



تدوین استراتژی معاملاتی بر اساس خوشه‌بندی داده‌های قیمت با الگوریتم یادگیری ماشین DBSCAN و یافتن داده‌های نويز

نسترن شریفی‌راد^۱

محمد رضا رستمی^۲

تاریخ دریافت مقاله: ۱۴۰۳/۰۹/۲۱ تاریخ پذیرش مقاله: ۱۴۰۴/۰۱/۱۴

چکیده

در این پژوهش، یک استراتژی معاملاتی مبتنی بر الگوریتم خوشه‌بندی DBSCAN و تحلیل داده‌های پرت ارائه شده است. هدف، شناسایی نقاط ورود و خروج بهینه در بازارهای مالی با استفاده از یادگیری ماشین و تحلیل تکنیکال است. داده‌های این پژوهش شامل اطلاعات قیمتی و حجمی ۵۰ سهم پرتراکنش بورس و فرابورس تهران در بازه ۱۳۹۸/۰۹/۱۵ تا ۱۴۰۳/۰۹/۱۵ است که از پایگاه TSETMC استخراج شده‌اند. پس از پیش‌پردازش داده‌ها، استانداردسازی و مهندسی ویژگی، الگوریتم DBSCAN برای خوشه‌بندی داده‌ها و شناسایی نقاط پرت استفاده شد. سپس، سیگنال‌های استخراج شده از DBSCAN با اندیکاتور MACD ترکیب شدند تا نقاط ورود و خروج بهینه معاملات مشخص شوند. به منظور ارزیابی عملکرد این استراتژی، نتایج آن با روش تحلیل انحراف معیار گاوسی و استراتژی خرید و نگهداری مقایسه شده است. همچنین، تأثیر پارامترهای DBSCAN بر دقت خوشه‌بندی بررسی شده و بهترین مقادیر برای تشخیص سیگنال‌های معاملاتی تعیین گردید. تحلیل‌ها در محیط Python با استفاده از کتابخانه‌های Algotik، NumPy، Pandas، Scikit-learn و Matplotlib انجام شده و نتایج با روش‌های آماری ارزیابی شده است. یافته‌ها نشان می‌دهد که استراتژی پیشنهادی در بسیاری از نمادهای بررسی شده بازدهی بالاتری نسبت به روش‌های سنتی داشته و توانسته است سیگنال‌های معاملاتی دقیق‌تری ارائه دهد. علاوه بر این، این رویکرد می‌تواند به بهبود دقت معاملات، کاهش سیگنال‌های کاذب و بهینه‌سازی تصمیم‌گیری‌های سرمایه‌گذاری منجر شود. این روش قابلیت گسترش به بازارهای دیگر را داشته و می‌تواند در توسعه معاملات الگوریتمی، سیستم‌های خودکار و تحلیل روندهای قیمتی بلندمدت نیز مورد استفاده قرار گیرد. یافته‌های این پژوهش نشان می‌دهد که ترکیب الگوریتم‌های یادگیری ماشین با تحلیل تکنیکال می‌تواند یک رویکرد نوین و کارآمد برای بهینه‌سازی استراتژی‌های معاملاتی فراهم کند.

کلمات کلیدی:

استراتژی معاملاتی، یادگیری ماشین، خوشه‌بندی، داده‌های پرت، تحلیل تکنیکال

۱- دانشجوی کارشناسی ارشد، گروه مدیریت، دانشکده علوم اجتماعی و اقتصاد، دانشگاه الزهراء، تهران، ایران. N.sharifirad@student.alzahra.ac.ir

۲- دانشیار، گروه مدیریت، دانشکده علوم اجتماعی و اقتصاد، دانشگاه الزهراء، تهران، ایران. (نویسنده مسئول) M.rostami@alzahra.ac.ir

مقدمه

بازار سهام نوسانات پویایی دارد که پیش‌بینی آن را دشوار می‌کند. طبق نظریه بازار کارآمد، قیمت‌ها بازتاب اطلاعات موجود هستند، اما شواهد نشان می‌دهد که برخی بازارها از این حالت فاصله دارند و الگوهایی در داده‌ها قابل‌شناسایی است. نویزهای ناشی از شایعات، اخبار و تصمیمات سرمایه‌گذاران، همراه با معاملات داخلی و دست‌کاری بازار، بر قیمت‌ها اثر گذاشته و دقت مدل‌های سنتی را کاهش می‌دهند. (نیپان^۱ و همکاران، ۲۰۲۱). حجم معاملات معیاری کلیدی در تحلیل بازار است که انتظارات سرمایه‌گذاران و تغییرات قیمت را نشان می‌دهد. پژوهش‌ها درباره‌ی ارتباط آن با بازده نتایج متفاوتی ارائه کرده‌اند؛ برخی این رابطه را مثبت و معنادار دانسته‌اند، درحالی‌که برخی دیگر به ارتباط منفی یا نامعنا اشاره کرده‌اند. این تفاوت‌ها به عواملی مانند نوع بازار، فرکانس داده‌ها و روش‌های آماری بستگی دارد. (نیپان و همکاران، ۲۰۲۱). علاوه بر این، حجم معاملات می‌تواند بر گشتاورهای مرتبه‌ی بالاتر بازده سهام، مانند چولگی^۲ و کشیدگی^۳، توزیع بازده نیز تأثیر بگذارد که در ارزیابی ریسک سرمایه‌گذاری اهمیت دارد (احدزی و جیاسریدهران^۴، ۲۰۲۰). داده‌های پرت^۵ که با الگوی کلی داده‌ها سازگار نیستند، می‌توانند ناشی از خطاهای اندازه‌گیری یا رویدادهای غیرمنتظره باشند. این داده‌ها گاهی اطلاعات ارزشمندی در مورد پویایی بازار ارائه می‌دهند اما درعین‌حال ممکن است منجر به تخمین‌های نادرست در مدل‌سازی مالی شوند (شهاده^۶ و همکاران، ۲۰۲۲). بررسی این داده‌ها در تحلیل‌های مالی اهمیت زیادی دارد، زیرا تأثیر آن‌ها بر مدیریت ریسک، تخصیص دارایی‌ها و مدل‌سازی اقتصادی قابل‌توجه است. در بازارهای مالی، سقوط‌های ناگهانی یکی از مهم‌ترین نمونه‌های داده‌های پرت هستند که شناخت آن‌ها برای بهبود پیش‌بینی‌های مالی و تصمیمات سرمایه‌گذاری ضروری است (شهاده و همکاران، ۲۰۲۲). در بازارهای مالی، وجود داده‌های پرت می‌تواند تأثیرات عمیقی بر تحلیل‌ها و مدل‌سازی‌ها داشته باشد. این داده‌های غیرعادی که ممکن است ناشی از خطاهای اندازه‌گیری یا رویدادهای غیرمنتظره باشند، می‌توانند باعث انحراف در نتایج و استنتاج‌های آماری شوند. یکی از حوزه‌هایی که داده‌های پرت در آن اهمیت ویژه‌ای دارند، تحلیل سری‌های زمانی مالی و اقتصادی است. این سری‌های زمانی اغلب دارای توزیع‌هایی با دم‌های چاق هستند که احتمال وقوع مقادیر پرت در آن‌ها بالاتر است (گوانز - سانچز^۷، ۲۰۲۱). یکی از نمونه‌های بارز داده‌های پرت در بازارهای مالی، سقوط‌های ناگهانی بازار است که می‌توانند تأثیرات شدیدی بر سرمایه‌گذاران و کل اقتصاد داشته باشند (شلتز و ترد^۸، ۲۰۰۸). یکی از زمینه‌های مهم در تحلیل داده‌های مالی، استفاده از الگوریتم‌های خوشه‌بندی برای شناسایی الگوهای معاملاتی و داده‌های پرت است. در مطالعه‌ای که

استراتژی معاملاتی بر اساس خوشه‌بندی داده‌های قیمت با الگوریتم... / شریفی‌راد و رستمی

توسط (صادقی-دهقان، ۱۳۹۵) ارائه شده است، روش‌های مختلف خوشه‌بندی مانند الگوریتم خوشه‌بندی میانگین^۹ و الگوریتم‌های چگالی‌محور برای تحلیل داده‌های مالی به کار گرفته شده‌اند. نتایج این پژوهش نشان داد که خوشه‌بندی می‌تواند در شناسایی الگوهای معاملاتی و خوشه‌بندی سهام مفید باشد.

در این پژوهش، ما نیز از داده‌های بازار سهام استفاده خواهیم کرد و با بهره‌گیری از یک الگوریتم خوشه‌بندی یادگیری ماشین^{۱۰} به عنوان یک روش خوشه‌بندی چگالی‌محور^{۱۱} برای شناسایی نقاط پرت و بهینه‌سازی استراتژی‌های معاملاتی و همچنین یک روش‌های آماری و تکنیک‌های داده‌کاوی، به شناسایی و تحلیل داده‌های پرت در حجم معاملات خواهیم پرداخت. این پژوهش به بررسی ارتباط داده‌های پرت و تغییرات قیمت سهام می‌پردازد تا میزان قابلیت پیش‌بینی روند بازار با استفاده از این اطلاعات ارزیابی شود. به جای تمرکز صرف بر داده‌های قیمت، بر ناهنجاری‌های حجم معاملات تمرکز شده است که می‌توانند نشانه‌ای از معاملات داخلی و دست‌کاری بازار باشند. تحلیل این ناهنجاری‌ها به کاهش تأثیرنویزهای قیمتی و ارائه پیش‌بینی دقیق‌تر از روند سهام کمک می‌کند. یافته‌های این مطالعه می‌تواند برای سرمایه‌گذاران، تحلیلگران و مدیران ریسک ارزشمند باشد و به توسعه مدل‌های پیش‌بینی و بهینه‌سازی استراتژی‌های معاملاتی کمک کند. همچنین، مرور پژوهش‌های پیشین در این حوزه، چارچوب نظری لازم برای استفاده از الگوریتم‌های خوشه‌بندی در شناسایی داده‌های پرت را فراهم می‌کند.

در پژوهش نخست، تأثیر داده‌های پرت و شکست‌ها بر مدل‌سازی نوسانات بازار سهام ایالات متحده بررسی شده و نشان داده شده است که نادیده گرفتن این عوامل منجر به سوگیری در تخمین پایداری نوسانات و پارامترهای مدل می‌شود (چاتزیکسنستانی^{۱۲}، ۲۰۱۷) پژوهشی دیگر با استفاده از نمودار جعبه‌ای، داده‌های پرت در ۱۴ شاخص سهام جهانی را تحلیل کرده و نشان داده است که داده‌های پرت منفی تأثیر قوی‌تری بر بازار دارند و تمایل به خوشه‌بندی دارند (شهاده و همکاران، ۲۰۲۲). همچنین، در مطالعه‌ای دیگر، رویکردی ترکیبی مبتنی بر مدل ARIMA و الگوریتم ژنتیک چندمنظوره برای تشخیص داده‌های پرت در سری‌های زمانی ارائه شده و دقت آن نسبت به روش‌های کلاسیک بهبود یافته است (شبانیه بوشهری et al. ۱۳۹۸). در پژوهشی دیگر، الگوریتم خوشه‌بندی برای تحلیل حجم معاملات در داده‌های فرکانس بالا به کار گرفته و نشان داده شده که ناهنجاری‌های توزیع حجم می‌توانند نشانه‌ای از ناکارآمدی بازار باشند (ژائو و وانگ، ۲۰۱۵). پژوهشی دیگر روشی آماری برای شناسایی داده‌های پرت در توزیع‌های دم‌چاق ارائه داده و سقوط بازار ۱۹۸۷ را از این منظر بررسی کرده است که نشان می‌دهد این سقوط یک داده پرت نبوده است (شلتر و ترد، ۲۰۰۸). در مطالعه‌ای دیگر، رابطه بین حجم معاملات و بازده سهام در بورس مصر مورد بررسی قرار گرفته و نشان داده شده

است که رابطه علیت دوطرفه‌ای بین این متغیرها وجود دارد (الانصاری و آتوئه^{۱۳}، ۲۰۱۲). پژوهشی دیگر یک روش بدون نظارت برای تشخیص داده‌های پرت بر پایه‌ی الگوریتم خوشه‌بندی میانگین ارائه داده و نشان داده است که حذف داده‌های پرت دقت خوشه‌بندی را بهبود می‌بخشد (احمد و محمود^{۱۴}، ۲۰۱۳). علاوه بر این، در مطالعه‌ی دیگر، رابطه بین بازده و حجم معاملات در بورس کراچی با استفاده از داده‌های هفتگی بررسی شده و نتایج نشان داده‌اند که افزایش حجم معاملات با افزایش بازده سهام همراه است (عطاری^{۱۵} و همکاران، ۲۰۱۲). پژوهشی دیگر اهمیت تحلیل حجم معاملات را در پیش‌بینی حرکات قیمتی بررسی کرده و نشان داده است که روش‌هایی مانند پروفایل بازار می‌توانند به بهینه‌سازی سود معاملاتی کمک کنند (چوتکا و ربتاک^{۱۶}، ۲۰۲۱). در مطالعه‌ی دیگر، داده‌های روزانه بازار سهام مورد بررسی قرار گرفته و مشخص شده است که حجم معاملات گذشته می‌تواند پیش‌بینی قابل قبولی از حرکات شاخص‌ها در دوره‌های بلندمدت ارائه دهد (آبوندانته^{۱۷}، ۲۰۱۰). در پژوهشی دیگر، ترکیب مدل مارکوف پنهان و الگوریتم خوشه‌بندی K-Means برای پیش‌بینی شاخص بازار سهام مورد استفاده قرار گرفته و نشان داده شده است که این روش دقت بالایی در شناسایی روندهای قیمتی دارد (عسگری و همکاران، ۲۰۱۷). همچنین، در مطالعه‌ی دیگر، مدل بازار سهام بررسی شده و مشخص شده است که معامله‌گران از حجم معاملات گذشته برای برآورد میزان در دسترس بودن اطلاعات خصوصی استفاده می‌کنند که این امر بر تصمیمات معاملاتی تأثیرگذار است (سومین^{۱۸}، ۲۰۰۱). همچنین در ادامه مقالاتی در جهت استفاده از روش‌های یادگیری ماشین مثل نزدیک‌ترین همسایه و پژوهشی دیگر به بررسی محدودیت‌های الگوریتم‌های خوشه‌بندی رایج پرداخته و روشی مبتنی بر پیوندهای داده‌ای برای کاهش پیچیدگی محاسباتی و افزایش دقت خوشه‌بندی ارائه داده است (احمد و محمود، ۲۰۱۳). در پژوهشی دیگر، نسخه‌ای اصلاح‌شده از الگوریتم DBSCAN معرفی شده که با استفاده از مفهوم «اشیاء هسته-قابل دسترسی»، خوشه‌بندی داده‌های پرتراکم را بهبود داده و وابستگی به ترتیب پردازش داده‌ها را کاهش داده است (تران^{۱۹} و همکاران، ۲۰۱۳).

در سال‌های اخیر، الگوریتم‌های یادگیری ماشین به‌ویژه خوشه‌بندی چگالی‌محور (DBSCAN) در تحلیل داده‌های مالی مورد توجه قرار گرفته‌اند، زیرا برخلاف روش‌های سنتی تحلیل تکنیکال، قادر به شناسایی الگوهای پنهان در داده‌های بازار هستند. در این پژوهش، مدلی ترکیبی ارائه شده که داده‌های پرت شناسایی شده توسط DBSCAN را با اندیکاتور MACD^{۲۰} ترکیب می‌کند تا سیگنال‌های معاملاتی دقیق‌تری تولید شود. دانش‌افزایی این پژوهش در چندین بُعد قابل بررسی است:

ترکیب نوآورانه الگوریتم DBSCAN با تحلیل تکنیکال: این پژوهش یکی از اولین مطالعاتی است

استراتژی معاملاتی بر اساس خوشه‌بندی داده‌های قیمت با الگوریتم... / شریفی‌راد و رستمی

که نقاط پرت شناسایی شده با DBSCAN را به عنوان سیگنال‌های احتمالی ورود و خروج در معاملات بررسی می‌کند.

مقایسه‌ی DBSCAN با روش‌های آماری سنتی: در این مطالعه، عملکرد DBSCAN در شناسایی داده‌های پرت با روش‌های انحراف معیار گاوسی و دیگر شاخص‌های آماری مقایسه شده است. این مقایسه نشان می‌دهد که روش‌های خوشه‌بندی چگالی‌محور می‌توانند نسبت به روش‌های کلاسیک در شرایط خاص بازار، عملکرد بهتری داشته باشند.

تمرکز بر بازار بورس ایران و در نظر گرفتن محدودیت‌های خاص: برخلاف بسیاری از مطالعات مشابه که بر بازارهای بین‌المللی انجام شده‌اند، این پژوهش شرایط خاص بورس و فرابورس ایران، از جمله دامنه نوسان، حجم‌مبنا و نوسانات محدود شده بازار را در نظر گرفته و استراتژی پیشنهادی را متناسب با این ساختار ارزیابی کرده است. این پژوهش تلاش دارد تا رویکردی نوین در بهینه‌سازی سیگنال‌های معاملاتی ارائه دهد که می‌تواند در بازارهای واقعی مالی کاربردی باشد.

سؤالات پژوهش

با توجه به اهمیت شناسایی داده‌های پرت در بازارهای مالی و کاربرد الگوریتم‌های خوشه‌بندی در بهینه‌سازی استراتژی‌های معاملاتی، این پژوهش به دنبال پاسخ به سؤالات زیر است:

۱. آیا الگوریتم DBSCAN می‌تواند نقاط پرت بازار را به‌طور مؤثر شناسایی کند و این نقاط در پیش‌بینی روند قیمت سهام قابل استفاده هستند؟

۲. چگونه عملکرد استراتژی پیشنهادی (DBSCAN + MACD) با روش‌های دیگر، از جمله تحلیل انحراف معیار گاوسی و استراتژی خرید و نگهداری مقایسه می‌شود؟

۳. عملکرد این استراتژی در بازار بورس ایران چگونه است و تا چه حد می‌تواند سودآوری معاملات را افزایش دهد؟

روش تحقیق

۱. پیش‌پردازش داده‌ها

پیش‌پردازش داده‌ها مرحله‌ای کلیدی در بهبود کیفیت و کاهش نویز داده‌های مالی است که شامل جمع‌آوری، پالایش، استخراج ویژگی‌های کلیدی، استانداردسازی، تحلیل همبستگی و حذف متغیرهای وابسته می‌شود. برای افزایش دقت مدل، از روش‌هایی مانند جنگل تصادفی برای انتخاب ویژگی‌های بهینه استفاده شده است. داده‌ها به مجموعه‌های آموزشی و آزمایشی تقسیم شده تا از بیش‌برازش

جلوگیری شود و دقت مدل در داده‌های جدید ارزیابی گردد. این فرآیندها پایه‌ای برای اجرای صحیح الگوریتم‌های شناسایی نویز و بهینه‌سازی عملکرد مدل پیشنهادی ایجاد کرده است

۲. یافتن داده‌های نویز با الگوریتم DBSCAN

در این پژوهش، برای شناسایی داده‌های نویز از الگوریتم خوشه‌بندی مبتنی بر چگالی استفاده شده است. این الگوریتم برخلاف روش‌هایی مانند K-Means که تنها قادر به شناسایی خوشه‌های کروی هستند، خوشه‌هایی با اشکال نامنظم و داده‌های پرت را نیز شناسایی می‌کند. مزیت دیگر این روش، عدم نیاز به تعیین تعداد خوشه‌ها از پیش و قابلیت تشخیص نویزها و نقاط مرزی است. باین‌حال، عملکرد DBSCAN به پارامترهای شعاع همسایگی^{۲۱} و حداقل تعداد نقاط^{۲۲} وابسته است که انتخاب نامناسب آن‌ها می‌تواند بر نتایج تأثیر منفی بگذارد DBSCAN خوشه‌ها را به‌عنوان مناطق پرتراکم داده‌ها تعریف کرده و نقاطی که در مناطق پراکنده قرار دارند، به‌عنوان نویز یا داده‌های پرت در نظر می‌گیرد. این الگوریتم نقاط را به سه دسته تقسیم می‌کند: نقاط هسته^{۲۳} که حداقل تعداد نقاط لازم را در همسایگی خود دارند، نقاط مرزی^{۲۴} که در محدوده‌ی یک خوشه قرار دارند اما خودشان هسته نیستند و نقاط نویز^{۲۵} که در هیچ خوشه‌ای قرار نمی‌گیرند.

فرمول چگالی محلی نقاط در این الگوریتم به‌صورت زیر تعریف می‌شود:

$$\text{density}(x_i) = \text{Count}(N_{\text{Eps}}(x_i))$$

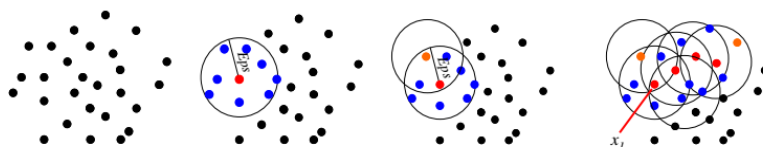
که در آن $N_{\text{Eps}}(x_i)$ نشان‌دهنده اشیاء همسایه در همسایگی با شعاع Eps است:

$$N_{\text{Eps}}(x_i) = \{x_j | \forall j, \text{distance}(x_j, x_i) < \text{Eps}\}$$

بر اساس این فرمول، نقاط به‌صورت زیر دسته‌بندی می‌شوند:

اگر $\text{density}(x_{\text{core}}) \geq \text{MinPts}$ (چگالی محلی آن بیشتر از حداقل تعداد نقاط است). یک هسته، x_{core} است اگر $\text{density}(x_{\text{border}}) < \text{MinPts}$ یعنی شیء مرزی x_{border} و چگالی محلی آن کمتر از حداقل تعداد نقاط است. شکل ۱ یک نمای کلی از نحوه‌ی عملکرد این الگوریتم را نشان می‌دهد.

شی قابل دسترسی با چگالی شی مرزی هسته یک خوشه



شکل ۱- جستجوی اشیاء با الگوریتم خوشه‌بندی مبتنی بر چگالی

استراتژی معاملاتی بر اساس خوشه‌بندی داده‌های قیمت با الگوریتم... / شریفی‌راد و رستمی

مراحل کلی اجرای این الگوریتم به این صورت است که ابتدا تمام نقاط به‌عنوان بازدید نشده علامت‌گذاری می‌شوند، سپس یک نقطه تصادفی انتخاب‌شده و همسایگی آن بررسی می‌شود، اگر تعداد همسایگان کمتر از MinPts باشد، نقطه به‌عنوان نویز ثبت می‌شود، در غیر این صورت، یک خوشه جدید تشکیل‌شده و نقطه به‌عنوان هسته خوشه ثبت می‌شود. تمام نقاط بازدید نشده در همسایگی شعاع خوشه بررسی و به‌عنوان هسته یا مرزی اضافه می‌شوند. این فرآیند تا زمانی که همه نقاط پردازش شوند، تکرار می‌شود. این الگوریتم نسبت به K-Means انعطاف‌پذیری بیشتری در شناسایی خوشه‌های نامنظم دارد و در حوزه‌هایی مانند تحلیل داده‌های مکانی، تشخیص تقلب در بازارهای مالی، پردازش تصویر و شناسایی ناهنجاری‌های داده‌ای کاربرد گسترده‌ای دارد.

۳. ترکیب خروجی الگوریتم DBSCAN با سیگنال‌های MACD

در این پژوهش، برای بهبود دقت سیگنال‌های معاملاتی، الگوریتم خوشه‌بندی DBSCAN با اندیکاتور MACD ترکیب‌شده است. ابتدا DBSCAN نقاط پرت را شناسایی می‌کند که ممکن است نشان‌دهنده تغییرات ناگهانی در روند بازار باشند. سپس، این نقاط با سیگنال‌های MACD مقایسه می‌شوند تا مشخص شود که آیا تغییرات قیمتی ناشی از یک‌روند واقعی است یا صرفاً نویز بازار.

در این رویکرد: اگر در یک نقطه پرت، خط MACD از زیرخط سیگنال عبور کند، سیگنال ورود به معامله صادر می‌شود. اگر در نقطه پرت، خط MACD از بالا خط سیگنال را قطع کند، سیگنال خروج از معامله ثبت می‌شود. اندیکاتور MACD از اختلاف میانگین متحرک نمایی ۱۲ و ۲۶ دوره‌ای تشکیل‌شده و خط سیگنال ۹ دوره‌ای به‌عنوان شاخص تأیید روند استفاده می‌شود. همچنین، هیستوگرام MACD میزان قدرت روند را مشخص می‌کند. در مدل پیشنهادی، نقاط پرت شناسایی‌شده توسط DBSCAN فقط در صورتی سیگنال معتبر محسوب می‌شوند که هم‌راستا با سیگنال MACD باشند:

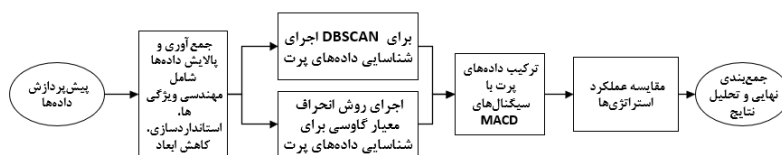
هم‌زمانی نقاط پرت با تقاطع صعودی MACD سیگنال ورود به معامله.

هم‌زمانی نقاط پرت با تقاطع نزولی MACD سیگنال خروج از معامله.

این ترکیب، دقت سیگنال‌های معاملاتی را افزایش داده، سیگنال‌های کاذب را کاهش داده و توانایی تشخیص روندهای واقعی را بهبود می‌بخشد. همچنین، امکان فیلتر کردن ناهنجاری‌های قیمتی و بهره‌برداری از تغییرات غیرعادی بازار را فراهم می‌کند.

۴. روش انحراف معیار گاوسی^{۲۶}

روش انحراف معیار گاوسی یک تکنیک آماری برای شناسایی داده‌های پرت در مجموعه داده‌های مالی است. این روش، با فرض توزیع نرمال داده‌ها، مقادیر پرت را بر اساس فاصله آن‌ها از میانگین شناسایی می‌کند. در این رویکرد، داده‌هایی که بیش از مقدار مشخصی از انحراف معیار فاصله دارند، به‌عنوان نقاط پرت در نظر گرفته می‌شوند. در مدل پیشنهادی، نقاط پرت شناسایی‌شده توسط انحراف معیار گاوسی با سیگنال‌های MACD ترکیب شده‌اند تا دقت معاملات افزایش یابد. اگر در نقطه پرت، MACD تقاطع صعودی داشته باشد، احتمال رشد قیمت افزایش یافته و سیگنال خرید تأیید می‌شود. برعکس، اگر MACD در نقطه پرت تقاطع نزولی داشته باشد، احتمال کاهش قیمت وجود دارد و سیگنال فروش صادر می‌شود. ترکیب MACD و انحراف معیار گاوسی موجب فیلتر کردن نویزهای بازار و کاهش سیگنال‌های اشتباه ناشی از نوسانات کوتاه‌مدت می‌شود. پس از اجرای مدل‌های DBSCAN + MACD و GSDM + MACD، عملکرد هر روش بررسی شده و با استراتژی خرید و نگهداری مقایسه شده است. این مقایسه شامل تحلیل بازدهی در بورس و فرابورس تهران و ارزیابی اثربخشی ترکیب الگوریتم‌های یادگیری ماشین و اندیکاتورهای تکنیکال بر بهینه‌سازی معاملات و کاهش ریسک سرمایه‌گذاری است.



شکل ۲- نمایش مراحل روش تحقیق و اجرای مدل پیشنهادی

یافته‌ها

در این پژوهش، سهم‌های بورس و فرابورس تهران برای بررسی روش‌های شناسایی داده‌های نویز انتخاب شدند تا کارایی مدل در شرایط مختلف بازار سنجیده شود. در گام پیش‌پردازش داده‌ها، ویژگی‌های کلیدی استخراج و استانداردسازی شده، متغیرهای وابسته با تحلیل همبستگی حذف و ویژگی‌های مهم با روش جنگل تصادفی انتخاب شدند تا داده‌ها برای اجرای الگوریتم‌های تحلیل آماده شوند.

در مرحله بعد، الگوریتم خوشه‌بندی DBSCAN برای شناسایی داده‌های نویز اجرا شد. این

استراتژی معاملاتی بر اساس خوشه‌بندی داده‌های قیمت با الگوریتم... / شریفی‌راد و رستمی

الگوریتم با تنظیم شعاع همسایگی و حداقل تعداد نقاط خوشه‌ها، داده‌های خارج از خوشه را به‌عنوان نویز شناسایی کرد. خروجی این مرحله شامل خوشه‌های اصلی و مجموعه‌ای از داده‌های پرت با رفتار غیرمعمولی در بازار بود. برای طراحی استراتژی معاملاتی، داده‌های نویز شناسایی‌شده توسط DBSCAN با سیگنال‌های اندیکاتور MACD ترکیب شد. در این فرآیند: نقاط نویز در داده‌های قیمتی علامت‌گذاری شدن MACD. برای شناسایی نقاط ورود و خروج بهینه معاملات استفاده شد. هم‌پوشانی نویزها و تغییرات MACD به‌عنوان سیگنال ورود و خروج به معامله در نظر گرفته شد. در ادامه، نتایج هر مرحله از تحلیل ارائه می‌شود.

۱. آماده‌سازی داده‌ها و مهندسی ویژگی‌ها

در ابتدا، داده‌های بازار مالی شامل حجم معاملات، قیمت و سایر متغیرهای مرتبط جمع‌آوری شد. سپس برای بهبود کیفیت داده‌ها و استخراج ویژگی‌های مفید، فرآیند مهندسی ویژگی انجام گرفت. ویژگی‌های جدید با ترکیب اطلاعات قیمت و حجم معاملات ساخته شدند تا الگوریتم بتواند الگوهای بهتری را شناسایی کند. ویژگی‌های ساخته‌شده در جدول ۱ نمایش داده‌شده است.

۲. استانداردسازی داده‌ها برای کاهش تأثیر مقیاس‌های مختلف داده‌های ورودی

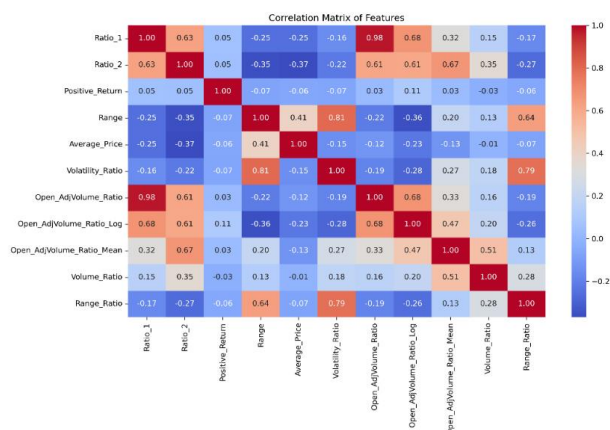
۳. شناسایی و حذف متغیرهای همبسته با استفاده از تحلیل همبستگی^{۲۷}

پس از ساخت ویژگی‌های مبتنی بر قیمت و حجم، ماتریس همبستگی برای هر سهم محاسبه شد تا میزان ارتباط میان فیچرها بررسی شود. این تحلیل کمک می‌کند تا ویژگی‌های با همبستگی بالا که اطلاعات مشابه ارائه می‌دهند، شناسایی‌شده و ویژگی‌های مستقل‌تر که ارزش‌افزوده بیشتری در مدل دارند، انتخاب شوند. به‌عنوان نمونه، در نماد فارس از بازار بورس تهران Volatility Ratio و Range Ratio همبستگی مثبت قوی (۰.۸۱) دارند Average Price با Ratio ۲ همبستگی منفی قابل‌توجهی (۰.۳۷-) دارد. ویژگی‌های همبسته مانند Ratio ۱ و Open Adj Volume Ratio قابل ساده‌سازی هستند. در ادامه، همبستگی ویژگی‌ها با بازدهی نیز بررسی شد. Positive Return بالاترین همبستگی مثبت را دارد، نشان‌دهنده تأثیر قوی آن بر بازدهی. Volatility Ratio و Range Ratio همبستگی منفی دارند، یعنی افزایش آن‌ها کاهش بازدهی را در پی دارد. ویژگی‌هایی مانند Average Price و Volume Ratio همبستگی ضعیفی داشته و تأثیر کمتری بر قیمت دارند. این نتایج نشان می‌دهد که در مدل‌سازی باید ویژگی‌های با همبستگی قوی‌تر (مثبت یا منفی) اولویت داشته باشند و متغیرهای کم اثر حذف یا ساده‌سازی شوند.

فصلنامه مهندسی مالی و مدیریت اوراق بهادار / دوره ۱۶ / شماره ۶۴ / پائیز ۱۴۰۴

جدول ۱: ویژگی‌های ساخته‌شده

نام ویژگی	فرمول محاسباتی
نسبت حجم معاملات تعدیل‌شده به فاصله بین قیمت باز و بسته شدن	$\text{Ratio}_1 = \frac{\text{Adj Volume}}{ \text{Open}-\text{Close} }$
نسبت حجم معاملات تعدیل‌شده به فاصله بین بیشترین و کمترین قیمت در روز	$\text{Ratio}_2 = \frac{\text{Adj Volume}}{ \text{High}-\text{Low} }$
بازده روزانه مثبت یا منفی	$\text{Positive Return} = \begin{cases} 1 & \text{if Return} > 0 \\ 0 & \text{else} \end{cases}$
فاصله بین بیشترین و کمترین قیمت در روز	$\text{Range} = \text{High} - \text{Low}$
میانگین قیمت روز	$\text{Average Price} = \frac{\text{Open} + \text{Close} + \text{High} + \text{Low}}{4}$
نسبت نوسان قیمت روزانه به قیمت بسته‌شدن	$\text{Volatility Ratio} = \frac{\text{Range}}{\text{Close}}$
نسبت حجم تعدیل‌شده به تفاوت قیمت باز و بسته شدن، وزن‌دار با قیمت باز شدن	$\text{Open Adj Volume Ratio} = \frac{\text{Open} \times \text{Adj Volume}}{ \text{Open} - \text{Close} }$
لگاریتم نسبت حجم تعدیل‌شده به تفاوت قیمت باز و بسته شدن، وزن‌دار با قیمت باز شدن	$\text{Log (Open Adj Volume Ratio)}$
نسبت حجم تعدیل‌شده وزن‌دار با میانگین قیمت باز و بسته شدن	$\text{Open Adj Volume Ratio Mean} = \frac{\text{Open} \times \text{Adj Volume}}{\frac{\text{Open} + \text{Close}}{2}}$
نسبت حجم تعدیل‌شده به میانگین حجم ۷ روز گذشته	$\text{Volume Ratio} = \frac{\text{Adj Volume}}{\text{Adj Volume (Rolling mean, window = 7)}}$
نسبت دامنه روزانه قیمت به میانگین دامنه ۷ روز گذشته	$\text{Range Ratio} = \frac{\text{Range}}{\text{Range (Rolling mean, window = 7)}}$



شکل ۲- ماتریس همبستگی ویژگی‌ها

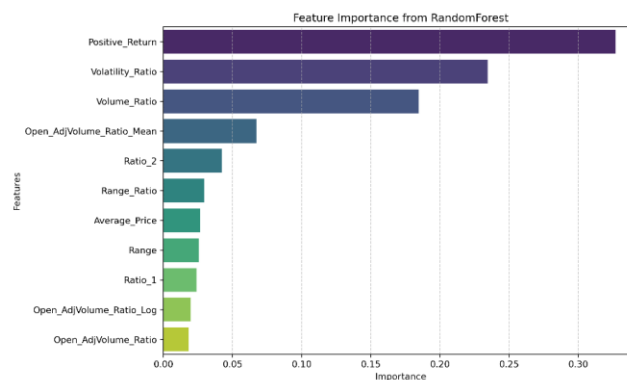
استراتژی معاملاتی بر اساس خوشه‌بندی داده‌های قیمت با الگوریتم... / شریفی‌راد و رستمی

۴. انتخاب ویژگی‌های کلیدی با استفاده از الگوریتم جنگل تصادفی^{۲۸}

نمودار اهمیت ویژگی‌ها^{۲۹} میزان تأثیر هر ویژگی را بر پیش‌بینی مدل نشان می‌دهد. این نمودار بر اساس مدل جنگل تصادفی محاسبه شده است و مشخص می‌کند کدام ویژگی‌ها بیشترین نقش را در پیش‌بینی هدف ایفا می‌کنند. ویژگی‌های با اهمیت بیشتر سهم بیشتری در تصمیم‌گیری مدل دارند. در این نمودار در که در مورد سهم فارس از بازار بورس تهران صورت گرفته و در شکل ۳ نشان داده شده، Positive Return, Volatility Ratio و Range Ratio از مهم‌ترین ویژگی‌ها هستند و ویژگی‌هایی مانند Range و Average Price در انتهای لیست قرار گرفته‌اند و نقش کم‌تری در مدل دارند. این اطلاعات می‌تواند برای بهینه‌سازی مدل و کاهش پیچیدگی داده‌ها مفید باشد، به این صورت که ویژگی‌های کم‌اهمیت حذف شوند و تمرکز بر ویژگی‌های مهم‌تر قرار گیرد.

تمام مراحل مربوط به ساخت و انتخاب ویژگی‌های مهم برای نمادهای دیگر نیز صورت گرفته و برای آموزش الگوریتم مورد استفاده قرار گرفته است؛ و پس از اتمام مراحل بالا ویژگی‌های زیر که برای تمام نمادهای مورد بررسی بهترین اثرگذاری را داشته‌اند، برای ورود به الگوریتم انتخاب شده است.

Positive Return, Volatility Ratio, Volume Ratio, Range Ratio, open AdjVolume, Ratio log, Ratio2.



شکل ۳- بررسی اهمیت ویژگی‌ها با الگوریتم جنگل تصادفی

۵. شناسایی خوشه‌ها و نقاط نويز با روش DBSCAN

در این پژوهش، مدل DBSCAN برای شناسایی خوشه‌های داده و نقاط نويز در بازار بورس و فرابورس تهران استفاده شد. داده‌های قیمتی از ۱۳۹۸/۰۹/۱۵ تا ۱۴۰۲/۰۹/۱۴ برای آموزش مدل و

۱۴۰۲/۰۹/۱۵ تا ۱۴۰۳/۰۹/۱۵ برای تست آن به کار گرفته شد. پارامترهای مدل با روش جستجوی شبکه‌ای بهینه‌سازی شدند و نقاط پرت شناسایی شده با اندیکاتور MACD ترکیب گردیدند. در این روش، نقاط پرت به عنوان سیگنال ورود یا خروج در نظر گرفته شدند؛ به طوری که اگر در نقطه نویز، خط MACD از زیر خط سیگنال عبور می‌کرد، سیگنال خرید صادر می‌شد و اگر از بالا قطع می‌کرد، سیگنال فروش ثبت می‌شد.

علاوه بر این، روش انحراف معیار گاوسی (GSDM) نیز برای شناسایی داده‌های پرت به کار گرفته شد و عملکرد آن با DBSCAN و استراتژی خرید و نگهداری مقایسه گردید. جدول‌های ۱ و ۲ بازدهی ۱۰ سهم پرتراکنش بورس و فرابورس را نشان می‌دهند و جدول‌های ۳ و ۴ نمادهایی با بیشترین اختلاف بازدهی نسبت به خرید و نگهداری را بررسی می‌کنند.

استراتژی DBSCAN + MACD : عملکرد برجسته‌ای در شناسایی فرصت‌های سود ده داشته است. برای مثال :و تجارت ۷۹.۷۸٪ بازدهی، ۶۵.۸۸٪ بیشتر از خرید و نگهداری، فجر ۵۱.۴۴٪ بازدهی، ۵۲.۵۲٪ بیشتر از خرید و نگهداری، های وب ۱۹.۳۳٪ بازدهی، ۴۷.۳۶٪ بیشتر از خرید و نگهداری

در نمادهای فرابورس نیز موفق عمل کرده است، از جمله :فرابورس ۳۸.۲۸٪ بازدهی، ۵۴.۳۳٪ بیشتر از خرید و نگهداری، عاليس ۲۹.۴۷٪ بازدهی، ۳۳.۸۸٪ بیشتر از خرید و نگهداری استراتژی GSDM + MACD : در برخی نمادها مانند ثنوسا (۸۶.۱۵٪) و حفارس (۷۰.۴۶٪) بهتر از DBSCAN عمل کرده، اما در مجموع پایداری کمتری داشته است.

استراتژی خرید و نگهداری: در بسیاری از نمادها بازدهی منفی داشته است، از جمله :های وب : ۲۸.۰۳٪-، شاهد ۲۴.۶۳٪-، وآتوس ۶۱.۹۹٪-، به طور کلی، DBSCAN و MACD یکی از بهترین استراتژی‌ها در این تحلیل بوده‌اند.

جدول ۲-مقایسه بازدهی سه استراتژی در نمادهای پرتراکنش بورس تهران (اعداد به درصد)

نماد	DBSCAN + MACD	GSDM + MACD	خرید و نگهداری	اختلاف بازدهی DBSCAN و خرید و نگهداری	اختلاف بازدهی GSDM و خرید و نگهداری
وتجارت	79.78	67.27	13.89	65.88	53.37
شاهد	29.64	-2.19	-24.63	54.27	22.43
فجر	51.44	23.58	-1.07	52.52	24.66
های وب	19.33	17.66	-28.03	47.36	45.69

استراتژی معاملاتی بر اساس خوشه‌بندی داده‌های قیمت با الگوریتم... / شریفی‌راد و رستمی

ذوب	54.89	45.49	8.56	46.32	36.921
برکت	3.19	-12.45	-40.70	43.89	28.24
خگستر	33.25	17.35	-6.428	39.67	23.78
اخابر	21.72	11.31	-17.65	39.38	28.97
شستا	50.20	60.31	16.72	33.47	43.58
شپنا	18.03	-0.62	-12.94	30.98	12.32

جدول ۳- مقایسه بازدهی سه استراتژی در نمادهای پرتراکنش فرابورس (اعداد به درصد)

نماد	DBSCAN + MACD	GSDM + MACD	خرید و نگهداری	اختلاف بازدهی DBSCAN و خرید و نگهداری	اختلاف بازدهی GSDM و خرید و نگهداری
فرابورس	38.28	12.03	-16.05	54.33	28.09
حآفرین	10.073	-25.60	-33.31	43.39	7.71
عالیس	29.47	13.69	-4.40	33.88	18.09
وسپهر	26.91	38.65	1.03	25.87	37.61
سرچشمه	27.47	17.59	1.74	25.73	15.84
گدنا	-4.04	19.80	-26.09	22.05	45.90
خاور	-29.54	-44.22	-50.64	21.09	6.41
سمگا	45.81	13.84	33.26	12.54	-19.42
فن افزار	19.93	34.51	14.18	5.741	20.33
داوه	17.54	-3.17	35.54	-18.00	-38.71

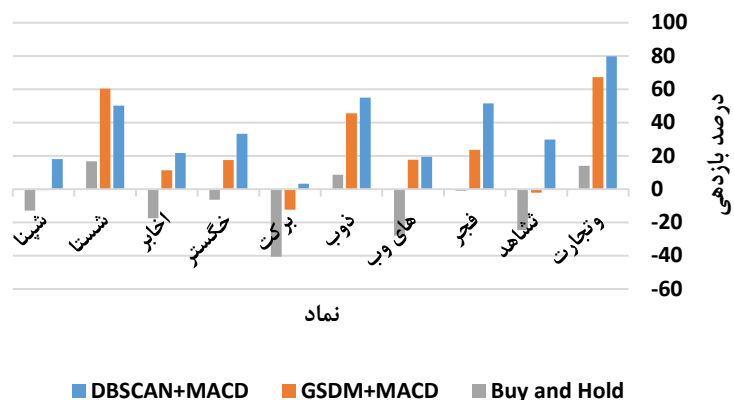
جدول ۴- نمادهایی با بیشترین اختلاف بازدهی استراتژی DBSCAN و خرید و نگهداری (اعداد به درصد)

نماد	DBSCAN + MACD	GSDM + MACD	خرید و نگهداری	اختلاف بازدهی DBSCAN و خرید و نگهداری	اختلاف بازدهی GSDM و خرید و نگهداری
واتوس	43.49	2.37	-61.99	105.48	64.36
دکوثر	122.58	44.49	28.73	93.86	15.76
خوساز	88.63	46.38	1.48	87.15	44.90
سدبیر	14.78	7.00	-53.88	68.66	60.88
خدیزل	38.22	-2.07	-30.16	68.39	28.09
وتجارت	79.78	67.28	13.90	65.89	53.38
غشهداب	31.96	15.57	-33.18	65.14	48.75
کیمیا	42.33	15.21	-21.57	63.90	36.79
کماسه	42.17	6.01	-20.93	63.10	26.94

جدول ۵- نمادهایی با بیشترین اختلاف بازدهی استراتژی GSDM و خرید و نگهداری (اعداد به درصد)

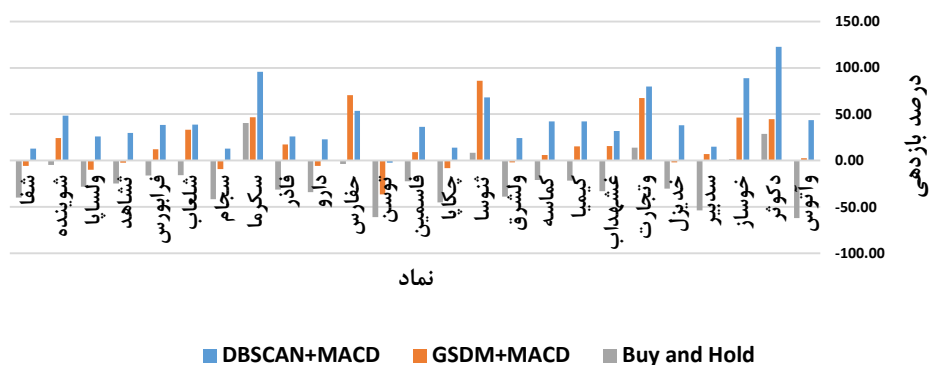
نماد	DBSCAN + MACD	GSDM + MACD	خرید و نگهداری	اختلاف بازدهی DBSCAN و خرید و نگهداری	اختلاف بازدهی GSDM و خرید و نگهداری
ثنوسا	67.97	86.15	8.44	59.53	77.72
حفارس	53.59	70.46	-3.64	57.23	74.10
وآتوس	43.49	2.37	-61.99	105.48	64.36
کسرا	39.43	51.94	-10.29	49.72	62.23
سدبیر	14.78	7.00	-53.88	68.66	60.88
ثامان	11.39	160.37	103.29	-91.90	57.08
وتجارت	79.78	67.28	13.90	65.89	53.38
خبهمن	43.37	43.61	-8.56	51.92	52.16
وتوس	0.00	52.32	0.83	-0.83	51.49
ولملت	-13.68	2.91	-48.47	34.79	51.38

نمودارهای ارائه شده در ادامه، مقایسه عملکرد استراتژی‌ها و نمودار قیمتی نماد ذوب با نقاط نويز را نشان می‌دهند که به نمایش تصویری عملکرد این روش‌ها در تعیین نقاط ورود و خروج کمک می‌کند.

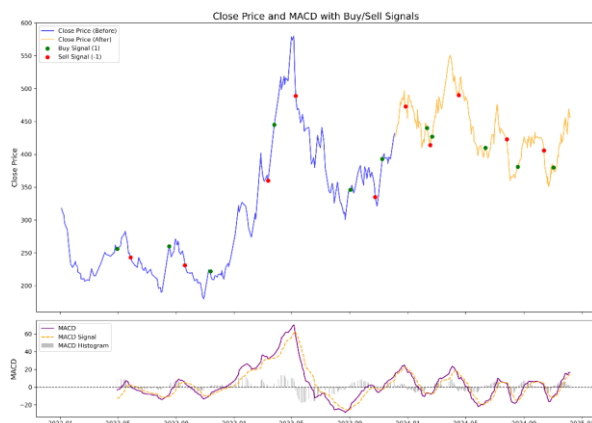


نمودار ۱-مقایسه بازدهی سه استراتژی در نمادهای پرتراکنش بورس تهران

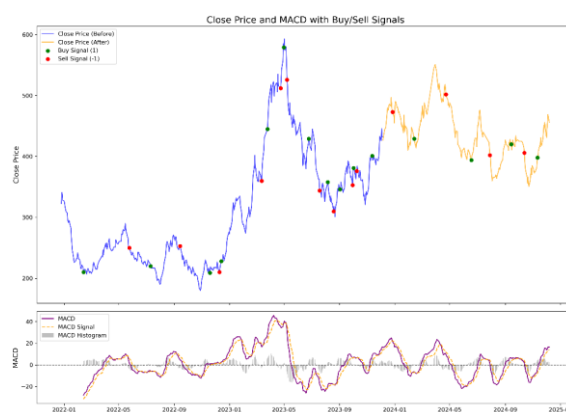
استراتژی معاملاتی بر اساس خوشه‌بندی داده‌های قیمت با الگوریتم... / شریفی‌راد و رستمی



نمودار ۲- نمادهایی با بیشترین اختلاف بازدهی استراتژی DBSCAN و خرید و نگهداری



شکل ۵- نمودار قیمتی نماد ذوب و داده‌های نویز با DBSCAN



شکل ۴- نمودار قیمتی نماد ذوب و داده‌های نویز با GSDM

بحث و نتیجه گیری

در این پژوهش، مدل DBSCAN برای شناسایی خوشه‌های داده و نقاط نویز در بازار بورس و فرابورس تهران به کار گرفته شد. پس از استخراج ویژگی‌های کلیدی از داده‌های قیمتی و حجمی و بهینه‌سازی پارامترهای الگوریتم، مدل برای تشخیص داده‌های پرت آموزش داده شد. سپس، در دوره‌ی آزمایشی، داده‌های جدید مورد بررسی قرار گرفتند و نقاط نویز شناسایی شده با سیگنال‌های MACD ترکیب شدند تا نقاط ورود و خروج بهینه معاملات مشخص شود. نتایج نشان داد که استراتژی DBSCAN + MACD در اکثر نمادهای بررسی شده عملکرد بهتری نسبت به روش‌های دیگر داشته و در سهم‌هایی مانند تجارت، فجر، های‌وب و شستا، اختلاف بازدهی بیش از ۵۰٪ نسبت به خرید و نگهداری ارائه کرده است.

در بازار فرابورس نیز، DBSCAN و MACD بازدهی مثبت و بالاتری نسبت به سایر روش‌ها داشتند. برای مثال، در نمادهای فرابورس، عالیس و وسپهر، اختلاف بازدهی مثبت این روش در مقایسه با خرید و نگهداری به ترتیب ۵۴.۳۳٪، ۸۸.۳۳٪ و ۸۷.۲۵٪ ثبت شده است. برخی نمادها نیز بازدهی چشمگیری داشتند، از جمله دکوثر (۱۲۲.۵۸٪)، خوساز (۶۳.۸۸٪) و وآتوس (۴۹.۴۳٪) که اختلاف بازدهی بالایی نسبت به خرید و نگهداری داشتند. در مقابل، روش GSDM + MACD در برخی نمادها بازدهی مثبتی داشت اما در مجموع عملکرد ضعیف‌تری نسبت به DBSCAN ارائه داد. در مقایسه‌ی روش‌ها، مدل انحراف معیار گاوسی در برخی نمادها عملکرد بهتری نسبت به DBSCAN نشان داد اما به طور کلی، پایداری کمتری داشت. برای مثال، در ثنوسا (۸۶.۱۵٪) و حفارس (۴۶.۷۰٪) این روش اختلاف بازدهی مثبتی با خرید و نگهداری داشت اما همچنان در مقایسه با DBSCAN ناپایدارتر بود. به طور کلی، استراتژی DBSCAN + MACD یکی از بهترین عملکردها را داشته و در نمادهایی مانند ثنوسا (۶۷.۹۷٪) و حفارس (۵۹.۵۳٪) اختلاف بازدهی بالاتری نسبت به خرید و نگهداری ارائه داده است. یکی از مزایای کلیدی DBSCAN، قابلیت شناسایی الگوهای پنهان و نقاط پرت در داده‌های بازار است که آن را از روش‌های سنتی تحلیل تکنیکال متمایز می‌کند. یافته‌های این مطالعه با نتایج پژوهش محمدزاده و همکاران (۱۴۰۱) که از الگوریتم ژنتیک و مدل ARIMA برای شناسایی داده‌های پرت در سری‌های زمانی استفاده کردند، همخوانی دارد. هر دو تحقیق نشان دادند که روش‌های یادگیری ماشین دقت بیشتری در شناسایی ناهنجاری‌های بازار نسبت به مدل‌های آماری دارند. با این حال، در تحقیق حاضر، از یک روش غیرپارامتریک (DBSCAN) استفاده شده است، در حالی که مطالعه‌ی محمدزاده (۱۴۰۱) بر مدل‌سازی توزیع نرمال داده‌ها تمرکز داشت. همچنین،

استراتژی معاملاتی بر اساس خوشه‌بندی داده‌های قیمت با الگوریتم... / شریفی‌راد و رستمی

نتایج این پژوهش با یافته‌های عسگری و همکاران (۱۳۹۸) که از مدل مارکوف پنهان و روش K-Means برای تحلیل روندهای بازار استفاده کرده بودند، قابل‌مقایسه است. هر دو مطالعه نشان دادند که روش‌های خوشه‌بندی می‌توانند به تحلیل بهتر روندهای قیمتی کمک کنند؛ اما در این پژوهش، DBSCAN به جای K-Means به کار رفته است که توانایی بیشتری در شناسایی خوشه‌های نامنظم و نقاط پرت دارد.

مقایسه‌ی این تحقیق با پژوهش صادقی و همکاران (۱۴۰۰) که از شاخص‌های تحلیل تکنیکال به‌تنهایی برای پیش‌بینی قیمت سهام استفاده کرده بودند، نشان می‌دهد که ترکیب شاخص‌های تکنیکال (MACD) با الگوریتم‌های یادگیری ماشین (DBSCAN) دقت بالاتری در شناسایی نقاط ورود و خروج معاملاتی ارائه می‌دهد.

به‌طور کلی، یافته‌های این پژوهش نشان می‌دهد که ترکیب DBSCAN با MACD می‌تواند استراتژی معاملاتی کارآمدتری نسبت به روش‌های آماری کلاسیک و مدل‌های تحلیل تکنیکال سنتی ارائه دهد. همچنین، استفاده از DBSCAN به‌عنوان یک ابزار برای تحلیل داده‌های مالی قابلیت بهینه‌سازی تصمیمات معاملاتی و کاهش ریسک را افزایش می‌دهد.

پیشنهادهای برای تحقیقات آینده

- ۱- ترکیب با الگوریتم‌های پیشرفته یادگیری ماشین
 - بررسی کارایی DBSCAN در ترکیب با الگوریتم‌های دیگر خوشه‌بندی مانند K-Means ، Gaussian Mixture Model (GMM) و Spectral Clustering برای بهبود شناسایی خوشه‌های قیمتی.
 - استفاده از الگوریتم‌های یادگیری عمیق مانند LSTM (شبکه‌های حافظه کوتاه‌مدت بلندمدت) و Transformer ها برای شناسایی روندهای پیچیده بازار.
- ۲- گسترش دامنه داده‌ها و بررسی بازارهای مالی مختلف
 - اعمال این استراتژی در بازارهای مالی دیگر مانند بازار رمز ارزها، فارکس و بازارهای دوطرفه بین‌المللی برای سنجش کارایی در شرایط مختلف معاملاتی.
 - استفاده از داده‌های بلندمدت‌تر برای بررسی تأثیرات دوره‌های رکود، رشد و شرایط ناپایدار اقتصادی.
- ۳- ادغام جنبه‌های زمانی و مدل‌های پیش‌بینی روند بازار

- ترکیب DBSCAN با مدل‌های سری زمانی مانند ARIMA ، GARCH یا زنجیره مارکوف برای پیش‌بینی بهتر رفتار قیمت در آینده.
- استفاده از تحلیل موج الیوت و شبکه‌های عصبی زمانی برای شناسایی الگوهای قیمتی پیچیده.
- ۴- بهینه‌سازی استراتژی با روش‌های فرا ابتکاری
- استفاده از الگوریتم‌های بهینه‌سازی مانند الگوریتم ژنتیک، بهینه‌سازی ازدحام ذرات و شبیه‌سازی تبرید برای یافتن پارامترهای بهینه DBSCAN و بهبود دقت سیگنال‌های معاملاتی.
- بررسی اثرات تغییر پارامترهای خوشه‌بندی DBSCAN بر روی عملکرد کلی استراتژی.
- ۵- استفاده عملی در سیستم‌های معاملاتی خودکار
- توسعه و پیاده‌سازی این استراتژی در پلتفرم‌های معاملاتی الگوریتمی و ربات‌های معامله‌گر.
- بررسی امکان اتصال این استراتژی به API های بازارهای مالی واقعی برای انجام معاملات خودکار و مدیریت ریسک.
- سنجش عملکرد مدل در حجم‌های بالای معاملات (High-Frequency Trading - HFT) برای بهینه‌سازی بهره‌وری.
- ۶- تحلیل تأثیر محدودیت‌های بازار و شرایط واقعی معاملاتی
- بررسی نحوه عملکرد استراتژی در شرایط واقعی بازار ایران، با محدودیت‌هایی مانند دامنه نوسان، حجم‌مبنا و نوسانات غیرقابل پیش‌بینی.
- مقایسه کارایی مدل در بازارهای دوطرفه و تأثیر حذف محدودیت‌های معاملاتی بر دقت استراتژی.

منابع

- ۱) شبانه بوشهری، د. امینی، ا. & محمدزاده، ج. (۱۳۹۸). تشخیص داده‌های پرت در سری‌های زمانی با استفاده از الگوریتم ژنتیک چندمنظوره سومین کنفرانس بین‌المللی محاسبات نرم، ۱۰۰۶۰۶۸ <https://civilica.com/doc/518221>
- ۲) صادقی، ح. ا. & دهقان، ا. (۱۳۹۵). کاربرد خوشه‌بندی در بازارهای مالی (مطالعه‌ای در بورس اوراق بهادار تهران) چهارمین کنفرانس ملی مدیریت، اقتصاد و حسابداری، ۵۱۸۲۲۱ <https://civilica.com/doc/518221>
- 3) Abbondante, P. (۲۰۱۰). Trading volume and stock indices: A test of technical analysis. researchgate.net. https://www.researchgate.net/profile/Paul-Abbondante-2/publication/49619617_Trading_Volume_and_Stock_Indices_A_Test_of_Technical_Analysis/links/5f170f6e4585151def3bfe47/Trading-Volume-and-Stock-Indices-A-Test-of-Technical-Analysis.pdf?_sg%2F.Bd%2FD=started_experiment_milestone&origin=journalDetail
- 4) Ahadzie, R. & Jeyasreedharan, N. A. A. & Ahadzie, R. & Jeyasreedharan, N. B. B. (۲۰۲۰). Trading volume and realized higher-order moments in the Australian stock market of the article. Journal of Behavioral and Experimental Finance. <https://www.sciencedirect.com/science/article/pii/S2214635020303403>
- 5) Ahmed, M. & Mahmood, A. A. A. & Ahmed, M. & Mahmood, A. B. B. (۲۰۱۳). A novel approach for outlier detection and clustering improvement of the article. ۲۰۱۳ IEEE ۸th Conference on Industrial Electronics and Applications. <https://ieeexplore.ieee.org/abstract/document/6566435>
- 6) Asgari, S. Yazdani, N. & Bokaei, M. N. A. A. & Asgari, S. Yazdani, N. & Bokaei, M. N. B. B. (۲۰۱۷). Predicting Index of Stock Exchange by Hidden Markov Model and K-Mean Algorithm of the article. Financial Knowledge of Securities Analysis, ۱۰(۳۶), ۸۲-۷۱. <http://sanad.iau.ir/fa/Article/803263>
- 7) Attari, M. Rafiq, S. & Awan, H. (۲۰۱۲). The dynamic relationship between stock volatility and trading volume. academia.edu. https://www.academia.edu/download/28/97406642_1097-201085.pdf
- 8) Chatzikonstanti, V. A. A. & Chatzikonstanti, V. B. B. (۲۰۱۷). Breaks and outliers when modelling the volatility of the US stock market of the article. Applied Economics. <https://doi.org/10.1080/09638185.2017.1293785>

- 9) Chutka, J. & Rebetak, F. A. A. & Chutka, J. & Rebetak, F. B. B. (۲۰۲۱). Analysis of trading volume and its use in prediction future price movements in the process of maximizing trading earnings of the article. SHS Web of Conferences. https://www.shs-conferences.org/articles/shsconf/abs/۰۳/۲۰۲۱/shsconf_glob۲۰_۰۲۰۱۰/shsconf_glob۲۰_۰۲۰۱۰.html
- 10) El-Ansary, O. & Atuea, M. (۲۰۱۲). The effect of stock trading volume on return in the Egyptian Stock Market. academia.edu. https://www.academia.edu/download/۴۱۷۲۵۲۲۹/The_Effect_of_Stock_Trading_Volume_on_Re۱-۱۵۲۲۵-۲۰۱۶۰۱۲۹nqsjow.pdf
- 11) González-Sánchez, M. A. A. & González-Sánchez, M. B. B. (۲۰۲۱). Is there a relationship between the time scaling property of asset returns and the outliers? Evidence from international financial markets of the article. Finance Research Letters. <https://www.sciencedirect.com/science/article/pii/S۱۵۴۴۶۱۲۳۱۹۳۰۶۷۷۴>
- 12) Nipane, S. Kale, P. Kapsekar, P. Agrawal, P. & Duddhure, O. (۲۰۲۱). Price trend prediction of stock market using outlier data mining algorithm.
- 13) Schluter, C. & Trede, M. A. A. & Schluter, C. & Trede, M. B. B. (۲۰۰۸). Identifying multiple outliers in heavy-tailed distributions with an application to market crashes of the article. Journal of Empirical Finance. <https://www.sciencedirect.com/science/article/pii/S۰۹۲۷۵۳۹۸۰۷۰۰۱۰۴۱>
- 14) Shehadeh, A. Alwadi, S ... & A. A. & Shehadeh, A. Alwadi, S. & ... B. B. (۲۰۲۲). Detecting and Analysing Possible Outliers in Global Stock Market Returns of the article. Cogent Economics & Finance. <https://doi.org/۲۳۳۲۲۰۳۹.۲۰۲۲.۲۰۶۶۷۶۲/۱۰.۱۰۸۰>
- 15) Suominen, M. A. A. & Suominen, M. B. B. (۲۰۰۱). Trading volume and information revelation in stock market of the article. Journal of Financial and Quantitative Analysis. <https://www.cambridge.org/core/journals/journal-of-financial-and-quantitative-analysis/article/trading-volume-and-information-revelation-in-stock-market/۴EE۷۱۲۰A۷۵۴۱D۱۰B۵۸۰FC۵۵۵DEEB۱BY۷>
- 16) Tran, T. Drab, K. & Daszykowski, M. A. A. & Tran, T. Drab, K. & Daszykowski, M. B. B. (۲۰۱۳). Revised DBSCAN algorithm to cluster data with dense adjacent clusters of the article. Chemometrics and Intelligent Laboratory Systems. <https://www.sciencedirect.com/science/article/pii/S۰۱۶۹۷۴۳۹۱۲۰۰۲۲۴۹>

17) Zhao, L. & Wang, L. A. A. & Zhao, L. & Wang, L. B. B. (۲۰۱۵). Price trend prediction of stock market using outlier data mining algorithm of the article. ۲۰۱۵ IEEE Fifth International Conference on Big Data and Cloud Computing. <https://ieeexplore.ieee.org/abstract/document/۷۳۱۰۷۲۲>

یادداشت‌ها:

-
- 1 Nipane
 - 2 Skewness
 - 3 Kurtosis
 - 4 Ahadzie & Jeyasreedharan
 - 5 Outlier data
 - 6 Shehadeh
 - 7 González-Sánchez
 - 8 Schluter & Trede
 - 9 K-Means
 - 10 Machine Learning - ML
 - 11 Density-Based Spatial Clustering of Applications with Noise - DBSCAN
 - 12 Chatzikonstanti
 - 13 El-Ansary & Atuea
 - 14 Ahmed & Mahmood
 - 15 Attari
 - 16 Chutka & Rebetak
 - 17 Abbondante
 - 18 Suominen
 - 19 Tran
 - 20 Moving Average Convergence Divergence
 - 21 Eps
 - 22 MinPts
 - 23 Core Points
 - 24 Border Points
 - 25 Noise Points
 - 26 Gaussian Standard Deviation Method - GSDM
 - 27 Correlation Analysis.
 - 28 Random Forest
 - 29 Feature Importance

Developing a Trading Strategy Based on Price Data Clustering Using the DBSCAN Machine Learning Algorithm and Identifying Noise Data

Nastaran Sharifirad¹

Receipt: 11/12/2024 Acceptance: 03/04/2025 Mohammadreza Reza Rostami²

Abstract

In this study, a trading strategy based on the DBSCAN clustering algorithm and outlier analysis is proposed. The objective is to identify optimal entry and exit points in financial markets using machine learning and technical analysis. The dataset includes price and volume information of 50 high-liquidity stocks from the Tehran Stock Exchange (TSE) and Iran's Over-the-Counter (OTC) market, covering the period from 15/09/1398 to 15/09/1403 extracted from the TSETMC database. After data preprocessing, standardization, and feature engineering, the DBSCAN algorithm was employed for data clustering and outlier detection. The extracted signals from DBSCAN were then combined with the MACD indicator to determine optimal trading entry and exit points. To evaluate the performance of this strategy, its results were compared with the Gaussian Standard Deviation (GSDM) analysis method and the buy-and-hold strategy. Additionally, the impact of DBSCAN parameters on clustering accuracy was analyzed, and the optimal values for trading signal detection were determined. The analyses were conducted in Python using libraries such as Algotik, NumPy, Pandas, Scikit-learn, and Matplotlib, and the results were assessed using statistical methods. The findings indicate that the proposed strategy achieved higher returns than traditional methods in most examined stocks and provided more precise trading signals. Furthermore, this approach can contribute to enhancing trade accuracy, reducing false signals, and optimizing investment decision-making. This method can be extended to other markets and applied to algorithmic trading systems, automated trading strategies, and long-term price trend analysis. The results suggest that integrating machine learning algorithms with technical analysis can provide a novel and efficient approach for optimizing trading strategies.

Keywords:

Trading Strategy, Machine Learning, Clustering, Outliers, Technical Analysis.

1-Master's Student, Department of Management, Faculty of Social Sciences and Economics, Alzahra University, Tehran, Iran. N.sharifirad@student.alzahra.ac.ir

2-Assistant Professor, Department of Management, Faculty of Social Sciences and Economics, Alzahra University, Tehran, Iran. (Corresponding Author) M.rostami@alzahra.ac.ir