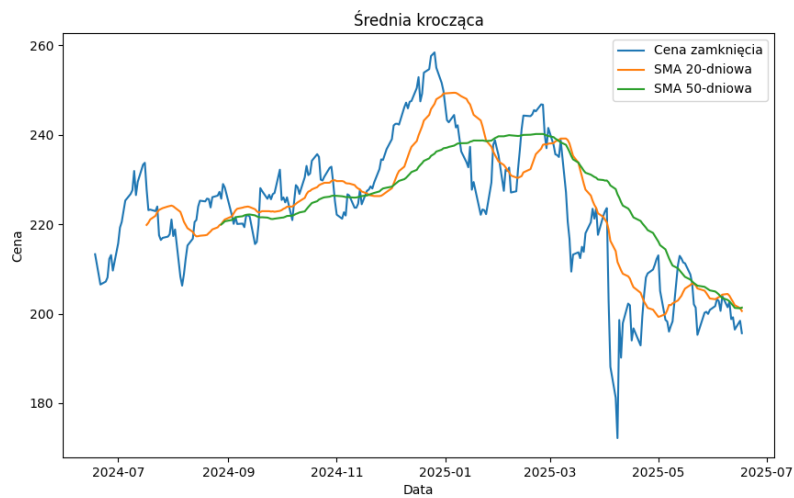
Model	Accuracy (%)
Linear Regression	97.5
Random Forest	87.5
Gradient Boosting	82.5

Tabela 2: Dokładność modeli predykcyjnych przy tolerancji 10 jednostek

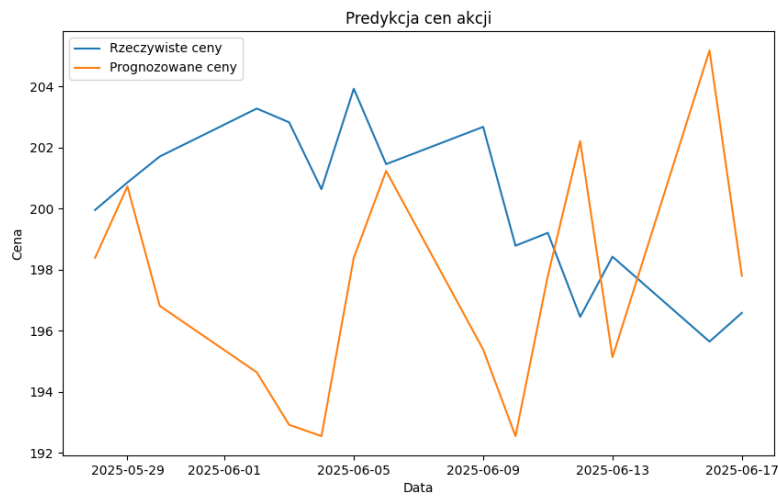
3.3 Wizualizacja wyników



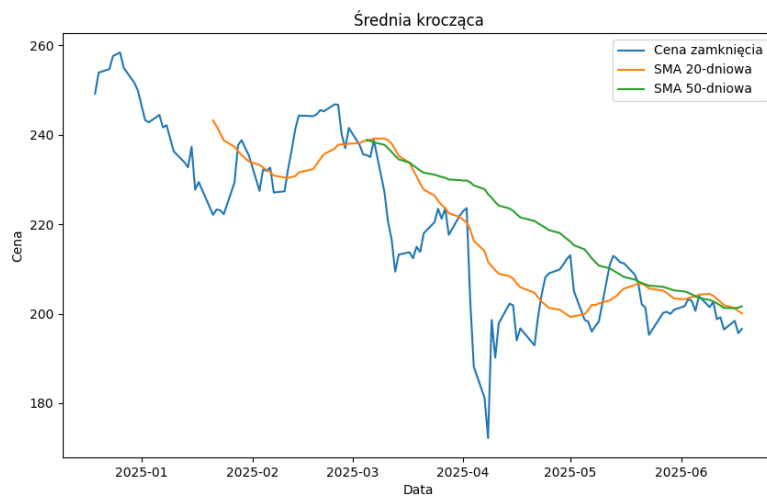
Rysunek 2: Porównanie predykcji z rzeczywistymi danymi – Linear Regression
najdokładniej odwzorowuje ceny



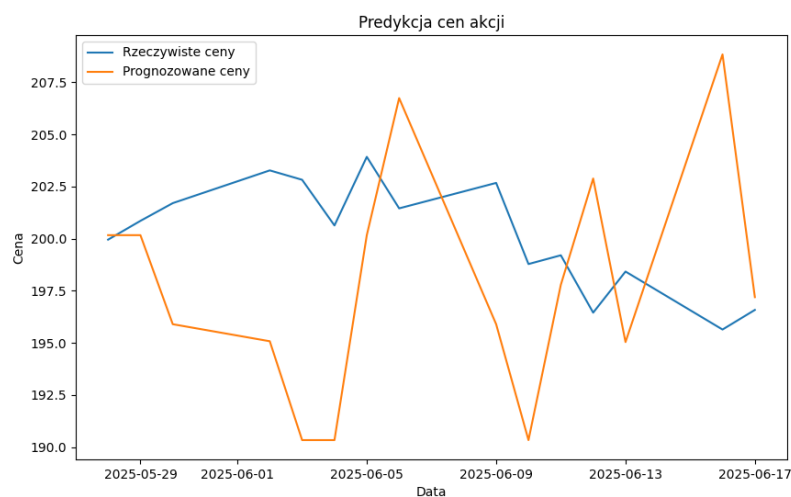
Rysunek 3: Wizualizacja średnich kroczących jako dodatkowego wskaźnika technicznego dla Linear Regression



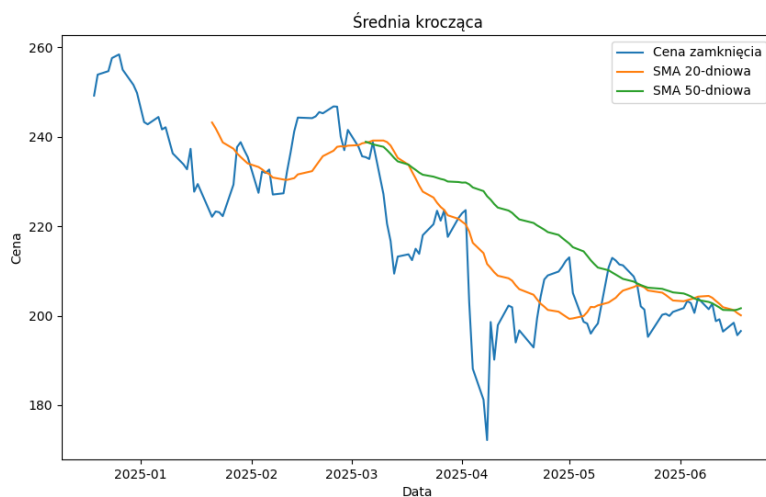
Rysunek 4: Porównanie predykcji z rzeczywistymi danymi – Random Forest Regressor



Rysunek 5: Wizualizacja średnich kroczących jako dodatkowego wskaźnika technicznego dla Random Forest Regressor

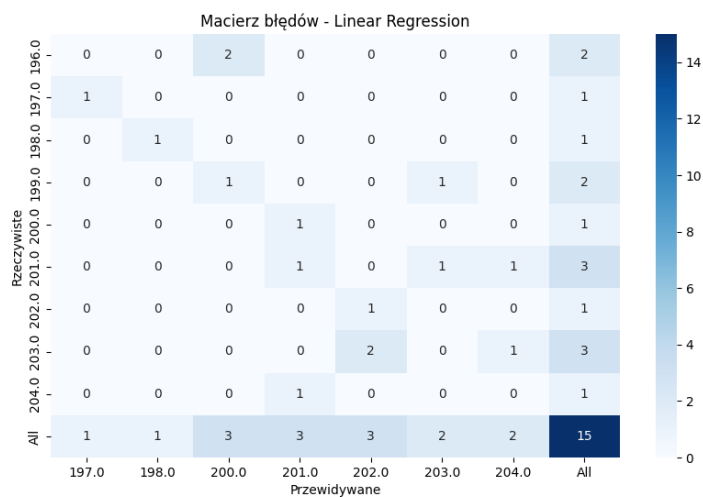


Rysunek 6: Porównanie predykcji z rzeczywistymi danymi – Gradient Boosting Regressor

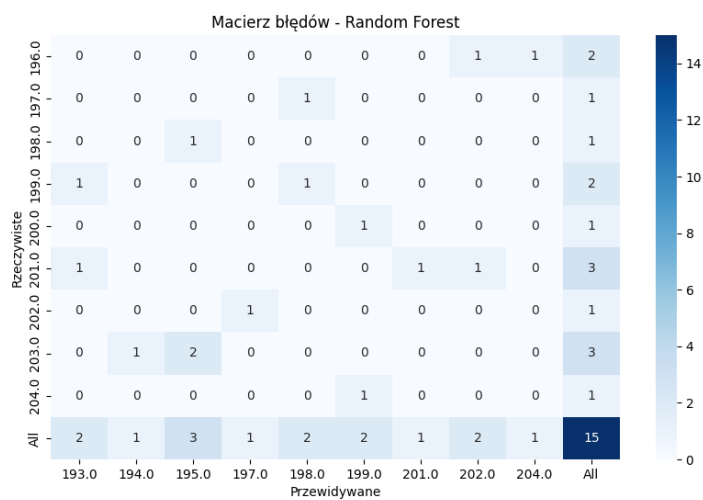


Rysunek 7: Wizualizacja średnich kroczących jako dodatkowego wskaźnika technicznego dla Gradient Boosting Regressor

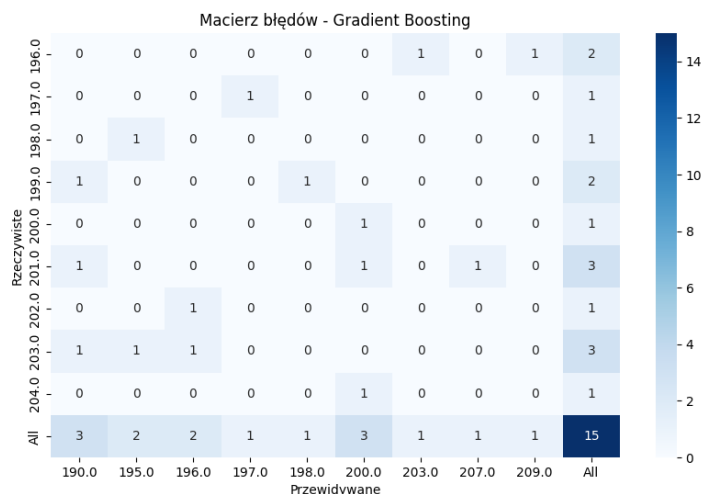
3.4 Macierze błędów



Rysunek 8: Macierz błędów dla Linear Regression



Rysunek 9: Macierz błędów dla Random Forest Regressor



Rysunek 10: Macierz błędów dla Gradient Boosting Regressor

4 Dyskusja

Wyniki analizy jednoznacznie wskazują, że model regresji liniowej (Linear Regression) przewyższył pozostałe metody zarówno pod względem dokładności, jak i stabilności prognoz. Wysoka wartość współczynnika R^2 , niskie błędy MAE i MSE oraz największa dokładność predykcji sugerują, że model ten dobrze radzi sobie z krótkoterminowym przewidywaniem cen giełdowych.

Zaskakująco słabe wyniki modeli Random Forest oraz Gradient Boosting mogą wynikać z nieprawidłowego ustawienia hiperparametrów lub też braku dostatecznych danych. Pokazuje to, że bardziej złożone modele niekoniecznie oferują lepszą skuteczność w analizie prostych i krótkoterminowych zależności cenowych.

5 Wnioski

W przeprowadzonej analizie model **Linear Regression** okazał się najbardziej efektywnym podejściem do prognozowania cen akcji. Jego przewaga wynika z prostego działania, małej podatności na przetrenowanie oraz dobrej zdolności do dopasowania się do krótkoterminowych danych akcyjnych. Wyniki te sugerują, że w kontekście krótkoterminowej analizy danych finansowych, klasyczne metody mogą być równie – jeśli nie bardziej – skuteczne niż bardziej złożone algorytmy uczenia maszynowego.