Raportu z projektu

06-DUMALI0 2024/SZ

# Cele projektu

* Przewidywanie ceny zamykającej BTC z następnego dnia na podstawie aktualnej ceny zamykającej, wolumenu oraz wskaźników technicznych EMA, SMA i RSI.
* Analiza ważności cech wejściowych.
* Sprawdzenie potencjalnej przydatności modelu w inwestowaniu.

# Dane

Dane dzienne dla BTCUSDT z okresu 08.09.2019 - 31.12.2024 zostały pobrane za pomocą biblioteki python-binance. Po przetworzeniu danych uzyskano 1941 przykładów, które podzielono na trzy zbiory:

* Zbiór treningowy: 70% (1358 przykładów)
* Zbiór walidacyjny: 10% (194 przykłady)
* Zbiór testowy: 20% (389 przykładów)

# Modele

W projekcie wykorzystano 4 modele:

* Random Forest (RF) - liczba drzew: **50**
* Gradient Boosting (GB) - liczba estymatorów: **50**, współczynnik uczenia się: **0.1**
* Long Short-Term Memory (LSTM) - liczba epok: **100**, rozmiar wsadu: **16**
* Gated Recurrent Unit (GRU) - liczba epok: **50, r**ozmiar wsadu: **16**

Hiperparametry modeli dobrano na podstawie zbioru walidacyjnego.

# Ewaluacja

Do ewaluacji wykorzystano metryki: RMSE, MSE, MAE, R2.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Model | RMSE | MSE | MAE | R2 |
| RF | 0.1189 | 0.0141 | 0.0615 | 0.3751 |
| GB | 0.1204 | 0.0145 | 0.0630 | 0.3601 |
| LSTM | 0.0250 | 0.0006 | 0.0177 | 0.9723 |
| GRU | 0.0199 | 0.0004 | 0.0144 | 0.9825 |

W celu dodatkowej ewaluacji zastosowano metryki klasyfikacyjne: accuracy, precision, recall i F1-score. Problem przewidywania ceny został przekształcony w klasyfikację, gdzie celem było określenie trendu wzrostowego lub spadkowego, a nie dokładna prognoza wartości.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Model | Accuracy | Precision | Recall | F1-score |
| RF | 0.50 | 0.54 | 0.48 | 0.51 |
| GB | 0.47 | 0.51 | 0.36 | 0.42 |
| LSTM | 0.47 | 0.50 | 0.51 | 0.51 |
| GRU | 0.46 | 0.49 | 0.48 | 0.49 |

# Wnioski

* Modele LSTM i GRU najlepiej radziły sobie z problemem regresji, uzyskując najmniejsze RMSE, MSE, MAE oraz najlepsze R2.
* Po przekształceniu problemu do klasyfikacji, modele ujawniły swój brak potencjału, np. w przypadku modelu Random Forest, czyli najdokładniejszego pod względem klasyfikacji efekt był równie przypadkowy jak rzucanie monetą. Żaden z modeli nie zapewnia istotnej przewagi w podejmowaniu decyzji inwestycyjnych.
* Modele Random Forest i Gradient Boosting nie radzą sobie z extrapolowaniem danych poza zakres, na którym były trenowane.
* Choć modele mogą wydawać się skuteczne przy prognozowaniu ceny zamknięcia następnej świecy na podstawie bieżącej ceny, predykcja ta często polega na "przesunięciu w czasie" bieżącej wartości, co nie ma realnej wartości w handlu.
* Krzywa predykcji może być niemal identyczna z rzeczywistą ceną, ale przesunięta o jedną świecę, co sugeruje, że model jedynie odwzorowuje ruchy rynku, zamiast je przewidywać.
* Modele korzystają głównie z wartości ceny zamknięcia, co widać po analizie ważności cech.
* Skuteczne przewidywanie ruchów cen w praktyce wymaga większej ilości danych i bardziej zaawansowanych metod decyzyjnych - sam model nie wystarczy.