

Собирање и Анализа на Финансиски Податоци

Вовед

Овој проект има за цел да и адресира потребите за следење на историски податоци како и нивна анализа и визуелизација. Нашата основна цел е да се создаде робусна платформа која ќе овозможи автоматизирано собирање на историски податоци од најголемите компании во светот, како и нивна ефикасна обработка, складирање, напредна анализа и интуитивна визуелизација. Системот е наменет за широк спектар на корисници, вклучувајќи и финансиски аналитичари, инвеститори, трговци и истражувачи кои имаат потреба од точни и ажурирани податоци.

Опис на системската архитектура

Нашиот систем е дизајниран како модуларна архитектура составена од пет основни компоненти кои комуницираат меѓу себе преку дефинирани интерфејси:

1. Сервис за собирање информации – Овој модул е одговорен за комуникацијата со надворешните извори на податоци. Се одлучивме да користиме податоци од “AlphaVantage”, како и “FMP”, податоците ги повлекуваме преку повеќе методи како REST-API повици, WebSocket врски и парсирање на JSON. Модулот вклучува механизми за обработка на грешки, проверка на дупликации.
2. Сервис за обработка на податоци во реално време – Овој модул го сметаме за јадрото на системот бидејќи тој е одговорен за обработката на податоците. Користиме Apache Kafka како broker за управување со податоците како и Apache Spark за извршување сложени трансформации.

3. Сервис за складирање на податоци – Тука имаме имплементирано хиерархиски пристап:
 - a. Cassandra – Оптимизирана за висок проток на податоци како и запишување и читање.
 - b. PostgreSQL – Користена за метаподатоци, кориснички профили и релациони податоци (Ќе биде имплементирана целосно во следната верзија)
4. Сервис за анализа – Тука користиме Machine Learning пристап преку Scikit-learn библиотеката со разновидни алгоритми за машинско учење. Некои од клучните алгоритми што ги користиме се:
 - a. LinearRegression– Базични предвидувања
 - b. RandomForestRegressor – Комплексни трендови
 - c. SVR – Нелинеарни односи
 - d. DecisionTreeRegressor – Референтен модел

Користиме и индикатори како влезни карактеристики

-ma_5: Показува 5 дневен просек

-ma_10: Показува 10 дневен просек

-volatility: Волатилност пресметана како отстапување

5. Сервис за визуелизација – Веб-базиран интерфејс со три нивоа

Детална спецификација на барањата

Функциски барања

1. Поддршка за повеќе API извори
2. Способност за обработка на 1000+ податоци паралелно
3. Историско пополнување минимум 15 години
4. Поддршка за повеќе временски прозорци
5. Пресметки користејќи технички индикатори
6. Гаранција за интегритет на податоците
7. Напредни механизми за компресија
8. Дневна рекалибрација на моделите
9. Интерактивни временски серии со zoom функционалност
10. Приказ на повеќекратни временски рамки

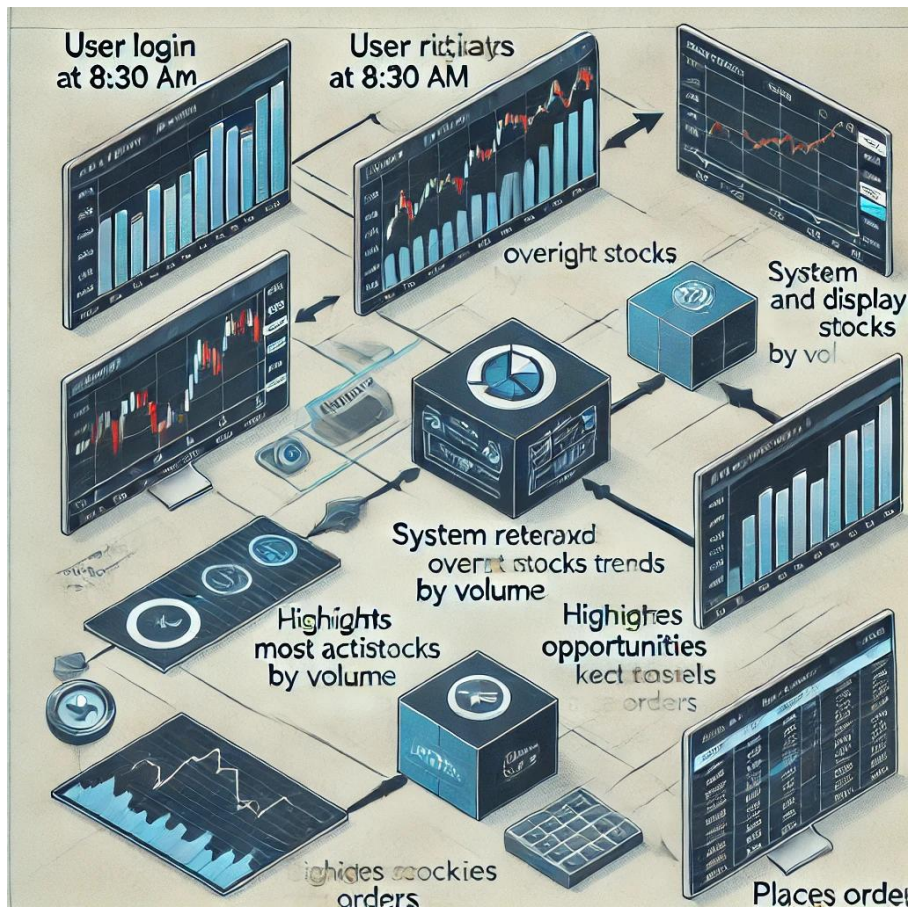
Нефункционални барања

1. Пропусен опсег: 10,000 трансакции во секунда
2. Латенција < 500ms
3. Време на опоравување при пад помало од 5 мин
4. Детални логови
5. Поддршка на повеќе јазици
6. Гео-редундантни копии
7. Безбедност на податоци

Кориснички сценарија

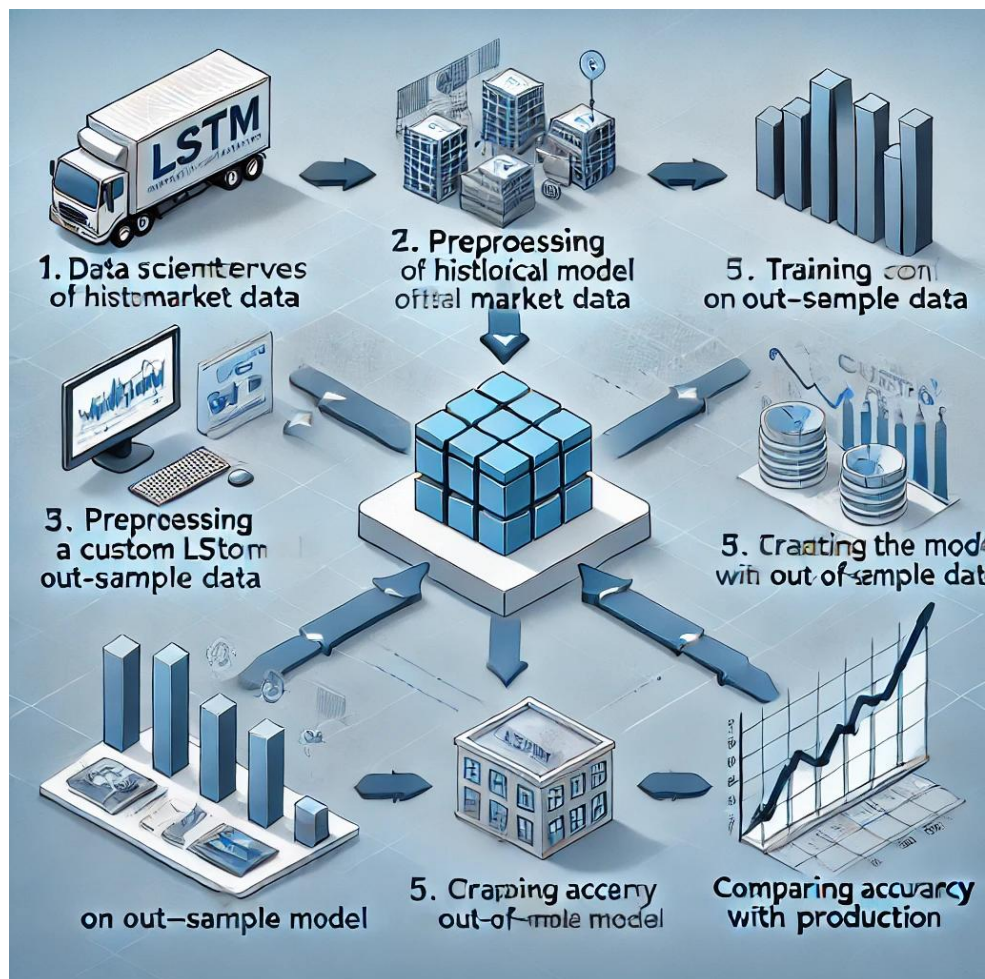
Сценарио 1 – Корисникот Наум е портфолио менаџер, се најавува на системот во 8:30 наутро. Системот автоматски му ги покажува:

- Преглед на пазарните движења преку ноќ
- Најголемите движења во цена од претходната ноќ
- Неговите компании коишто ги следи



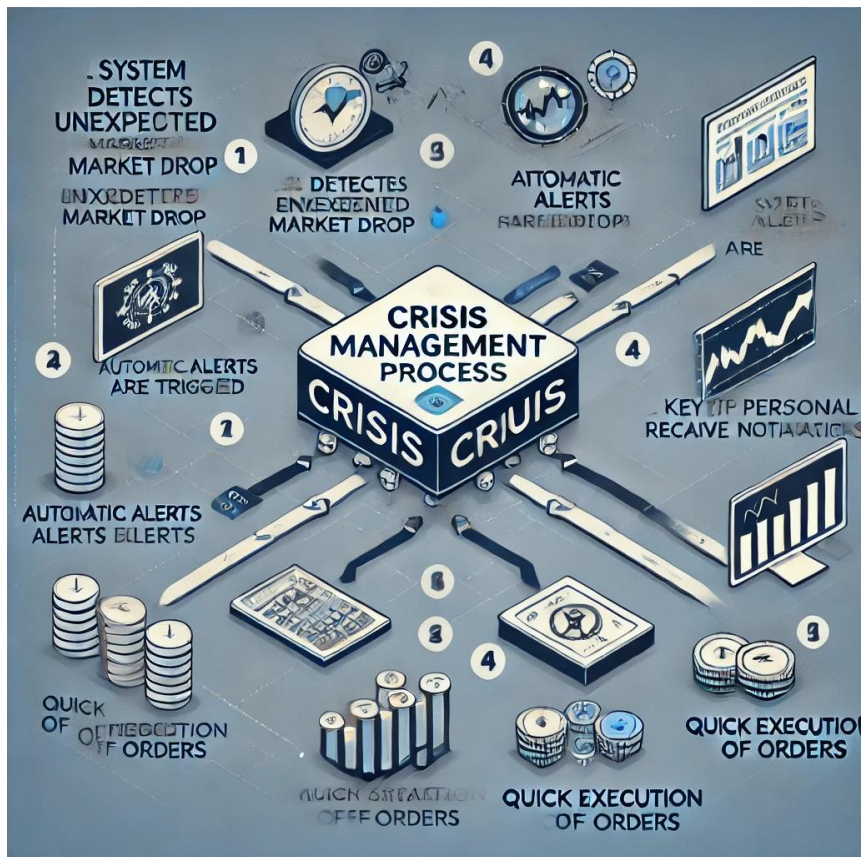
Сценарио 2 – Корисничката Марија која работи како data-scientist го користи моделот да:

- Земе историски податоци
- Тренира custom LSTM модел
- Тестира
- Ја спореди точноста со вградените модели



Сценарио 3 – При детекција на неочекуван пад на пазарот системот:

- Активира аларми
- Ги известува корисниците
- Генерира аналитички извештај
- Ги прикажува релевантните податоци



Планирани насоки и развој:

Краток Рок:

- Интеграција со криптовалути
- Sentiment анализа од вести
- Мобилна апликација

Среден Рок:

- AI- асистирани препораки
- Пазарни симулации
- Алго-трејдинг интерфејс

Долг Рок:

- Интеграција со DeFi платформи
- Предиктивна анализа на макро економски трендови
- Децентрализирана архитектура