

Predikcia Cien Kryptomien

1. Úvod

Cieľom tohto projektu je predpovedať ceny kryptomien pomocou rôznych modelov strojového učenia, ako je lineárna regresia, Random Forest a XGBoost. Pre tento projekt bol použitý historický dataset o kryptomene Aave (AAVE), ktorý obsahuje atribúty ako otvorenie, najvyššia, najnižšia, close price, objem obchodov a trhovú kapitalizácia.

Cieľ Projektu:

Predikcia cien kryptomien na základe historických dát.

2. Načítanie a príprava dát

Prvým krokom bolo načítanie datasetu a jeho príprava na modelovanie. Dataset obsahuje údaje o kryptomene Aave, s viacerými stĺpcami, ktoré sú relevantné pre predikciu cien.

Zdroj dát

<https://www.kaggle.com/datasets/sudalairajkumar/cryptocurrencypricehistory>

Dataset obsahuje nasledujúce atribúty:

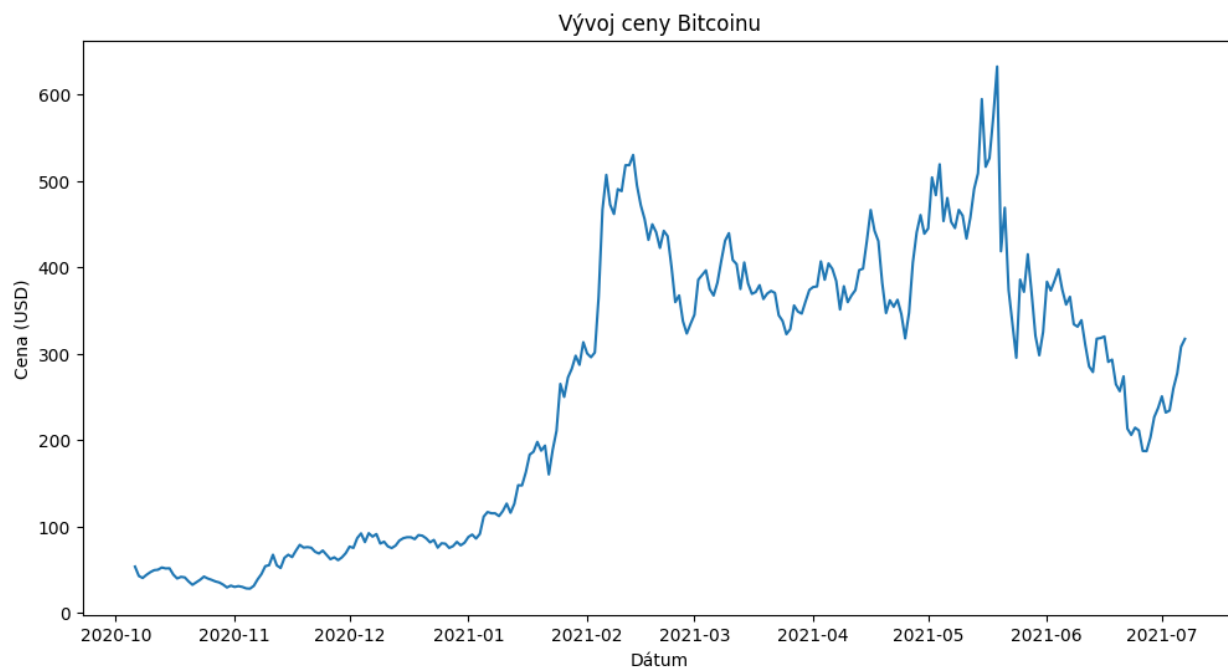
- **Name:** Názov kryptomeny
- **Symbol:** Symbol kryptomeny
- **High:** Najvyššia cena za deň
- **Low:** Najnižšia cena za deň
- **Open:** Otváracia cena za deň
- **Close:** Zavretá cena za deň
- **Volume:** Objemy obchodov
- **Marketcap:** Trhovú kapitalizácia
- **Daily Return:** Denný výnos

Príprava dát:

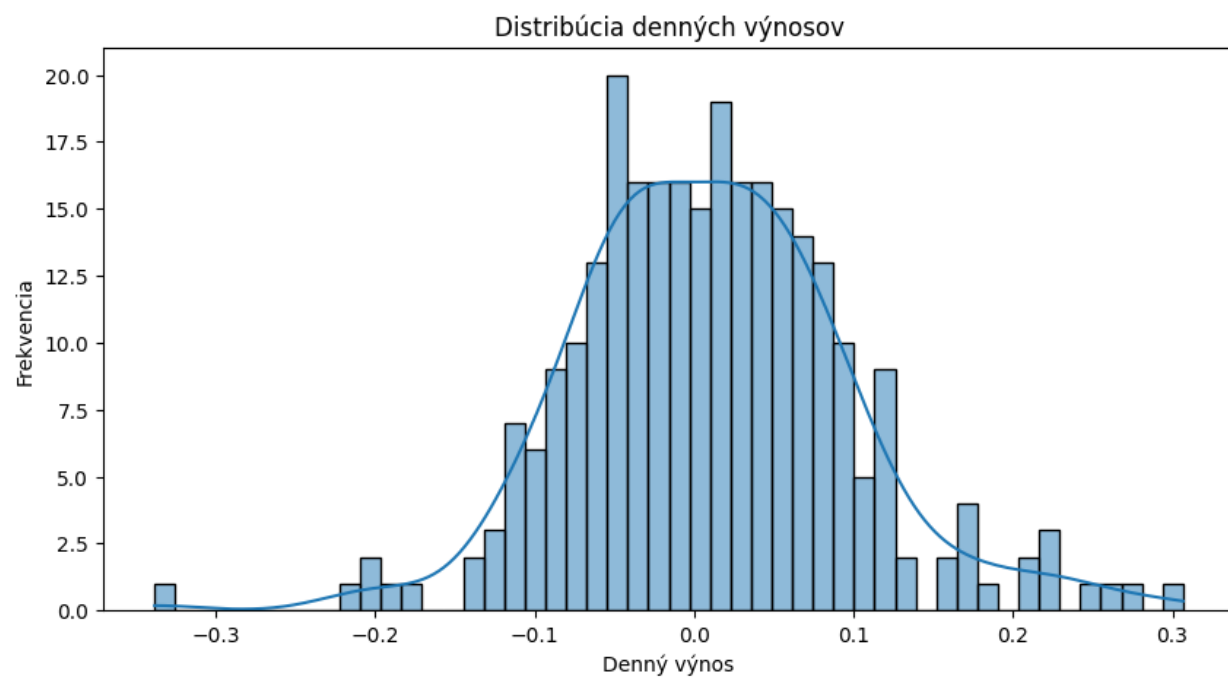
Pre modelovanie sme sa rozhodli predikovať cenu (stĺpec Close) na základe ostatných atribútov. Pred modelovaním sme odstránili chýbajúce hodnoty (NaN).

3. Exploratívna analýza dát

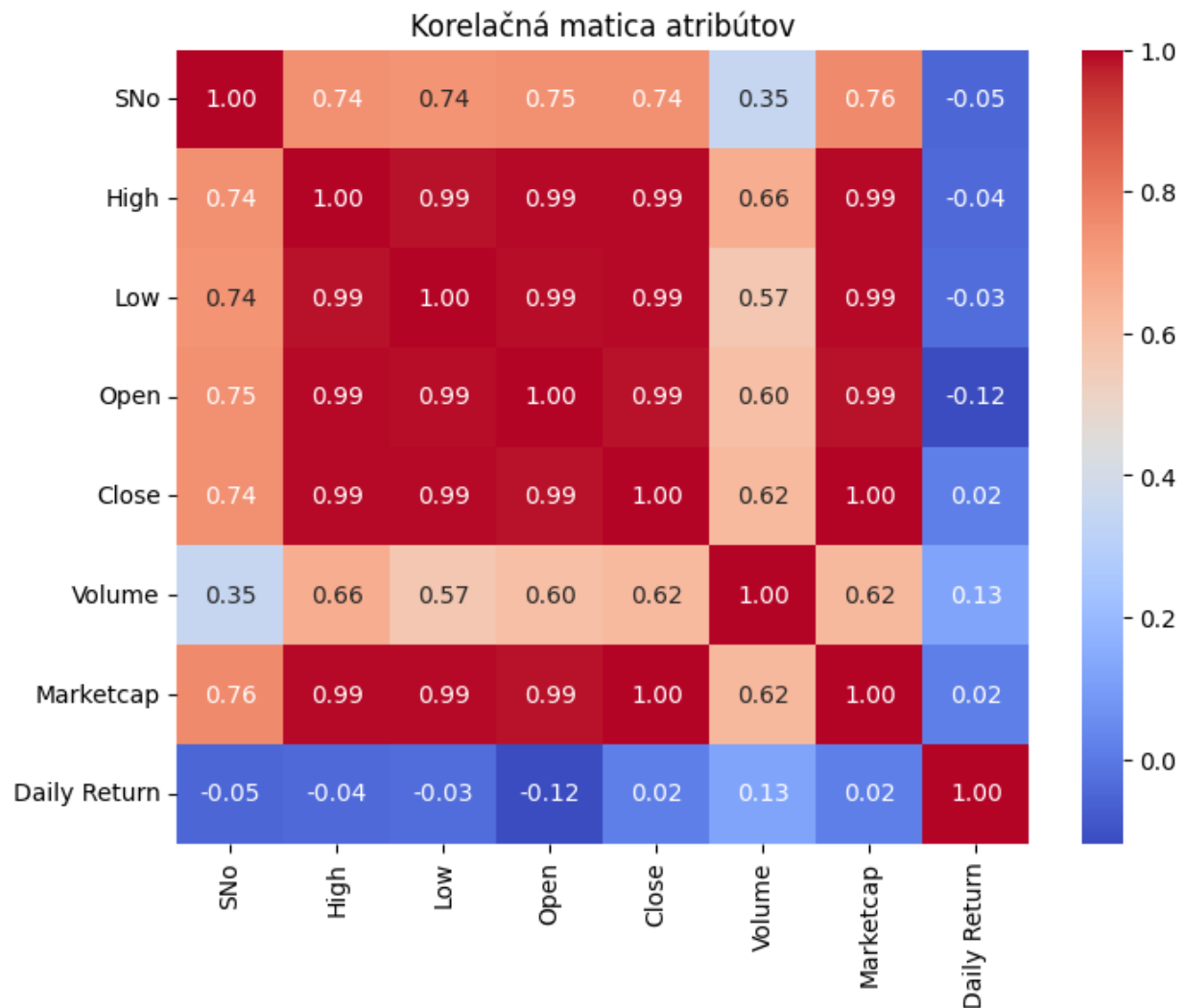
Historický vývoj cien Aave



Distribúcia denných výnosov



Korelačná matica:



Zobrazenie korelácie medzi jednotlivými atribútmi nám pomohlo pochopiť vzťahy medzi nimi.

4. Predspracovanie dát

Dataset sme rozdelili na tréningovú a testovaciu množinu (80% / 20%).

5. Modely a ich tréning

Lineárna regresia

Prvým modelom, ktorý sme použili, bola lineárna regresia, ktorá slúži ako základný model.

Výsledky:

- **MSE:** 18.54
- **RMSE:** 4.31
- **R²:** 0.999

Random Forest

Ďalším modelom bol Random Forest, ktorý je vhodný na prácu s nelineárnymi vzťahmi v dátach.

Výsledky:

- **MSE:** 49.40
- **RMSE:** 7.03
- **R²:** 0.998

XGBoost

Posledným modelom bol XGBoost, ktorý je vhodný pre jeho rýchlosť a presnosť.

Výsledky:

- **MSE:** 32.66
- **RMSE:** 5.71
- **R²:** 0.999

6. Porovnanie výkonnosti modelov

Model	MSE	RMSE	R ²
Lineárna regresia	18.54	4.31	0.999
Random Forest	49.40	7.03	0.998
XGBoost	32.66	5.71	0.999

Výsledky analýzy a modelovania

- **Lineárna regresia** poskytla veľmi dobré výsledky s vysokým R^2 (0.999), čo naznačuje, že model veľmi dobre zachytáva vzťahy v dátach. Tento model však môže byť príliš jednoduchý pre zložitejšie nelineárne vzťahy v kryptomenách.
- **Random Forest** model, aj keď dosiahol nižší R^2 (0.998) a vyšší MSE (49.4), stále poskytol veľmi dobré predikcie. Tento model je schopný modelovať nelineárne vzťahy medzi atribútmi.
- **XGBoost**, ktorý je založený na gradientnom zosilňovaní, dosiahol výsledky veľmi podobné lineárnej regresii, ale s mierne lepšími výsledkami v porovnaní s Random Forest. Tento model využíva techniku, ktorá sa zameriava na optimalizáciu modelu a minimalizáciu chýb, čo ho robí výkonným pri spracovaní komplexných dát.

Všetky tri modely poskytli predikcie, ktoré sa veľmi blížili reálnym hodnotám, čo svedčí o kvalite a presnosti použitých techník. Z hľadiska výkonnosti sa modely **XGBoost** a **lineárna regresia** ukázali ako najlepšie vo vzťahu k predikciám kryptomien, pričom XGBoost mal malý náskok pred Random Forest.

Odporúčania a ďalší vývoj:

- **Ladenie modelov:** V budúcnosti by bolo možné vykonať ďalšie ladenie hyperparametrov, aby sa ešte zlepšila výkonnosť modelov, najmä pre Random Forest a XGBoost.
- **Zvýšenie počtu atribútov:** Pridanie ďalších relevantných atribútov, ako napríklad indikátory technickej analýzy (RSI, Moving Average, atď.), by mohli zlepšiť predikciu cien kryptomien.
- **Testovanie s rôznymi kryptomenami:** Tento prístup by bolo možné aplikovať aj na ďalšie kryptomeny, čo by umožnilo porovnanie a možno lepšie porozumenie vzorcov správania cien kryptomien.
- **Zohľadnenie makroekonomických faktorov:** Pre ďalšie vylepšenia by bolo možné pridať aj externé faktory, ako sú ekonomické zmeny, regulačné zmeny, a iné, ktoré môžu ovplyvniť trh s kryptomenami.

Projekt ukázal, že aj relatívne jednoduché modely, ako je lineárna regresia, môžu byť veľmi účinné v oblasti predikcie cien kryptomien, avšak pokročilejšie modely ako XGBoost ponúkajú ešte lepšie výsledky. Tento prístup môže byť základom pre budúce aplikácie a zlepšovanie predikcií.

Vizualizácie porovnania jednotlivých modelov

