

پیش‌بینی رفتار روندی سهام مبتنی بر تشخیص الگوهای تکراری

زهرة صدرنژاد^۱، مجید وفايي جهان^۲

^۱ کارشناسی ارشد نرم‌افزار کامپیوتر- دانشگاه آزاد اسلامی مشهد، Sadrnezhad@yahoo.com

^۲ گروه کامپیوتر- نرم‌افزار، دانشگاه آزاد اسلامی مشهد، VafaeiJahan@mshdiau.ac.ir

چکیده- در این مطالعه، رفتار روندی سهام با استخراج الگوهای مشابه تاریخی بر روی نمودار شاخص قدرت نسبی پیش‌بینی می‌شود؛ این نمودار سرعت و میزان تغییرات قیمت را نسبت به یک دوره‌ی زمانی اندازه‌گیری می‌کند و سطوح مختلفی دارد که با قرار گرفتن در آن سطوح رفتار خاصی برای آینده خواهد داشت. همچنین یک الگوی مشابه در هر سطحی احتمالاً روند خاصی را برای آینده دنبال می‌کند. در الگوریتم پیشنهادی، شباهت سنجی در دو مرحله بر اساس فاصله همینگ صورت می‌گیرد، در مرحله اول، الگوهای مشابه بر روی نمودار شاخص قدرت نسبی به همراه الگوهای متناظر بر روی نمودار قیمت استخراج می‌شوند و در مرحله دوم، شباهت سنجی بر روی الگوهای قیمت استخراج شده انجام می‌شود، زیرا نوسانات مشابه بر روی نمودار شاخص قدرت نسبی، ممکن است دارای نوسانات مشابه قیمت نباشند. در نهایت طبق بازه زمانی که روند آینده سهام پیش‌بینی می‌شود، نوسانات آینده در الگوهای استخراج شده در نظر گرفته می‌شود، سپس یک نمودار بر اساس تابع میانگین‌گیری در بین آن‌ها برآزش می‌شود تا از تأثیر تمام الگوها در پیش‌بینی رفتار روندی سهام استفاده شود. برای ارزیابی روش پیشنهادی از مجموعه داده سهام فولاد اصفهان در بازار بورس اوراق بهادار ایران استفاده شده است، همچنین برای پیش‌بینی میان‌مدت رفتار روندی سهام، روش پیشنهادی با روش‌های میانگین متحرک یکپارچه‌ی خود کاهنده و اتو رگرسیون مقایسه می‌شود، نتایج نشان می‌دهد روش پیشنهادی برای پیش‌بینی روند دوهفته‌ای دارای ۷۸٪، روند ماهانه ۸۰٪ و روند دوماهه ۷۳٪ صحت است. کلیدواژه- پیش‌بینی، نوسانات مشابه، شاخص قدرت نسبی، فاصله همینگ.

۱- مقدمه

تحلیلگران حرفه‌ای معمولاً برای پیش‌بینی سهام حوادث تاریخی یا الگوهایی که دوباره تکرار می‌شوند و در حرکت بازار سهام منعکس شده است را در نظر می‌گیرند [۹، ۱۰]، از جمله روش‌های تجزیه تحلیل برای پیش‌بینی قیمت سهام استفاده از برخی شاخص‌های فنی و تشخیص الگو است [۱۱، ۱۲]، در این روش‌ها از تطبیق الگوهای زمانی و استخراج الگوهای تاریخی مشابه برای پیش‌بینی قیمت سهام استفاده شده است و شباهت بین دو الگو زمانی با استفاده از اندازه‌گیری فاصله مشخص می‌شود. معیارهای مختلفی برای تعیین فاصله بین الگوهای زمانی وجود دارد مانند فاصله اقلیدسی، فاصله همینگ، Hausdorff distance، HMM-based distance [۱۳].

در این پژوهش، برای پیش‌بینی روند قیمت سهام از روش استخراج الگوهای مشابه تاریخی استفاده شده است. در الگوریتم پیشنهادی، الگوهای مشابه بر روی نمودار شاخص قدرت نسبی استخراج شده، زیرا این نمودار سطوح مختلفی دارد که با قرار گرفتن در آن سطوح رفتار خاصی برای آینده خواهد داشت، همچنین یک الگوی مشابه در هر سطحی احتمالاً روند خاصی را برای آینده دنبال می‌کند، اما در نمودار قیمت تفاوتی بین الگوهای مشابه استخراج شده وجود ندارد. همچنین برای

امروزه سرمایه‌گذاری به بخش جداناپذیر از زندگی روزمره تبدیل شده و بسیاری از مردم تلاش دارند با استفاده از دارایی‌های شخصی خود سود بیشتری کسب کنند. سرمایه‌گذاری در سهام با وجود ریسک بالا، یکی از محبوب‌ترین سرمایه‌گذاری‌ها است و می‌تواند منجر به سوددهی بالا شود، اما بازار سهام تحت تأثیر عواملی از جمله اقتصاد کلان و بین‌المللی، نوسانات بازار سهام و سیاست دولت است و پیش‌بینی قیمت را دشوار می‌سازد [۱، ۲].

تجزیه و تحلیل اطلاعات موجود در بازار سهام به منظور پیش‌بینی مؤثر روند آینده سهام انجام می‌شود. در بازار سهام، تجزیه و تحلیل فنی برای پیش‌بینی قیمت سهام در کوتاه‌مدت استفاده می‌شود [۳، ۴]. فرضیه بازار کارآمد معتقد است قیمت بازار به طور کامل منعکس‌کننده تمام اطلاعات موجود است و نوسانات در قیمت سهام کاملاً تصادفی است [۵]؛ این فرضیه ارزش تجزیه و تحلیل فنی را از بین می‌برد، اما مطالعات انجام‌شده در سال‌های اخیر نشان می‌دهد تجزیه و تحلیل فنی برای پیش‌بینی سهام مفید است [۶-۸].

محاسبه میزان فاصله، از روش‌های مختلفی از جمله فاصله اقلیدسی، فاصله همینگ و ... در کارهای مرتبط استفاده شده است؛ در این مطالعه، از فاصله همینگ برای محاسبه شباهت استفاده می‌شود، زیرا پیاده‌سازی آن آسان است و آن را به عنوان یک معیار شباهت پایه می‌توان در نظر گرفت.

ساختار مقاله به این شرح است: در بخش ۲، به بررسی کارهای گذشته پرداخته می‌شود، بخش ۳، شامل تعریف مفاهیم پایه، بخش ۴، تشریح الگوریتم پیشنهادی، بخش ۵، ارزیابی الگوریتم پیشنهادی و در بخش ۶، نتیجه‌گیری کلی از این مطالعه بیان می‌شود.

۲- کارهای گذشته

در طول سال‌های اخیر روش‌های مختلفی برای پیش‌بینی قیمت سهام از جمله مدل‌های آماری مانند اتو رگرسیون^۱ (AR)، میانگین متحرک خودگردان^۲ (ARMA)، و به تازگی روش‌های یادگیری ماشین مانند مدل مخفی مارکوف [۱۳-۱۵]، شبکه‌های عصبی [۱۶، ۱۷] و ماشین بردار پشتیبانی [۱۸] ارائه شده است. علاوه بر این، برخی از محققان از تجزیه تحلیل فنی و استخراج اطلاعات از گزارش‌های مالی برای پیش‌بینی حرکت قیمت سهام استفاده کرده‌اند [۱۹-۲۱].

تجزیه تحلیل فنی به دو دسته کمی و گرافیکی تقسیم شده است [۲۲]. تجزیه و تحلیل کمی با استفاده از شاخص‌های فنی مانند شاخص قدرت نسبی^۴ (RSI)، میانگین متحرک همگرایی / واگرایی و ... به پیش‌بینی روند سهام می‌پردازد، در حالی که تجزیه و تحلیل گرافیکی با استفاده از نمودار پیش‌بینی را انجام می‌دهد. در این بخش چند مطالعه که از تشخیص الگو و نمودارهای تجزیه تحلیل فنی در پیش‌بینی بازار سهام استفاده کرده‌اند، بررسی می‌شود.

Richi Nayak در سال ۲۰۰۷ [۳]، برای پیش‌بینی قیمت سهام از تطبیق الگو زمانی استفاده کرد و یک خوشه بر اساس روش شبه نزدیک‌ترین همسایه^۵ (KNN)، در سرتاسر مجموعه داده تشکیل داد و مجموعه‌های متوالی از اطلاعات قیمت سهام را استخراج کرد، سپس شبیه‌ترین الگو را به عنوان خروجی در نظر گرفت، همچنین از فاصله اقلیدسی برای اندازه‌گیری فاصله استفاده شد. Dattashann و همکارانش در سال ۲۰۰۸ [۲۳]، بر اساس مفاهیم رویداد، بخش‌بندی و استراتژی تجارت جفت، به

پیش‌بینی قیمت سهام پرداختند و داده‌های سری زمانی را کدگذاری و بخش‌های مختلف داده‌ها را در یک جهت پی‌درپی زمان پیدا کردند، همچنین فاصله بین سری‌های کدگذاری شده با فاصله لوناشتاین و فاصله Jaro-Winkler اندازه‌گیری شد. Hon Leon Lee و همکارانش در سال ۲۰۱۱ [۲۴]، تصمیمات سرمایه‌گذاری را با شناسایی الگو در نمودار شمعی گرفتند و با استفاده از متغیرهای زبانی- فازی الگوهای نمودار شمعی را استخراج و از الگوریتم ژنتیک برای انتخاب مجموعه‌ای از الگوهای استخراج‌شده، استفاده کردند. Chih-Fong Tsai و همکارش در سال ۲۰۱۴ [۲۲]، از موجک گسسته^۶ (DWT) که یک الگوریتم پردازش تصویر است در تجزیه و تحلیل محتوای تصویری نمودار شمعی استفاده کردند. ابتدا هفت ویژگی مختلف از نمودار شمعی، سپس دوره‌های مشابه در نمودار استخراج شد و در مرحله بازایی تصویر، از فاصله اقلیدسی برای اندازه‌گیری شباهت میان نمودارهای شمعی استخراج‌شده، استفاده شده است، در نهایت، نموداری که بیشترین شباهت را داشت به عنوان خروجی انتخاب شده است.

جدول (۱)، خلاصه‌ای از مطالعات بررسی شده در این بخش است؛ با توجه به این جدول، معمولاً الگوهای مشابه بر روی نمودار قیمت و نمودار شمعی استخراج می‌شود، اما از آن جهت که تفاوتی بین محل قرارگیری الگوهای مشابه در نمودار قیمت و نمودار شمعی وجود ندارد، در این مطالعه از نمودار شاخص قدرت نسبی استفاده می‌شود، این نمودار سطوح مختلفی دارد و الگوهای مشابه در هر سطح احتمالاً روند خاصی را برای آینده دنبال می‌کنند، بنابراین برای تعیین الگوهای مشابه علاوه بر فاصله، محل قرارگیری الگو در نمودار نیز مهم است.

۳- تعریف مفاهیم پایه

۳-۱- شاخص قدرت نسبی (Relative Strength Index)

شاخص قدرت نسبی (RSI)، یک شاخص فنی است و در تحلیل بازارهای مالی استفاده می‌شود. این شاخص سرعت و میزان تغییرات قیمت را نسبت به یک دوره‌ی زمانی اندازه‌گیری می‌کند (معمولاً دوره‌ی زمانی ۱۴ روز در نظر گرفته می‌شود) و برای تعیین زمان خرید و فروش سهام استفاده می‌شود. شاخص RSI بین دو مقدار ۰ تا ۱۰۰ نوسان می‌کند و در هر سطحی رفتاری متفاوت دارد، دو سطح مهم این شاخص، سطوح ۳۰ و ۷۰ است که به ترتیب نشان‌دهنده سطوح "بیش فروش" و

^۴ Relative Strength Index

^۵ K-Nearest Neighbor

^۶ Discrete wavelet transform

^۱ Autoregressive

^۲ Autoregressive Moving Average

^۳ Autoregressive Integrated Moving Average

مفهوم "بیش خرید"، زمانی بکار می‌رود که یک سهم به‌صورت افراطی و غیرعادی خریداری می‌شود، در این زمان شاخص RSI به سمت سطوح بالای ۷۰ متمایل می‌شود و احتمال کاهش قیمت‌ها وجود دارد. همچنین مفهوم "بیش فروش"، زمانی بکار می‌رود که یک سهم به‌صورت افراطی و غیرعادی فروخته می‌شود، در این زمان شاخص RSI به سمت سطوح زیر ۳۰ متمایل می‌شود و احتمال افزایش قیمت‌ها وجود دارد [۲۵].

۲-۳- فاصله همینگ

فاصله‌ی همینگ یک معیار اندازه بر روی فضای برداری است و در تئوری اطلاعات برای دو الگو با طول مساوی، برابر کمترین فاصله‌ی موردنیاز است که یک الگو به الگوی دیگر تبدیل شود. برای محاسبه‌ی فاصله‌ی همینگ از رابطه (۳) استفاده می‌شود:

$$HD = \frac{1}{n} \sum_{i=1}^n |x_i - y_i| \quad (3)$$

در رابطه (۳)، n طول الگو، x_i مقدار روزانه (RSI/ قیمت) نوسانات جاری سهام و y_i مقدار روزانه (RSI/ قیمت) الگوهای استخراج‌شده است. $HD=0$ ، نشان‌دهنده‌ی شباهت ۱۰۰٪ هست، اما هر چه مقدار HD از صفر دورتر و به سمت اعداد مثبت میل کند، شباهت کمتر می‌شود.



شکل ۱: سطوح بیش خرید و بیش فروش در نمودار شاخص قدرت نسبی

برای محاسبه شاخص RSI، ابتدا مقدار RS مطابق رابطه (۱) محاسبه می‌شود، در این رابطه $EMA(U,N)$ میانگین تغییرات مثبت قیمت‌های پایانی، $EMA(D,N)$ میانگین تغییرات منفی قیمت‌های پایانی در یک دوره زمانی N روزه است، سپس مقدار نهایی RSI مطابق رابطه (۲) محاسبه می‌شود:

$$RS = \frac{EMA(U, N)}{EMA(D, N)} \quad (1)$$

$$RSI = 100 - \frac{100}{1 + RS} \quad (2)$$

جدول ۱: بررسی چند مقاله که از تکنیک الگوهای مشابه در نوسانات گذشته، برای پیش‌بینی بازارهای مالی استفاده کرده‌اند

مقاله	روش پیش‌بینی	روش‌های مقایسه	روش اندازه‌گیری شباهت	نمودار استفاده‌شده	معیارهای مقایسه
[۳]	روش شبه نزدیک‌ترین همسایه	-	فاصله اقلیدسی	نمودار قیمت	مقایسه با مقادیر واقعی - نرخ خطا
[۲۳]	استفاده از مفاهیم رویداد، بخش‌بندی و استراتژی تجارت جفت	-	فاصله لوناشتاین - Jaro-Winkler	نمودار قیمت	ضریب همبستگی - فاصله
[۲۴]	متغیرهای زمانی - فازی، الگوریتم ژنتیک	الگوهای غیر فازی	-	نمودار شمعی قیمت	مقایسه بین مجموعه الگوهای متفاوت
[۲۲]	استفاده از الگوریتم موجک گسسته (DWT)	GARCH- SVR- BPN	فاصله اقلیدسی	نمودار شمعی قیمت	Accuracy, Precision, Recall, F-measure- MAPE- Theil's U

ضریب تیل (Coefficient THEIL)، واریانس ناهمسانی شرطی اتو رگرسیون (GARCH)، رگرسیون بردار پشتیبانی (SVR)، شبکه پس انتشار (BPN)، میانگین درصد خطای مطلق (MAPE).

الگوهای مشابه استفاده شده است. در الگوریتم پیشنهادی، شباهت سنجی در دو مرحله انجام می‌شود، مرحله اول بر روی نمودار شاخص قدرت نسبی و در مرحله دوم بر روی نمودار قیمت الگوهای استخراج‌شده، همچنین برای محاسبه‌ی شباهت

۴- الگوریتم پیشنهادی

در این مطالعه، برای پیش‌بینی میان مدت (دو هفته، ماهانه و دوماهه) رفتار روندی سهام از تجزیه تحلیل فنی و استخراج

بین نوسانات جاری سهام با الگوهای مختلف، از فاصله‌ی همینگ استفاده شده است.

۴-۱- مجموعه داده

برای ارزیابی روش پیشنهادی از مجموعه داده سهام فولاد اصفهان در بازار بورس اوراق بهادار ایران از تاریخ ۲۰۰۷/۳/۱۱ تا ۲۰۱۴/۱۱/۲۲ استفاده شده [۲۶]، همچنین برای آموزش مدل از ۲۰۰۷/۳/۱۱ تا ۲۰۱۲/۱/۱۰ و تست مدل از ۲۰۱۲/۱/۱۱ تا ۲۰۱۴/۱۱/۲۲ استفاده شده است.

۴-۲- مراحل الگوریتم پیشنهادی

در این بخش مراحل الگوریتم پیشنهادی برای پیش‌بینی میان‌مدت رفتار روندی سهام بیان شده است.

مرحله ۱: ابتدا مجموعه داده قیمت سهام به تابع شاخص قدرت نسبی داده می‌شود (مقادیر شاخص RSI با دوره‌ی تناوب ۱۴ روز محاسبه می‌شود)، سپس نوسانات جاری RSI، به همراه الگوهای مختلف RSI در نوسانات گذشته سهام (طول الگوهای استخراج‌شده با طول الگوی نوسانات جاری سهام برابر است)، استخراج می‌شوند.

مرحله ۲: در این مرحله، میزان شباهت بین نوسانات جاری RSI با الگوهایی که در مرحله‌ی قبل استخراج‌شده، بر اساس فاصله همینگ محاسبه و الگوهایی که بیشترین شباهت را دارند به همراه الگو قیمت متناظرشان استخراج می‌شوند، اما ممکن است همیشه شبیه‌ترین الگو در سطح نوسانات جاری RSI نباشد، بنابراین میزانی نويز در نظر گرفته می‌شود تا الگویی مشابه در سطح نوسانات جاری RSI استخراج شود (شکل ۲).

مرحله ۳: در این مرحله آزمون شباهت سنجی بر روی نمودار قیمت الگوهای استخراج‌شده در مرحله‌ی قبل اعمال می‌شود، زیرا نوسانات مشابه RSI ممکن است دارای نوسانات مشابه قیمت نباشند، بنابراین طبق مرحله‌ی قبل، شبیه‌ترین الگوها بر روی نمودار قیمت استخراج می‌شوند (شکل ۳).

مرحله ۴: در نهایت، برای پیش‌بینی میان‌مدت رفتار روندی سهام، مطابق بازه زمانی (۲ هفته، ماهانه و ۲ ماهه) که پیش‌بینی انجام می‌شود، نوسانات آینده در الگوهای مشابه استخراج‌شده، در نظر گرفته می‌شود، سپس مقادیر در یک رنج عددی قرار می‌گیرد و یک نمودار بر اساس تابع میانگین‌گیری (رابطه ۴) در بین آن‌ها برازش می‌شود تا از تأثیر تمام الگوها در پیش‌بینی روند سهام استفاده شود، نمودار حاصل، پیش‌بینی

روند آینده سهام را نشان می‌دهد (شکل ۴).

$$\text{Average} = \frac{\sum_{i=1}^n x_i}{n} \quad (4)$$

در رابطه فوق مقادیر متناظر نمودارها باهم جمع و بر تعداد نمودارها تقسیم می‌شوند. n ، تعداد الگوهای نهایی استخراج‌شده و x_i ، قیمت سهام در هر روز است.

۴-۳- تعیین مقادیر اولیه الگوریتم

در الگوریتم پیشنهادی همان‌طور که گفته شد، نیاز است طول دوره‌ی الگوهای مشابه مشخص شود، بنابراین از آزمون صحت، میانگین درصد خطای مطلق $(MAPE)^y$ و ریشه میانگین مربع خطا $(RMSE)^x$ بر روی مجموعه داده‌ی سهام فولاد اصفهان استفاده می‌شود. در این آزمون‌ها طول دوره‌ی الگوهای مشابه ۱۴، ۳۰، ۴۵ و ۶۰ روز در نظر گرفته شده است. جدول (۲)، نتایج حاصل از آزمون Accuracy بر روی مجموعه داده‌ی سهام فولاد اصفهان، جدول (۳) نتایج حاصل از آزمون MAPE و جدول (۴) نتایج حاصل از آزمون RMSE بر روی مقادیر مشخص‌شده برای طول دوره‌ی الگوهای مشابه است. نتایج نشان‌دهنده‌ی این است که مقادیر MAPE، RMSE در طول الگو ۶۰ روز کمترین مقدار را دارد و مقادیر پیش‌بینی شده به مقادیر واقعی نزدیک‌تر است. بنابراین طول الگو ۶۰ روز در نظر گرفته می‌شود.

۵- ارزیابی الگوریتم پیشنهادی

۵-۱- معیارهای ارزیابی

معیارهایی که روش پیشنهادی با آن‌ها مورد سنجش قرار داده شده است، به شرح زیر می‌باشند:

۵-۱-۱- میانگین درصد خطای مطلق (MAPE)

میانگین درصد خطای مطلق (MAPE)، معیاری برای اندازه‌گیری دقت و صحت است، همچنین میزان خطا در پیش‌بینی را مشخص می‌کند و به صورت درصد بیان می‌شود. زمانی که تناسب کامل است، $MAPE=0$ و هر چه مقدار آن از صفر بیشتر می‌شود خطا در پیش‌بینی افزایش و دقت کاهش می‌یابد. برای محاسبه‌ی MAPE از رابطه (۵) استفاده می‌شود، $actual(t)$ نشان‌دهنده‌ی مقدار واقعی، $forecast(t)$ مقدار پیش‌بینی شده در زمان t و n طول الگو است.

[^] Root Mean Square Error

^y Mean Absolute Percentage Error

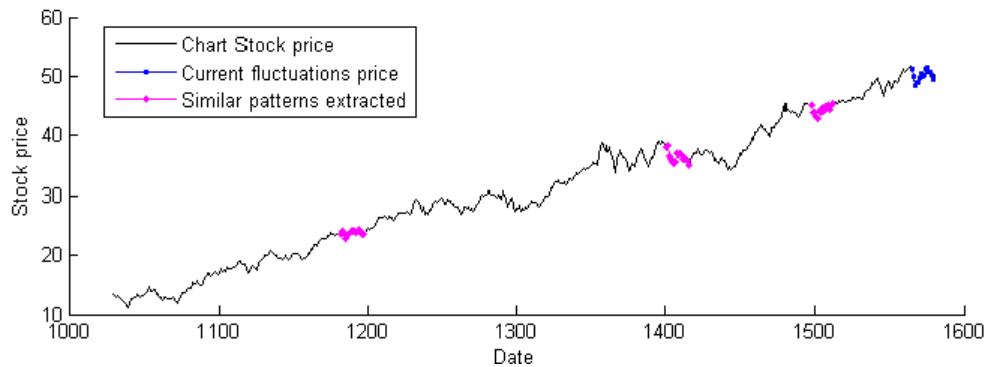
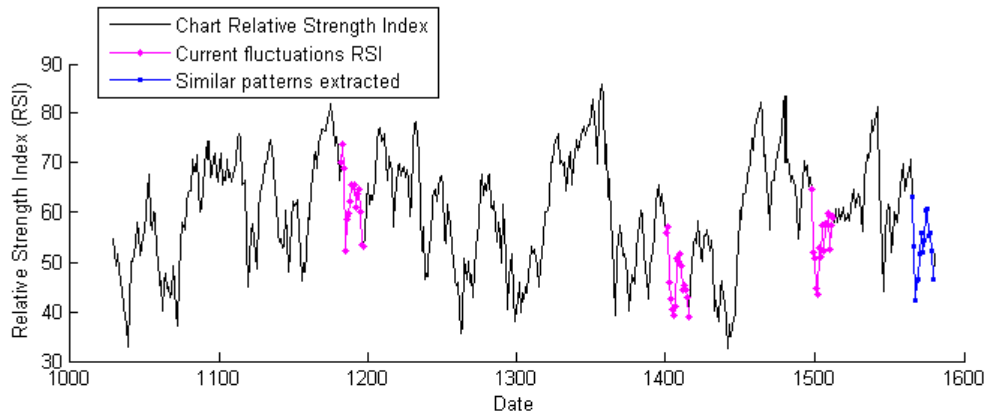
نشان‌دهنده انحراف نمونه استاندارد از تفاوت بین آن‌ها است. برای محاسبه‌ی RMSE از رابطه (۶) استفاده می‌شود، $actual(t)$ نشان‌دهنده‌ی مقدار واقعی، $forecast(t)$ مقدار پیش‌بینی شده در زمان t و n طول الگو است.

$$RMSE = \sqrt{\frac{\sum_{t=1}^n (actual(t) - forecast(t))^2}{n}} \quad (۶)$$

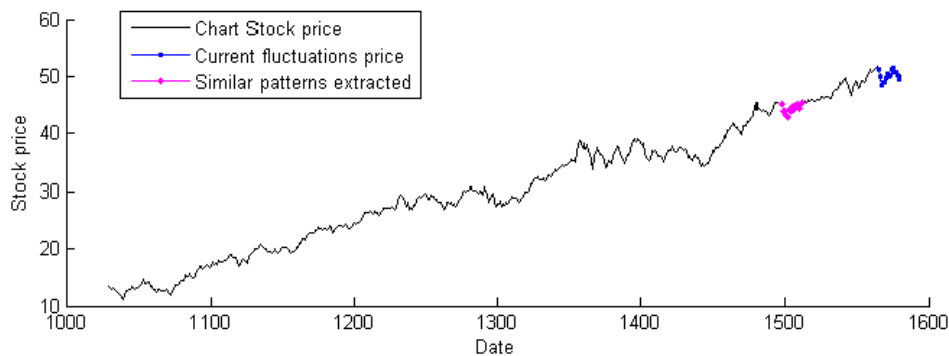
$$MAPE = \frac{\sum_{t=1}^n \frac{|actual(t) - forecast(t)|}{actual(t)}}{n} * 100 \quad (۵)$$

۲-۱-۵- ریشه میانگین مربع خطا (RMSE)

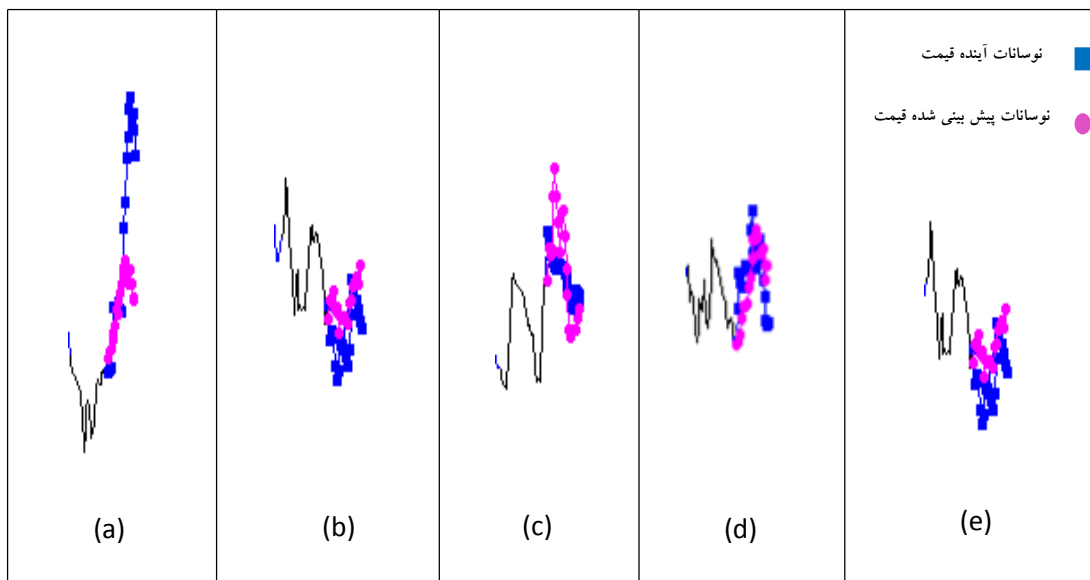
ریشه میانگین مربع خطا (RMSE)، برای تعیین تفاوت بین مقادیر واقعی با مقادیر پیش‌بینی شده استفاده می‌شود و



شکل ۲: الگوهای مشابه نوسانات جاری شاخص RSI، به همراه الگوهای متناظر قیمت



شکل ۳: الگوهای مشابه نوسانات جاری قیمت سهام



شکل ۴: پنج نمونه از پیش‌بینی‌های میان‌مدت روند سهام فولاد اصفهان بر اساس الگوریتم پیشنهادی

جدول ۲: نتایج حاصل از آزمون Accuracy برای پیش‌بینی‌های میان‌مدت بر روی مجموعه داده‌ی سهام فولاد اصفهان

طول الگو	مدت‌زمان پیش‌بینی میان‌مدت											
	دو هفته				۱ ماه				دو ماه			
	۲۰۱۲	۲۰۱۳	۲۰۱۴	میانگین صحت	۲۰۱۲	۲۰۱۳	۲۰۱۴	میانگین صحت	۲۰۱۲	۲۰۱۳	۲۰۱۴	میانگین صحت
۱۵	۰.۷۰	۰.۸۰	۰.۶۵	۰.۷۱	۰.۷۰	۰.۷۵	۰.۷۰	۰.۷۱	۰.۶۵	۰.۸۵	۰.۶۰	۰.۷۰
۳۰	۰.۷۰	۰.۷۰	۰.۶۵	۰.۶۸	۰.۶۵	۰.۸۰	۰.۶۵	۰.۷۰	۰.۷۰	۰.۸۵	۰.۵۰	۰.۶۸
۴۵	۰.۷۰	۰.۶۵	۰.۶۰	۰.۶۵	۰.۷۰	۰.۶۵	۰.۶۵	۰.۶۶	۰.۶۵	۰.۷۵	۰.۵۰	۰.۶۳
۶۰	۰.۹۰	۰.۷۵	۰.۷۰	۰.۷۸	۰.۸۵	۰.۸۰	۰.۷۵	۰.۸۰	۰.۸۰	۰.۸۰	۰.۶۰	۰.۷۳

جدول ۳: نتایج حاصل از آزمون MAPE برای پیش‌بینی‌های میان‌مدت بر روی مجموعه داده‌ی سهام فولاد اصفهان

طول الگو	مدت‌زمان پیش‌بینی میان‌مدت											
	دو هفته				۱ ماه				دو ماه			
	۲۰۱۲	۲۰۱۳	۲۰۱۴	میانگین MAPE	۲۰۱۲	۲۰۱۳	۲۰۱۴	میانگین MAPE	۲۰۱۲	۲۰۱۳	۲۰۱۴	میانگین MAPE
۱۵	۱۶	۲۵	۱۵	۱۸	۲۲	۳۰	۲۰	۲۴	۲۹	۳۹	۳۰	۳۲
۳۰	۱۴	۲۳	۱۷	۱۸	۲۰	۲۸	۲۳	۲۳	۲۸	۳۳	۲۹	۳۰
۴۵	۲۰	۱۹	۲۰	۱۹	۲۴	۲۵	۲۴	۲۴	۳۳	۳۰	۳۰	۳۱
۶۰	۱۵	۱۸	۱۶	۱۶	۲۵	۲۶	۲۰	۲۳	۲۹	۲۸	۲۴	۲۷

جدول ۴: نتایج حاصل از آزمون RMSE برای پیش‌بینی‌های میان‌مدت بر روی مجموعه داده‌ی سهام فولاد اصفهان

طول الگو	مدت‌زمان پیش‌بینی میان‌مدت											
	دو هفته				۱ ماه				دو ماه			
	۲۰۱۲	۲۰۱۳	۲۰۱۴	میانگین RMSE	۲۰۱۲	۲۰۱۳	۲۰۱۴	میانگین RMSE	۲۰۱۲	۲۰۱۳	۲۰۱۴	میانگین RMSE
۱۵	۱۹	۲۸	۱۸	۲۱	۲۶	۳۲	۲۴	۲۷	۲۹	۳۹	۳۳	۳۳
۳۰	۱۷	۲۶	۲۱	۲۱	۲۴	۳۰	۲۵	۲۶	۲۸	۳۳	۳۲	۳۱
۴۵	۲۳	۲۲	۲۳	۲۲	۲۵	۲۹	۲۸	۲۷	۳۳	۳۰	۳۴	۳۲
۶۰	۱۹	۲۸	۱۸	۲۱	۲۶	۳۲	۲۴	۲۷	۲۹	۳۹	۳۳	۳۳

۲-۱-۵- آزمون‌های Accuracy, Precision, Recall و F-measure

برای محاسبه‌ی آزمون‌های Accuracy, Precision, Recall و F-measure به متغیرهایی که در جدول (۵) نشان داده شده است، نیاز است.

آزمون صحت (Accuracy)، درصد دقت در پیش‌بینی روند را نشان می‌دهد و مطابق رابطه (۷) محاسبه می‌شود:

$$\text{Accuracy} = \frac{TP + TN}{P + N} \quad (7)$$

آزمون Precision، میزان دقت در یک کلاس خاص از پیش‌بینی را نشان می‌دهد و مطابق رابطه (۸) محاسبه می‌شود:

$$\text{Precision} = \frac{TP}{TP + FP} \quad (8)$$

آزمون Recall، توانایی مدل در پیش‌بینی یک کلاس خاص از مجموعه داده‌ها را اندازه‌گیری می‌کند و مطابق رابطه (۹) محاسبه می‌شود:

$$\text{Recall} = \frac{TP}{TP + FN} \quad (9)$$

آزمون F-measure یک معیار اندازه‌گیری برای تعیین دقت و جامعیت، همچنین شاخص خوبی برای تعیین عملکرد مدل است و برای محاسبه‌ی آن، مطابق رابطه (۱۰)، از مقادیر Precision و Recall استفاده می‌شود:

$$F - \text{measure} = 2 * \frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}} \quad (10)$$

جدول ۵: متغیرهای موردنیاز برای محاسبه‌ی آزمون‌ها

متغیرها	شرح متغیر
TP	روند افزایشی، پیش‌بینی هم‌روند افزایشی است.
FN	روند افزایشی، پیش‌بینی روند کاهشی است.
TN	روند کاهشی، پیش‌بینی هم‌روند کاهشی است.
FP	روند کاهشی، پیش‌بینی روند افزایشی است.
P	تعداد روندهای افزایشی
N	تعداد روندهای کاهشی

۲-۵- روش‌های مقایسه شده با روش پیشنهادی

در این مطالعه روش پیشنهادی با روش‌های میانگین متحرک یکپارچه‌ی خود کاهنده (ARIMA) و اتو رگرسیون (AR) مقایسه شده است.

۱-۲-۵- میانگین متحرک یکپارچه‌ی خود کاهنده (ARIMA)

مدل میانگین متحرک یکپارچه‌ی خود کاهنده (ARIMA)، عمومی‌ترین مدل برای پیش‌بینی سری‌های زمانی است و درجایی که داده‌ها غیر ایستا باشند برای فهم بهتر مدل یا

پیش‌بینی آینده به کار می‌رود. این مدل در اکثر موارد به صورت ARIMA(p,q,d) نشان داده می‌شود که در آن p، q و d اعداد حقیقی غیر منفی هستند و درجه خودگردانی، یکپارچگی و میانگین متحرک را مشخص می‌کنند.

۲-۲-۵- اتو رگرسیون برداری (AR)

اتو رگرسیون برداری (AR)، یک مدل آماری است و وابستگی خطی میان چند سری زمانی را بیان می‌کند، همچنین آینده سری زمانی بر اساس مجموع وزنی ارزش گذشته، برآورد می‌شود. مدل (۲) AR، ارزش فعلی را بر اساس دو مقدار قبلی محاسبه می‌کند.

۳-۵- نتایج آزمایش‌ها

جدول (۶) نتایج حاصل از مقایسه‌ی روش پیشنهادی با روش‌های ARIMA و AR، بر اساس آزمون Precision، جدول (۷) نتایج آزمون Recall، جدول (۸) نتایج آزمون F-measure، جدول (۹) نتایج آزمون Accuracy، جدول (۱۰) نتایج آزمون MAPE و جدول (۱۱) نتایج حاصل از آزمون RMSE بر روی مجموعه داده‌ی سهام فولاد اصفهان است. نتایج آزمایش‌ها نشان می‌دهد، روش پیشنهادی برای پیش‌بینی‌های دوهفته‌ای دارای ۷۸٪، پیش‌بینی ماهانه ۸۰٪ و پیش‌بینی دوماهه ۷۳٪ صحت است، همچنین نشان‌دهنده‌ی برتری روش پیشنهادی نسبت به روش‌های AR و ARIMA است.

۶- نتیجه‌گیری

پیش‌بینی بازارهای مالی یک موضوع جذاب برای پژوهشگران است و روش‌های متفاوتی برای آن ارائه شده، از جمله استخراج الگوهای مشابه در نوسانات گذشته که قابلیت پیش‌بینی احتمالی را دارد. در این مقاله، با استخراج الگوهای مشابه تاریخی بر روی نمودار شاخص قدرت نسبی، به پیش‌بینی رفتار روندی سهام فولاد اصفهان در بازار بورس اوراق بهادار ایران پرداخته شده است. در الگوریتم پیشنهادی، شباهت سنجی در دو مرحله انجام شد، مرحله اول بر روی نمودار شاخص قدرت نسبی و مرحله دوم بر روی نمودار قیمت، همچنین برای محاسبه‌ی شباهت بین نوسانات جاری سهام با الگوهای مختلف، از فاصله‌ی همینگ استفاده شده است. روش پیشنهادی با روش‌های میانگین متحرک یکپارچه‌ی خود کاهنده (ARIMA) و اتو رگرسیون (AR)، بر اساس معیارهای Accuracy, F-measure, Recall, Precision و MAPE، RMSE مقایسه شد. نتایج نشان داد، روش پیشنهادی برای

طولانی‌تر یا کوتاه‌تر منجر به ایجاد الگوی مشابه شود. در مطالعه بعد، می‌خواهیم الگوهای مشابه را با طول‌های مختلف استخراج کنیم و بر اساس آن‌ها رفتار روندی سهام را مورد پیش‌بینی قرار دهیم.

پیش‌بینی‌های ۲ هفته‌ای دارای ۷۸٪، پیش‌بینی ماهانه ۸۰٪ و پیش‌بینی دوماهه ۷۳٪ صحت است. الگوهایی که مشابه نوسانات جاری سهام است، ممکن است در نوساناتی با همان طول الگوی مشخص رخ ندهد و در بازه‌ای

جدول ۶: مقایسه روش پیشنهادی با روش AR و ARIMA بر اساس آزمون Precision در پیش‌بینی‌های میان‌مدت

روش‌ها	مدت‌زمان پیش‌بینی کوتاه‌مدت								
	دو هفته			۱ ماه			دو ماه		
	۲۰۱۲	۲۰۱۳	۲۰۱۴	۲۰۱۲	۲۰۱۳	۲۰۱۴	۲۰۱۲	۲۰۱۳	۲۰۱۴
AR (۲)	۰.۵۰	۰.۴۰	۰.۳۵	۰.۴۰	۰.۵۷	۰.۳۳	۰.۴۲	۰.۴۴	۰.۶۰
ARIMA (۰۰۰۰۲)	۰.۳۳	۰.۷۵	۰.۶۶	۰.۴۲	۰.۴۴	۰.۴۱	۰.۵۰	۰.۴۴	۰.۲۵
روش پیشنهادی	۰.۷۷	۰.۷۲	۰.۷۵	۰.۸۷	۰.۹۰	۰.۷۰	۰.۹۰	۰.۸۵	۰.۶۵

جدول ۷: مقایسه روش پیشنهادی با روش AR و ARIMA بر اساس آزمون Recall در پیش‌بینی‌های میان‌مدت

روش‌ها	مدت‌زمان پیش‌بینی کوتاه‌مدت								
	دو هفته			۱ ماه			دو ماه		
	۲۰۱۲	۲۰۱۳	۲۰۱۴	۲۰۱۲	۲۰۱۳	۲۰۱۴	۲۰۱۲	۲۰۱۳	۲۰۱۴
AR (۲)	۰.۵۵	۰.۵۰	۰.۴۵	۰.۵۰	۰.۳۶	۰.۳۳	۰.۳۰	۰.۳۳	۰.۵
ARIMA (۰۰۰۰۲)	۰.۲۵	۰.۶۰	۰.۵۴	۰.۳۳	۰.۴۰	۰.۵۰	۰.۴۲	۰.۴۰	۰.۳۵
روش پیشنهادی	۰.۷۷	۰.۶۶	۰.۸۰	۰.۷۰	۰.۸۰	۰.۷۱	۰.۸۱	۰.۸۰	۰.۶۵

جدول ۸: مقایسه روش پیشنهادی با روش AR و ARIMA بر اساس آزمون F-measure در پیش‌بینی‌های میان‌مدت

روش‌ها	مدت‌زمان پیش‌بینی کوتاه‌مدت								
	دو هفته			۱ ماه			دو ماه		
	۲۰۱۲	۲۰۱۳	۲۰۱۴	۲۰۱۲	۲۰۱۳	۲۰۱۴	۲۰۱۲	۲۰۱۳	۲۰۱۴
AR (۲)	۰.۵۲	۰.۴۴	۰.۳۹	۰.۴۴	۰.۴۵	۰.۳۳	۰.۳۵	۰.۳۸	۰.۵۴
ARIMA (۰۰۰۰۲)	۰.۲۸	۰.۶۶	۰.۵۹	۰.۳۷	۰.۴۱	۰.۴۵	۰.۴۶	۰.۴۱	۰.۲۹
روش پیشنهادی	۰.۷۷	۰.۶۹	۰.۷۷	۰.۷۷	۰.۸۴	۰.۷۰	۰.۸۶	۰.۸۲	۰.۶۵

جدول ۹: مقایسه روش پیشنهادی با روش AR و ARIMA بر اساس آزمون Accuracy در پیش‌بینی‌های میان‌مدت

روش‌ها	مدت‌زمان پیش‌بینی کوتاه‌مدت								
	دو هفته			۱ ماه			دو ماه		
	۲۰۱۲	۲۰۱۳	۲۰۱۴	۲۰۱۲	۲۰۱۳	۲۰۱۴	۲۰۱۲	۲۰۱۳	۲۰۱۴
AR (۲)	۰.۵۵	۰.۴۵	۰.۵۵	۰.۵۰	۰.۵۰	۰.۵۰	۰.۴۵	۰.۳۵	۰.۵۵
ARIMA (۰۰۰۰۲)	۰.۴۵	۰.۷۰	۰.۶۰	۰.۵۰	۰.۴۵	۰.۵۵	۰.۳۰	۰.۴۰	۰.۳۵
روش پیشنهادی	۰.۹۰	۰.۷۵	۰.۷۰	۰.۸۵	۰.۸۰	۰.۷۵	۰.۸۰	۰.۸۰	۰.۶۰

جدول ۱۰: مقایسه روش پیشنهادی با روش AR و ARIMA بر اساس آزمون MAPE در پیش‌بینی‌های میان‌مدت

روش‌ها	مدت‌زمان پیش‌بینی کوتاه‌مدت								
	دو هفته			۱ ماه			دو ماه		
	۲۰۱۲	۲۰۱۳	۲۰۱۴	۲۰۱۲	۲۰۱۳	۲۰۱۴	۲۰۱۲	۲۰۱۳	۲۰۱۴
AR (۲)	۲۲	۲۵	۱۵	۳۲	۳۳	۲۵	۶۳	۴۷	۳۲
ARIMA (۰۰۰۰۲)	۱۹	۲۳	۱۸	۲۹	۳۵	۳۱	۳۵	۳۷	۲۸
روش پیشنهادی	۱۵	۱۸	۱۶	۲۵	۲۶	۲۰	۲۹	۲۸	۲۴

روش‌ها	مدت‌زمان پیش‌بینی کوتاه‌مدت								
	دو هفته			۱ ماه			دو ماه		
	۲۰۱۲	۲۰۱۳	۲۰۱۴	۲۰۱۲	۲۰۱۳	۲۰۱۴	۲۰۱۲	۲۰۱۳	۲۰۱۴
AR (۲)	۲۵	۲۸	۳۳	۳۶	۳۸	۳۷	۶۹	۵۲	۴۰
ARIMA (۰۰۰۰۲)	۲۸	۲۹	۳۰	۳۲	۳۲	۲۸	۳۷	۳۲	۳۹
روش پیشنهادی	۱۸	۲۱	۱۹	۲۵	۲۹	۲۴	۲۹	۲۸	۳۶

مراجع

- [۱۸] S. Han and R.-C. Chen, "Using svm with financial statement analysis for prediction of stocks," *Communications of the IIMA*, vol. ۷, p. ۶۳, ۲۰۰۷.
- [۱۹] S. W. Chan and J. Franklin, "A text-based decision support system for financial sequence prediction," *Decision Support Systems*, vol. ۵۷, pp. ۱۹۸-۱۸۹, ۲۰۱۱.
- [۲۰] M.-C. Lin, A. J. Lee, R.-T. Kao, and K.-T. Chen, "Stock price movement prediction using representative prototypes of financial reports," *ACM Transactions on Management Information Systems (TMIS)*, vol. ۲, p. ۱۹, ۲۰۱۱.
- [۲۱] R. P. Schumaker and H. Chen, "Textual analysis of stock market prediction using breaking financial news: The AZFin text system," *ACM Transactions on Information Systems (TOIS)*, vol. ۲۷, p. ۱۲, ۲۰۰۹.
- [۲۲] C.-F. Tsai and Z.-Y. Quan, "Stock prediction by searching for similarities in candlestick charts," *ACM Transactions on Management Information Systems (TMIS)*, vol. ۵, p. ۹, ۲۰۱۴.
- [۲۳] A. Dattasharma, P. K. Tripathi, and S. Gangadharpalli, "Identifying stock similarity based on episode distances," in *Computer and Information Technology, ۲۰۰۸. ICCIT ۲۰۰۸. ۱۱th International Conference on*, ۲۰۰۸, pp. ۳۵-۲۸.
- [۲۴] C.-H. L. Lee, Y.-C. Liaw, and L. Hsu, "Investment decision making by using fuzzy candlestick pattern and genetic algorithm," in *Fuzzy Systems (FUZZ), ۲۰۱۱. IEEE International Conference on*, ۲۰۱۱, pp. ۲۷۰۱-۲۶۹۶.
- [۲۵] https://en.wikipedia.org/wiki/Relative_strength_index.
- [۲۶] <http://new.tse.ir/Instrument.html?IRO\FOLD۰۰۱>.
- [۱] R. d. A. Araujo, "Hybrid intelligent methodology to design translation invariant morphological operators for Brazilian stock market prediction," *Neural Networks*, vol. ۲۳, pp. ۱۲۵۱-۱۲۳۸, ۲۰۱۰.
- [۲] K. Park and H. Shin, "Stock price prediction based on a complex interrelation network of economic factors," *Engineering Applications of Artificial Intelligence*, vol. ۲۶, pp. ۱۵۶۱-۱۵۵۰, ۲۰۱۳.
- [۳] R. Nayak and P. Te Braak, "Temporal pattern matching for the prediction of stock prices," in *Proceedings of the ۱nd international workshop on Integrating artificial intelligence and data mining-Volume ۸۴, ۲۰۰۷*, pp. ۱۰۳-۹۵.
- [۴] P. Barekati, M. Jalali, M. V. Jahan, and V. A. Mahboob, "Extracting edge centrality from social networks using heat diffusion algorithm," *International Journal of Knowledge-based and Intelligent Engineering Systems*, vol. ۱۹, pp. ۸۲, ۲۰۱۵.
- [۵] E. F. Fama, "Efficient capital markets: A review of theory and empirical work*," *The Journal of Finance*, vol. ۲۵, pp. ۴۱۷-۳۸۳, ۱۹۷۰.
- [۶] W. Brock, J. Lakonishok, and B. LeBaron, "Simple technical trading rules and the stochastic properties of stock returns," *Journal of finance*, pp. ۱۷۶۴-۱۷۳۱, ۱۹۹۲.
- [۷] P. R. Hansen and A. Lunde, "A forecast comparison of volatility models: does anything beat a GARCH (", *Journal of applied econometrics*, vol. ۲۰, pp. ۸۸۹-۸۷۳, ۲۰۰۵.
- [۸] T. Levy and J. Yagil, "The week-of-the-year effect: Evidence from around the globe," *Journal of Banking & Finance*, vol. ۳۶, pp. ۱۹۷۴-۱۹۶۳, ۲۰۱۲.
- [۹] R. D. Edwards, J. Magee, and W. Bassetti, *Technical analysis of stock trends*: CRC Press, ۲۰۱۲.
- [۱۰] J. J. Murphy, *Technical analysis of the financial markets: A comprehensive guide to trading methods and applications*: Penguin, ۱۹۹۹.
- [۱۱] G. C. Friesen, P. A. Weller, and L. M. Dunham, "Price trends and patterns in technical analysis: A theoretical and empirical examination," *Journal of Banking & Finance*, vol. ۳۳, pp. ۱۱۰۰-۱۰۸۹, ۲۰۰۹.
- [۱۲] Y. Zhu and G. Zhou, "Technical analysis: An asset allocation perspective on the use of moving averages," *Journal of Financial Economics*, vol. ۹۲, pp. ۵۴۴-۵۱۹, ۲۰۰۹.
- [۱۳] A. Gupta and B. Dhingra, "Stock market prediction using hidden Markov models," in *Engineering and Systems (SCES), ۲۰۱۲ Students Conference on*, ۲۰۱۲, pp. ۴-۱.
- [۱۴] M. R. Hassan, "A combination of hidden Markov model and fuzzy model for stock market forecasting," *Neurocomputing*, vol. ۷۲, pp. ۳۴۴۶-۳۴۳۹, ۲۰۰۹.
- [۱۵] M. R. Hassan and B. Nath, "Stock market forecasting using hidden Markov model: a new approach," in *Intelligent Systems Design and Applications, ۲۰۰۵. ISDA'۰۵. Proceedings. ۵th International Conference on*, ۲۰۰۵, pp. ۱۹۶-۱۹۲.
- [۱۶] X.-h. LIU, M.-l. LI, and B.-y. TANG, "Forecast of Stocks Index Based on Investor Sentiment and Neural Network," *Journal of Donghua University, Natural Science*, vol. ۲, p. ۰۱۳, ۲۰۰۴.
- [۱۷] P. K. H. Phua, D. Ming, and W. Lin, "Neural network with genetically evolved algorithms for stocks prediction," *Asia-Pacific Journal of Operational Research*, vol. ۱۸, p. ۱۰۳, ۲۰۰۱.