



دانشگاه تهران
پردیس دانشکده‌های فنی
دانشکده مهندسی برق و
کامپیوتر

راه‌اندازی و آماده‌سازی سامانه‌ی پویش و پردازش داده‌های بورس تهران

پایان‌نامه برای دریافت درجه کارشناسی
در رشته مهندسی کامپیوتر گرایش فناوری اطلاعات

نام
علی قانع
شماره دانشجویی
810194379

استاد راهنما:
دکتر هشام فیلی

شهریورماه ۱۴۰۰

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ


تعهدنامه اصالت اثر

باسمه تعالی

اینجانب علی قانع تأیید می‌کنم که مطالب مندرج در این پایان‌نامه حاصل کار پژوهشی اینجانب است و به دستاوردهای پژوهشی دیگران که در این نوشته از آنها استفاده شده است مطابق مقررات ارجاع گردیده است. این پایان‌نامه قبلاً برای احراز هیچ مدرک هم سطح یا بالاتر ارائه نشده است. کلیه حقوق مادی و معنوی این اثر متعلق به دانشکده فنی دانشگاه تهران می‌باشد.

نام و نام خانوادگی دانشجو : علی قانع

امضای دانشجو :



چکیده

در سال‌های اخیر، با تغییرات اقتصادی به‌وجود آمده و ورود شرکت‌های بزرگ، بازار بورس رونق دوچندانی یافته و از استقبال عمومی بسیاری برخوردار گشته است. هر چند به دلیل عدم دانش فنی مناسب کاربران و پیچیدگی تحلیل‌های بورسی، امکان مبادله و سوددهی برای همه‌ی کاربران مهیا نیست اما این مورد از استقبال عمومی و سرمایه‌گذاری در بورس نکاسته است. از این رو نیاز به سامانه‌های گزارش و تحلیل بورس بیش از پیش احساس می‌شود که می‌توان گفت بستر کنونی وب‌سایت مدیریت فناوری بورس تهران¹، از ارائه‌ی آن عاجز است. اطلاعات مؤثر این بازار عمدتاً به صورت دسته‌بندی‌نشده و غیرقابل تحلیل و تنها به صورت داده‌ی خام از طریق وب‌سایت قابل دسترسی است. پیش‌بینی قیمت و ارزش سهام یکی از مسائل مهمی است که در صورت امکان استفاده از آن می‌توان تصمیم بر خرید، فروش و یا حفظ یک سهم کرد. هدف این پروژه آماده‌سازی و فراهم کردن بستر مناسبی برای استفاده از روش‌های مختلف تحلیل اطلاعات با استفاده از اطلاعات با ارزش گردآوری‌شده از بازار بورس تهران در ساختاری مناسب است. در اختیار داشتن این داده‌ها زمینه‌ی استفاده از انواع روش‌ها برای پیش‌بینی آینده‌ی بازار را میسر می‌سازد.

کلمات کلیدی: جمع‌آوری داده‌های بورس تهران، مدل داده بورس، پیاده‌سازی خزنده

¹ Tsetmc.com

فهرست مطالب

Contents

فصل 1: مقدمه و بیان مساله	1
1-1- مقدمه	2
1-2- تاریخچه‌ای از موضوع تحقیق	2
1-3- شرح مسئله تحقیق	2
1-4- تعریف موضوع تحقیق	3
1-5- اهداف و آرمان‌های کلی تحقیق	3
1-6- روش انجام تحقیق	4
1-7- ساختار پایان‌نامه	4
فصل 2: مفاهیم اولیه و پیش زمینه‌ی خزنده و مدل داده	5
2-1- بررسی ساختار سایت شرکت مدیریت فناوری بورس تهران	6
2-2- ساختار کلی خزنده	7
2-3- روش توسعه‌ی خزنده‌ی جدید	7
2-4- خلاصه و جمع‌بندی	8
فصل 3: ساختار خزنده، معماری اجزا و روش کارکرد	10
3-1- مقدمه	11
3-2- جداول اصلی پایگاه داده‌ی پروژه	12
3-2-2- جدول Instrument	12
3-2-3- جدول Shareholder	13
3-2-4- جدول Share	13

14.....	DayPriceThreshold	جدول 3-2-5-
15.....	DayTradeSummary	جدول 3-2-6-
16.....	ClientTradeInfo	جدول 3-2-7-
16.....	TableOfInterest10	جدول 3-2-8-
17.....	Bourse_ChangeCapital	جدول 3-2-9-
18.....		3-3- دریافت اطلاعات
19.....		3-3-1- روش دریافت اطلاعات
19.....		3-3-2- ثبت اطلاعات اولیه
20.....		3-3-3- ثبت سایر اطلاعات نمادها
21.....	Get_ShareHolders	3-3-3-1- تابع
21.....	Get_day_price_threshold	3-3-3-2- تابع
21.....	Get_day_trade_summaries	3-3-3-3- تابع
21.....	Get_client_trade_info	3-3-3-4- تابع
22.....	Create_table_of_interest	3-3-3-5- تابع
22.....		3-3-4- افزایش سرمایه‌ها
22.....		3-3-4-1- فرمت اطلاعات
24.....		3-3-5- توابع بازخوانی دوره‌ای اطلاعات
24.....		3-3-5-1- دریافت اطلاعات نمادها به صورت روزانه
24.....		3-3-5-2- دریافت اطلاعات افزایش سرمایه و تعدیل قیمت‌ها
25.....		3-4- ذخیره‌سازی اطلاعات
25.....		3-5- خلاصه و جمع‌بندی
26.....		فصل 4: راه‌اندازی و اجرا خزنده و رابط کاربری
27.....		4-1- آماده‌سازی سیستم

- 27..... 2-4- اجرای خزنده‌ی کدها
- 27..... 3-4- اجرای توابع اصلی
- 28..... 4-4- رابط گرافیکی
- 29..... 2-4-4- هدر
- 29..... 3-4-4- منوی کناری
- 30..... 4-4-4- محتوای اصلی
- 30..... 5-4- خلاصه و جمع‌بندی
- 31..... فصل 5: جمع‌بندی، نتیجه‌گیری و پیشنهادها
- 32..... 1-5- جمع‌بندی
- 32..... 2-5- دستاوردها
- 32..... 3-5- پیشنهادها
- 33..... فصل 6: مراجع

فهرست شکل‌ها

- شکل (2-1) بخشی از صفحه‌ی اصلی نماد شستا در tsetmc.com 6
- شکل (2-2) بخشی از صفحه‌ی تاریخچه‌ی نماد شستا در cdn.tsetmc.com 6
- شکل (3-2) بخشی از صفحه‌ی اطلاعات نمادها 7
- شکل (4-2) نمونه‌ای از پاسخ API ها 8
- شکل (3-1) مدل داده‌ای قبلی پایگاه داده 11
- شکل (3-2) مدل داده‌ای پایگاه داده‌ی جدید 12
- شکل (3-3) شمای جدول نماد 13
- شکل (4-3) نمونه‌ی داده‌های جدول نماد 13
- شکل (3-5) شمای جدول سهام‌داران 13
- شکل (3-6) شمای جدول سهام 14
- شکل (3-7) نمونه‌ی داده‌های جدول سهام 14
- شکل (3-8) شمای جدول قیمت اطلاعات روزانه‌ی نماد 14
- شکل (3-9) نمونه‌ی داده‌های جدول قیمت اطلاعات روزانه‌ی نماد 15
- شکل (3-10) شمای جدول مبادلات روز نماد 15
- شکل (3-11) نمونه‌ی داده‌های جدول مبادلات روز نماد 16
- شکل (3-12) نمونه‌ی داده‌های جدول مبادلات روز نماد 16
- شکل (3-13) شمای جدول اطلاعات میانگین ده‌روزه‌ی معاملات 17
- شکل (3-14) شمای جدول اطلاعات افزایش سرمایه 18
- شکل (3-15) نمونه‌ی داده‌های جدول اطلاعات افزایش سرمایه 18
- شکل (3-16) خروجی API اطلاعات اولیه‌ی نماد به صورت JSON 20
- شکل (3-17) صفحه‌ی افزایش سرمایه‌ها 23
- شکل (3-18) اطلاعات صفحه‌ی افزایش سرمایه‌ها 23
- شکل (4-1) پیاده‌سازی رابط گرافیکی 29

فصل 1:

مقدمه و بیان مساله

در این فصل نخست به تعریف و شرح مسأله و روش کلی تحقیق پرداخته، سپس مساله و موضوع مورد بررسی در این پایان نامه و اهداف و آرمان های کلی تحقیق بیان می شود و در انتها ساختار پایان نامه ی پیش رو ذکر شده است.

1-1- مقدمه

امروزه روش‌های تحلیل داده‌های بورسی بسیار گسترده‌اند و بهره‌گیری از این روش‌ها می‌تواند تضمینی برای سرمایه‌گذاری مطمئن و مبتنی بر اطلاعات باشد. در این میان اهمیت پیوستگی و دقت داده‌ها غیرقابل چشم‌پوشی و نیازمند جمع‌آوری و دسته‌بندی داده‌ها به سرعت و همراه با پیوستگی است. وجود یک بانک اطلاعاتی جامع و قدرتمند می‌تواند ابزاری قابل اتکا برای سرمایه‌گذاران این حوزه باشد و به کاربران در سرمایه‌گذاری مناسب کمک شایانی کند.

1-2- تاریخچه‌ای از موضوع تحقیق

این پروژه در امتداد پروژه‌ی «ایجاد مدل داده برای نماد‌های بورس تهران به منظور پیش‌بینی قیمت سهام» که توسط سرکار خانم مریم کریمی زیر نظر دکتر ه‌شام فیلی در مهرماه ۱۳۹۹ در دانشکده‌ی برق و کامپیوتر دانشگاه تهران ارائه شده، نوشته شده است. از این رو لازم به ذکر است قسمتی از مدل‌سازی‌ها و ساختار داده‌ای بر اساس این تحقیق طراحی و اجرا شده که به تفصیل در فصول آتی به شرح آن پرداخته خواهد شد.

1-3- شرح مسئله تحقیق

پیش‌بینی بازار بورس به معنی سعی در تشخیص ارزش سهام و موارد مورد مبادله در گذر زمان است. این تلاش، در جهت به‌دست‌آوردن سود از طریق خرید و فروش در این بازار صورت می‌گیرد. این موضوع از ابتدای پیدایش بازار بورس مورد توجه افراد و سرمایه‌گذاران مختلف بوده و هم‌اکنون نیز، توجه ویژه‌ای به آن می‌شود. برای پیش‌بینی بازار بورس روش‌های مختلفی به کار گرفته شده است. با رونق و پیش‌رفت دنیای اطلاعات در عصر نوین نیز این روش‌ها، که بیش‌تر به تحلیل‌های بورسی شناخته می‌شوند، از تأثیر فناوری مصون نمانده‌اند. از این رو نیاز به بستر مناسبی برای استفاده از روش‌های مختلف تحلیل اطلاعات با استفاده از اطلاعات بازارش گردآوری‌شده از بازار بیش از پیش احساس می‌شود.

4-1- تعریف موضوع تحقیق

در این پروژه، هدف فراهم‌سازی بستر ابتدایی برای استفاده از روش‌های تحلیل آماری و کامپیوتری با جمع‌آوری داده‌های پراهمیت بازار بورس تهران در ساختاری مناسب است. برای این منظور نیازمند پیاده‌سازی یک خزنده¹ جهت جمع‌آوری داده‌های مورد نیاز است که طرح اولیه‌ی این خزنده توسط خانم مریم کریمی طراحی و پیاده‌سازی شده است. یکی از مسائل مهم در این خزنده، دسترسی سریع و دقیق اطلاعات در مقابل چالش‌های مختلفی است که برای آن وجود دارد. با استفاده از این خزنده می‌توان داده‌های مورد نیاز را جمع‌آوری کرده و با استفاده از آن‌ها امکان بهره‌گیری از روش‌های مختلف تحلیل برای پیش‌بینی بازار را فراهم آورد. مدل داده‌ای این خزنده و نحوه‌ی ذخیره‌سازی اطلاعات مؤثر و پراستفاده، دقت داده‌ای و کاربردی بودن آن می‌تواند کمک شایانی در کاربردپذیری این خزنده داشته باشد. در این رابطه بررسی و استقرار خزنده‌های قبلی و راه‌اندازی سامانه‌ی پویش و ذخیره‌سازی اطلاعات دریافتی با چالش‌های غیرقابل پیش‌بینی مختلفی مواجه بوده که یکی از اهداف اساسی این پروژه، رفع این چالش‌ها است. از سمت دیگر، ایجاد نمایشی مناسب و کاربردپذیر نیازمند طراحی و توسعه‌ی رابط کاربری گرافیکی مناسب با امکانات مختلف است.

4-1- اهداف و آرمان‌های کلی تحقیق

اهداف کلی این پروژه شامل رفع ایرادات و ایجاد تغییرات ساختاری مورد نیاز روی خزنده‌های قبلی و به‌روزرسانی روش‌های واکنشی اطلاعات است. همچنین به تبع این تغییرات ساختاری، بازسازی مدل و ساختار پایگاه داده و ذخیره‌سازی داده‌ها نیز ضروری به‌نظر می‌رسد. با در دست داشتن اطلاعات پایه و تاریخچه‌ای و همچنین روند مبادلات سهام، می‌توان از ابزارهای قدرت‌مند تحلیل داده بهره گرفت و الگوهای متفاوتی را تشخیص داد.

¹ Crawl

6-1- روش انجام تحقیق

روش انجام این پروژه در گام‌های زیر خلاصه می‌شود:

- مطالعه‌ی اولیه روی مفاهیم اولیه‌ی بازار بورس
- کسب اطلاعات درباره‌ی نحوه‌ی پیاده‌سازی خزنده‌ی قبلی
- یافتن ایرادات خزنده‌ی قبلی و ایجاد خزنده‌ی جدید
- بازطراحی مدل داده‌ای و پایگاه داده
- راه‌اندازی سرور و استخراج منظم داده‌ها
- طراحی و پیاده‌سازی رابط گرافیکی و استقرار آن

7-1- ساختار پایان‌نامه

فصل دوم، شامل بررسی تعاریف اساسی مربوط به حوزه‌ی بورس، مفاهیم اولیه و اجزای اساسی خزنده و مدل داده، مروری بر پیشینه‌ی تحقیق و پیش‌زمینه‌های مورد نیاز برای درک هرچه بهتر ساختار آن است.

فصل سوم در برگیرنده‌ی توضیحات مربوط به ساختار خزنده، معماری و اجزای آن و روش کارکرد آن است.

فصل چهارم به دستورالعمل اجرای مجدد پروژه روی سرور آماده‌سازی شده و یا روی سرور جدید اختصاص یافته است. هم‌چنین در این فصل به توضیحاتی در مورد رابط کاربری نیز خواهیم پرداخت.

در نهایت، در فصل پنجم، نتیجه‌گیری‌های کلی حاصل‌شده در این پروژه مورد بحث قرار می‌گیرد و پیشنهادهایی برای ادامه‌ی مسیر به علاقمندان این حوزه ارائه خواهد شد.

فصل 2: مفاهیم اولیه و پیش زمینه‌ی خزنده و مدل داده

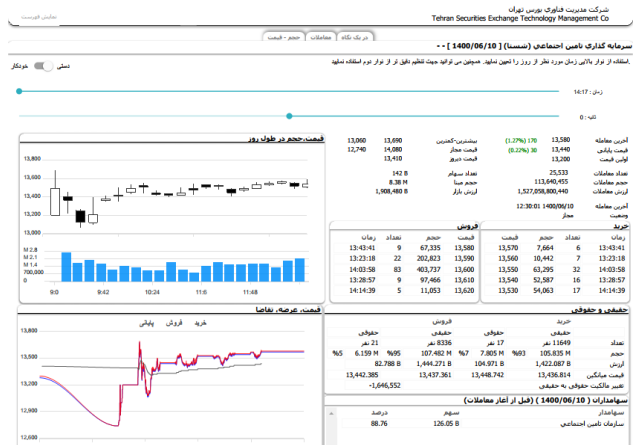
فصل پیش رو شامل تعاریف اساسی مربوط به حوزه‌ی بورس، مفاهیم اولیه و اجزای اساسی خزنده و مروری بر خزنده‌ی قبلی است.

1-2- بررسی ساختار سایت شرکت مدیریت فناوری بورس تهران

تمامی اطلاعات مربوط به بازار بورس تهران در وب‌سایت به آدرس tsetmc.com توسط شرکت مدیریت فناوری بورس تهران در اختیار عموم کاربران قرار می‌گیرد. این وب سایت اطلاعات مربوط به تمامی نمادها، معاملات انجام‌شده، عرضه و تقاضا، سهام‌داران و خریداران حقیقی و حقوقی را شامل می‌شود. (شکل ۱-۲) هم‌چنین تاریخچه‌ی تمامی رویدادهای مربوط به هر روز از هر سهام در صفحه‌ای اختصاصی در cdn.tsetmc.com قابل دسترسی است. (شکل ۲-۲)



شکل (1-2) بخشی از صفحه‌ی اصلی نماد شستا در tsetmc.com



شکل (2-2) بخشی از صفحه‌ی تاریخچه‌ی نماد شستا در cdn.tsetmc.com

برای تمامی نمادها یک کد منحصر به فرد وجود دارد که در تمامی لینک‌های دسترسی به آن نماد از آن کد استفاده می‌شود. لیستی از تمامی کدهای نمادها در سایت مدیریت فناوری بورس تهران موجود است که از طریق خواندن این لیست¹ می‌توان تمامی کدها را دریافت کرد. (شکل ۲-۳)

لیست همه نمادهای بازار عادی									
کد نماد	گروه	گروه های صنعت	نماد	نام لاتین	نماد	نام	نماد	نام	نام
IRB5IKC08751	N2	خودرو و ساخت قطعات	IKCQ1	Iran Khodro-D	18719101	اوراق مشارکت ایران خودرو	ا	ا	ا
IRO1NBA0001	N2	عرضه برقی، گاز، بخار و آب گرم	NBAB1	Abadan PG		تولید نیروی برقی آبادان	ا	ا	ا
IRO1APPE0001	N2	رایانه و فعالیت‌های وابسته به آن	APPE1	Asan Pardakht Pers		آسان پرداخت پرسین	ا	ا	ا
IRO1ASIA0001	N1	بیمه و صندوق بازنشستگی به جز بیمه اجتماعی	ASIA1	Asia Bime		بیمه آسیا	ا	ا	ا
IRO1AST0001	N2	اطلاعات و ارتباطات	ASTC1	Asiatech		انتقال داده های آسیاتک	ا	ا	ا
IRO1CONT0001	N1	امرازیزشنگی، آئینگی و اندازوگری	CONT1	Iran Counter		کنترساری ایران	ا	ا	ا
IRR1CONT0101	N1	امرازیزشنگی، آئینگی و اندازوگری	CONX1	Iran Counter-R		کنترساری ایران	ا	ا	ا
IRO1OPAL0001	N2	استخراج کانه های فلزی	OPAL1	Opal Kani Pars		فرآوری معدنی ایل کانی پارس	ا	ا	ا
IRO1ETKA0001	N2	بیمه و صندوق بازنشستگی به جز بیمه اجتماعی	ETKA1	Amin Company		بیمه انگانی امین	ا	ا	ا
IRR1ETKA0101	N2	بیمه و صندوق بازنشستگی به جز بیمه اجتماعی	ETKC1	Amin Company-R		بیمه انگانی امین	ا	ا	ا
IRO1ZBAL0001	N2	زراعت و خدمات وابسته	ZBAL1	Ajdad Zarbal Co		مرغ اجداد زربال	ا	ا	ا
IRO1MKBT0001	N1	مخابرات	MKBT1	.Iran Tele. Co		مخابرات ایران	ا	ا	ا

شکل (2-3) بخشی از صفحه‌ی اطلاعات نمادها

2-2- ساختار کلی خزنده

در پروژه‌ی «ایجاد مدل داده برای نمادهای بورس تهران به منظور پیش‌بینی قیمت سهام» طراحی و اجرا شده تو سط سرکار خانم مریم کریمی زیر نظر دکتر ه‌شام فیلی، از یک خزنده با فریم‌ورک Scrapy استفاده شده است. «اسکرپی» فریم‌ورکی از پایتون برای خزش در مقیاس بالا روی داده‌های وب است. ساختار این خزنده مبتنی بر خواندن اطلاعات از طریق تگ‌های HTML در صفحات وب است. در این پروژه با توجه به اصلاحات ساختاری صورت‌گرفته در وب‌سایت مدیریت فناوری بورس تهران، ناچار به اصلاحات ساختاری روی خزنده شدیم که در نهایت به طراحی و توسعه‌ی مجدد خزنده از ابتدا انجامید.

2-3- روش توسعه‌ی خزنده‌ی جدید

در میانه‌ی انجام این پروژه، اصلاحات ساختاری روی وب‌سایت مدیریت فناوری بورس تهران صورت گرفت. پیش از این اطلاعات نمادها تو سط اسکرپیتی درون بدنه‌ی HTML این صفحات بارگذاری می‌شد اما طی تغییرات اخیر، این اسکرپیت جای خود را به API‌هایی² داده است که در این پروژه از آن‌ها بهره‌گیری می‌شود. (شکل ۲-۴)

¹ <http://www.tsetmc.com/Loader.aspx?ParTree=111C1417>

² Application Programming Interface

```

{
  instrumentHistory: {insCode: 0, lVal30: "سرمایه گذاری تامین اجتماعی", lVal18AFC: "فستا", flow: 1, cIsin: "IRO1TAMN0006",...}
  baseVol: 8379888
  cComVal: "5"
  cIsin: "IRO1TAMN0006"
  cgrValCot: "91"
  cgrValCotTitle: "-"
  flow: 1
  flowTitle: "بازار بورس"
  insCode: 0
  instrumentID: "IRO1TAMN0001"
  lVal18AFC: "فستا"
  lVal30: "سرمایه گذاری تامین اجتماعی"
  zTitad: 142000000000
}

```

شکل (2-4) نمونه‌ای از پاسخ APIها

همان‌طور که در شکل ۲-۴ مشاهده می‌شود، پاسخ تمامی APIها به صورت یک فایل JSON¹ است. بدین ترتیب با ارسال درخواست به هر کدام از این APIها یک فایل JSON دریافت و واکشی اطلاعات از طریق آن انجام شده و به پایگاه داده منتقل می‌شود. یک فایل JSON شامل جفت‌های ویژگی-مقدار² است.

خزنده‌ی جدید تحت فریم‌ورک جنگو طراحی و توسعه یافته است. جنگو بر اساس معماری MVT بوده که مخفف عبارت مدل، نمایش، الگواست. وقتی درخواستی به یک سرور وب می‌رسد، ابتدا این درخواست به جنگو منتقل می‌شود و جنگو سعی می‌کند بفهمد چه چیزی درخواست می‌شود. سپس جنگو لیستی از الگوها را می‌گیرد و سعی می‌کند URL را با آنها تطبیق دهد. جنگو الگوها را از بالا به پایین بررسی می‌کند و اگر درخواست با الگوی خاصی مطابقت داشته باشد، درخواست را به عملکرد مرتبط منتقل می‌کند. هم‌چنین قابلیت ORM³ جنگو بستر را برای هماهنگ‌سازی با پایگاه داده‌ی MySQL فراهم می‌آورد تا حداقل تغییرات نسبت به پایگاه داده‌ی قبلی را داشته باشیم.

2-4- خلاصه و جمع‌بندی

در این فصل با مفاهیم اولیه و پیش‌زمینه‌های پروژه که برای ایجاد درکی مناسب برای فهم مدل داده‌ای و خزنده مورد نیاز است، آشنا شدیم؛ به بررسی ساختار سایت tsetmc پرداختیم و صفحاتی که اطلاعات مورد نیاز در آن‌ها قرار گرفته بود را شناختیم. و نهایتاً با توجه به

¹ JavaScript Object Notation

² Attribute-Value

³ Object-Relational-Mapping

نیازمندی‌های ذکرشده و تغییرات صورت‌گرفته، پایگاه داده و فریم‌ورک مناسب برای استخراج داده‌ها را دریافتیم.

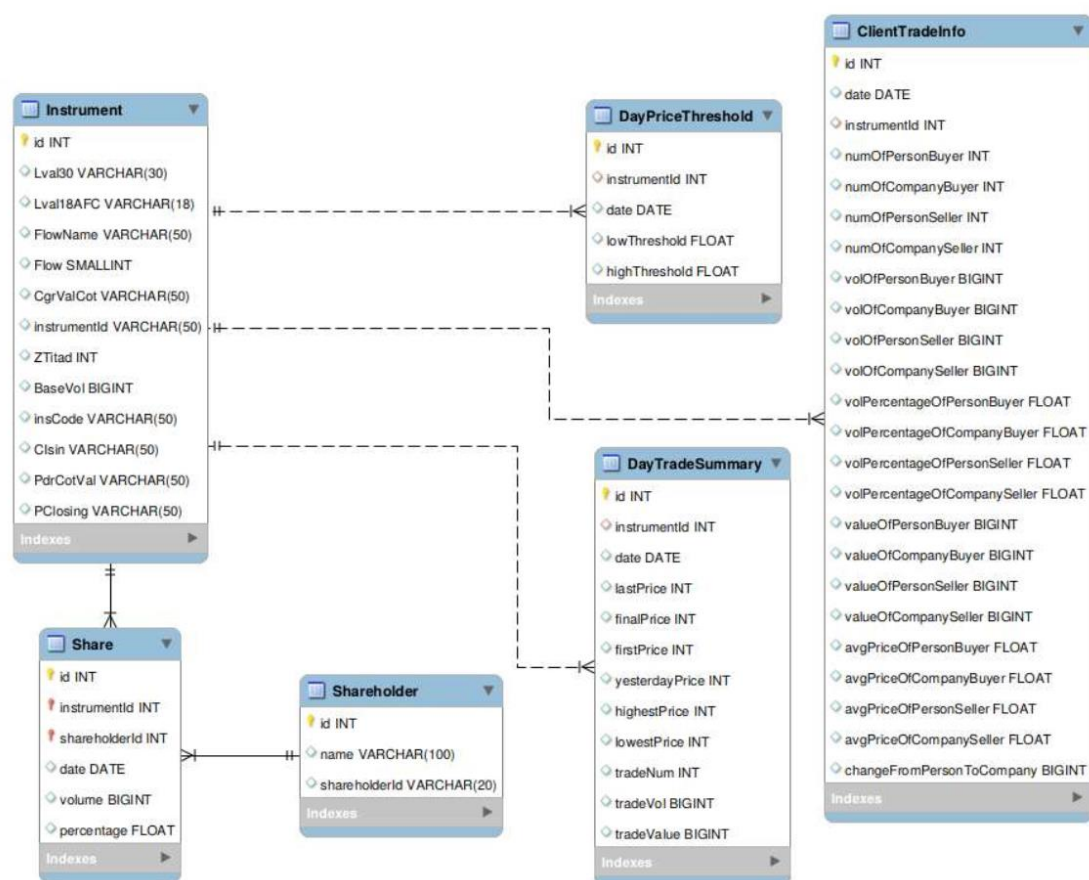
فصل 3: ساختار خزنده ، معماری اجزا و روش کارکرد

فصل سوم در برگیرنده‌ی توضیحات مربوط به مدل داده‌ای پیشنهادی و ساختار خزنده مربوطه است.

1-3- مقدمه

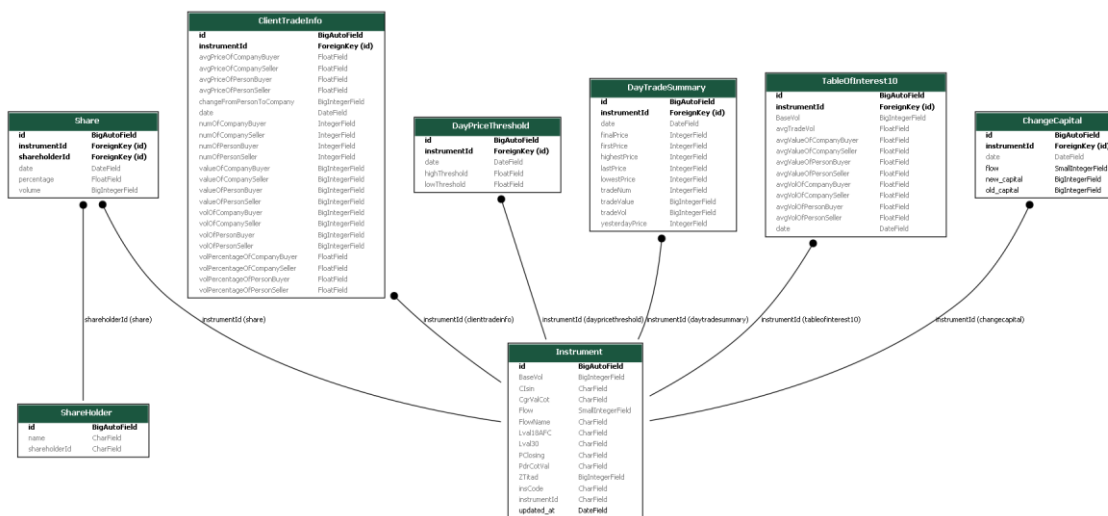
پایگاه داده‌ی پیاده‌سازی شده برای این پروژه از روی پایگاه داده‌ی قبلی (شکل ۱-۳) تهیه شده توسط خانم مریم کریمی) طراحی و بهینه‌سازی شده است؛ لیکن پایگاه داده‌ی جدید به دلیل استفاده از جنگو شامل جداول پیش‌فرض این فریم‌ورک نیز هست. بنابراین می‌توان این پایگاه داده را به دو بخش تقسیم نمود؛ الف) جداول پیش‌فرض جنگو ب) جداول اصلی. در این گزارش از آوردن اطلاعات بخش «الف» چشم‌پوشی می‌کنیم و به توضیح بخش «ب» می‌پردازیم.

اطلاعات تمامی جداول به تفصیل در گزارش پایانی پروژه‌ی «ایجاد مدل داده برای نمادهای بورس تهران به منظور پیش‌بینی قیمت سهام» قابل دسترسی است و در این گزارش تنها به ذکر تغییرات به‌وجودآمده اکتفا می‌شود.



شکل (3-1) مدل داده‌ای قبلی پایگاه داده

3-2- جداول اصلی پایگاه داده‌ی پروژه



شکل (3-2) مدل داده‌ای پایگاه داده‌ی جدید

3-2-2- جدول Instrument

این جدول اطلاعات نمادها را در خود نگهداری می‌کند. در واقع تمامی اطلاعات پایگاه داده از طریق دسترسی به این جدول و خواندن از روی آن تکمیل می‌شود. به دلیل استفاده از پیاده‌سازی قبلی، این جدول تنها جدولی است که اطلاعات آن در دو مرحله واکشی و تکمیل می‌شود. برای ساخت این جدول ابتدا با اجرای کد خزنده‌ی قبلی اطلاعات شماره‌ی نمادها باید در یک فایل با فرمت csv ذخیره‌سازی شود. البته می‌توان به صورت دستی نیز فایل را تکمیل و در اختیار خزنده گذاشت. این جدول در شکل ۳-۳ و ۴-۳ به نمایش گذاشته شده. تغییر به‌وجودآمده در این جدول تنها شامل اضافه‌شدن ستون زیر است:

Updated_at: روزی که برای آن آخرین قیمت نماد ثبت شده. این ستون برای اطمینان از به‌روز بودن آخرین قیمت ضروری است.

Instrument	
id	BigAutoField
BaseVol	BigIntegerField
Cisin	CharField
CgrValCot	CharField
Flow	SmallIntegerField
FlowName	CharField
Lval18AFC	CharField
Lval30	CharField
PClosing	CharField
PdrCotVal	CharField
ZTitad	BigIntegerField
insCode	CharField
instrumentId	CharField
updated_at	DateField

شکل (3-3) شمای جدول نماد

id	instrumentID	Lval30	Lval18AFC	FlowName	Flow	CgrValCot	Ztitad	BaseVol	Cisin	PdrCotVal	Pclosing	Updated_at
2820	204092872752957	ح - سالیانین	خانیج	بازار بورس	2	71	700000000	317280	IRO3SDFZ0007	39800	40339	8/31/2021
2822	408934423224097	فرآوری مواد معدنی ایران	فرآور	بازار بورس	1	N1	660000000	421230	IRO1FRVR0008	30970	31230	8/31/2021
2824	611986653700161	سیمان لار سینزوار	سینوا	بازار فرابورس	2	Z1	3757797579	3838281	IRO3SBZZ0005	3631	3658	8/31/2021
2825	655060129740445	فرآوری معدنی ایال کانی یارس	ایال	بازار بورس	1	91	20000000000	4854369	IRO1OPAL0000	22690	22750	8/31/2021
2828	778253364357513	بانک ملت	ویملت	بازار بورس	1	N1	207042862000	26773762	IRO1BMLT0007	4049	4059	8/31/2021

شکل (3-4) نمونه‌ی داده‌های جدول نماد

3-2-3- جدول Shareholder

این جدول (شکل ۳-۵) شامل اطلاعات سهام‌داران تمامی نمادها و شامل اطلاعات زیر است :

ShareHolder	
id	BigAutoField
name	CharField
shareholderId	CharField

شکل (3-5) شمای جدول سهام‌داران

3-2-4- جدول Share

در این جدول (شکل ۳-۶ و ۳-۷) اطلاعات مربوط به سهامی تمامی نمادها در هر روز نگهداری می‌شود. به ازای هر نماد در هر روز به تعداد سهام‌داران این نماد در جدول سطر وجود دارد. تغییرات این جدول نسبت به طراحی قبلی تنها شامل تغییر نام دو ستون زیر است:

- InstrumentID_id: شناسه ی نماد.

- ShareholderID_id: شناسه ی سهام دار.

لازم به ذکر است پسوند id_ در دو ستون آخر توسط خود فریم ورک جنگو برای کلیدهای خارجی اعمال می شود.

Share	
id	BigAutoField
instrumentId	ForeignKey (id)
shareholderId	ForeignKey (id)
date	DateField
percentage	FloatField
volume	BigIntegerField

شکل (3-6) شمای جدول سهام

id	date	Vol	Percentage	InstrumentID	ShareHolderID
318	8/30/2021	357650000	51.09	2820	137
319	8/30/2021	112000000	16	2820	1272
320	8/30/2021	81993367	11.71	2820	139
321	8/30/2021	7577038	1.08	2820	1273
322	8/30/2021	342252360	51.85	2822	125
323	8/30/2021	101565073	15.38	2822	339
324	8/30/2021	15079056	2.28	2822	401

شکل (3-7) نمونه ی داده های جدول سهام

3-2-5- جدول DayPriceThreshold

این جدول (شکل ۳-۸ و ۳-۹) شامل بازه ی قیمت معاملات هر روز برای هر نماد است.

DayPriceThreshold	
id	BigAutoField
instrumentId	ForeignKey (id)
date	DateField
highThreshold	FloatField
lowThreshold	FloatField

شکل (3-8) شمای جدول قیمت اطلاعات روزانه ی نماد

id	date	low_thrshold	high_threshold	InstrumentID
927	8/30/2021	2402	2934	2816
928	8/30/2021	920	1080	2817
929	8/30/2021	688700	761194	2818
930	8/30/2021	5842	6328	2819
931	8/30/2021	41639	46021	2820
932	8/30/2021	509	621	2821
933	8/30/2021	32580	36000	2822
934	8/30/2021	3885	4747	2823

شکل (3-9) نمونه ی داده های جدول قیمت اطلاعات روزانه ی نماد

3-2-6- جدول DayTradeSummary

این جدول (شکل ۳-۱۰ و ۳-۱۱) شامل اطلاعات مربوط به معاملات یک نماد برای یک روز مشخص است. با توجه به اضافه شدن اطلاعات افزایش سرمایه ی نمادها، ستون های زیر برای نگهداری از قیمت های تعدیل شده ی نمادها به این جدول اضافه شده اند:

- **BalancedPrice**: قیمت پایانی تعدیل شده برای نماد
- **BalancedHigh**: مقدار تعدیل شده ی بالاترین قیمت نماد
- **BalancedLow**: مقدار تعدیل شده ی کم ترین قیمت نماد
- **BalancedFirst**: مقدار تعدیل شده ی قیمت باز شدن نماد

DayTradeSummary	
id	BigAutoField
instrumentId	ForeignKey (id)
date	DateField
finalPrice	IntegerField
firstPrice	IntegerField
highestPrice	IntegerField
lastPrice	IntegerField
lowestPrice	IntegerField
tradeNum	IntegerField
tradeValue	BigIntegerField
tradeVol	BigIntegerField
yesterdayPrice	IntegerField

شکل (3-10) شمای جدول مبادلات روز نماد

id	date	lastPrice	Final Price	FirstPrice	YesterdayPrice	HighestPrice	Lowest Price	TradeNum	TradeVol	TradeValue	InstrumentID
122	5/16/2021	33100	31977	33100	31524	33100	33100	708	137329	4545589900	2820
124	5/16/2021	32340	32340	32340	34040	32340	32340	37	532963	17236023420	2822
126	5/16/2021	3134	3275	3134	3298	3134	3134	67	626232	1962611088	2824
127	5/16/2021	17910	17640	17600	16760	17980	17500	5361	6648680	117516042290	2825
130	9/1/2020	18785	17914	18785	17891	18785	18785	3520	72947	1370309395	2820
132	9/1/2020	27390	27110	25000	26090	27390	25000	2293	9796053	265572618270	2822

شکل (3-11) نمونه‌ی داده‌های جدول مبادلات روز نماد

3-2-7- جدول ClientTradeInfo

این جدول (شکل ۳-۱۲) مشروح رویدادهای خرید و فروش مشتریان حقیقی و حقوقی است.

ClientTradeInfo	
id	BigAutoField
instrumentId	ForeignKey (id)
avgPriceOfCompanyBuyer	FloatField
avgPriceOfCompanySeller	FloatField
avgPriceOfPersonBuyer	FloatField
avgPriceOfPersonSeller	FloatField
changeFromPersonToCompany	BigIntegerField
date	DateTimeField
numOfCompanyBuyer	IntegerField
numOfCompanySeller	IntegerField
numOfPersonBuyer	IntegerField
numOfPersonSeller	IntegerField
valueOfCompanyBuyer	BigIntegerField
valueOfCompanySeller	BigIntegerField
valueOfPersonBuyer	BigIntegerField
valueOfPersonSeller	BigIntegerField
volOfCompanyBuyer	BigIntegerField
volOfCompanySeller	BigIntegerField
volOfPersonBuyer	BigIntegerField
volOfPersonSeller	BigIntegerField
volPercentageOfCompanyBuyer	FloatField
volPercentageOfCompanySeller	FloatField
volPercentageOfPersonBuyer	FloatField
volPercentageOfPersonSeller	FloatField

شکل (3-12) نمونه‌ی داده‌های جدول مبادلات روز نماد

3-2-8- جدول TableOfInterest10

این جدول اطلاعات مربوط به میانگین‌های ده‌روزه‌ی معاملات هر سهام را شامل می‌شود. لازم به ذکر است این اطلاعات در بازه‌ی ده روز اول کراول قابل استناد نیست. (شکل ۳-۱۳)

- Id: ردیف هر سطر از اطلاعات

- Date: تاریخ میلادی که این اطلاعات مربوط به آن است.

- instrumentID_id: شناسه‌ی نماد

- BaseVol: حجم مبنای نماد
- AvgTradeVol: میانگین دهروزه‌ی حجم کل معاملات
- AvgVolOfPersonBuyer: میانگین دهروزه‌ی حجم خرید خریداران حقیقی.
- AvgVolOfCompanyBuyer: میانگین دهروزه‌ی حجم خرید خریداران حقوقی.
- AvgVolOfPersonSeller: میانگین دهروزه‌ی حجم فروش فروشندگان حقیقی.
- AvgVolOfCompanySeller: میانگین دهروزه‌ی حجم فروش فروشندگان حقوقی.
- AvgValueOfPersonBuyer: میانگین دهروزه‌ی ارزش معاملات خریداران حقیقی.
- AvgValueOfCompanyBuyer: میانگین دهروزه‌ی ارزش معاملات خریداران حقوقی.
- AvgValueOfPersonSeller: میانگین دهروزه‌ی ارزش معاملات فروشندگان حقیقی.
- AvgValueOfCompanySeller: میانگین دهروزه‌ی ارزش معاملات فروشندگان حقوقی.

TableOfInterest10	
id	BigAutoField
instrumentId	ForeignKey (id)
BaseVol	BigIntegerField
avgTradeVol	FloatField
avgValueOfCompanyBuyer	FloatField
avgValueOfCompanySeller	FloatField
avgValueOfPersonBuyer	FloatField
avgValueOfPersonSeller	FloatField
avgVolOfCompanyBuyer	FloatField
avgVolOfCompanySeller	FloatField
avgVolOfPersonBuyer	FloatField
avgVolOfPersonSeller	FloatField
date	DateField

شکل (3-13) شمای جدول اطلاعات میانگین دهروزه‌ی معاملات

3-2-9- جدول Bourse_ChangeCapital

این جدول شامل تاریخچه و اطلاعات مربوط به افزایش سرمایه‌ی نمادها است (شکل ۱۴-۳ و ۳-۱۵)

- Id: ردیف هر سطر از اطلاعات
- Date: تاریخ میلادی که افزایش سرمایه در آن صورت گرفته است.
- instrumentID_id: شناسه‌ی نماد

- Flow: کد بازاری است که نماد در آن قرار دارد و می‌تواند از این اعداد باشد: ۱- بورس ۲- فرابورس
- Old_capital: میزان سرمایه‌ی قبلی نماد.
- New_capital: میزان سرمایه‌ی جدید نماد.

ChangeCapital	
id	BigAutoField
instrumentId	ForeignKey (id)
date	DateField
flow	SmallIntegerField
new_capital	BigIntegerField
old_capital	BigIntegerField

شکل (3-14) شمای جدول اطلاعات افزایش سرمایه

id	date	new_capital	old_capital	Flow	InstrumentID
1400	8/25/2021	5141234000	1468924000	1	3328
1401	8/24/2021	2959000000	1802040245	1	3916
1402	8/23/2021	70000000000	25000000000	1	3894
1403	8/15/2021	11000000000	4350000000	1	3296
1404	8/11/2021	16880000000	1200000000	1	3706
1405	8/11/2021	4350000000	11000000000	1	3296
1406	8/8/2021	70100000	24100000	1	3906
1407	8/8/2021	12771022000	2936422000	1	3140
1408	8/2/2021	2000000000	1000000000	1	3448
1409	8/1/2021	900000000	600000000	1	3770

شکل (3-15) نمونه‌ی داده‌های جدول اطلاعات افزایش سرمایه

3-3- دریافت اطلاعات

دریافت اطلاعات از طریق API‌هایی که سایت tsetmc تعبیه کرده صورت می‌پذیرد. در سایت cdn.tsetmc.com برای هر نماد در هر روز صفحه‌ای وجود دارد که شامل اطلاعات آن نماد در آن روز است. پیش از این اطلاعات نمادها توسط اسکریپتی درون بدنه‌ی HTML این صفحه بارگذاری می‌شد اما طی تغییرات اخیر، این اسکریپت جای خود را به API‌هایی داده که در این پروژه از آنها بهره‌گیری می‌شود. لازم به ذکر است دلیل بازنویسی خزنده‌ی قبلی و همچنین تغییر فریم‌ورک پایه از scrapy به django، همین تغییر رویه در وب‌سایت مدیریت فناوری بورس است. بدین ترتیب با

ار سال درخواست به هر کدام از این API ها یک فایل Json دریافت و واکنشی اطلاعات از طریق آن انجام و به پایگاه داده منتقل می شود.

خزنده ی جدید شامل دو فایل instrument_services و others است که فایل اول به واکنشی اطلاعات نمادها پرداخته و باقی اطلاعات توسط فایل دوم استخراج می شود. دلیل آن که این خزنده به دو فایل تقسیم شده، تفاوت کاربردی این دو و استفاده از فایل instrument تنها یک بار در ابتدای واکنشی اطلاعات است.

1-3-3- روش دریافت اطلاعات

در cdn برای هر نماد در هر روز، فرمت لینک صفحه به شکل زیر است:

`Cdn.tsetmc.com/history/[instrumentID]/[Date]`

که در آن InstrumentID همان InstrumentID است که در فایل csv ذخیره شده و Date تاریخ میلادی به فرمت YMD است. از آن جا که اطلاعات دیگر مستقیماً در فایل این صفحات بارگذاری نمی شود می توانیم از API برای دریافت اطلاعات از هر بخش استفاده کنیم.

2-3-3- ثبت اطلاعات اولیه

در فایل instrument_services دو تابع `get_instrument_ids` و `get_instruments` وجود دارد. تابع `get_instrument_ids` از طریق فایل csv کد نماد را برای تابع دیگر واکنشی می کند و در اختیار آن قرار می دهد. از طرف دیگر، تابع `get_instruments` به ازای هر خط از اطلاعات، یک بار از طریق API اطلاعات را واکنشی و در دیتابیس قرار می دهد. این API از طریق لینک زیر قابل دسترسی است:

`Cdn.tsetmc.com/api/Instrument/GetInstrumentHistory/[id]/[date]`

```
{
  "instrumentHistory": {
    "insCode": 0,
    "lVal30": "سرمایه گذاری تامین اجتماعی",
    "lVal18AFC": "فست",
    "flow": 1,
    "cIsin": "IRO1TAMN0006",
    "baseVol": 8379888,
    "cComVal": "5",
    "cIsin": "IRO1TAMN0006",
    "cgrValCot": "91",
    "cgrValCotTitle": "-",
    "flow": 1,
    "flowTitle": "بازار بورس",
    "insCode": 0,
    "instrumentID": "IRO1TAMN0001",
    "lVal18AFC": "فست",
    "lVal30": "سرمایه گذاری تامین اجتماعی",
    "zTitad": 14200000000
  }
}
```

شکل (3-16) خروجی API اطلاعات اولیه‌ی نماد به صورت JSON

طبیعتاً برای خواندن اطلاعات نماد از به‌روزترین تاریخ ممکن، یعنی تاریخ روزی که در حال اجرای این تابع هستیم، استفاده می‌کنیم. بدین ترتیب با ارسال درخواست به این API داده‌های زیر استخراج و ثبت می‌گردد: نام فارسی نماد، نام فارسی خلاصه‌شده‌ی نماد، نام و کد بازار نماد، کد گروه نماد، تعداد سهام، حجم مبنا، کد شرکت، آخرین قیمت و قیمت نهایی امروز. مشخصاً آخرین قیمت و قیمت نهایی یک سهم که در جدول Instrument وجود دارد، در طول زمان امکان تغییر دارد که در ادامه به آن نیز خواهیم پرداخت.

3-3-3- ثبت سایر اطلاعات نمادها

در فایل `others.py` توابع دیگری با کارکردی بسیار شبیه به تابع `get_instruments` در فایل `instruments_services` وجود دارد که هر کدام یکی از جداول پایگاه داده را پر می‌کنند. در این فایل یک تابع `period_generator` برای تولید تاریخ‌های مورد نیاز برای واکنشی اطلاعات نیز وجود دارد. تابع `period_generator` به ازای دو ورودی تاریخ‌های شروع و پایان، تمامی تاریخ‌های درون آن بازه را تولید و به تابع اصلی ارسال می‌کند. تابع اصلی ما در این فایل تابع `crawl` است که به ازای هر روزی که در `period_generator` تولید می‌شود، تمامی توابع دیگر را برای آن تاریخ صدا می‌زند. این توابع، بر خلاف تابع `crawl` نمادها، در صورت تکرار `crawl` (هم‌پوشانی بازه‌ی تاریخی در دوبار اجرای خزنده) اطلاعات را دوباره تولید نکرده و در نتیجه مشکلی ایجاد نمی‌کنند. این توابع شامل موارد زیر است:

3-3-3-1 تابع Get_ShareHolders

این تابع اطلاعات تمامی سهام داران تمامی نمادها را بررسی کرده و اگر اطلاعات سهام داری در پایگاه داده (جدول Shareholders) موجود نباشد، آن را ایجاد می کند. سپس اطلاعات مربوط به این سهام داران را در پایگاه داده (جدول share) ذخیره می نماید.

- `cdn.tsetmc.com/api/shareholder/[instrumentID]/[date]`

3-3-3-2 تابع Get_day_price_threshold

این تابع از طریق API تمامی اطلاعات مربوط به جدول DayPriceThreshold را واکشی و ذخیره می کنند.

- `cdn.tsetmc.com/api/MarketData/GetStaticThreshold/[InstrumentID]/[Date]`

3-3-3-3 تابع Get_day_trade_summaries

این تابع ابتدا اطلاعات جدول dayTradeSummary را تکمیل کرده و بعد در صورتی که تاریخ تکمیل (ذخیره شده در ستون Updated_at در جدول نماد) ستون های قیمت نهایی و آخرین قیمت جدول Instrument از این تاریخ قدیمی تر باشد، آن را نیز به روزرسانی می کند.

- `cdn.tsetmc.com/API/ClosingPrice/GetClosingPriceDaily/[InstrumentID]/[Date]`

3-3-3-4 تابع Get_client_trade_info

این تابع مقادیر جدول ClientTradeInfo را واکشی و در پایگاه داده ذخیره سازی می کند.

- `cdn.tsetmc.com/api/ClientType/GetClientTypeHistory/[InstrumentID]/[Date]`

3-3-3-5- Create_table_of_interest تابع

این تابع از طریق API عمل نکرده و با انجام محاسبه بر روی اطلاعات ده روز گذشته ی نماد در دو جدول DayTradeSummary و ClientTradeInfo اطلاعات جدول TableOfInterest10 را تولید می کند.

3-3-4- افزایش سرمایه ها

دو تابع دیگر نیز در این فایل وجود دارند که مخصوص خواندن اطلاعات مربوط به افزایش سرمایه ها است. اطلاعات مربوط به افزایش سرمایه ها از ابتدا تا کنون در دو صفحه (تصویر ۳-۱۷) نگهداری می شود. آدرس این صفحات به شکل زیر است:

- Tsetmc.com/loader.aspx?partree=151310&flow=1
- Tsetmc.com/loader.aspx?partree=151310&flow=2

که اولی به بازار بورس و دومی به نمادهای فرابورس مربوط می شود. لازم به ذکر است سطر اول هر کدام از جداول این صفحات، بی استفاده است و باید از سطر دوم (ایندکس ۱) شروع به خواندن کنیم.

3-3-4-1- فرمت اطلاعات

اطلاعات مربوط به افزایش سرمایه ها در جدولی نگهداری می شوند که هر ردیف، مربوط به یک افزایش سرمایه است. با بررسی کد HTML صفحه (تصویر ۳-۱۸) در می یابیم که اطلاعات به صورت زیر نگهداری می شود:

- در هر ردیف کد نماد در انتهای href دیتای اول قرار دارد. دو td اول شامل نام فارسی و نام اختصاری فارسی نیز هست که می دانیم پیش از این در جدول نماد ذخیره شده و نیازی به واکشی آنها نیست.
- تاریخ مربوط به این افزایش سرمایه درون td سوم به فرمت جلالی Y/M/D قرار دارد.
- عدد دقیق سرمایه ی جدید و سرمایه ی قبلی به ترتیب درون td چهارم و پنجم و به فرمت متنی با جداکننده ی سه رقم قرار گرفته است.

افزایش سرمایه				
نماد	نام	تاریخ	سهام جدید	سهام قبلی
ولیز	لیزینگ ایران	1400/6/17	4.000 B	600.000 M
اتکام	بیمه اتکایی امین	1400/6/16	5.000 B	3.900 B
فمراد	آلومراد	1400/6/14	200.000 M	49.535 M
مدار	س. مدیریت سرمایه مدار	1400/6/14	4.600 B	2.300 B
کحافظ	کاشی و سرامیک حافظ	1400/6/3	5.141 B	1.469 B
سایک	سیمان آبیک	1400/6/2	2.959 B	1.802 B
پترول	گروه پتروشیمی س. ایرانیان	1400/6/1	70.000 B	25.000 B
تیبیکو	سرمایه گذاری دارویی تامین	1400/5/24	11.000 B	4.350 B
کراز	کارخانجات تولیدی شیشه رازی	1400/5/20	16.880 B	1.200 B
تیبیکو	سرمایه گذاری دارویی تامین	1400/5/20	4.350 B	11.000 B

شکل (17-3) صفحه‌ی افزایش سرمایه‌ها

```

1 <!doctype html><html><head><title>.TSETMC. :: افزایش سرمایه </title><!-- Global site tag (gtag.js) - Google Analytics --><script async src="https://www.googletagmanager.com/gtag/js?id=UA-63076938-1"></script><script>
2 <div class="asplethidden">
3 <input type="hidden" name="__VIEWSTATE" id="__VIEWSTATE" value="" />
4 </div>
5 <span id="PureData">
6 <div class="box1 yellow tbl">
7 <div class="header">افزایش سرمایه</div>
8 <div class="content">
9
10 <table class="table">
11 <thead><tr><th>نام</th><th>تاریخ</th><th>سهام جدید</th><th>سهام قبلی</th></tr></thead>
12 <tbody>
13
14 <tr><td><a target="_blank" href="Loader.aspx?ParTree=151311&i=20946530370609828" title="لیزینگ ایران">ولیز</td>
15 <td><a target="_blank" href="Loader.aspx?ParTree=151311&i=20946530370609828" title="بیمه اتکایی ایران">اتکام</td>
16 <td><a target="_blank" href="Loader.aspx?ParTree=151311&i=20946530370609828" title="آلومراد">فمراد</td>
17 <td><a target="_blank" href="Loader.aspx?ParTree=151311&i=20946530370609828" title="س. مدیریت سرمایه مدار">مدار</td>
18 <td><a target="_blank" href="Loader.aspx?ParTree=151311&i=20946530370609828" title="کاشی و سرامیک حافظ">کحافظ</td>
19 <td><a target="_blank" href="Loader.aspx?ParTree=151311&i=20946530370609828" title="سیمان آبیک">سایک</td>
20 <td><a target="_blank" href="Loader.aspx?ParTree=151311&i=20946530370609828" title="گروه پتروشیمی س. ایرانیان">پترول</td>
21 <td><a target="_blank" href="Loader.aspx?ParTree=151311&i=20946530370609828" title="سرمایه گذاری دارویی تامین">تیبیکو</td>
22 <td><a target="_blank" href="Loader.aspx?ParTree=151311&i=20946530370609828" title="کارخانجات تولیدی شیشه رازی">کراز</td>
23 <td><a target="_blank" href="Loader.aspx?ParTree=151311&i=20946530370609828" title="سرمایه گذاری دارویی تامین">تیبیکو</td>
24 </tbody></table>
25
26 <tr><td><a target="_blank" href="Loader.aspx?ParTree=151311&i=27405735172634593" title="لیزینگ ایران">ولیز</td>
27 <td><a target="_blank" href="Loader.aspx?ParTree=151311&i=27405735172634593" title="بیمه اتکایی ایران">اتکام</td>
28 <td><a target="_blank" href="Loader.aspx?ParTree=151311&i=27405735172634593" title="آلومراد">فمراد</td>
29 <td><a target="_blank" href="Loader.aspx?ParTree=151311&i=27405735172634593" title="س. مدیریت سرمایه مدار">مدار</td>
30 <td><a target="_blank" href="Loader.aspx?ParTree=151311&i=27405735172634593" title="کاشی و سرامیک حافظ">کحافظ</td>
31 <td><a target="_blank" href="Loader.aspx?ParTree=151311&i=27405735172634593" title="سیمان آبیک">سایک</td>
32 <td><a target="_blank" href="Loader.aspx?ParTree=151311&i=27405735172634593" title="گروه پتروشیمی س. ایرانیان">پترول</td>
33 <td><a target="_blank" href="Loader.aspx?ParTree=151311&i=27405735172634593" title="سرمایه گذاری دارویی تامین">تیبیکو</td>
34 <td><a target="_blank" href="Loader.aspx?ParTree=151311&i=27405735172634593" title="کارخانجات تولیدی شیشه رازی">کراز</td>
35 <td><a target="_blank" href="Loader.aspx?ParTree=151311&i=27405735172634593" title="سرمایه گذاری دارویی تامین">تیبیکو</td>
36 </tr>
37

```

شکل (18-3) اطلاعات صفحه‌ی افزایش سرمایه‌ها

به این ترتیب برای تکمیل اطلاعات مربوط به افزایش سرمایه نیاز است تا برای هر flow:

- ابتدا کد HTML صفحه را خوانده و trها را جداسازی کنیم.
- برای هر tr از td اول کد نماد را استخراج کرده و شناسه‌ی نماد را به دست بیاوریم. با توجه به ثابت بودن طول href می‌دانیم کد نماد از ایندکس ۲۹ شروع می‌شود.
- سپس با استفاده از مبدل تاریخ جلالی به میلادی td سوم را به فرمت استاندارد پایگاه داده که میلادی است، تبدیل کنیم.
- سپس با استفاده از یک تابع کمکی (تابع get_num_from_str) اطلاعات متنی td چهارم و پنجم را به عدد تبدیل کنیم.
- در نهایت این اطلاعات را در پایگاه داده ذخیره کنیم.

کتابخانه‌های استفاده‌شده برای این عملیات، BeautifulSoup و Jdatetime است که اولی برای خواندن اطلاعات HTML با استفاده از تگ‌ها و دومی برای تبدیل تاریخ جلالی به میلادی کاربرد دارد. با خواندن هر ردیف از جدول، اطلاعات یک افزایش سرمایه به دست می‌آید.

3-3-5- توابع بازخوانی دوره‌ای اطلاعات

در انتها برای بازخوانی دوره‌ای اطلاعات، دو تابع job_crawl و job_capital به همراه توابع کمکی start_crawling و start_change_capitals اجرا شده‌اند. این توابع با استفاده از توابعی که پیش از این شرح داده شده‌اند، به ذخیره‌سازی اطلاعات در پایگاه داده می‌پردازند. این توابع را می‌توان با استفاده از ساخت ترمینال‌های مجازی توسط screen همیشه در حال اجرا نگاه داشت.

3-3-5-1- دریافت اطلاعات نمادها به صورت روزانه

تابع job_crawl در هر ساعت یکبار اجرا می‌شود و در ساعتی از شبانه‌روز که از قبل در آن تنظیم شده، با فراخوانی تابع start_crawling اطلاعات دو روز قبل را واکشی و ذخیره‌سازی می‌کند. این رویکرد برای اطمینان از صحت دریافت اطلاعات پیاده‌سازی شده است. در صورتی که ساعت اجرا با ساعت مقرر یکسان نباشد، این تابع با استفاده از متد به‌خواب‌رفتن پایتون یک ساعت توقف می‌کند.

3-3-5-2- دریافت اطلاعات افزایش سرمایه و تعدیل قیمت‌ها

تابع job_capital در هر روز یکبار اجرا می‌شود. (با استفاده از همان متد به‌خواب‌رفتن) این تابع در روز مشخصی از هفته، با اجرای تابع کمکی start_change_capitals ابتدا اطلاعات افزایش سرمایه را بازخوانی و سپس قیمت تعدیل‌شده‌ی تمامی نمادها را با توجه به افزایش سرمایه‌های جدید اصلاح می‌کند.

4-3- ذخیره سازی اطلاعات

اطلاعات در دیتابیس ی با عنوان `Tehran_stock_exchange` ذخیره می شوند. برای تغییر این نام می توان در فایل `settings.py` نام دیتابیس را عوض کرد. همچنین می توان با استفاده از تغییر فایل `Models.py`، تغییرات دلخواه را در ساختار مدل داده ای ایجاد نمود؛ لازم به ذکر است پس از انجام این عملیات، باید `migration` های لازم دوباره صورت پذیرد.

5-3- خلاصه و جمع بندی

فصل سوم به طور عمده شیوه ی عملی اجرای پروژه را در بر می گیرد که شامل مستندات پایگاه داده و پیاده سازی خزنده به همراه اطلاعات فنی پروژه است.

فصل 4: راه اندازی و اجرا خزنده و رابط کاربری

پس از توضیحات فنی و اجرایی، در این فصل در مورد نحوه‌ی راه‌اندازی و اجرای خزنده بر روی سرور میزبان و استفاده از آن بر روی سایر سیستم‌ها صحبت خواهیم کرد. در انتها نیز به بررسی رابط کاربری این پروژه می‌پردازیم.

1-4- آماده‌سازی سیستم

برای اجرای خزنده بر روی هر سیستم عامل، نیازمند نصب پایتون و کتابخانه‌های استفاده‌شده در پروژه و پایگاه داده‌ی MySQL هستیم. با فعال‌سازی environment اجرا، پکیج‌ها را نصب و آماده‌سازی کنید.

2-4- اجرای خزنده‌ی کدها

برای اجرای خزنده‌ی کدها می‌توانید از دستورالعمل خزنده‌ی قبلی که به اختصار در اینجا آورده می‌شود استفاده کنید.

ابتدا پروژه را از GitLab دانلود کرده و با استفاده از فایل requirements پکیج‌های مورد نیاز را نصب کنید.

با اجرای دستور

```
Scrapy Crawl InstrumentCode
```

فایل instrumentCode.csv ساخته‌شده را در کنار فایل manage.py در پروژه‌ی اصلی کپی کنید. با ساختن یک پایگاه داده به نامی که در settings.py مشخص شده، می‌توانید به اجرای توابع پروژه بپردازید.

3-4- اجرای توابع اصلی

خزنده با استفاده از هر گونه shell در پایتون قابل اجرا است. پس از جای‌گذاری فایل csv ابتدا در ترمینال به دایرکتوری اصلی پروژه وارد شوید. فایل manage.py در این دایرکتوری قرار دارد. با استفاده از دستور زیر می‌توان shell را باز کرد:

```
Python manage.py shell
```

سپس باید سرویس‌های پروژه را با استفاده از دستورهای زیر ایمپورت کنیم:

```
From bourse.services.instrument_services import*
```

From bourse.services.others import*

برای اجرای تابع crawl نیازمند دو تاریخ به فرمت استاندارد هستیم. پس با دستور

From datetime import date

تابع تولید تاریخ را از کتابخانه‌ی مربوطه ایمپورت می‌کنیم؛ سپس می‌توانیم دو تاریخ برای شروع و پایان (اختیاری) کراول وارد کنیم. فرمت استاندارد تاریخ به صورت سال، ماه، روز و بدون صفر قبل اعداد تکرریمی است، به عنوان مثال برای شروع و پایان آرگومان‌های زیر را می‌توان وارد کرد:

Start= date(2020, 1, 1)

End= date(2020, 2, 2)

سپس با استفاده از تابع اصلی کراول، تمامی توابع اجرا می‌شوند:

Crawl(Start, End)

در صورت نیاز می‌توان برای موازی‌سازی استخراج اطلاعات، توابع را به صورت جداگانه نیز فراخوانی کرد که در این صورت باید هر روز دستور فراخوانی تابع را با آرگومان آن روز اجرا کنیم. به عنوان مثال:

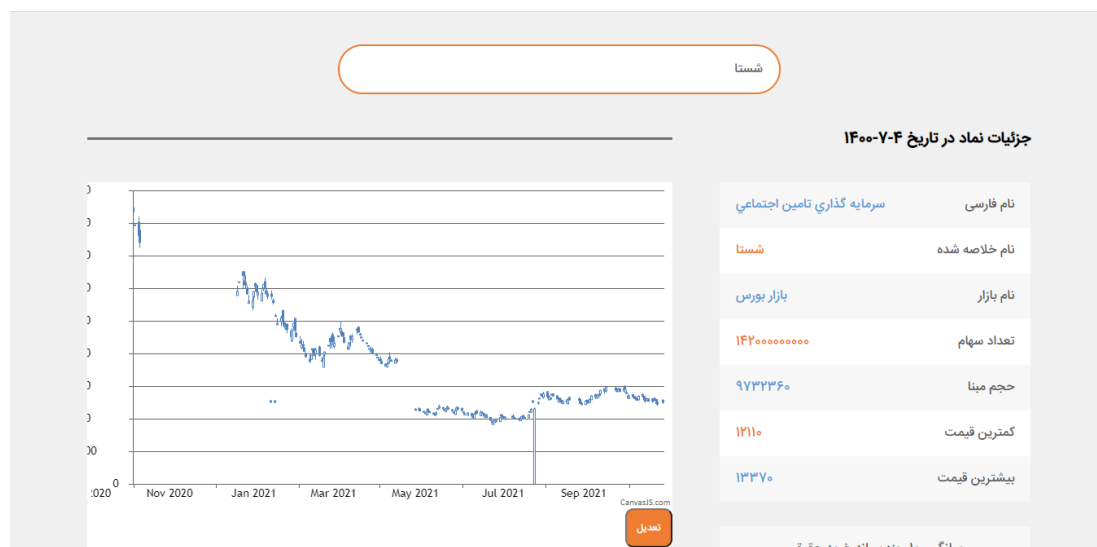
Get_day_Price_Threshold(Start)

یا

Get_day_Price_Threshold(date(2020, 1, 1))

4-4- رابط گرافیکی

رابط گرافیکی مورد نیاز برای نظارت بر پروژه شامل یک صفحه‌ی پیاده‌سازی‌شده توسط react است که با استفاده از API های PHP از دیتابیس کار می‌کند. این وب‌سایت هم‌اکنون از طریق آی‌پی 185.126.200.119 قابل دسترسی است. این صفحه به سه بخش Header، Sidebar و Content تقسیم شده است.



شکل (4-1) پیاده‌سازی رابط گرافیکی

4-4-2- هدر

هدر شامل یک ماژول جستجو برای یافتن نمادها است که با نوشتن در آن، لیستی از نمادهایی که نام فارسی اختصاری آنها شامل آن متن هستند را به نمایش می‌گذارد. با انتخاب هر کدام از این نمادها اطلاعات آن به نمایش در خواهد آمد. لازم به ذکر است از آنجا که در پایگاه داده‌ی شرکت مدیریت فناوری بورس تهران، نام اختصاری نمادها با حروف عربی ذخیره‌سازی شده است، یک نرمالایزر برای تبدیل حروف فارسی به عربی نیز در این ماژول به کار گرفته شده است.

4-4-3- منوی کناری

این منو شامل اطلاعات اولیه‌ی نماد، میانگین ده‌روزه‌ی خرید حقیقی و حقوقی و همچنین حجم معاملات است.

4-4-4- محتوای اصلی

این بخش شامل نمودار تغییر قیمت و اطلاعات خرید و فروش روزانه‌ی نماد در یک قسمت و در قسمت پایین نمودارهای مقایسه‌ای سرانه‌ی معاملات، حجم خرید و فروش و در صد سهم خرید و فروش برای سرمایه‌گذاران حقیقی و حقوقی است. نمودار تغییر قیمت اطلاعات کامل تهیه‌شده را به شکل شمعی که استاندارد نمودارهای مالی است به نمایش می‌گذارد. توسط دکمه‌ی تعدیل در زیر این نمودار می‌توان اطلاعات قیمت تعدیل‌شده را نیز مشاهده نمود.

4-5- خلاصه و جمع‌بندی

در این فصل به توضیح فرایند آماده‌سازی محیط اجرای خزنده‌ها و دستورات لازم برای اجرای هر دو خزنده پرداخته‌شد. هم‌چنین به شرح ویژگی‌های رابط کاربری نیز پرداختیم.

فصل 5: جمع‌بندی، نتیجه‌گیری و پیشنهادها

در این فصل نتیجه‌گیری‌های کلی حاصل‌شده در این تحقیق بیان می‌شود و محدودیت‌ها مورد بحث قرار می‌گیرد. همچنین پیشنهادهایی برای ادامه‌ی مسیر به علاقمندان این حوزه ارائه خواهد شد.

1-5- جمع‌بندی

در این تحقیق در گام نخست ما به مطالعه‌ی اولیه روی مفاهیم اولیه‌ی بازار بورس پرداختیم و با ساختار صفحات و معماری اطلاعات وب سایت مدیریت فناوری بورس تهران آشنا شدیم. سپس با مطالعه درباره‌ی نحوه‌ی پیاده‌سازی خزنده‌ی قبلی متوجه اصلاحات ساختاری این وب سایت شدیم. بدین ترتیب با یافتن ایرادات خزنده‌ی قبلی، به این نتیجه رسیدیم که نیازمند اصلاحات ساختاری هستیم و در نتیجه، خزنده‌ی جدیدی را پیاده‌سازی نمودیم. در بازطراحی مدل داده‌ای و پایگاه داده سعی بر آن کردیم که حتی‌الامکان ساختار مدل داده‌ای قبل را حفظ کنیم. پس از پایان این فرایند با راه‌اندازی سرور و استخراج منظم داده‌ها از صحت این اطلاعات، اطمینان حاصل کردیم و سپس رابط کاربری ساده و مناسبی پیاده‌سازی و مستقر کردیم.

2-5- دستاوردها

ایجاد یک مدل داده‌ای کامل و به روز با قابلیت تغییرات گسترده از جمله حذف یا اضافه کردن اطلاعات دل‌خواه (با توجه به API‌هایی که در اختیار داریم) مهم‌ترین دستاورد این پروژه است. از دیگر دستاوردهای این پروژه می‌توان به رابط گرافیکی ساده و مناسب برای بررسی صحت داده‌ها و همچنین استفاده‌های آتی در تحلیل‌های مختلف اشاره کرد.

3-5- پیشنهادها

با توجه به تغییرات صورت گرفته، در ادامه‌ی روند این پروژه می‌توان اطلاعات دقیق‌تر و کامل‌تری از تمامی تغییرات و ویژگی‌های نمادها و سهام به دست آورد. اجرای منظم این خزنده و افزودن اطلاعات تمام معاملات می‌تواند در دسترسی به داده‌هایی دقیق‌تر و مفیدتر کمک شایانی کند. با استفاده از جداول کمکی مانند جدول میانگین‌های ده‌روزه و با گسترش این جداول می‌توان داده‌های آماده برای اجرای روش‌های مختلف مبتنی بر هوش مصنوعی و یادگیری ماشین فراهم نمود. این روش‌ها در کنار روش‌های تحلیل بنیادین و تکنیکال می‌توانند محاسبات بهتری بر روی کاهش و افزایش ارزش سهام و در نتیجه راهنمای مناسبی برای مبادله‌ی هوشمند ارائه دهند.

فصل 6: مراجع

مراجع

- [1] S. Madge, S. Bhatt, "Predicting stock price direction using support vector machines," Independent work report spring, 2015.
- [2]Rubio D. (2017) Django Models. In: Beginning Django. Apress, Berkeley, CA.
https://doi.org/10.1007/978-1-4842-2787-9_7
- [3] Yatskov, A.K., Varlamov, M.I. & Turdakov, D.Y. Extraction of Data from Mass Media Web Sites. Program Comput Soft 44, 344–352 (2018).
<https://doi.org/10.1134/S0361768818050092>
- [4] Hetland M.L. (2017) Python and the Web. In: Beginning Python. Apress, Berkeley, CA.
https://doi.org/10.1007/978-1-4842-0028-5_15



University of
Tehran



College of Engineering

School of Electrical and Computer Engineering

Preparing and Launching the Tehran Stock Exchange Data Crawling and Processing System

A thesis submitted to the Undergraduate Studies Office

In partial fulfillment of the requirements for

The degree of Master in

Computer Engineering

By:

Ali Ghane

Supervisor:

Hesham Faili PhD