



سیستم تحلیل مالی و معاملات بلادرنگ با هوش مصنوعی

پروژه نهایی درس تحلیلها و سیستمهای داده حجیم





استاد درس: دکتر حسن نادری

دانشجویان: علی احمدی، امیر کریمی، مهدی شکاری و اهورا امینی

بهمن ۱۴۰۳

فهرست مطالب

2	ىقدمە
2	هدف يروژه
2	·
2	·
3	معماری سیستم
4	
4	Data Ingestion Service
4	Apache Kafka
4	Stream Data Processing Service (Apache Spark)
5	محاسبه شاخصها
5	تولید سیگنالها
6	پردازش میکروبچها
6	خواندن و نوشتن با Spark
6	مديريت استريم
6	سرویس میانی Kafka-to-QuestDB
6	پردازش پیامها
6	جایگزینی nan و تبدیل مقادیر
7	ساخت قالب ILP و ارسال دادهها به QuestDB
7	QuestDB
8	Aggregation Service
8	Aggregate API
9	Summarize API
9	Summarize Multiple API
10	Visualization Service (Grafana)

مقدمه

هدف بروژه

پــروژه تحلیــل دادههــای مــالی در بازارهــای مــالی ماننــد رمــزارز و نمادهــای ســهام مختلــف بــه منظــور تجزیــه و تحلیــل دادههــا در زمــان واقعــی (Real-time) طراحــی و پیادهســازی شـــده اســـت. ایـــن سیســـتم بــا اســـتفاده از تکنولوژیهــای توزیــع شـــده و پــردازش بلادرنــگ دادههــا، امکــان تحلیل سریع و تصمیمگیری به موقع برای معامله گران بازارهای مالی را فراهم میکند.

بازارهـای مـالی بـه دلیـل نوسـانات شـدید و لعظـهای قیمتهـا، نیـاز بـه ابزارهـایی دارنـد کـه بتواننـد بـه سـرعت و بـه صـورت دقیـق، اطلاعـات را پـردازش کـرده و سـیگنالهای مناسـب بـرای خریـد و فـروش ارائـه دهنـد. در ایـن پـروژه، هـدف اصـلی ایـن اسـت کـه بـا اسـتفاده از پردازشهـای پیچیده و بلادرنگ، سیگنالهای دقیق و به موقع برای معاملهگران فراهم شود تا بتوانند تصمیمات بهتری اتخاذ نمایند.

اهمیت پروژه

تحلیـل دادههـای مـالی از اهمیـت بسـیار بـالایی برخـوردار اسـت. بـا افـزایش حجـم دادههـا و نیـاز بـه پــردازش بلادرنـگ، اسـتفاده از سیسـتمهای مقیاسپــذیر و توزیــع شــده ماننــد Apache Spark ،Apache Kafka و QuestDB بـه ویــژه در تحلیلهــای مــالی اهمیـت پیــدا میکنــد. ایــن پــروژه نه تنها بر مقیاسپذیری تمرکز دارد، بلکه توانایی پردازش دادهها در لحظه را نیز فراهم میآورد.

به طور سـاده، ایـن پـروژه یـک سیسـتم بلادرنـگ اسـت کـه بـه صـورت خودکـار، دادههـا را از بازارهـای مـالی دریافـت کـرده، پـردازش کـرده و ســپس سیگنالهای مناسب خرید و فروش و همچنین با استفاده از هوش مصنوعی پیشبینی روند را نیز به کاربران نمایش میدهد.

این توانایی میتوانـد بـرای کسـانی کـه در بازارهـای مـالی فعالیـت میکننــد، بـه ویـژه بـرای کسـانی کـه بـه دنبـال تصـمیمگیریهای سـریع و بهینــه هستند، بسیار ارزشمند باشد.

جالشها

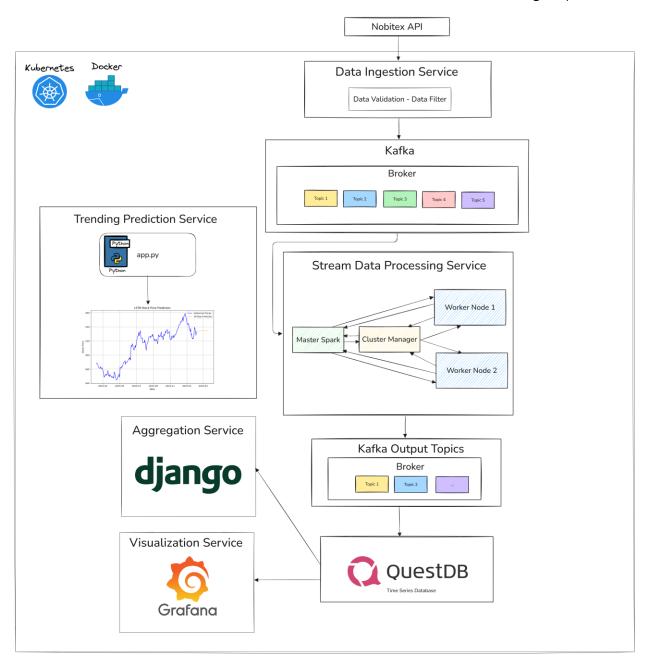
در این پروژه، برخی از چالشها به شرح زیر هستند:

- حجم بالای دادهها: با توجه بـه حجـم عظـیم دادههـای موجـود در بازارهـای مـالی، پـردازش سـریع دادههـا بـه صـورت بلادرنـگ نیازمنــد استفاده از سیستمهای توزیع شده است.
- تاخیر در پردازش (Latency): سیستم باید قادر باشدتا دادهها را با کمترین تاخیر پردازش کرده و سیگنالهای دقیق و به موقع برای کاربران ارسال نماید.
- مقیاس پذیری: با توجه به اینکه دادهها به صورت لعظهای وارد سیستم می شوند، مقیاس پذیری سیستم برای مدیریت تعداد زیاد درخواستها و دادهها ضروری است.
- پیشبینی رونـد بـا اسـتفاده از هـوش مصـنوعی: بـه دلیـل پیچیـدگی ذاتـی دادههـای مـالی و همچنـین پیشـبینی آنهـا، اسـتفاده از الگوریتم مناسب برای تحلیل روند در آینده بسیار چالش برانگیز و محل بحث است.

هــدف از ایــن پــروژه ایــن اســت کــه بــا اســتفاده از معماریهــای مقیاسپــذیر، پردازشهــای توزیــع شــده و اســتفاده از هــوش مصــنوعی (یــادگیری عمیق) این چالشها را پشت سر گذاشته و سیستمی کارآمد و سریع برای تحلیل دادههای بازارهای مالی ارائه دهد.

معماري سيستم

معماری سیستم به شرح زیر است:



این معماری یک سیستم توزیـع شـده بـرای پـردازش دادههـا بـه صـورت بلادرنـگ اسـت کـه شـامل اجـزای مختلفـی میباشــد. در ادامـه توضـیحات معماری ارائه شده به تفصیل شرح داده خواهد شد.

API Nobitex

ایـن API بـه عنـوان منبـع دادههـای رمـز ارز عمـل میکنــد. دادههـا شـامل اطلاعـات معـاملات، قیمتهـا، حجـم بـازار و سـایر معیارهـای مـالی مرتبط با رمزارزها میباشد. ساختار دادههای دریافتی به شرح زیر است:

توضيحات	نوع متغیر	نام متغیر
نماد سهام	object	stock_symbol
تاريخ و ساعت دقيق دريافت داده	timestamps	local_time
قیمت باز شدن (بر اساس کندل)	float	open_prices
بالاترين قيمت	float	high_prices
پایین ترین قیمت	float	low_prices
قيمت بسته شدن	float	close_prices
حجم معامله بازار	int	volumes

Data Ingestion Service

وظیفے ایےن سےرویس جمع آوری و انتقال دادہ از API Nobitex اسےت. در ایےن فرآینید، دادہ ها مےورد اعتبارسینجی (Validation) و فیلترگذاری قـرار میگیرنید تیا فقیط دادہ هـای معتبـر بـه مرحلـه بعـدی ارسـال شـوند. در صـورت صـحت دادہ هـا، توسـط کـد producer هـر دادہ جدیـد بـر اساس فیلد stock_symbol به یک Topic در Kafka ارسال میشود.

برای مثال دادههای مربوط به BTCIRT به تاییک btcirt_topic ارسال خواهد شد.

```
2025-01-25 18:35:44 INFO:root:Data sent to topic=shibirt_topic, partition=0, offset=7, data={\stock_symbol': \sHiBIRT', \local_time': \l2025-01-25 18:34:00', \open': 1680.2, \high': 1684.6, \low': 1680.1, \lclose': 1680.1, \volume': 811.0, \tag{\stock_symbol}: \shibirt_topic'} 2025-01-25 18:35:44 INFO:root:payload: (\stock_symbol': \shibirt_topic': \shibirt_topic \shibirt_topi
```

Apache Kafka

دادههـای تصـحیح شــده توســط ســرویس Data Ingestion بــه عنــوان یــک سیســتم مقیاسپــذیر و پایــدار وارد میشــوند. Kafka امکان انتقال دادهها به اجزای مختلف را به شیوهای بسیار کارآمد فراهم میکند.

Broker: مســئول توزیــع دادههــا در قالــب Topicهــای مختلــف اســت. در معمــاری سیســتم چنــدین Topic نمــایش داده شــده اســت کــه هــر کدام ممکن است یک جزئی از داده را نمایش دهند.

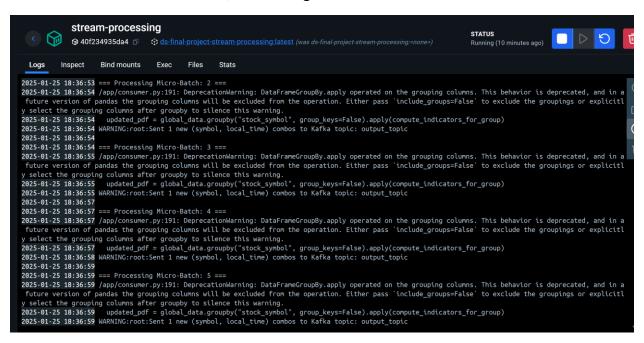
Stream Data Processing Service (Apache Spark)

دادههای ارسال شده از Kafka در اینجا پـردازش میشـوند. Apache Spark بـه عنـوان یـک محـیط پـردازش مـوازی بـرای انجـام عملیـات پیچیـده روی دادهها عمل میکند.

Master Spark و Cluster Spark: مــديريت عمليــات پــردازش توزيـــع شــده را برعهــده دارنــد کــه دادههــا ميــان Worker Node1 و Worker Spark و Node2 تقسيم مىشوند.

Kafka Output Topics: دادههـای پــردازش شــده مجــددا بــه Kafka ارســال میشــوند، جــایی کــه موضــوعات خروجــی جدیــد بــرای ایــن دادههــا ایجاد میشود.

ایـن سـرویس شــامل چنــدین بخــش اســت کــه بــه طــور کلــی بــرای پــردازش و دادههــای اســتریم از Kafka، محاســبه شــاخصهای معــاملاتی یــا همان اندیکاتورها (SMA, EMA, RSI) و ارسال داده به Kafka و مدیریت میکروبچها با Spark طراحی شده است.



محاسبه شاخصها

این بخش برای هر گروه از دادهها شاخصهای زیر را محاسبه میکند:

- SMA (5) میانگین ساده در دورههای ۵ دقیقهای
- EMA (10): میانگین نمایی دورههای ۱۰ دقیقهای
- (10) RSI: شاخص قدرت نسبی در دورههای ۱۰ دقیقهای
- Average gain & Average Loss: در دورههای ۱۰ دقیقهای

تولید سیگنالها

در این بخش سیگنالهای خرید، فروش یا نگهداشتن بر اساس شرایط زیر تولید میشود:

- سیگنال خرید: اگر (10) EMA (5) > EMA باشد و (10) RSI نیز کمتر از ۷۰ باشد.
- سیگنال فروش: اگر (10) SMA (5) < EMA باشد و (10) RSI بیشتر از ۳۰ باشد.
 - در غیر این صورت سیگنال «نگه داشتن» یا همان hold است.

پردازش میکروبچها

هر میکروبچ دادههای زیر را پردازش میکند:

- تبدیل دادههای Spark به Pandas
- محاسبه شاخصها و سیگنالها
- فیلتر کردن دادههایی که قبلا ارسال شدهاند.
 - ارسال دادههای جدید به Kafka

خواندن و نوشتن با Spark

- دادهها از Kafka با فرمت Ison خوانده میشوند و با استفاده از foreachBatch به صورت دستهای پردازش میشوند.
 - یک اسکیمای مشخص برای ستونهای داده تعریف شده است.

مديريت استريم

دادهها به صورت استریم از Kafka خوانده شده، شاخصها محاسبه شده و دادههای نهایی مجددا به output_topic ارسال میشوند.

سرویس میانی Kafka-to-QuestDB

ایــن ســرویس بــرای دریافــت دادههــا از Kafka، پــردازش آنهــا و ارســال آنهــا بـه QuestDB از طریــق Influx Line Protocol یــا ILP طراحــی شـــده است. در ادامه بخشهای مختلف آن توضیح داده میشود.



- Kafka از کریافت بیامها از Kafka Consumer
- Socket: برای ارتباط با QuestDB از طریق پروتکل LP
- کتابخانههای Pandas و math: برای پردازش دادهها

پردازش پیامها

- - اگر پیام شامل دادههای نامعتبر باشد، خطا گزارش میشود.

جایگزینی nan و تبدیل مقادیر

- مقدار "NaN" با مقدار ('nan') جایگزین میشود.
- مقادیر عددی مانند volume و سایر شاخصها به نوع مناسب تبدیل میشوند.

ساخت قالب LP و ارسال دادهها به LP

- دادهها در قالب Influx Line Protocol برای QuestDB قالببندی میشوند.
- هر رکورد شامل اطلاعاتی مثل قیمتها (...,open, close) که پیشتر در ابتدا آمده بود و شاخصهای محاسبه شده است.

QuestDB

دادههـای پــردازش شــده در QuestDB ذخیــره میشــوند. QuestDB یـک پایگــاه داده ســری زمــانی اســت کــه بــرای انجــام تحلیلهـــای پیچیــده و بلادرنــگ اســتفاده میشــوند. ایــن پایگــاه داده بــرای دادههــای بازارهــای مــالی نیــز بســـیار مناســب اســـت و stock_symbol را بــه صــورت خــاص index گذاری میکند که به صورت سریع بتواند دادههای مربوط به هر نماد سهام را در کوئریها برگرداند.

استفاده از این دیتابیس بسیار بهینه تر از ترکیب یک دیتابیس دیگر مثل PostgreSQL با Redis است زیرا:

- عملکرد بالا: QuestDB برای پـردازش دادههـای سـری زمـانی بهینـه شـده و میتوانـد حجـم عظیمـی از دادههـا را بـا سـرعت بسـیار بـالا وارد کوئری کند.
 - کاهش پیچیدگی معماری: با استفاده از QuestDB دیگر نیازی به ترکیب چندین ابزار نیست.
- مقیاس پـذیری: QuestDB در مـدیریت حجـم عظیمـی از دادههـای سـری زمـانی مقیاس پـذیری بـالایی دارد، در حـالی کـه Redis بـرای
 دادههای بسیار بزرگ نیاز به حافظه بیشتر و تنظیمات پیچیده تری دارد.
 - پشتیبانی از ILP: امکان ارسال دادهها با فرمت ILP باعث سادهتر شدن ارسال دادههای بازار سهام میشود.
- جســتجو و انجــام کوئریهــای پیشــرفته SQL: قابلیــت اســتفاده از SQL اســتاندارد بــرای کوئریهــای پیچــده ماننــد محاســبه میــانگین، محاسبه شاخصهای معاملاتی (... , EMA, SMA) و جستجوی مقادیر خاص در بازههای زمانی
- پشـتیبانی از دادههای بلادرنـگ (Realtime): مناسـب بـرای سیسـتمهای بلادرنـگ کـه دادههای بـازار سـهام را در لحظـه پـردازش و تحلیل میکنند.
 - مقیاس پذیری بالا: مناسب برای ذخیره و پردازش دادههای حجیم مانند دادههای تاریخی بازار سهام
 - پشتیبانی از چارتهای مختلف برای تحلیل بصری داده

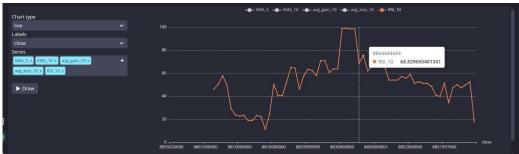
در زیر تصویری از محیط دیتابیس نمایش داده شده است:



در زیر تصویری از چارتهای QuestDB نمایش داده شده است:





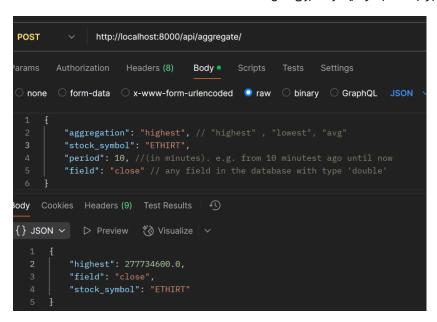


Aggregation Service

در این سرویس از یک اپلیکیشن جنگو استفاده شده است که شامل ۳ نوع API است.

Aggregate API

در ایــن API بـا نوشــتن اســم ســهام مــورد نظــر، نــوع aggregation شــامل میــانگینگیری، بیشــترین یــا کمتــرین مقــدار، بــازه زمــانی مــورد نظــر و فیلد انتخابی مورد نظر جواب مناسب خواهید گرفت. برای مثال:



Summarize API

خلاصه عملکرد یک سهام در بازه زمانی مشخص را برای معیارهای SMA (5), EMA (10), RSI (10), loss, gain, close تولید میکند.

برای مثال ورودی:

و خروجی آن:

```
'summary": {
    "avg": 277430236.3636364,
    "highest": 277800000.0,
    "lowest": 277170000.0
"SMA_5": {
    "avg": 277150965.8181818,
    "highest": 277502506.0,
    "lowest": 276346944.0
"EMA_10": {
    "avg": 277018267.9226054,
    "highest": 277289505.1079296,
    "lowest": 276704324.598024
"RSI_10": {
    "avg": 65.07231402887675,
    "highest": 72.10420259192908,
    "lowest": 50.96143555559214
"gain_loss": {
    "highest_gain": 1734569.0,
    "highest_loss": 569950.0
```

Summarize Multiple API

لیستی از چند نماد سهام را به عنوان ورودی گرفته و خلاصه عملکرد را برای هر کدام در بازه زمانی مشخص تولید میکند.

Visualization Service (Grafana)

در ســرویس گرافانــا کــه بــرای نمــایش نمودارهــا و چارتهــا اســتفاده شــده اســت ایــن قابلیــت وجــود دارد کــه بــه صــورت socket-based بــرای دریافت سیگنال و همچنین query-based برای انجام عملیات aggregation استفاده شود.

در این سرویس در گرافانا یـک داشـبورد بـرای هـر سـهام سـاخته شــده اسـت کـه در یـک نگـاه خلاصـهای از عملکـرد سـهام و نمودارهـای آن نشــان داده شده است که به شرح زیر است:

