

در این پروژه داده‌های شاخص کل بازار بورس تهران و داده‌های چند سهم مختلف این بازار در اختیار شما قرار داده شده است. در ادامه خلاصه‌ای از فرایند خرید و فروش سهام در بورس و سپس قسمت‌های مختلف پروژه توضیح داده شده است.

### توضیحات مقدماتی:

**ساختار کلی بازار بورس** به این صورت است که هر شرکت در قالب تعدادی سهم با قیمت مشخص در بازار بورس عرضه می‌شود. سپس سهام‌های عرضه شده توسط مشتریان خرید و فروش می‌شوند. بدین ترتیب خریداران و فروشندگان یک سهم در هر روز، پیشنهادات خود را در بازه مجاز مشخص شده برای قیمت سهام آن شرکت در آن روز اعلام می‌کنند. میزان تغییرات مجاز قیمت اکثر سهام‌ها در هر روز حداقل منفی ۵ درصد قیمت پایانی روز قبل و حداکثر مثبت ۵ درصد قیمت پایانی روز قبل است، ولی برخی دیگر از سهام‌ها اجازه تغییر در بازه کمتر یا بیشتری را دارند. در صورت تطبیق قیمت یک خریدار و یک فروشنده معامله به میزان حجم کمینه دو درخواست انجام می‌شود. الگوریتم‌های تطبیق‌گر مختلفی برای مدیریت معاملات با توجه به صف خرید و فروش وجود دارد. هر سهم یک حجم مبنای مشخصی نیز دارد که برابر است با حداقل تعداد سهام یک شرکت که باید در طول یک روز معاملاتی، خرید و فروش شود تا قیمت آن سهم بتواند بر اساس سقف دامنه نوسان، تغییر کند.

در هر روز (یا هر بازه زمانی دیگری) چهار قیمت اصلی برای هر سهم تعریف می‌شود که نشان دهنده تغییرات قیمت در قالب نمودار شمعی هستند:

**قیمت باز شدن (opening):** قیمت سهم در زمان شروع معاملات در روز مشخص را بیان می‌کند.

**قیمت کمینه (low):** کمترین قیمت معامله شده سهم در روز مشخص را بیان می‌کند.

**قیمت بیشینه (high):** بیشترین قیمت معامله شده سهم در روز مشخص را بیان می‌کند.

**قیمت بسته شدن (closing):** قیمت سهم در آخرین معامله انجام شده در آن روز را نشان می‌دهد.

علاوه بر چهار عدد فوق، قیمت پایانی سهم در هر روز از اهمیت ویژه ای برخوردار است و تعیین کننده قیمت پایه سهم در روز بعدی معاملات خواهد بود:

قیمت پایانی: قیمت سهم در پایان آن روز مشخص را بیان می کند که بر اساس میانگین قیمت معاملات انجام شده و متناسب با حجم مبنا محاسبه می شود.

یکی دیگر از ویژگی های مهم بورس، **شاخص کل بورس** است که بیانگر ارزش نسبی سهام کلیه شرکت های موجود در بازار است. شاخص بورس نشانگر یا نموداری برای آگاهی از وضعیت تغییرات بورس نسبت به روزهای قبل است. رشد یا افت شاخص به ترتیب نمایانگر رشد یا افت بازدهی کل بورس است.

برای پیش بینی رفتار آینده یک سهم خاص از اطلاعات سهم های مختلف می توان استفاده کرد. علاوه بر این می توان از سایر اطلاعات مرتبط نظیر اخبار، قیمت مسکن، طلا، دلار و ..... نیز بهره برد.

### فرمت گزارش:

گزارش نهایی می بایست به زبان فارسی یا انگلیسی در قالب فایل PDF باشد. در کنار آن پوشه ای با نام Code قرار داشته باشد و همه فایل های یاد شده، در یک فایل فشرده با فرمت مشخص جهت ارزیابی ارسال گردند. در گزارش نیاز است روی خروجی ها و نمودارهای درج شده **حتما** گزارش تحلیلی داشته باشید و استدلال خود را از تحلیل ها نشان دهید. نتیجه گیری خود راجع به هر قسمت را به صورت حداقل یک پاراگراف بیان نمایید.

فایل گزارش را به فرمت SML\_report\_StdNum.pdf نام گذاری نمایید (همانند SML\_report\_98131.pdf).

### فرمت کد نویسی:

برای هر قسمت باید فایل کد جداگانه با زبان های MATLAB یا Python نوشته شود. علاقه مندان به پایتون حتما از محیط Jupyter notebook استفاده نمایند تا خروجی در فایل مذکور نیز مشهود باشد. کامنت گذاری در حد لازم نیز انجام پذیرد.

## نحوه تحویل:

فایل‌های کد و گزارش خود را مطابق فرمت‌های فوق آماده و در قالب یک فایل فشرده با نام SML\_Project\_StdNum.zip تهیه نمایید.

در صورت وجود سوال به آدرس زیر ایمیل بزنید:

sml.ce.aut@gmail.com

## داده‌ها:

داده‌های شاخص کل بورس تهران و قیمت پایانی چند سهم زیر از طریق سایت TSETMC که لینک‌های آن در پایین داده شده است قابل دریافت است. از بازه دو ساله برای انجام آزمایش‌ها استفاده کنید.

[فولاد](#)، [شستا](#)، [کخاک](#)، [صندوق اطلس](#)، [خودرو](#)

جدول اطلاعات سهم قابل دریافت از سایت فوق به شکل زیر است:

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	<TICKER>	<DTYYYYM>	<FIRST>	<HIGH>	<LOW>	<CLOSE>	<VALUE>	<VOL>	<OPENINT>	<PER>	<OPEN>	<LAST>		
2	ToseAtlas	20230103	310996	312900	303002	306968	1.28E+11	417524	1227	D	309108	305000		
3	ToseAtlas	20230102	306266	311210	304001	309108	1.66E+11	536568	1318	D	300261	309500		
4	ToseAtlas	20230101	300510	311001	294001	300261	2.62E+11	873546	1531	D	307097	305995		
5	ToseAtlas	20221231	317001	317001	300002	307097	4.84E+11	1576967	2268	D	326194	303400		
6	ToseAtlas	20221228	311010	340452	310005	326194	5.99E+11	1835702	2326	D	309502	328011		
7	ToseAtlas	20221226	305100	313993	303000	309502	2.43E+11	786030	1214	D	306832	310101		
8	ToseAtlas	20221225	305001	310001	300010	306832	1.64E+11	535450	1140	D	301381	308020		
9	ToseAtlas	20221224	292995	304985	292900	301381	2E+11	663790	1047	D	288344	304675		
10	ToseAtlas	20221221	286514	290000	286514	288344	5.86E+10	203378	447	D	287756	288968		

<DTYYYYMMDD> تاریخ

<FIRST> قیمت اولین معامله

<HIGH> بالاترین قیمت

<LOW> پایین ترین قیمت

<CLOSE> قیمت بسته شدن (پایانی)

<VALUE> ارزش معاملات

<VOL> حجم معاملات

<OPEN> قیمت باز شدن

<LAST> قیمت آخرین معامله

## پیاده‌سازی:

### (۱) بررسی توزیع احتمالاتی مقادیر بازده

مقدار بازده (return) یک سهم در یک روز طبق رابطه زیر محاسبه می‌شود. در این رابطه  $P_i$  قیمت پایانی سهم در روز  $i$  و  $r_i$  مقدار بازده در این روز است.

$$r_i = \frac{P_i - P_{i-1}}{P_{i-1}}$$

الف) توزیع احتمالاتی مقادیر بازده شاخص کل را به دو روش پارامتری (با استفاده از یک توزیع نرمال) و غیرپارامتری (هیستوگرام گسسته) تخمین زده و نتایج را تحلیل کنید.

ب) مقارن یا نامتقارن بودن توزیع را با استفاده از آزمون‌های مربوطه بررسی کرده و آن را تحلیل کنید.

ج) میانگین و واریانس بازده هر یک از سهم‌ها در کل بازه مشخص شده را محاسبه کرده و آنها را با یکدیگر مقایسه کنید. نتایج را بر اساس وضعیت تغییرات قیمت سهم‌های مورد نظر در بازه مذکور تفسیر کنید.

د) تغییرات میانگین و واریانس بازده شاخص کل و هر یک از سهم‌ها در پنجره‌های زمانی مشخص (۱۰، ۲۰ و ۵۰ روزه) در طول زمان را بررسی کنید.

### (۲) بررسی همبستگی‌ها

الف) میزان خودهمبستگی (autocorrelation) بازده شاخص کل لگ‌های زمانی مختلف (۱ تا ۵ روز) را محاسبه و تحلیل کنید.

ب) میزان همبستگی (Correlation) بازده شاخص کل با بازده هر یک از سهم‌های مورد نظر را محاسبه کنید.

ج) همبستگی بازده شاخص کل با بازده طلا و دلار را مقایسه کنید (اختیاری)

۳) پیش‌بینی مقدار بازده

در آزمایش‌های این بخش از داده‌های 0 تا  $t$  برای آموزش و از داده‌های  $t$  تا  $T$  (اندیس آخرین روز) برای آزمون استفاده کنید.

الف) یک مدل رگرسیون خطی آموزش دهید که بر اساس اطلاعات چند روز بازده شاخص کل، مقدار بازده روز بعد را پیش‌بینی کند. در مورد نحوه انتخاب اطلاعات چند روز قبل از نتایج قسمت قبل کمک بگیرید.

ب) مدلی دسته‌بندی‌کننده آماری آموزش دهید که تنها روند تغییرات شاخص کل (یعنی مثبت یا منفی بودن بازده شاخص) را پیش‌بینی کند. دقت مدل را در پیش‌بینی یک، دو، سه و چهار روز بعد بررسی کنید.

ج) یک مدل ترکیبی آموزش دهید که با استفاده از مقادیر قبلی بازده شاخص کل و سهم‌های مد نظر، مقدار بازده شاخص کل و مثبت یا منفی بودن آن را پیش‌بینی کند.

د) مدل رگرسیون Lasso را برای پیش‌بینی بازده شاخص کل با استفاده از اطلاعات ده روز قبل بازده شاخص کل به کار بگیرید. توانایی این مدل در انتخاب ویژگی‌های مطلوب را بررسی کنید.

در پیاده‌سازی موارد فوق، به نکات زیر توجه کنید:

- برای ارزیابی مدل پیش‌بینی مقدار بازگشت می‌توانید از MSE استفاده کنید.
- برای ارزیابی مدل پیش‌بینی روند، می‌توانید از Accuracy استفاده کنید.
- می‌توانید از کتاب زیر نیز استفاده کنید.

Severini, Thomas A. *Introduction to statistical methods for financial models*. CRC Press, 2017.

موفق باشید