

(عنوان مقاله : تحلیل داده‌های بازار سهام و پیش‌بینی قیمت‌ها با استفاده از مدل‌های یادگیری ماشین)

نام و نام خانوادگی نویسنده اول (غلامحسین مرادی)

وابستگی سازمانی نویسنده (دانشجوی کارشناسی ارشد دانشگاه جامع امام حسین (ع))

نام و نام خانوادگی نویسنده دوم (محمدعلی جوادزاده)

وابستگی سازمانی نویسنده (استادیار دانشگاه جامع امام حسین (ع))

نام و نام خانوادگی نویسنده سوم (حسین حسینی)

وابستگی سازمانی نویسنده (دانشجوی دکتری دانشگاه جامع امام حسین (ع))

چکیده

در این پروژه به تحلیل داده‌های بازار سهام و پیش‌بینی قیمت بسته شدن سهام با بهره‌گیری از مدل‌های یادگیری ماشین پرداخته شده است. داده‌های به کار رفته شامل قیمت باز شدن، بالاترین قیمت، پایین‌ترین قیمت، قیمت بسته شدن و حجم معاملات می باشد که از یک فایل اکسل استخراج شده است. و پس از پیش‌پردازش به منظور تحلیل‌های بیشتر، تغییر نام ستون‌ها، مرتب‌سازی داده‌ها بر اساس تاریخ، مدیریت مقادیر گمشده و تقسیم داده‌ها به مجموعه‌های آموزشی و تست انجام شد. مدل‌سازی با استفاده از الگوریتم رگرسیون خطی انجام گرفت که هدف آن پیش‌بینی قیمت بسته شدن سهام بر اساس ویژگی‌هایی نظیر قیمت باز شدن، بالاترین قیمت، پایین‌ترین قیمت و حجم معاملات بود. ارزیابی مدل با استفاده از معیارهای خطای میانگین مربعات (MSE) و ضریب تعیین (R^2) نشان داد که مدل به خوبی توانسته است عملکرد دقیق و قابل قبولی ارائه دهد. میانگین خطای مربعات برابر با (99773.93) و ضریب تعیین برابر با (0.9976) به دست آمد که نشان از دقت بالای مدل دارد. علاوه بر این، برای بررسی دقیق تر روند تغییرات قیمت و حجم معاملات، نمودارهایی از جمله میانگین 30 و 90 روزه ترسیم شد که اطلاعات مفیدی درباره رفتار بازار ارائه کردند. نتایج کلی پروژه حاکی از آن است که مدل رگرسیون خطی به همراه تحلیل‌های تکمیلی می‌تواند ابزاری قدرتمند برای پیش‌بینی قیمت سهام و ارائه اطلاعات مفید برای سرمایه‌گذاران باشد. همچنین، این پروژه نشان داد که استفاده از داده‌های گذشته و مدل‌های یادگیری ماشین می‌تواند به بهبود تصمیمات مالی کمک شایانی کند.

واژگان کلیدی: تحلیل داده‌های بازار سهام، رگرسیون خطی، مدل‌سازی قیمت سهام، مدل‌های پیش‌بینی مالی



مقدمه

بازارهای سهام از جمله پیچیده‌ترین و جذاب‌ترین سیستم‌های مالی در جهان محسوب می‌شوند. تحلیل دقیق داده‌های بازار سهام و پیش‌بینی قیمت‌ها اهمیت بالایی دارد، زیرا می‌تواند به سرمایه‌گذاران در اتخاذ تصمیم‌های بهتر کمک کند. با رشد روزافزون حجم داده‌های موجود در این حوزه، نیاز به استفاده از مدل‌های پیشرفته برای تحلیل و استخراج اطلاعات از این داده‌ها بیش از پیش احساس می‌شود. در این مقاله، با استفاده از مدل‌های یادگیری ماشین، سعی شده است روند تغییرات قیمت سهام و عوامل مؤثر بر آن بررسی شود. داده‌های مورد استفاده شامل ویژگی‌های مهمی مانند قیمت باز شدن، قیمت بسته شدن، بالاترین قیمت، پایین‌ترین قیمت، و حجم معاملات است که از یک دیتاست استخراج شده‌اند. یکی از اهداف اصلی این پروژه، پیش‌بینی دقیق قیمت بسته شدن سهام بر اساس ویژگی‌های ذکر شده است. همچنین، تحلیل روند بازار و شناسایی الگوهای رفتاری از طریق ابزارهای تحلیل آماری و الگوریتم‌های خوشه‌بندی انجام شده است. این مقاله می‌تواند به بهبود تصمیم‌گیری‌های مالی کمک کرده و به عنوان یک ابزار پشتیبان برای تحلیل‌گران بازارهای مالی مفید باشد.

روش تحقیق

در این مقاله از داده‌های موجود در دیتاست (مهراد آریا، ۲۰۲۰) ویژگی‌های مورد نیاز استخراج شد این دیتاست شامل اطلاعاتی مانند قیمت باز شدن، قیمت بسته شدن، بالاترین قیمت، پایین‌ترین قیمت، و حجم معاملات سهام در طول زمان است.

ساختار داده‌ها

- تعداد نمونه‌ها: ۲۷۴۶ نمونه داده.

- تعداد ویژگی‌ها: ۶ ویژگی اصلی شامل:

Date	Open	Close	Low	High	Volume
تاریخ معاملات	قیمت باز شدن	قیمت بسته شدن	پایین‌ترین قیمت در روز	بالاترین قیمت در روز	حجم معاملات

این داده‌ها به بررسی رفتار بازار و پیش‌بینی قیمت سهام کمک می‌کنند. داده‌ها شامل ویژگی‌هایی است که نشان دهنده روند قیمتی و حجم معاملات در بازه‌های زمانی مختلف می‌باشند.



پیش پردازش داده‌ها

مراحل آماده سازی داده ها به شرح زیر می باشد:

۱. خواندن و بارگذاری داده‌ها: داده‌ها از فایل اکسل بارگذاری شدند.
۲. تغییر نام ستون‌ها: ستون‌ها به نام‌هایی با معانی واضح تر مانند **Open**, **High**, **Low**, **Close**, و **Volume** تغییر یافتند.
۳. تبدیل فرمت تاریخ: ستون تاریخ به فرمت **DateTime** تبدیل شد تا امکان مرتب‌سازی و فیلتر داده‌ها بر اساس تاریخ را داشته باشیم.
۴. مرتب‌سازی داده‌ها: داده‌ها بر اساس ستون تاریخ به ترتیب صعودی مرتب شدند.
۵. مدیریت مقادیر گمشده: داده‌ها برای شناسایی مقادیر گمشده بررسی شدند و در صورت وجود، کارهای لازم برای جایگزینی یا حذف آن‌ها انجام شد.
۶. استانداردسازی داده‌ها: برای خوشه‌بندی، ویژگی‌ها استانداردسازی شدند تا مقیاس داده‌ها بر الگوریتم‌ها تأثیر نگذارد.
۷. تقسیم داده‌ها: داده‌ها به دو مجموعه آموزشی (۸۰٪) و تست (۲۰٪) تقسیم شدند.

تشریح مسئله

مسئله

هدف از این مقاله، پیش‌بینی قیمت بسته شدن سهام در بازار بورس تهران است. این پیش‌بینی به سرمایه‌گذاران کمک می‌کند تا تصمیمات بهتری در خرید و فروش سهام داشته باشند. علاوه بر این، تحلیل رفتار سهام‌ها و پیدا کردن الگوهای نوسانی از اهداف دیگر این تحقیق است.

اهداف

۱. پیش‌بینی قیمت بسته شدن سهام با استفاده از الگوریتم‌های یادگیری ماشین.
۲. بررسی نوسانات قیمتی و شناسایی الگوهای رفتاری سهام‌ها.
۳. دسته‌بندی سهام‌ها به دسته‌های پایدار، نوسان متوسط و نوسان بالا.

محدودیت‌ها

۱. حجم داده‌ها: برخی از داده‌های تاریخی احتمال دارد ناقص یا محدود باشند.
۲. نویز در داده‌ها: به دلیل رویدادهای خاص بعضی داده‌ها احتمال دارد باعث ایجاد انحراف در مدل شوند.
۳. پیچیدگی محاسباتی: این احتمال وجود دارد که پردازش داده‌های مالی نیاز به منابع محاسباتی زیادی داشته باشد.
۴. پیش‌فرض‌های مدل: برخی از مدل‌ها مانند رگرسیون خطی، پیش‌فرض‌هایی مثل رابطه خطی میان متغیرها دارند که ممکن است همیشه این مورد برقرار نباشد.

روش انجام حل مسئله

الگوریتم یا مدل

برای حل مسئله پیش‌بینی قیمت بسته شدن سهام و بررسی الگوهای قیمتی، از مدل‌ها و روش‌های زیر استفاده شده است: رگرسیون خطی: این مدل برای پیش‌بینی قیمت بسته شدن سهام با استفاده از متغیرهای ورودی مانند قیمت باز شدن، بالاترین قیمت، پایین‌ترین قیمت و حجم معاملات استفاده شده است.. خوشه‌بندی (K-Means) : از این الگوریتم برای بررسی نوسانات قیمتی و گروه‌بندی سهام تحت تاثیر رفتار نوسانات، استفاده شد.

ویژگی‌های انتخاب شده

ویژگی‌های قیمت باز شدن (Open)، بالاترین قیمت (High)، پایین‌ترین قیمت (Low)، حجم معاملات (Volume)

به عنوان ورودی برای مدل‌ها انتخاب شدند.

پارامترهای مدل

برای هر مدل، پارامترهای زیر در نظر گرفته شدند:

رگرسیون خطی: این مدل از پارامترهای پیش‌فرض خود استفاده کرده و پارامتر قابل تنظیم خاصی نداشت.

K-Means: تعداد خوشه‌ها برابر با ۳ انتخاب شد، چون سه گروه متفاوت از سهام (پایدار، نوسان متوسط، نوسان بالا) برای ما مهم بودند.



روش ارزیابی

برای ارزیابی مدل‌ها، معیارهای زیر در نظر گرفته شد:

میانگین خطای مربعات (MSE): این معیار برای سنجش میزان خطای پیش‌بینی مدل رگرسیون خطی استفاده شد.

ضریب تعیین (R^2): این شاخص نشان‌دهنده میزان دقت مدل در توضیح تغییرات داده‌ها بود.

بصری‌سازی نتایج K-Means: نتایج خوشه‌بندی با نمودارهای پراکندگی و حرارتی بررسی شدند.

تشریح فایل سورس کد

تشریح کد

برای پیاده‌سازی مدل‌ها و انجام تحلیل‌ها، از زبان برنامه‌نویسی پایتون و کتابخانه‌های زیر استفاده شده است:

pandas	برای پردازش داده‌ها و مدیریت دیتافریم‌ها
Sklearn	برای مدل‌سازی، تقسیم داده‌ها، و ارزیابی مدل‌ها.
Seaborn و Matplotlib	برای بصری‌سازی داده‌ها و نتایج

کد به بخش‌های زیر تقسیم شده است:

وارد کردن کتابخانه‌ها: در ابتدای کد، کتابخانه‌های مورد نیاز برای تحلیل و مدل‌سازی وارد شدند.

بارگذاری داده‌ها: داده‌ها از فایل اکسل بارگذاری شده و ستون‌ها نام‌گذاری مجدد شدند.

پیش‌پردازش: برای تحلیل و مدل‌سازی داده‌ها اقدام شد که شامل مدیریت مقادیر گمشده، استانداردسازی و تبدیل تاریخ می باشد..

مدل‌سازی: مدل رگرسیون خطی برای پیش‌بینی قیمت بسته شدن سهام و الگوریتم K-Means برای خوشه‌بندی نوسانات پیاده‌سازی شدند.

بصری‌سازی و تحلیل: نتایج مدل‌ها با استفاده از سنج‌ها و نمودارها ارزیابی و تحلیل شدند.

نتایج

نتایج حاصل از مدل سازی و تحلیل داده ها به شرح زیر است:

۱. مدل رگرسیون خطی:

- میانگین خطای مربعات (MSE): مقدار ۹۹۷۷۳.۹۳، که نشان دهنده خطای به صورت نسبی کم مدل در پیش بینی قیمت بسته شدن سهام است.

- ضریب تعیین (R^2): مقدار ۰.۹۹۷۶، که بیانگر دقت خوب مدل در توضیح تغییرات داده ها است.

۲. خوشه بندی K-Means: سه گروه از سهام مشخص شدند:

- گروه پایدار: شامل سهام هایی که نوسانات کم دارند.

- گروه نوسان متوسط: شامل سهام هایی که تغییرات قابل پیش بینی دارند.

- گروه نوسان بالا: شامل سهام هایی که نوسانات زیاد دارند.

تفسیر نتایج

۱. عملکرد مدل رگرسیون خطی: مدل رگرسیون خطی با دقت خوبی، پیش بینی های مناسبی کرده است. این عملکرد به دلیل ارتباط میان متغیرهای ورودی و متغیر هدف (قیمت بسته شدن) می باشد.

۲. عملکرد الگوریتم K-Means: خوشه بندی سهام نشان داد که نوسانات قیمتی به سه گروه تقسیم می شوند. این اطلاعات برای تصمیم گیری در مورد سرمایه گذاری بسیار کار آیی دارد.

۳. نمودارهای روند قیمت و نوسانات: نمودارهای میانگین متحرک ۳۰ و ۹۰ روزه نشان می دهند که روند کلی قیمت ها در بازه های زمانی کوتاه مدت و بلندمدت چگونه است. نمودار رابطه حجم معاملات و قیمت بسته شدن، وجود همبستگی و ارتباط و تاثیر بین این دو متغیر را تأیید کرد.

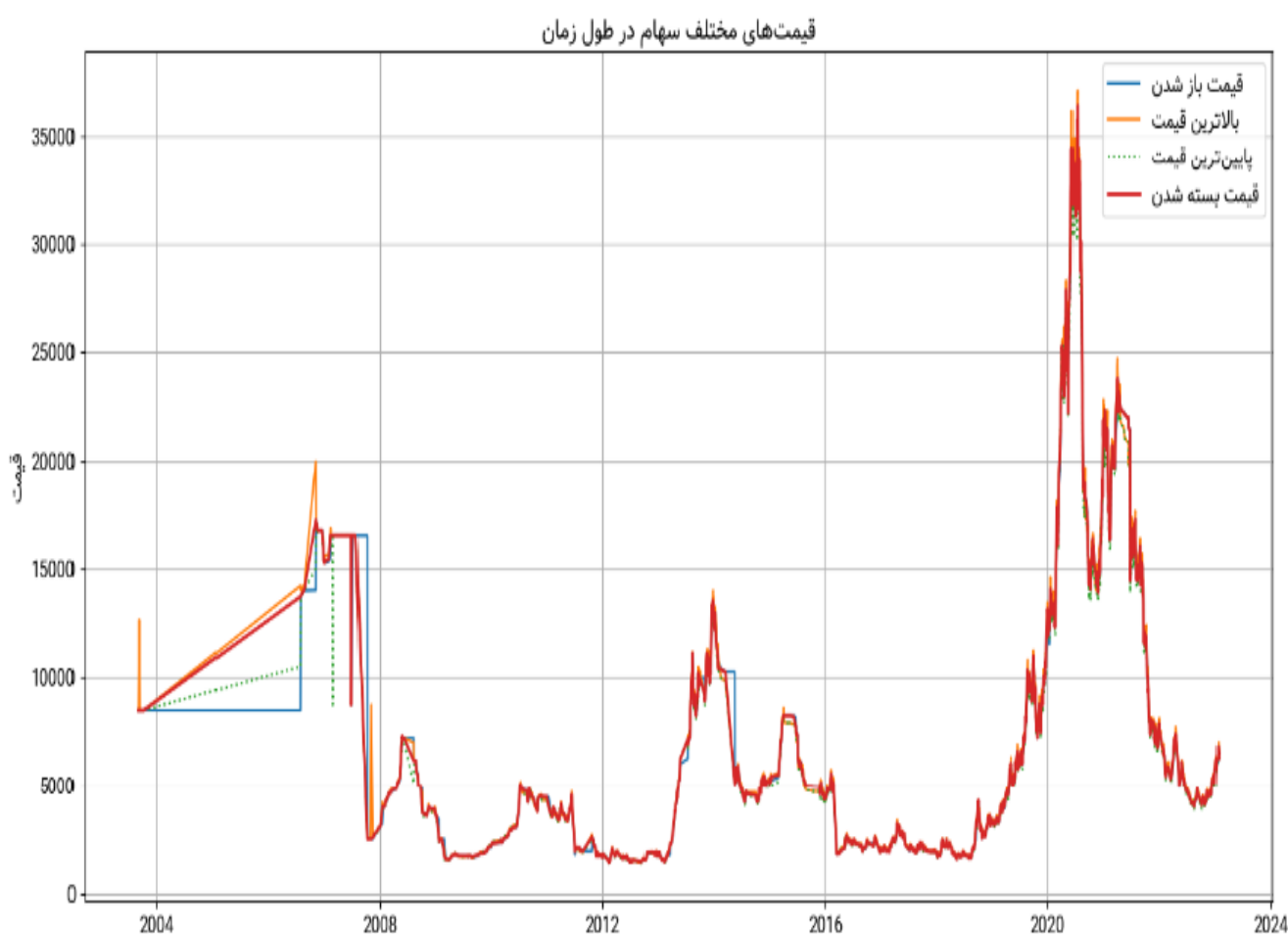
این تحقیق سعی دارد با استفاده از مدل رگرسیون خطی، قیمت بسته شدن سهام را بر اساس ویژگی هایی مانند قیمت باز، بالا، پایین و حجم معاملات پیش بینی کند. سپس مدل روی داده های آموزشی تنظیم شده و برای مجموعه تست پیش بینی صورت می گیرد. در نهایت، عملکرد مدل با معیارهایی مانند میانگین خطای مربعات (MSE) و ضریب تعیین (R^2) ارزیابی شده و مدل برای استفاده های بعدی ذخیره می شود. نتایج حاصل به شرح زیر استخراج می شوند:

(MSE) میانگین خطای مربعات	99773.93233232641
(R^2) ضریب تعیین	0.9975912751530526

جداول، شکل ها و نمودارها

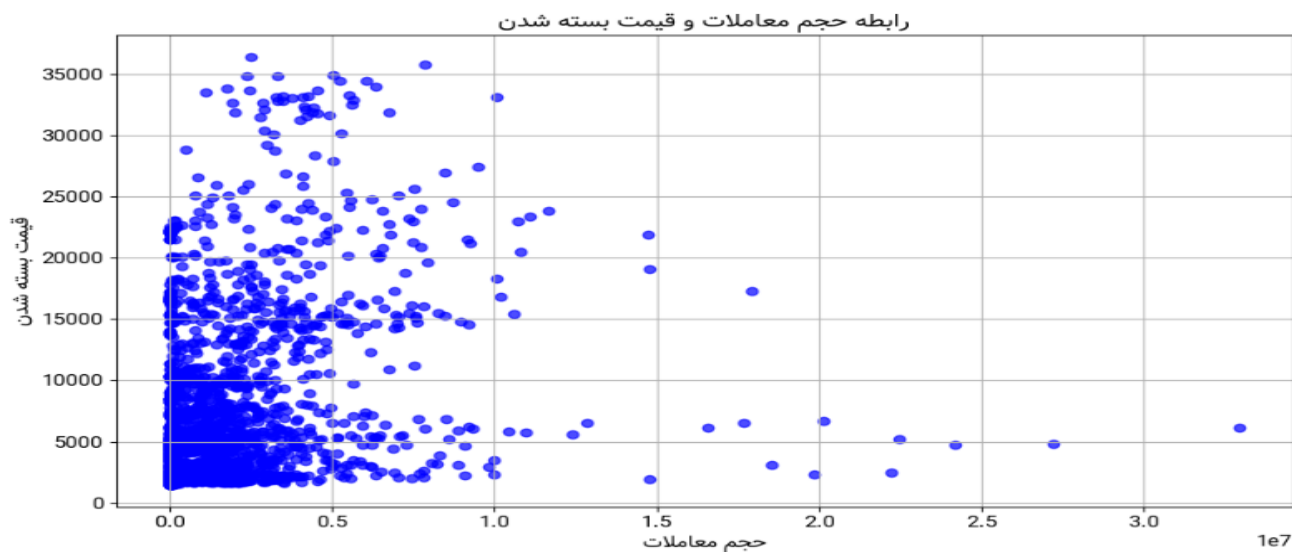
در ادامه نمودار تغییرات قیمت سهام در طول زمان را بررسی و رسم می‌کند. که در این نمودار، قیمت‌های باز، بالا، پایین و بسته شدن به صورت خطوط جداگانه و رنگی نمایش داده می‌شوند تا روند کلی بازار مشخص شود. محور افقی نشان‌دهنده تاریخ و محور عمودی نشان‌دهنده قیمت‌ها است. با استفاده از راهنما (Legend) و شبکه‌بندی (Grid)، نوشته زیر نمودار تنظیم می‌شود.

بر اساس خروجی که از اجرای کدها می‌گیریم نمودارهای زیر حاصل شد:



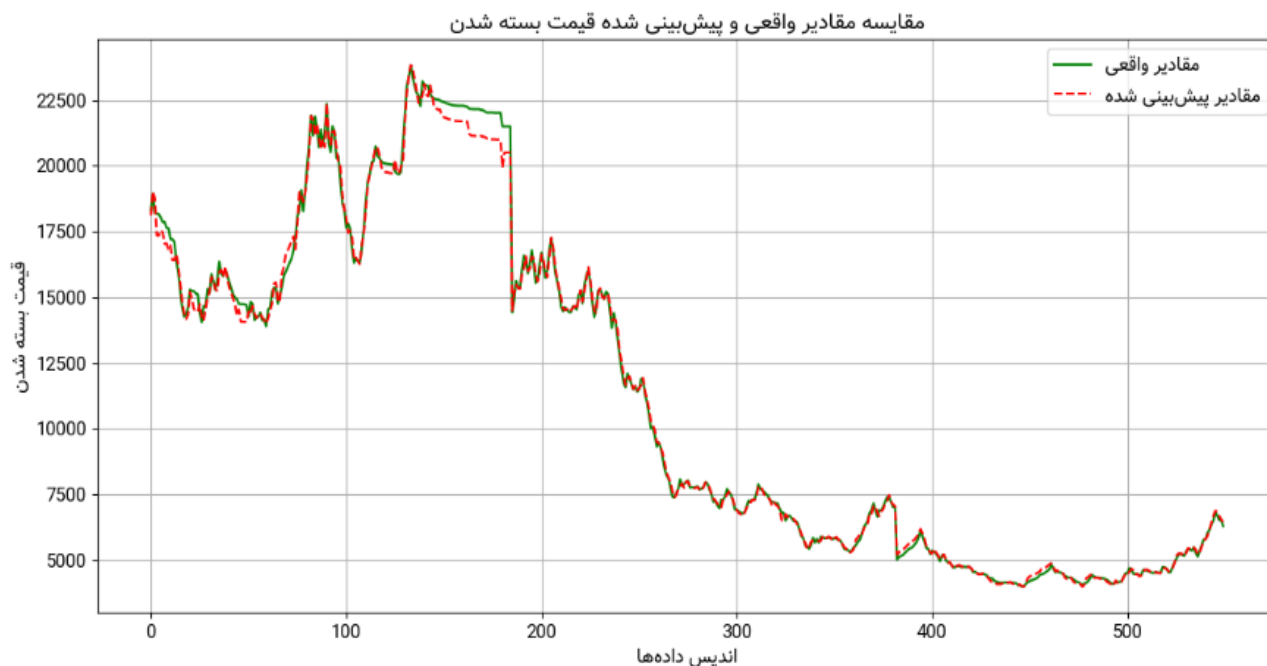
این نمودار روند قیمت‌ها (Open, High, Low, Close) را در طول زمان نشان می‌دهد.

در ادامه نمودار پراکندگی (Scatter Plot) را بررسی می‌کند و رابطه بین حجم معاملات و قیمت بسته شدن را رسم می‌کند. هر نقطه نشان‌دهنده یک نمونه از داده‌ها است. این نمودار به تحلیل همبستگی بین این دو متغیر کمک می‌کند و نشان می‌دهد که آیا حجم معاملات تأثیری بر قیمت بسته شدن دارد یا ندارد.



این نمودار نشان دهنده رابطه حجم معاملات با قیمت بسته شدن می باشد.

در بخش بعدی نموداری برای مقایسه مقادیر واقعی و پیش بینی شده قیمت بسته شدن سهام را بررسی و رسم می کند. مقادیر واقعی با خط سبز و مقادیر پیش بینی شده با خط قرمز نقطه چین نمایش داده می شوند. این مقایسه نشان می دهد که مدل رگرسیون خطی تا چه حد توانسته روند قیمت ها را پیش بینی کند.



این نمودار مقایسه ای از مقادیر واقعی و پیش بینی شده قیمت بسته شدن ارائه می دهد.

جدول آماری:

در بخش بعدی خروجی ها جدول مقادیر آماری داده ها را نمایش می دهیم. این جدول شامل اطلاعاتی مانند میانگین، انحراف معیار، کمینه، بیشینه و بخش های مختلف برای هر متغیر عددی در مجموعه داده است. این اطلاعات به تحلیل توزیع داده ها و شناسایی الگوهای کلی در ویژگی های سهم کمک می کند.

جدول مقادیر آماری داده ها

	Date	Open	High
count	2746	2746.000000	2746.000000
mean	2016-03-01 15:46:32	6719.612163	6838.208303
min	2003-09-02 00:00:00	1433.000000	1411.000000
25%	2012-06-25 06:00:00	2195.250000	2249.250000
۵۰%	2016-12-29 12:00:00	4407.000000	4471.000000
۷۵%	2020-01-04 18:00:00	8201.750000	8264.750000
max	2023-01-25 00:00:00	36410.000000	37100.000000
std	NaN	6470.364580	6596.702854

جدول مقادیر آماری داده ها

	Low	Close	Volume
count	2746.000000	2746.000000	2.746000e+03
mean	6579.183904	6725.935543	1.329340e+06
min	1407.000000	1433.000000	1.000000e+00
25%	2165.000000	2206.500000	1.960350e+04
۵۰%	4320.000000	4410.500000	3.270620e+05
۷۵%	7924.000000	8184.000000	1.791527e+06
max	35300.000000	36410.000000	3.293313e+07
std	6311.642342	6479.142907	2.336565e+06

جدول مقایسه:

در بخش بعدی جدولی برای مقایسه مقادیر واقعی و مقادیر پیش بینی شده قیمت بسته شدن ایجاد کردیم. در این جدول، مقدار واقعی و مقدار محاسبه شده توسط مدل رگرسیون خطی در کنار هم قرار می گیرند. نمایش چند سطر ابتدایی این جدول به تحلیل میزان دقت مدل در پیش بینی کمک کرده و امکان بررسی اختلافات بین مقدار واقعی و مقدار پیش بینی شده را فراهم می کند.

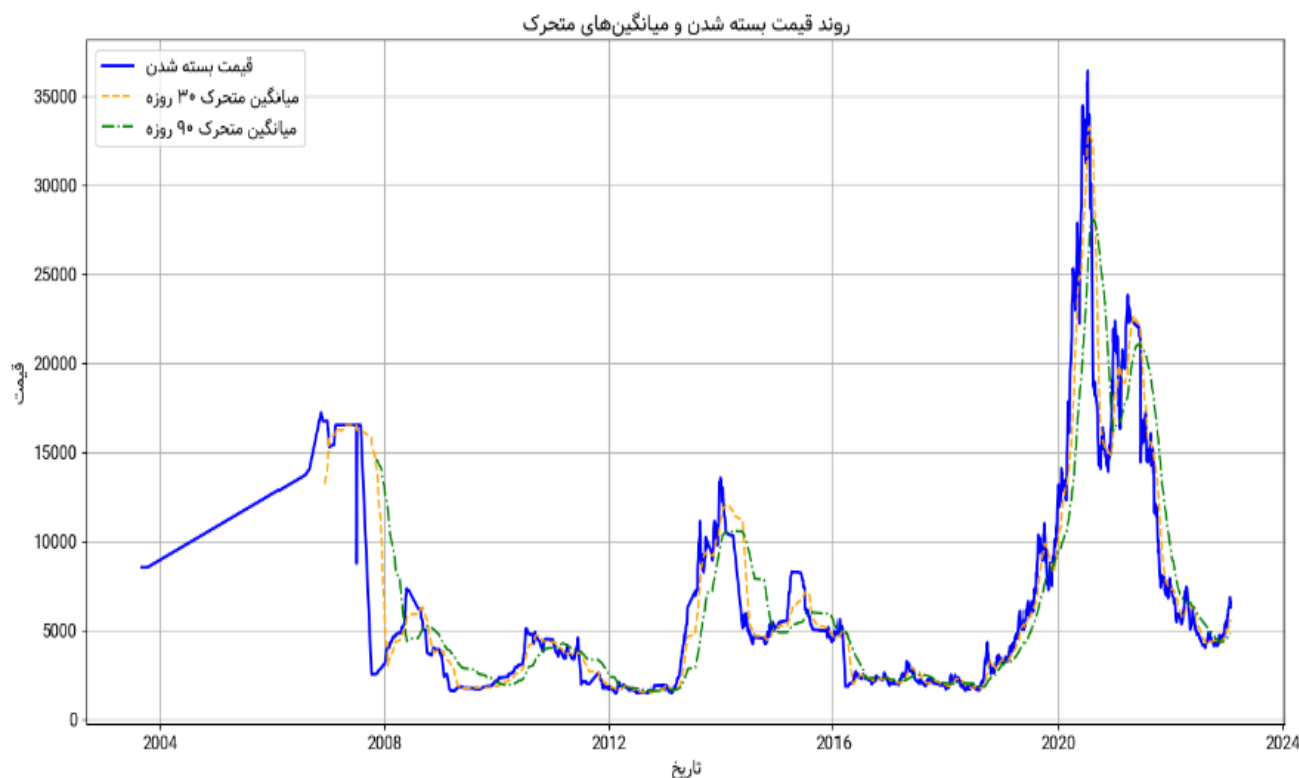
	واقعی	پیش بینی شده
۰	18290	18108.148263
۱	18940	19019.709225
۲	18200	18763.916429

۳	18180	17357.095751
۴	18160	17346.656215

میانگین متحرک (Moving Average): با استفاده از میانگین متحرک ۳۰ روزه و ۹۰ روزه مقایسه روند کوتاه مدت و بلندمدت قیمت بسته شدن محاسبه می شوند. این میانگین کمک می کند که ما روند کلی داده ها را بدون تأثیر نوسانات کوچک بررسی کنیم.

روند حجم معاملات: نمودار حجم معاملات در طول زمان نشان دهنده فعالیت سرمایه گذاران در بازار است. نوسانات روزانه (Volatility): تغییرات روزانه قیمت بسته شدن به درصد محاسبه شده و نمودار نوسانات رسم می شود. که نشان دهنده میزان پایداری یا بی ثباتی قیمت هاست.

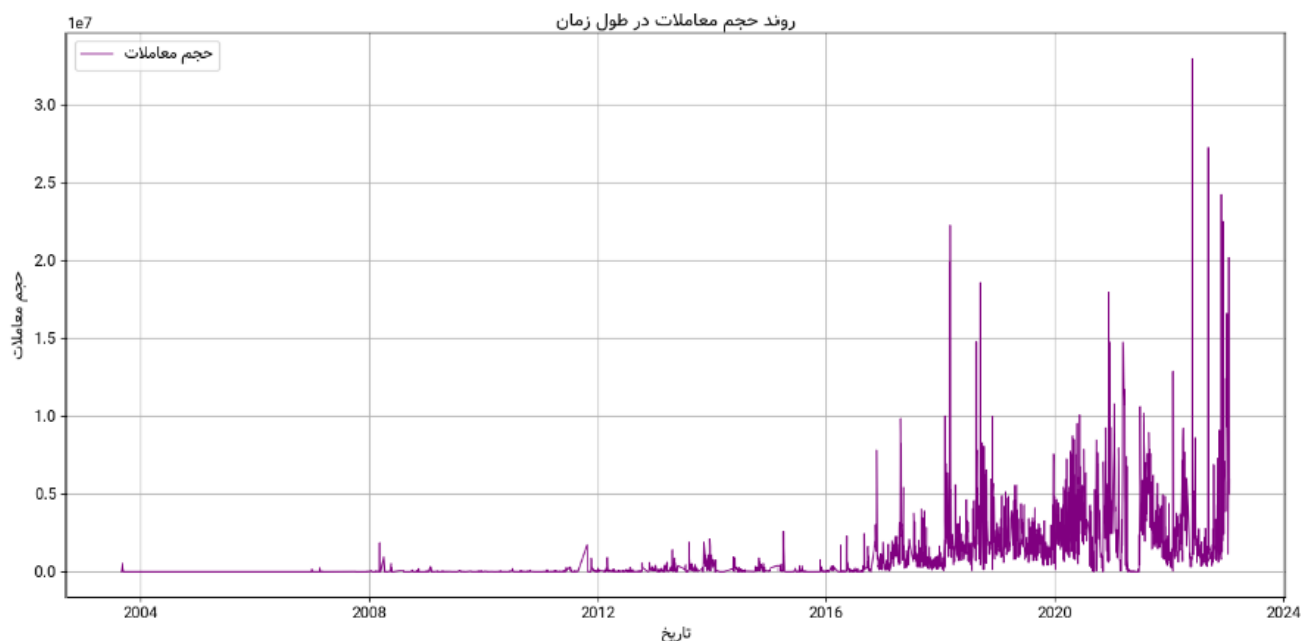
در بخش بعدی ما میانگین های متحرک ۳۰ و ۹۰ روزه را برای قیمت بسته شدن سهام محاسبه و رسم می کنیم. همانطور که گفته شد میانگین های متحرک به عنوان شاخص های تحلیل تکنیکال، روند کلی بازار را نشان داده و به شناسایی نقاط ضعف و قوت می پردازد. در نمودار، قیمت بسته شدن با خط آبی، میانگین متحرک ۳۰ روزه با خط نارنجی نقطه چین و میانگین متحرک ۹۰ روزه با خط سبز خط چین نمایش داده می شود.



نمودار روند قیمت بسته شدن: نمایش قیمت واقعی همراه با میانگین های متحرک ۳۰ و ۹۰ روزه

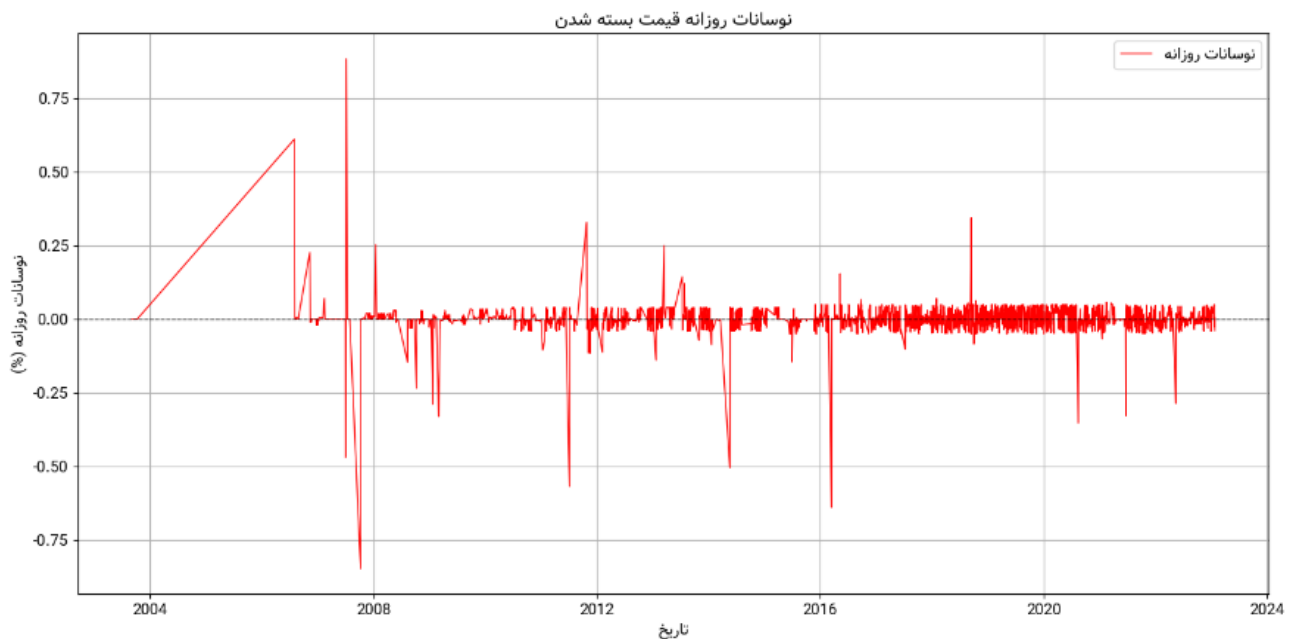
در ادامه کار روند حجم معاملات در طول زمان را بررسی و نمودار آن را رسم می کنیم. این نمودار تغییرات حجم معاملات را نمایش می دهد و به تحلیل رفتار بازار در بازه های زمانی مختلف کمک می کند. بررسی این روند می تواند نشان دهد که آیا

افزایش یا کاهش حجم معاملات با تغییرات قیمت ارتباط دارد یا خیر. در این نمودار، حجم معاملات با خط بنفش نمایش داده شده است.



نمودار حجم معاملات نمایش تغییرات حجم معاملات در بازه زمانی را نشان می دهد.

در ادامه به بررسی درصد تغییرات روزانه قیمت بسته شدن سهام پرداخته و آن را محاسبه کرده و نمودار آن را رسم می کنیم. این تغییرات نشان دهنده نوسانات روزانه بازار است و می تواند برای تحلیل میزان پایداری یا بی ثباتی قیمت ها استفاده شود. در نمودار، خط سیاه نقطه چین مقدار صفر را نشان می دهد که از آن برای مقایسه تغییرات مثبت و منفی استفاده می شود.



نمودار نوسانات قیمت به بررسی میزان تغییرات روزانه قیمت بسته شدن می پردازد.

در ادامه به شناسایی همبستگی بین حجم معاملات، ارزش معاملات و قیمت بسته شدن پرداخته ایم.

ماتریس همبستگی: نشان می دهد آیا تغییرات حجم معاملات و ارزش معاملات با قیمت بسته شدن همبستگی دارد یا خیر.

نمودار پراکندگی: مشخص می کند آیا افزایش حجم معاملات به تغییر قیمت بسته شدن منجر می شود یا خیر.

اگر ضریب همبستگی بین **Volume** و **Close** مثبت باشد، افزایش حجم معاملات با افزایش قیمت همراه خواهد بود.

اگر ضریب منفی باشد، افزایش حجم معاملات با کاهش قیمت همراه است.

در ادامه کار ابتدا ماتریس همبستگی بین قیمت بسته شدن و حجم معاملات را محاسبه می کند. همچنین، ارزش معاملات (حجم \times قیمت بسته شدن) به عنوان یک ویژگی جدید اضافه شده و همبستگی بین آن ها محاسبه می شود. سپس ماتریس همبستگی به صورت جدول نمایش داده می شود. وبعد از آن ، نمودار حرارتی (**Heatmap**) این ماتریس با استفاده از رنگ های مختلف برای نمایش شدت همبستگی رسم می شود.

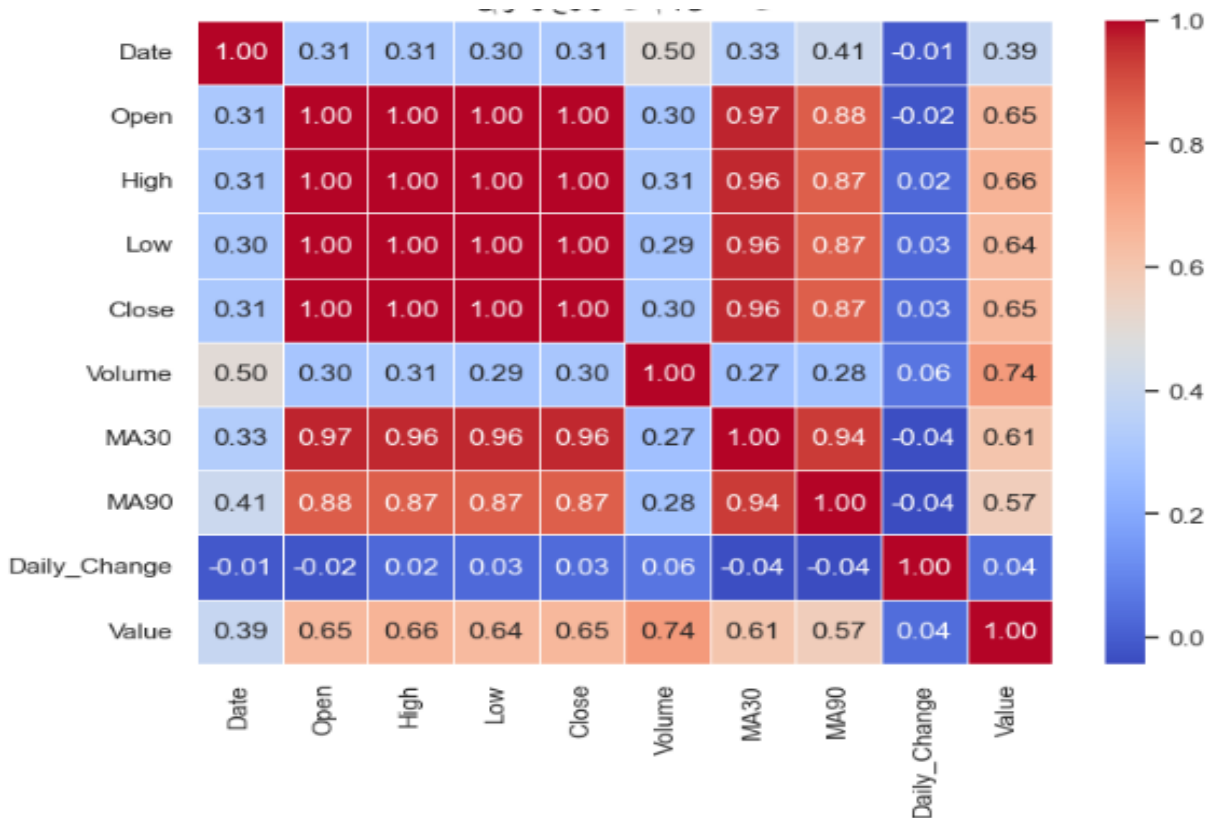


ماتریس همبستگی بین متغیرها

	Date	Open	High	Low	Close	Volume
Date	1.000000	0.308160	0.310960	0.303360	0.308451	0.496705
Open	0.308160	1.000000	0.996991	0.996854	0.997250	0.296133
High	0.310960	0.996991	1.000000	0.998054	0.998603	0.308781
Low	0.303360	0.996854	0.998054	1.000000	0.999128	0.294154
Close	0.308451	0.997250	0.998603	0.999128	1.000000	0.299257
Volume	0.496705	0.296133	0.308781	0.294154	0.299257	1.000000
MA30	0.333502	0.967859	0.964049	0.963444	0.964932	0.274878
MA90	0.411579	0.875806	0.870766	0.870771	0.872835	0.283548
Daily_Change	-0.012811	-0.022562	0.024105	0.029126	0.034147	0.057188
Value	0.393873	0.646398	0.662404	0.644439	0.650025	0.735769

	MA30	MA90	Daily_Change	Value
Date	0.333502	0.411579	-0.012811	0.393873
Open	0.967859	0.875806	-0.022562	0.646398
High	0.964049	0.870766	0.024105	0.662404
Low	0.963444	0.870771	0.029126	0.644439
Close	0.964932	0.872835	0.034147	0.650025
Volume	0.274878	0.283548	0.057188	0.735769
MA30	1.000000	0.938678	-0.036428	0.606786
MA90	0.938678	1.000000	-0.042982	0.570386
Daily_Change	-0.036428	-0.042982	1.000000	0.035781
Value	0.606786	0.570386	0.035781	1.000000

نمودار حرارتی همبستگی



این نمودار به شناسایی روابط خطی میان متغیرها کمک می‌کند.

خوشه‌بندی سهام بر اساس نوسانات قیمتی

برای خوشه‌بندی سهام، از الگوریتم‌های یادگیری ماشین بدون نظارت مانند (K-Means) استفاده می‌کنیم.

محاسبه نوسانات: از انحراف معیار قیمت بسته شدن در بازه ۳۰ روزه برای هر سهم استفاده شده است.

استانداردسازی داده‌ها: برای بهبود عملکرد الگوریتم خوشه‌بندی، مقادیر نوسانات استانداردسازی شده‌اند.

K-Means: به سه خوشه تقسیم شده‌اند:

خوشه ۰: سهام پایدار (نوسانات کم).

خوشه ۱: سهام با نوسانات متوسط.

خوشه ۲: سهام پرنوسان.

تحلیل:

سهام در خوشه پایدار (Cluster 0) برای سرمایه‌گذاران ریسک‌پذیر جذاب نیست.

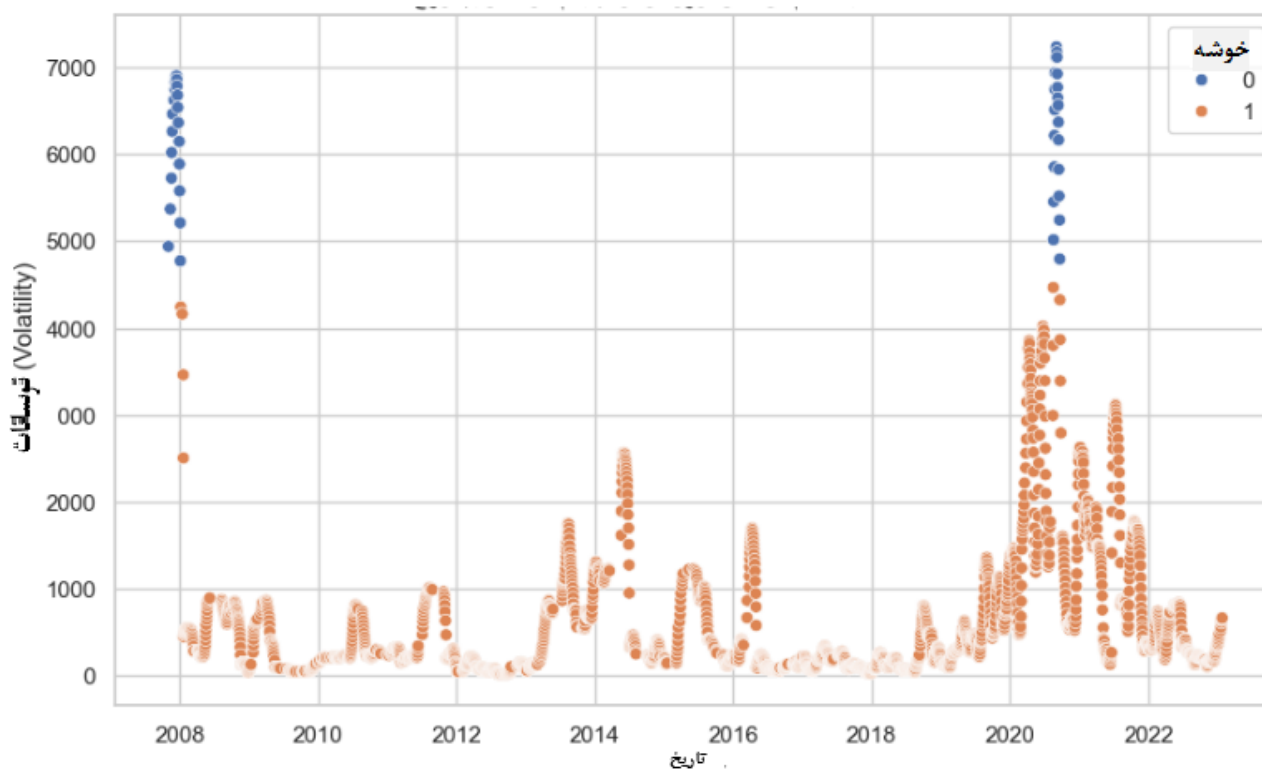
سهام در خوشه متوسط (Cluster 1) تعادلی از ریسک و بازده دارند.

سهام در خوشه پرنوسان (Cluster 2) ممکن است برای سرمایه‌گذاران ریسک‌پذیر مناسب باشد.

جدول خوشه‌بندی با DBSCAN: مشخص می‌کند هر سهم در کدام خوشه قرار دارد.

	Date	Close	Volatility	Cluster_DBSCAN
89	2007-11-04	2514	4940.008712	0
90	2007-11-14	2564	5369.934736	0
91	2007-11-20	2615	5725.736257	0
92	2007-11-21	2667	6020.930757	0
93	2007-11-24	2720	6264.449248	0

نمودار پراکندگی قوسان ها با رنگ‌بندی خوشه ها



نمودار پراکندگی نمایش تصویری خوشه‌های سهام بر اساس نوسانات قیمتی می‌باشد.

محاسبه میانگین قیمت بسته شدن در شروع و پایان هفته:



در بخش بعدی میانگین قیمت بسته شدن را در سه بازه زمانی مختلف محاسبه می کند:

۱. شروع هفته (دوشنبه): میانگین قیمت بسته شدن در روزهای شروع هفته.

۲. پایان هفته (یکشنبه): میانگین قیمت بسته شدن در روزهای پایان هفته.

۳. سایر روزها: میانگین قیمت بسته شدن در روزهای میانه هفته (غیر از دوشنبه و یکشنبه).

این مقایسه می تواند کمک کند تا الگوهای احتمالی در نوسانات قیمت در روزهای مختلف هفته شناسایی شود.

میانگین قیمت بسته شدن در شروع هفته: 6682.275362318841

میانگین قیمت بسته شدن در پایان هفته: 6727.794495412844

میانگین قیمت بسته شدن در سایر روزها: 6739.936325045482

محاسبه میانگین قیمت بسته شدن در شروع و پایان ماه:

به مانند بخش قبلی، این بار برای شروع و پایان ماه، میانگین قیمت بسته شدن و حجم معاملات را محاسبه می کند:

۱. میانگین قیمت بسته شدن در شروع ماه: محاسبه قیمت بسته شدن در روزهای اول ماه.

۲. میانگین قیمت بسته شدن در پایان ماه: محاسبه قیمت بسته شدن در روزهای آخر ماه.

۳. میانگین قیمت بسته شدن در سایر روزها: محاسبه قیمت بسته شدن در روزهای میانه ماه.

همچنین میانگین حجم معاملات برای همین بازه های زمانی محاسبه می شود که می تواند کمک کند تا روند حجم معاملات در ابتدا، انتها و وسط ماه را تحلیل کنید.

میانگین قیمت بسته شدن در شروع ماه: 7562.283950617284

میانگین قیمت بسته شدن در پایان ماه: 6598.066666666667

میانگین قیمت بسته شدن در سایر روزها: 6704.096310679612

میانگین حجم معاملات در شروع ماه: 1206774.2839506173

میانگین حجم معاملات در پایان ماه: 1725839.7

میانگین حجم معاملات در سایر روزها: 1319336.9250485436

پیش پردازش داده ها

و در پایان و در بخشی از پیش پردازش داده ها، تاریخ ها به فرمت مناسب تبدیل شده اند. سپس نوسانات هر روز از طریق اختلاف بین قیمت های بالاترین و پایین ترین محاسبه می شود. بر اساس این نوسانات، یک ویژگی به نام نوسانات بالاترین شده که نشان دهنده روزهایی با نوسان بیش از ۲٪ است. همچنین، میانگین های متحرک ۳۰ روزه و ۹۰ روزه برای قیمت بسته شدن محاسبه شده اند که برای تحلیل روندهای بلندمدت مفید است. در بخش دیگری، داده ها به دو مجموعه آموزشی و تست تقسیم شده اند. سپس مدل **Classifier Forest Random** آموزش داده شده و برای پیش بینی نوسانات بالا در مجموعه تست استفاده شده است. ارزیابی مدل با استفاده از گزارش طبقه بندی (**Classification Report**) و ماتریس (**Confusion Matrix**) انجام می شود تا عملکرد آن بررسی شود. همچنین، اهمیت ویژگی ها با استفاده از ویژگی های مدل استخراج شده و در یک نمودار نمایش داده می شود تا تاثیرگذاری هر ویژگی در پیش بینی ها دیده شود.

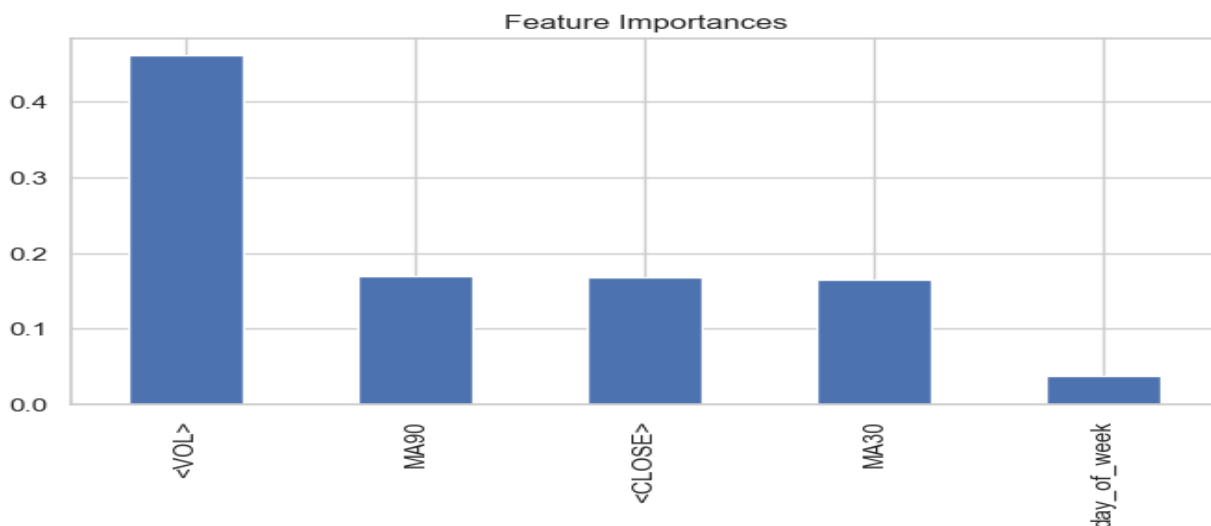
Classification Report:

	precision	recall	f1-score	support
0	0.85	0.77	0.81	285
1	0.88	0.92	0.90	513
accuracy			0.87	798
macro avg	0.86	0.85	0.85	798
weighted avg	0.87	0.87	0.87	798

Confusion Matrix:

```
[[219 66]
 [ 40 473]]
```

نمودار اهمیت ویژگی ها



پیشنهادهای برای بهبود نتایج



۱. استفاده از مدل‌های پیشرفته‌تر مانند شبکه‌های عصبی عمیق برای پیش‌بینی دقیق‌تر.

۲. افزودن داده‌های خارجی مانند گزارش‌های اطلاعاتی مالی و اقتصادی به مدل برای افزایش دقت.

۳. بررسی در مورد حساسیت و تأثیر ویژگی‌های مختلف بر عملکرد مدل.

منابع

دیتاست مورد استفاده در این مقاله مربوط به بورس اوراق بهادار تهران بوده و از طریق لینک زیر قابل دسترسی است:

[Tehran Stock Exchange Dataset on Kaggle]

(<https://www.kaggle.com/datasets/mehradaria/tehran-stock-exchange/code>)

این دیتاست توسط Mehrad Aria در سال ۲۰۲۰ منتشر شده و دارای DOI زیر است:

Mehrad Aria, "Tehran Stock Exchange Market Data." Kaggle, 2020,

DOI: 10.34740/kaggle/ds/652083.

Patel, J., Shah, S., Thakkar, P., & Kotecha, K. (2015). Predicting stock and stock price index movement using trend deterministic data preparation and machine learning techniques. *Expert Systems with Applications*, 42(1), 259-268.

حسینی، م.، رضایی، الف. (۱۴۰۰). پیش‌بینی قیمت سهام با استفاده از الگوریتم‌های رگرسیون خطی و جنگل تصادفی. کنفرانس بین‌المللی مدیریت و مهندسی صنایع، تهران، ایران.

کریمی، س.، محمدی، ن. (۱۳۹۹). پیش‌بینی روند قیمت در بازار سهام با استفاده از الگوریتم جنگل تصادفی. *فصلنامه پژوهش‌های مهندسی مالی و مدیریت اوراق بهادار*، دوره ۱۸، شماره ۲، صفحات ۴۵-۶۰.

شریفی، ب.، احمدی، ی. (۱۳۹۸). پیش‌بینی جهت حرکت قیمت سهام شرکت‌های بیمه بورسی با استفاده از الگوریتم‌های رده‌بندی. *مجله علمی پژوهشی اقتصاد و مدیریت مالی*، دوره ۲۲، شماره ۳، صفحات ۷۵-۹۰.