

Опис на проект

Проект: CryptoInfo

Тим:

- Димитар Арсов
- Филип Гавриловски
- Андреј Ристиќ

1. Вовед и цели на проектот

Целта на овој проект е да се дизајнира, развие и имплементира целосно функционална веб апликација за детална анализа на крипто берзата. Апликацијата ќе им овозможи на корисниците да пристапат, визуелизираат и анализираат историски податоци за највредните 1000 активни криптовалути. Основниот фокус е ставен на собирање и обработка на податоци на дневно ниво кои се протегаат најмалку во последните 10 години (или максималниот достапен период за секоја валута), со што ќе се создаде основа за техничка анализа и следење на пазарните трендови.

2. Истражување и избор на извор на податоци

Клучен предизвик за овој проект е набавката на комплетни, историски податоци за широк опсег на средства. Нашиот тим спроведе истражување на неколку пристапи:

1. **Јавно достапни API (на пр. Binance, Kraken, CoinGecko):** Иако овие API се моќни, утврдивме дека нивните бесплатни верзии се несоодветни за нашите барања. Тие наметнуваат строги ограничувања на бројот на повици (rate limits) и, што е најважно, драстично го ограничуваат историскиот опсег на податоци.
2. **Web Scraping (директно):** Разгледавме и директен scraping на веб-страници како Yahoo Finance. Ја користиме табелата што ја нуди овој сајт за да собереме барем 1000 криптовалути што потоа ќе ги филтрираме според дадените барања.

Нашиот пристап ќе биде:

- Да го користиме Yahoo Finance за да преземеме ажурирана листа на **топ 1000 криптовалути** според пазарна капитализација. Потоа ги филтрираме истите.
- Користиме open-source библиотека yfinance што ни помага да собереме информации за валутите, т.е. побараните информации како OHLCV.

3. Технологии

- **Обработка на податоци:** За обработка на податоците го одбравме Python поради неговата зрелост во екосистемот за наука за податоци, одличните библиотеки како `requests` (за мрежни повици), `Pandas` (за трансформација на податоци) и `BeautifulSoup` (за парсирање html).
- **Backend:** Java Spring Boot, како водечка рамка за работа на серверска страна.
- **Складирање на податоци:** За потребите на првата домашна, ќе користиме PostgreSQL база на податоци.
- **Frontend (Идни фази): React.js** во комбинација со потребните библиотеки за визуелизација на податоци.

4. Обработка и складирање на податоци

Процесот на ETL (Extract, Transform, Load) ќе биде имплементиран преку **Pipe and Filter** архитектура:

- **Филтер 1:** Scraping на Yahoo Finance за да се добие листата на топ 1000 валути и да се исфилтрираат тие што не задоволуваат одредени услови.
- **Филтер 2:** За секоја валута, паралелно повикување на yFinance за собирање на додатните податоци (OHLCV, пазарна капитализација, волумен). При иницијалното пополнување на базата се преземаат сите податоци за топ 1000-те валути, додека при последователните полнења, се ставаат податоци во базата само за новите (досега непостоечки во базата) валути. Дополнително се ажурира датумот кога последните информации за секоја валута биле земени и тие датуми се праќаат на следниот филтер.
- **Филтер 3:** 3 Проверка на последниот датум кога се земени информации за секоја валута, за да може да ги преземе само податоците што недостасуваат.
- **Филтер 4 :** Зачувување на прочистените податоци во нашата база (PostgreSQL).

Ќе ги обработуваме и складираме следните податоци за секоја валута на дневно ниво: **Датум, Open, High, Low, Close, Volume (волумен).**

5. Очекувани резултати и придобивки

Примарниот резултат од овој проект ќе биде **сеопфатна, чиста и ажурирана база на податоци** за историското движење на крипто-пазарот. Придобивките од овој проект за **крајните корисници:** Обезбедување на моќна алатка за анализа која е бесплатна и базирана на комплетни податоци, што овозможува информирано донесување одлуки за тргување и инвестирање.