

ارائه مدل‌های پیش‌بینی از شاخص‌های مالی با استفاده از رگرسیون و شبکه‌های عصبی

علی نقاش اسدی*

استادیار - دانشکده فنی فومن - دانشکدگان فنی دانشگاه تهران - فومن - ایران
پست الکترونیکی: naghashasadi@ut.ac.ir

زهرا میرزائی

کارشناسی ارشد - دانشکده مهندسی کامپیوتر - دانشگاه صنعتی شریف - تهران - ایران
پست الکترونیکی: zahra.mirzaei@ce.sharif.edu

سپهر نیکوکار

کارشناسی - دانشکده فنی فومن - دانشکدگان فنی دانشگاه تهران - فومن - ایران
پست الکترونیکی: sepehr.nikookar@ut.ac.ir

چکیده

مالی داخلی (کشور ایران) احساس می‌شود. در این مقاله، ابتدا داده‌های شش شاخص مالی داخلی شامل طلای ۱۸ عیار، سکه تمام بهار، نفت (ریال)، نفت (دلار)، دلار، و یورو از تاریخ ۱۳۹۸/۰۱/۰۱ الی ۱۴۰۲/۰۲/۳۱ (در مجموع ۱۵۲۲ رکورد) جمع‌آوری شده و بعد از پاک‌سازی و نرمال‌سازی، با استفاده از روش‌های رگرسیون و پرسپترون چندلایه مورد تحلیل قرار گرفته و مدل‌های پیش‌بینی از آنها استخراج می‌شود. نتایج معیارهای ارزیابی (شامل ضریب همبستگی، میانگین خطای مطلق و غیره) از مقایسه نتایج مدل‌های استخراج‌شده با نتایج واقعی، اختلاف بسیار کمی را نشان می‌دهند (برای مثال، معیار ضریب همبستگی در روش رگرسیون و پرسپترون چندلایه به ترتیب برابر با ۰,۹۹۶۴ و ۰,۹۹۹۹ است) که نشان‌دهنده انتخاب درست شاخص‌های مالی، تأثیر مستقیم و اساسی آنها بر یکدیگر، و دقت بالای مدل‌های پیش‌بینی به دست آمده است. واژه‌های کلیدی: داده‌کاوی، شاخص‌های مالی، مدل‌های پیش‌بینی، رگرسیون، شبکه‌های عصبی.

فرآیند داده‌کاوی در حوزه اقتصاد و مسائل مالی در مقالات بسیاری انجام شده است. در هر کدام از این تحقیقات، معمولاً ابتدا تعدادی عوامل اقتصادی و غیراقتصادی انتخاب شده و در ادامه پس از جمع‌آوری داده‌ها و انجام فرآیند داده‌کاوی، الگوی ارتباطی بین آنها کشف می‌شود. به عبارت دیگر، در این تحقیقات معمولاً تلاش می‌شود که با ارائه یک مدل، میزان تأثیر هر شاخص مالی از تغییر شاخص‌های مالی دیگر به دست آید. با این حال در این تحقیقات معمولاً داده‌ها یا خیلی قدیمی هستند که باعث می‌شود نتایج آنها برای وضعیت فعلی کاربردی نداشته باشد؛ و یا مربوط به شاخص‌های مالی بین‌المللی هستند که در بسیاری از مواقع نمی‌توان نتایج آنها را به شرایط داخلی تعمیم داد. در معدود تحقیقات انجام‌شده نیز، عوامل اقتصادی زیادی مورد بررسی قرار نگرفته است و بنابراین خلأ وجود یک تحقیق جامع و جدید از وضعیت شاخص‌های

* نویسنده مسئول

۱. مقدمه

بودن داده‌ها نسبت به تاریخ روز (خیلی قدیمی نبودن آنها) می‌تواند بر اعتبار نتایج به دست آمده و امکان استفاده از آن برای پیش‌بینی رویدادهای آینده تأثیرگذار باشد. علاوه بر این، در اکثر تحقیقات انجام‌شده، شاخص‌های مالی بین‌المللی مورد بررسی قرار گرفته‌اند که در بسیاری از مواقع نمی‌توان نتایج آنها را به شرایط داخلی تعمیم داد. به عبارت دیگر، بعضی از شاخص‌ها (برای مثال قیمت سکه تمام بهار آزادی) صرفاً در ایران موضوعیت داشته و به عنوان یک شاخص مالی مهم محسوب می‌شوند؛ ولی در کشورهای دیگر تعریف نشده هستند. سایر شاخص‌ها (برای مثال، قیمت نفت، دلار، طلای ۱۸ عیار و غیره) به عنوان شاخص‌های مالی بین‌المللی شناخته می‌شوند، ولی اگر قیمت ریالی آنها مورد بررسی قرار گیرد، می‌توان گفت که این شاخص‌های مالی داخلی هستند چون علاوه بر تغییر قیمت بین‌المللی این شاخص‌ها، تغییر قیمت ریالی آنها به ارزش پول ملی نیز وابسته است. برای مثال تغییر قیمت جهانی نفت به لیر ترکیه را نمی‌توان به قیمت همین شاخص به ریال ایران تعمیم داد چون ممکن است با کاهش قیمت جهانی نفت و ثابت ماندن ارزش لیر، قیمت نفت به لیر کاهشی باشد ولی به دلیل کاهش ارزش پول ملی ایران، قیمت جهانی نفت به ریال ایران، افزایشی باشد. در بعضی از تحقیقات انجام‌شده نیز شاخص‌های مالی داخلی در کشورهای مختلف مورد بررسی قرار گرفته است ولی نتایج آن قابل تعمیم به شاخص‌های مالی ایران نیست [۴] [۵]. برای مثال، افزایش یا کاهش تولید ناخالص داخلی کشور هند نمی‌تواند بر تغییر تولید ناخالص ایران تعمیم یابد. بنابراین نمی‌توان به مدل پیش‌بینی آنها استناد کرده و از نتایج ارائه شده توسط آنها استفاده کنیم. در معدود تحقیقات انجام شده‌ای که کمبودهای ذکر شده را نداشتند، عوامل اقتصادی زیادی مورد بررسی قرار نگرفته، و یا عوامل اقتصادی بررسی شده توسط آنها با عوامل بررسی شده در این مقاله متفاوت هستند. بنابراین خلأ یک تحقیق جامع و جدید از وضعیت شاخص‌های مالی داخلی (کشور ایران) وجود دارد. در این

امروزه در اکثر سازمان‌ها، داده‌های زیادی در حال جمع‌آوری و ذخیره شدن هستند ولی با وجود حجم زیادی از داده‌ها، این سازمان‌ها با فقر دانش در تصمیم‌گیری روبرو هستند [۱]. برای رفع این مشکل معمولاً از فرآیند داده‌کاوی استفاده می‌شود. داده‌کاوی یک فرآیند مهم در مهندسی علوم داده و هوش مصنوعی است. فرآیند داده‌کاوی شامل مجموعه‌ای از روش‌های قابل اعمال بر داده‌های بزرگ و پیچیده است که به منظور کشف الگوهای پنهان در میان داده‌ها استفاده می‌شود [۲]. پژوهش‌ها در زمینه داده‌کاوی در گستره وسیعی از حوزه‌ها شامل علوم کامپیوتر، اقتصاد، پزشکی، مهندسی و غیره دنبال می‌شود. استفاده از فرآیند داده‌کاوی در حوزه اقتصاد می‌تواند نقش بسیار مهمی ایفا کند. به عبارت دیگر، همواره تمایل داریم میزان تأثیر هر شاخص مالی را از تغییر شاخص‌های مالی دیگر بدانیم. حوزه اقتصاد دارای شاخص‌های مالی بسیاری است که عوامل اقتصادی و غیراقتصادی (سیاسی، فرهنگی و غیره) بسیاری وجود دارند که می‌توانند بر آنها تأثیرگذار باشند. برای مثال، تغییر قیمت نفت و دیگر حامل‌های انرژی و همچنین نرخ تورم می‌توانند بر تغییر قیمت طلا تأثیر مستقیم داشته باشند. البته تأثیرگذاری شاخص‌ها بر یکدیگر محرز است ولی با انجام فرآیند داده‌کاوی می‌توان با استخراج مدل پیش‌بینی، میزان تأثیرگذاری هر شاخص بر شاخص‌های دیگر را به دست آورد. بنابراین باید داده‌های شاخص‌های مالی جمع‌آوری شده و الگوی تأثیرگذاری آنها بر یکدیگر شناسایی و مدلی به منظور پیش‌بینی مقادیر آنها از سایر شاخص‌ها، استخراج شود. استخراج این مدل می‌تواند میزان تأثیرگذاری هر شاخص را بر شاخص‌های دیگر مشخص کرده و امکان پیش‌بینی مقادیر هر شاخص را با تغییر مقدار شاخص‌های دیگر فراهم کند. با این حال در اکثر تحقیقات انجام‌شده، داده‌های بررسی‌شده بسیار قدیمی هستند که باعث می‌شود نتایج آنها برای وضعیت فعلی کاربردی نداشته باشد [۳]. جدید

داده‌کاوی پیش‌بینی کننده، از داده‌های گذشته و فعلی برای پیش‌بینی رویدادهای آینده استفاده می‌شود. از مهم‌ترین زیردسته‌های داده‌کاوی پیش‌بینی کننده، طبقه‌بندی^۳، رگرسیون^۴ و شبکه‌های عصبی مصنوعی^۵ است. در داده‌کاوی توصیفی، هدف خلاصه‌سازی و کنار هم قرار دادن داده‌های مشابه است. از مهم‌ترین زیردسته‌های داده‌کاوی توصیفی، خوشه‌بندی^۶ است. در ادامه بعضی از روش‌های داده‌کاوی که در این مقاله از آنها استفاده خواهد شد، معرفی می‌شوند.

تحلیل رگرسیون

یکی از بهترین و ساده‌ترین روش‌ها برای به دست آوردن رابطه بین متغیرها، تحلیل رگرسیون است. از نظر آماری، تحلیل رگرسیون فرآیندی است که معمولاً برای شناسایی و تحلیل رابطه بین متغیرها و در نتیجه پیش‌بینی اهداف استفاده می‌شود. همچنین رگرسیون می‌تواند در درک تغییر مقدار یک متغیر در صورت تغییر متغیرهای دیگر کمک کند. مطابق با رابطه ۱، تحلیل رگرسیون با استخراج یک رابطه از پارامترهای ورودی، پارامتر خروجی را محاسبه می‌کند. به عبارت دیگر، فرض کنید n متغیر (X_1 الی X_n) وجود دارند که می‌خواهیم به کمک آنها، مقدار متغیر Y را محاسبه کنیم. تحلیل رگرسیون با تعیین ضرایب این متغیرها، این رابطه را فراهم می‌کند. البته مطابق با رابطه ۱، همیشه ارتباط بین متغیرها خطی نیست و ممکن است غیرخطی باشد که مدل‌های پیشرفته‌تر رگرسیون می‌توانند روابط غیرخطی بین متغیرها را نیز استخراج کنند.

$$Y = F(X_i, \alpha_i) + e = \alpha_1 X_1 + \alpha_2 X_2 + \dots + \alpha_n X_n + e \quad (1)$$

پرسپترون چندلایه

شبکه عصبی مصنوعی روشی برای پردازش

مقاله، چهار گام فرآیند داده‌کاوی شامل «انتخاب شاخص‌ها و جمع‌آوری داده»، «پاک‌سازی و نرمال‌سازی داده‌ها»، «انتخاب روش داده‌کاوی و کشف الگوها» و «تفسیر و ارائه نتایج» به منظور استخراج مدل‌های پیش‌بینی انجام می‌شود. برای این منظور، داده‌های شش شاخص مالی شامل طلای ۱۸ عیار، سکه تمام بهار، نفت (ریال)، نفت (دلار)، دلار، و یورو از تاریخ ۱۳۹۸/۰۱/۰۱ الی ۱۴۰۲/۰۲/۳۱ (در مجموع ۱۵۲۲ رکورد) جمع‌آوری می‌شوند. سپس این داده‌ها بعد از پاک‌سازی و نرمال‌سازی، با استفاده از روش‌های رگرسیون و پرسپترون چندلایه مورد تحلیل قرار گرفته و مدل‌های پیش‌بینی از آنها استخراج می‌شود. نتایج معیارهای ارزیابی به دست آمده (شامل ضریب همبستگی، میانگین خطای مطلق و غیره) از مقایسه نتایج مدل‌های استخراج شده با نتایج واقعی، اختلاف بسیار کمی را نشان می‌دهند که نشان‌دهنده انتخاب درست شاخص‌های مالی، تأثیر مستقیم و اساسی آنها بر یکدیگر، و دقت بالای مدل‌های پیش‌بینی به دست آمده است.

در ادامه این مقاله، ابتدا در مورد مفاهیم اصلی استفاده‌شده در مقاله شامل روش‌های داده‌کاوی و معیارهای ارزیابی، توضیحاتی ارائه می‌شود. سپس کارهای مرتبط با مقاله معرفی شده و در ادامه، فرآیند داده‌کاوی انجام‌شده به منظور استخراج مدل‌های پیش‌بینی ارائه می‌شود. همچنین در این مرحله، نتایج به دست آمده از مدل‌های استخراج شده با نتایج واقعی بر اساس معیارهای ارزیابی مختلف مقایسه می‌شوند. در انتها نیز، نتیجه‌گیری و کارهای آینده ارائه می‌شود.

۲. مفاهیم اصلی

۲-۱. روش‌های داده‌کاوی

روش‌های مختلفی برای داده‌کاوی داده‌ها وجود دارند که می‌توان آنها را به دو دسته داده‌کاوی پیش‌بینی کننده^۱ و داده‌کاوی توصیفی^۲ تقسیم‌بندی کرد [۶] [۷]. در

3- Classification

4- Regression

5- Artificial Neural Network

6- Clustering

1- Predictive Data Mining

2- Descriptive Data Mining

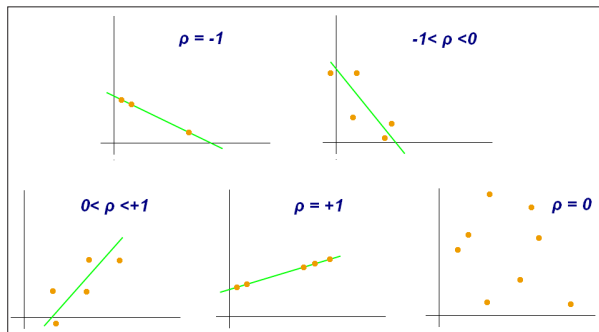
۲-۲. معیارهای ارزیابی

معیارهای مختلفی برای ارزیابی مدل‌های به دست آمده از فرآیند داده‌کاوی وجود دارد. در ادامه بعضی از معیارهای ارزیابی که در این مقاله از آنها استفاده خواهد شد، معرفی می‌شوند.

ضریب همبستگی

ضریب همبستگی (CC) میزان همبستگی خطی بین دو متغیر تصادفی را نشان می‌دهد. مطابق با رابطه ۳، به منظور ارزیابی مدل به دست آمده از فرآیند داده‌کاوی، مقادیر خروجی به دست آمده از مدل (y_i) با مقادیر واقعی (x_i) مقایسه می‌شوند. مقدار این ضریب می‌تواند بین منفی یک تا یک تغییر می‌کند؛ مطابق با شکل ۲، مقدار یک به معنای همبستگی مثبت کامل، مقدار صفر به معنی نبود همبستگی، و مقدار منفی یک به معنی همبستگی منفی کامل است.

$$CC = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2 \sum(y_i - \bar{y})^2}} \quad (3)$$



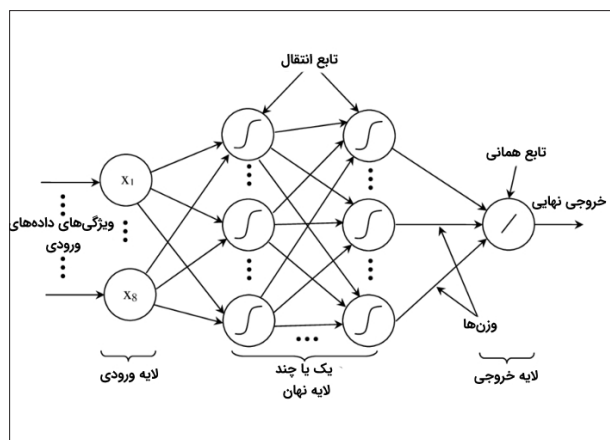
شکل ۲. مقادیر مختلف ضریب همبستگی

میانگین خطای مطلق

میانگین خطای مطلق (MAE) میانگین تفاوت بین مقدار واقعی (x_i) و مقدار پیش‌بینی شده (y_i) از مدل بدست آمده از فرآیند داده‌کاوی را بر روی تمام نمونه‌های آموزش محاسبه می‌کند. برای مثال برای n نمونه آزمایش، برای هر مقدار واقعی x و مقدار پیش‌بینی شده y میانگین خطای مطلق مطابق با رابطه ۴ است. هر چقدر این مقدار به صفر نزدیک‌تر باشد، نشان‌دهنده کمترین فاصله بین نتایج واقعی و پیش‌بینی شده است.

اطلاعات است که از سیستم عصبی زیستی الهام گرفته و مانند مغز به پردازش اطلاعات می‌پردازد. این روش از تعدادی عناصر پردازشی به نام نورون^۷ تشکیل شده است که برای حل یک مسئله با هم به صورت هماهنگ عمل می‌کنند. پرسپترون چندلایه^۸، نمونه‌ای از شبکه‌های عصبی مصنوعی است. مطابق با شکل ۱، پرسپترون چندلایه شامل حداقل سه لایه گره است (یک لایه ورودی، یک یا چند لایه پنهان، و یک لایه خروجی). در این روش، از خروجی‌های لایه ورودی، به عنوان ورودی‌های لایه‌های بعدی استفاده می‌شود؛ این کار به همین شکل ادامه پیدا می‌کند، تا زمانی که پس از تعداد خاصی از لایه‌ها، خروجی‌های آخرین لایه نهان به عنوان ورودی‌های لایه خروجی مورد استفاده قرار می‌گیرد. به جز گره‌های لایه ورودی، سایر گره‌ها، نورون‌ها هستند که از یک تابع محاسباتی استفاده می‌کنند. مطابق با رابطه ۲، اگر نورون i دارای m ورودی (I_1 الی I_m) باشد، خروجی تابع آن از مجموع حاصل ضرب مقدار هر ورودی در وزن آن (w_i) به دست می‌آید که در انتها با یک مقدار ثابت (حد آستانه^۹)، جمع می‌شود. در نهایت نورون نهایی در لایه خروجی، خروجی شبکه عصبی را ایجاد می‌کند.

$$O_i = \sum_{i=1}^m w_i I_i + t \quad (2)$$



شکل ۱. مثالی از پرسپترون چندلایه

7- Neurons

8- Multilayer Perceptron

9- Threshold

10- Correlation Coefficient

11- Mean Absolute Error

مربعات مطابق با رابطه ۷ است. هر چقدر این مقدار به صفر نزدیکتر باشد، نشان‌دهنده کمترین فاصله بین نتایج واقعی و پیش‌بینی شده است.

$$RRSE = \sqrt{\frac{\sum_{i=1}^n (x_i - y_i)^2}{\sum_{i=1}^n (x_i - \bar{x})^2}} \times 100 \quad (7)$$

۳. کارهای مرتبط

فرآیند داده‌کاوی در حوزه اقتصاد و مسائل مالی در مقالات بسیاری انجام شده است. در هر کدام از مقالات، معمولاً ابتدا تعدادی عوامل اقتصادی و غیراقتصادی انتخاب شده و در ادامه پس از جمع‌آوری داده‌ها و انجام فرآیند داده‌کاوی، الگوی ارتباطی بین آنها کشف شده است. همچنین در برخی از مقالات، الگوریتم‌های جدیدی که به صورت خاص منظوره برای تحلیل داده‌های مالی مورد استفاده قرار می‌گیرند، ارائه شده است. برای مثال در [۸]، از الگوریتم‌های یادگیری ماشینی^{۱۵} برای تحلیل داده‌های سری زمانی و بهبود تحلیل بصری اقتصاد منطقه‌ای، استفاده شده است. در [۹]، با استفاده از فرآیند داده‌کاوی سعی شده است که رشد تولید ناخالص داخلی که معیاری برای تولید اقتصادی و تعیین سطح استاندارد زندگی مردم است، در ایالتی از کشور هند پیش‌بینی شود. در [۹]، اطلاعات افراد، شغل‌ها و شرکت‌ها در اتحادیه اروپا که مرتبط با کلید واژه اقتصاد دورانی^{۱۶} بودند، از بُن‌سازۀ لینکدین جمع‌آوری شده و سپس با استفاده از تحلیل سری زمانی، الگوی روابط میان داده‌ها، استخراج شده است. در [۵]، از فرآیند داده‌کاوی برای اندازه‌گیری شاخص توسعه اقتصاد دیجیتال در کشور چین به منظور تأیید سطح واقعی آن استفاده شده است. در [۳]، اثربخشی روش‌های داده‌کاوی در تشخیص رفتارهای متقلبانه شرکت‌هایی که صورت‌های مالی متقلبانه گزارش نموده‌اند، بررسی شده تا عوامل مؤثر بر این گونه رفتارها شناسایی شود. در [۱۰]، از فرآیند داده‌کاوی برای پیش‌بینی تلاطم بازار به

$$MAE = \frac{1}{n} \sum_{i=1}^n |x_i - y_i| \quad (4)$$

خطای جذر میانگین مربعات

خطای جذر میانگین مربعات (RMSE)^{۱۲} تفاوت میان مقدار پیش‌بینی شده از مدل بدست آمده از فرآیند داده‌کاوی (y_i) و مقدار واقعی (x_i) است. برای مثال برای n نمونه آزمایش، برای هر مقدار واقعی x و مقدار پیش‌بینی شده y ، خطای جذر میانگین مربعات مطابق با رابطه ۵ است. هر چقدر این مقدار به صفر نزدیکتر باشد، نشان‌دهنده کمترین فاصله بین نتایج واقعی و پیش‌بینی شده است.

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (x_i - y_i)^2}{n}} \quad (5)$$

درصد خطای مطلق نسبی

درصد خطای مطلق نسبی (RAE)^{۱۳} نیز درصد تفاوت نسبی بین مقدار پیش‌بینی شده از مدل بدست آمده از فرآیند داده‌کاوی (y_i) و مقدار واقعی (x_i) است. برای مثال برای n نمونه آزمایش، برای هر مقدار واقعی x و مقدار پیش‌بینی شده y ، درصد خطای مطلق نسبی مطابق با رابطه ۶ است. هر چقدر این مقدار به صفر نزدیکتر باشد، نشان‌دهنده کمترین فاصله بین نتایج واقعی و پیش‌بینی شده است.

$$RAE = \frac{\sqrt{\sum_{i=1}^n (x_i - y_i)^2}}{\sqrt{\sum_{i=1}^n (x_i)^2}} \times 100 \quad (6)$$

درصد خطای نسبی جذر مربعات

درصد خطای نسبی جذر مربعات (RRSE)^{۱۴} نیز درصد تفاوت جذر نسبی بین مقدار پیش‌بینی شده از مدل بدست آمده از فرآیند داده‌کاوی (y_i) و مقدار واقعی (x_i) است. برای مثال برای n نمونه آزمایش، برای هر مقدار واقعی x و مقدار پیش‌بینی شده y ، درصد خطای نسبی جذر

15- Machine Learning

16- Circular Economy

12- Root Mean Squared Error

13- Relative Absolute Error

14- Root Relative Squared Error

جدول ۱. مقایسه بین کارهای مرتبط

مقاله	سال انتشار	داده‌های مالی ایران	عوامل بررسی شده
[۸]	۲۰۲۲	خیر	تحلیل داده‌های سری زمانی و بهبود تحلیل بصری اقتصاد منطقه‌ای
[۴]	۲۰۱۳	خیر	تولید ناخالص داخلی، سطح استاندارد زندگی، رشد اقتصادی، کشور هند
[۹]	۲۰۲۳	خیر	اقتصاد دورانی، اطلاعات افراد و شغل‌ها و شرکت‌ها، ثبات اقتصادی
[۵]	۲۰۲۲	خیر	روند توسعه اقتصاد دیجیتال در چین
[۳]	۱۳۹۱	خیر	کشف تقلب در صورت‌های مالی، شرکت‌های تولیدی کشور یونان، پارامترهای مالی شرکت شامل میزان سود، دارایی، مالیات، بدهی و غیره
[۱۰]	۱۳۹۹	بله	قیمت جهانی نفت، قیمت جهانی طلا، تلاطم بازار بورس، تولید ناخالص داخلی، سود بلند مدت بانکی، حجم پول در گردش، نرخ تورم کشور ایران
[۱۱]	۱۴۰۰	خیر	بررسی روش‌های داده‌کاوی موجود در حوزه‌های رتبه‌بندی اعتباری، پیش‌بینی پیش‌فرض وام، پول‌شویی، پیش‌بینی سهام و پیش‌بینی سری‌های مالی
[۱۲]	۱۴۰۱	خیر	نرخ بهره، نرخ ارز، شاخص بازار سهام، کالا، سایر شاخص‌های اقتصادی در کره جنوبی، سایر شاخص‌های اقتصادی در ایالات متحده
[۱۳]	۲۰۱۹	بله	استفاده از درخت تصمیم برای پیش‌بینی درماندگی مالی شرکت‌های پذیرفته‌شده در «بورس اوراق و بهادار تهران» طی سال‌های ۱۳۸۹ تا ۱۳۹۴

نگرفته، و یا عوامل اقتصادی بررسی شده توسط آنها با عوامل بررسی شده در این مقاله متفاوت هستند.

۴. فرآیند داده‌کاوی انجام‌شده

مطابق با شکل ۳، فرآیند داده‌کاوی شامل چندین مرحله است. این فرآیند از جمع‌آوری داده‌های خام شروع شده و تا ایجاد و تفسیر دانش جدید و ارائه آن ادامه دارد. در این بخش، ما هر یک از مراحل داده‌کاوی را در مورد مطالعاتی خود انجام داده و نتایج به دست آمده را ارائه می‌کنیم. برای این منظور از ابزار Weka استفاده شده است [۱۴].



شکل ۳. تغییر قیمت شاخص‌های مالی جمع‌آوری شده

۴-۱. انتخاب شاخص‌ها و جمع‌آوری داده

مطابق با جدول ۲، در اولین گام از فرآیند داده‌کاوی شاخص‌های مالی، پنج شاخص طلا، ۱۸ عیار، سکه تمام بهار، نفت، دلار، و یورو انتخاب و قیمت آنها از سایت www.tgju.org جمع‌آوری شده است. البته قیمت نفت به دلار جمع‌آوری شده است ولی با ضرب مقدار آن با قیمت

عنوان یک عامل مؤثر در تعیین ریسک سرمایه‌گذاری استفاده شده است. در [۱۱]، کاربردهای روش‌های داده‌کاوی در شش حوزه مالی مختلف شامل، رتبه‌بندی اعتباری، پیش‌بینی پیش‌فرض وام، پول‌شویی، پیش‌بینی سهام و پیش‌بینی سری‌های مالی مورد بررسی قرار گرفته است. از آنجایی که عامل زمان در معاملات مالی و کسب سود حداکثری نقش مهمی ایفا می‌کند، در [۱۲]، یک چارچوب داده‌کاوی بر اساس دانش تخصصی برای کاهش هزینه و کسب سود در بازارهای مالی ارائه شده است. در [۱۳]، از مدل‌های درخت تصمیم و مدل تحلیل ممیزی خطی برای پیش‌بینی درماندگی مالی شرکت‌ها که همواره مورد توجه سهامداران، اعتباردهندگان و مدیران واحدهای تجاری است، استفاده شده است. در جدول ۱، خلاصه‌ای از مقایسه بین کارهای مرتبط ارائه شده است. همان‌طور که از این جدول قابل مشاهده است، در اکثر مقالات، معمولاً داده‌های بررسی شده یا به زمان‌های گذشته تعلق دارند (خیلی قدیمی هستند) که باعث می‌شود نتایج آنها برای وضعیت فعلی کاربردی نداشته باشد؛ و یا مربوط به شاخص‌های مالی بین‌المللی هستند که در بسیاری از مواقع نمی‌توان نتایج آنها را به شرایط داخلی تعمیم داد. در معدود مقالاتی که کمبودهای ذکر شده را نداشتند، عوامل اقتصادی زیادی مورد بررسی قرار

روز دلار بر مبنای ریال، قیمت نفت به ریال نیز به عنوان یک جمع آوری شده است. شاخص در نظر گرفته شده است. بنابراین شش شاخص مورد بررسی قرار گرفته است که قیمت آنها از تاریخ ۱۳۹۸/۰۱/۰۱ الی ۱۴۰۲/۰۲/۳۱ (در مجموع ۱۵۲۲ رکورد) بر واحد ریال است و فقط قیمت جهانی نفت به دلار نیز

جدول ۲. جزئیات آماری از مقادیر شاخص های مالی جمع آوری شده

عنوان شاخص (واحد)	اطلاعات آماری	نمودار فراوانی
طلا ۱۸ عیار (ریال) Gold	حداقل مقدار: ۳۹۳۲۰۰۰ حداکثر مقدار: ۲۷۴۸۰۰۰۰ مقدار میانگین: ۱۱۰۴۸۰۰۰ انحراف معیار: ۵۴۱۵۵۹۴	
سکه تمام بهار (ریال) Coin	حداقل مقدار: ۳۸۰۹۰۰۰۰ حداکثر مقدار: ۳۰۸۰۱۰۰۰۰ مقدار میانگین: ۱۱۲۹۱۲۵۶۸ انحراف معیار: ۵۹۹۰۸۴۳۵	
نفت (دلار) Oil_D	حداقل مقدار: ۱۶,۵۹ حداکثر مقدار: ۱۲۵,۶۴ مقدار میانگین: ۵۸,۷۰۶ انحراف معیار: ۲۷,۹۲	
نفت (ریال) Oil_R	حداقل مقدار: ۲۵۷۱۹۴۸ حداکثر مقدار: ۴۲۳۳۲۶۲۵ مقدار میانگین: ۱۶۴۳۸۵۸۵ انحراف معیار: ۱۱۲۹۸۳۶۰	

ادامه جدول ۲. جزئیات آماری از مقادیر شاخص‌های مالی جمع‌آوری شده

<p>حدافل مقدار: ۱۱۱۰۰۰ حداکثر مقدار: ۵۵۰۵۱۰ مقدار میانگین: ۲۴۸۰۸۳ انحراف معیار: ۱۰۰۰۴۵</p>	<p>دلار (ریال) Dollar</p>
<p>حدافل مقدار: ۱۲۲۹۴۰ حداکثر مقدار: ۶۱۱۵۱۰ مقدار میانگین: ۲۷۸۶۷۰ انحراف معیار: ۱۰۹۷۱۹</p>	<p>یورو (ریال) Euro</p>

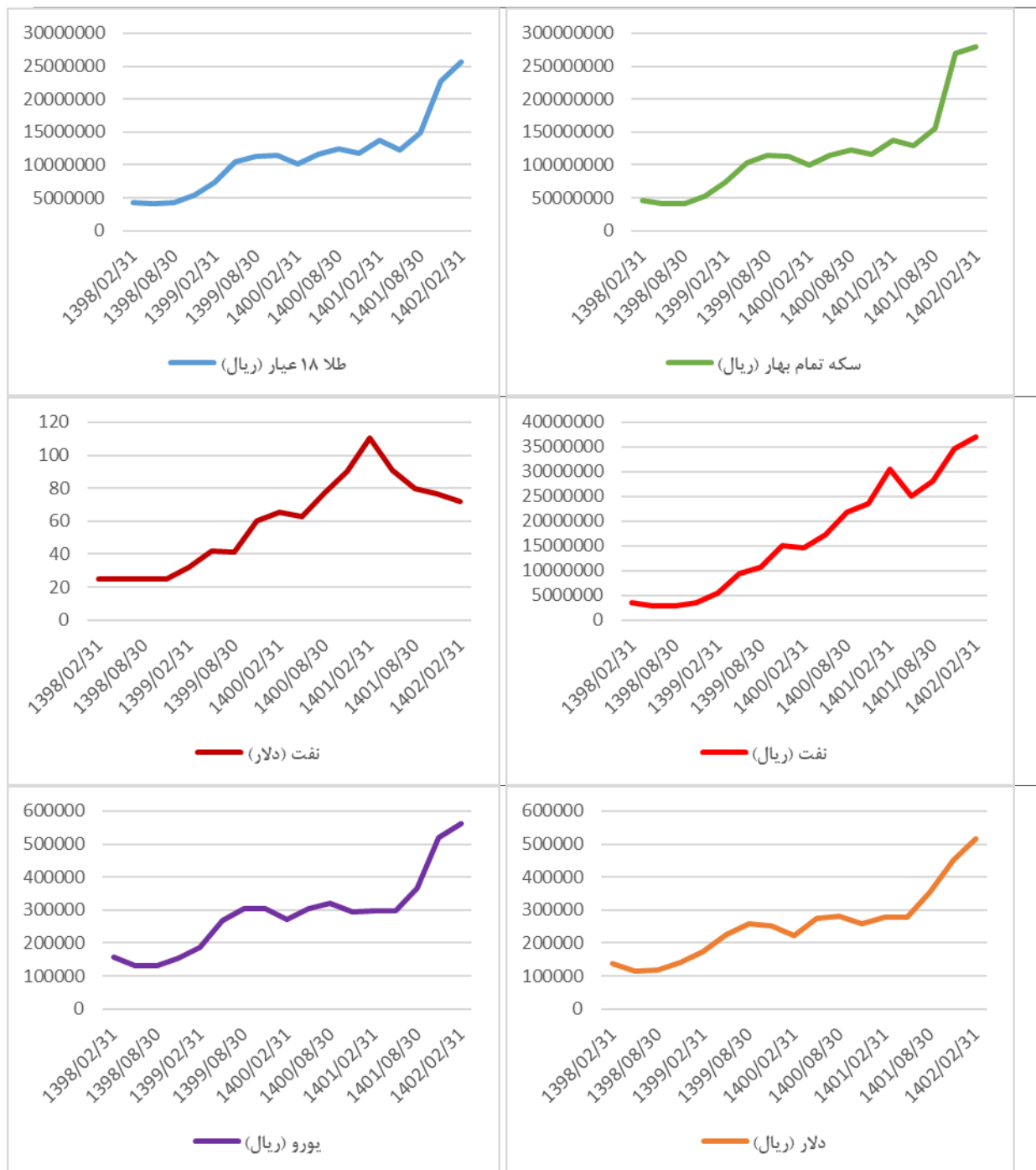
سال دارای مقدار نیستند، از روش میانگین مقادیر قبلی و بعدی استفاده شده است. برای مثال، فرض کنید قیمت اعلام شده برای شاخص نفت در روز شنبه برابر با ۵۰ دلار و برای روز دوشنبه برابر با ۵۲ دلار باشد. ولی برای روز یکشنبه مقداری برای این شاخص ثبت نشده باشد. با محاسبه میانگین قیمت نفت در روز شنبه و دوشنبه، قیمت نفت برای روز یکشنبه، با ۵۱ دلار مقداردهی می‌شود. همچنین اگر اختلاف بین مقادیر قبلی و بعدی یک شاخص، چند روز باشد، مقدار میانگین به تمامی آن روزهایی که برای آن شاخص مقداری اعلام نشده است، داده می‌شود. برای مثال، فرض کنید قیمت اعلام شده برای شاخص نفت در روز شنبه برابر با ۵۰ دلار و برای روز چهارشنبه برابر با ۵۲ دلار باشد. ولی برای روزهای یکشنبه الی سه‌شنبه مقداری برای این شاخص ثبت نشده باشد. با محاسبه میانگین قیمت نفت در روز شنبه و چهارشنبه، قیمت نفت برای روزهای یکشنبه الی سه‌شنبه، با ۵۱ دلار مقداردهی می‌شود.

پس از حل مسئله فوق، باید داده‌ها نرمال‌سازی شوند. از

ارائه شده است. همان طور که از شکل ۴ مشاهده می‌شود، قیمت طلای ۱۸ عیار، سکه تمام بهار، دلار و یورو تقریباً مشابه یکدیگر افزایش پیدا می‌کنند، ولی تغییر قیمت نفت به دلار و ریال تا حدودی متفاوت با دیگر شاخص‌های مالی است. قیمت نفت به دلار و ریال تا تاریخ ۱۴۰۱/۰۲/۳۱ تقریباً مشابه هستند ولی از این تاریخ به بعد به دلیل کاهش ارزش ریال نسبت به دلار، با وجود کاهش قیمت جهانی نفت به دلار، قیمت نفت به ریال افزایش پیدا کرده است.

۴-۲. پاک‌سازی و نرمال‌سازی داده‌ها

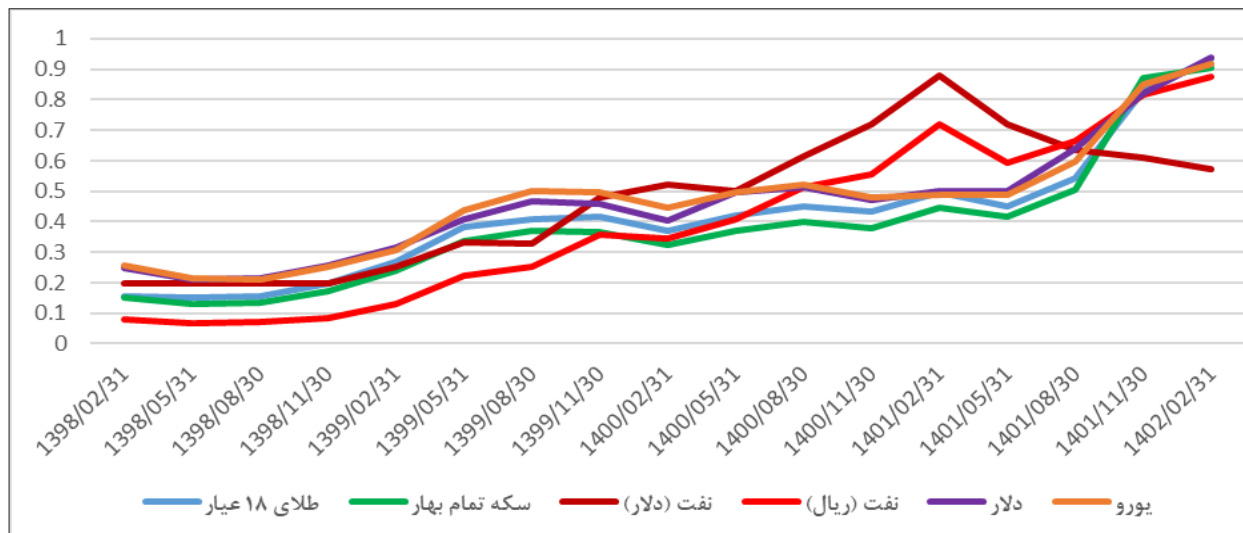
در دومین گام از فرآیند داده‌کاوی شاخص‌های مالی، ابتدا باید مسئله عدم وجود داده در برخی از روزها حل شود. در هنگام جمع‌آوری داده‌ها این مسئله وجود دارد که ممکن است به هر دلیلی برای برخی از شاخص‌ها در برخی از روزها، داده‌ای اعلام نشده باشد. برای مثال، ممکن است برای یک روز از سال، قیمت نفت اعلام نشده باشد ولی سایر شاخص‌ها دارای مقدار باشند. به منظور پاک‌سازی داده‌ها و مقداردهی به شاخص‌هایی که در روزهایی از



شکل ۴. تغییر قیمت شاخص‌های مالی جمع‌آوری‌شده

صفر و یک نرمال‌سازی شده است. برای این منظور ابتدا حداکثر مقدار هر شاخص در بازه زمانی جمع‌آوری داده مشخص شده و سپس با تقسیم مقادیر هر شاخص در زمان‌های مختلف بر حداکثر مقدار شاخص، مقادیر آن بین صفر و یک نرمال‌سازی می‌شود. این فرآیند برای

آنجایی که در مورد مطالعاتی، واحد بعضی از شاخص‌های مالی متفاوت است (ریال و دلار) و مقدار عددی هر شاخص با شاخص دیگر متفاوت است (زیر ۱۰۰، هزار و میلیون)، به منظور تأثیر یکسان هر شاخص بر شاخص‌های دیگر و در نتیجه داده‌کاوی بهتر، مقادیر تمامی شاخص‌ها بین



شکل ۵. تغییر قیمت شاخص‌های مالی پس از نرمال‌سازی

رابطه بین متغیرها (در اینجا شاخص‌های مالی) رگرسیون است. از نظر آماری، تحلیل رگرسیون فرآیندی است که معمولاً برای شناسایی و تحلیل رابطه بین متغیرها و در نتیجه پیش‌بینی اهداف استفاده می‌شود. همچنین رگرسیون می‌تواند در درک تغییر مقدار یک متغیر در صورت تغییر متغیرهای دیگر کمک کند. با انجام تحلیل رگرسیون خطی با استفاده از ابزار Weka روی داده‌های جمع‌آوری شده و نرمال‌سازی شده شاخص‌های مالی، رابطه‌ای که به وسیله آن بتوان قیمت دلار را بر حسب دیگر شاخص‌های مالی محاسبه کرد، به صورت رابطه ۸ به دست آمده است. در انجام این تحلیل، داده‌های جمع‌آوری شده به ۱۰ قسمت تقسیم شدند و در یک فرآیند تکرار، هر بار ۹ قسمت از آن به عنوان داده آموزش^{۱۷} و یک قسمت از آن به عنوان داده آزمون^{۱۸} مورد استفاده قرار گرفته است. این فرآیند تکرار تا زمانی ادامه پیدا می‌کند که مدل بدست آمده دارای بیشترین دقت باشد. نتایج ارزیابی بدست آمده از این رابطه (مدل) خطی استخراج شده در جدول ۲ ارائه شده است. همان طور که از این جدول مشاهده می‌شود، نتایج به دست آمده از رابطه خطی استخراج شده و مقایسه آن با نتایج واقعی، اختلاف بسیار کمی را نشان می‌دهد که نشان‌دهنده دقت مدل استخراج شده است.

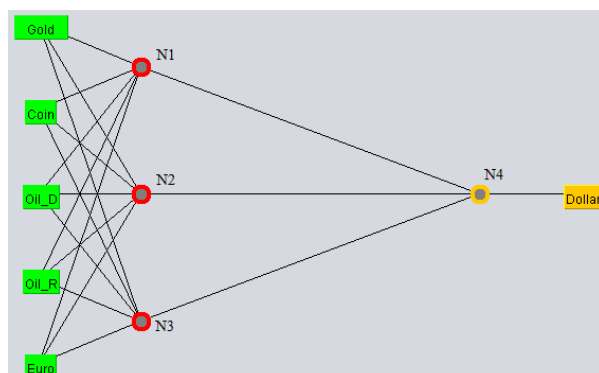
مقادیر همه شاخص‌ها تکرار می‌شود. شکل ۵ تغییر قیمت شاخص‌های مالی جمع‌آوری شده را پس از انجام مرحله پاک‌سازی و نرمال‌سازی نشان می‌دهد. همان طور که در شکل ۵ مشاهده می‌شود، بعد از نرمال‌سازی داده‌ها، امکان نمایش همه شاخص‌ها در یک نمودار فراهم شده است.

۴-۳. انتخاب روش داده‌کاوی و کشف الگوها

در سومین گام از فرآیند داده‌کاوی شاخص‌های مالی، باید متناسب با داده‌های جمع‌آوری و نرمال‌سازی شده، روش داده‌کاوی انتخاب شده و الگوها کشف شوند. در بخش ۱-۲، انواع روش‌های داده‌کاوی معرفی شد. هدف ما در این مقاله این است که با استفاده از داده‌های موجود، رابطه‌ای به دست آوریم که به وسیله آن بتوان قیمت شاخص‌های مالی را در آینده پیش‌بینی کرد. به همین دلیل باید از روش‌های داده‌کاوی پیش‌بینی کننده استفاده کنیم. همان طور که اشاره شد، انتخاب روش داده‌کاوی باید متناسب با داده‌های جمع‌آوری و نرمال‌سازی شده باشد. با بررسی شکل ۵ می‌توان مشاهده کرد که تغییر قیمت شاخص‌های مالی، مشابه یکدیگر هستند و تقریباً خطی تغییر می‌کنند و بنابراین می‌توان با یک رابطه خطی، قیمت یک شاخص مالی را از شاخص‌های مالی دیگر محاسبه کرد. یکی از بهترین و ساده‌ترین روش‌ها برای به دست آوردن

جدول ۳. نتایج ارزیابی مدل‌های استخراج شده از داده‌های شاخص‌های مالی

RRSE	RAE	RMSE	MAE	CC	روش داده‌کاوی
۸,۵۳۳۲ درصد	۷,۷۷۵۷ درصد	۰,۰۱۵۵	۰,۰۱۰۵	۰,۹۹۶۴	رگرسیون
۱,۴۹۳۴ درصد	۱,۵۴۱۳ درصد	۰,۰۰۲۷	۰,۰۰۲۱	۰,۹۹۹۹	پرسپترون چندلایه



شکل ۶. مدل پرسپترون چندلایه استخراج شده از داده‌های شاخص‌های مالی

$$N2 = 0.0506 \times Gold - 0.0401 \times Coin - 2.5351 \times Oil_D + 1.4842 \times Oil_R + 0.2406 \times Euro - 1.6169 \quad (10)$$

$$N3 = -0.0243 \times Gold - 0.4306 \times Coin - 0.9790 \times Oil_D - 1.3818 \times Oil_R - 0.3365 \times Euro - 3.7620 \quad (11)$$

$$N4 = -1.6778 \times N1 + 2.2786 \times N2 - 2.7643 \times N3 + 0.0751 \quad (12)$$

۴-۴. تفسیر و ارائه نتایج

در آخرین گام از فرآیند داده‌کاوی شاخص‌های مالی باید الگوهای کشف شده، مورد تفسیر و ارزیابی قرار گیرند. همان‌طور که پیشتر توضیح داده شد، در گام اول، داده‌های شش شاخص مالی شامل طلای ۱۸ عیار، سکه تمام بهار، نفت (ریال)، نفت (دلار)، دلار، و یورو از تاریخ ۱۳۹۸/۰۱/۰۱ الی ۱۴۰۲/۰۲/۳۱ (در مجموع ۱۵۲۲ رکورد) جمع‌آوری شدند. فرضیه اولیه ما این بود که تأثیرگذاری این شاخص‌ها بر یکدیگر بسیار زیاد است. بنابراین پس از پاک‌سازی و نرمال‌سازی داده‌ها، از رگرسیون و پرسپترون چندلایه به عنوان روش‌های داده‌کاوی برای استخراج مدل از داده‌های جمع‌آوری شده استفاده شد. همان‌طور که مشاهده شد، مدل استخراج شده از روش رگرسیون، یک رابطه خطی بین شاخص‌های مالی را ارائه

(۸)

$$Dollar = 0.2838 \times Gold - 0.2548 \times Coin - 0.2617 \times Oil_D + 0.4252 \times Oil_R + 0.6204 \times Euro + 0.1044$$

با وجود دقت بالای مدل استخراج شده از رگرسیون، با استفاده از روش پرسپترون چندلایه نیز داده‌های شاخص‌های مالی مورد ارزیابی قرار گرفت. همان‌طور که اشاره شد، یکی دیگر از بهترین روش‌های به دست آوردن رابطه بین متغیرها (در اینجا شاخص‌ها مالی) پرسپترون چندلایه است. با انجام تحلیل پرسپترون چندلایه با استفاده از ابزار Weka روی داده‌های جمع‌آوری شده و نرمال‌سازی شده شاخص‌های مالی، شبکه‌ای که به وسیله آن بتوان قیمت دلار را بر حسب دیگر شاخص‌های مالی محاسبه کرد، به صورت شکل ۶ به دست آمده است. این شبکه از ۳ لایه تشکیل شده است که در لایه اول (ورودی)، پنج شاخص مالی جمع‌آوری شده به غیر از دلار قرار دارد و در لایه سوم (خروجی)، یک نرون با شاخص مالی دلار تنظیم شده است. در لایه دوم (پنهان) که می‌تواند از چندین لایه نیز تشکیل شود، یک لایه با سه نرون وجود دارد که در روابط ۹ الی ۱۲، توابع محاسباتی هر نرون ارائه شده است. نتایج ارزیابی به دست آمده از این رابطه (مدل) استخراج شده در جدول ۳ ارائه شده است. همان‌طور که از این جدول مشاهده می‌شود، نتایج به دست آمده از مدل استخراج شده و مقایسه آن با نتایج واقعی، اختلاف بسیار کمی را نشان می‌دهد که نشان‌دهنده دقت مدل استخراج شده است. حتی نتایج ارزیابی مدل استخراج شده از روش پرسپترون چندلایه از نتایج مدل رگرسیون نیز بهتر است.

$$N1 = 0.2448 \times Gold - 0.0660 \times Coin + 1.2426 \times Oil_D - 2.2464 \times Oil_R + 0.2733 \times Euro - 1.0002 \quad (9)$$

نتایج واقعی، معیارهای ارزیابی مختلف اختلاف بسیار کمی را نشان می‌دهند که نشان‌دهنده انتخاب درست شاخص‌های مالی، تأثیر مستقیم و اساسی آنها بر یکدیگر، و دقت بالای مدل‌های پیش‌بینی بدست آمده است. به عنوان کار آینده می‌توان عوامل اقتصادی و غیراقتصادی (سیاسی، فرهنگی و غیره) دیگر را نیز به مجموعه شاخص‌ها اضافه کرده و با جمع‌آوری داده و تحلیل آنها، مدل‌های دقیق‌تر و جامع‌تری ارائه کرد. برای مثال می‌توان نرخ تورم، وضعیت بورس، میزان نقدینگی، نرخ سود بانکی، تولید ناخالص ملی و غیره را به مجموعه شاخص‌ها اضافه کرده و مدل پیش‌بینی جامع‌تری ارائه کرد. همچنین در صورت امکان می‌توان بازه زمانی جمع‌آوری داده را افزایش داد. این موضوع باعث می‌شود که مدل استخراج شده به بازه زمانی محدودی تعلق نداشته باشد و در نتیجه اعتماد به نتایج ارائه شده توسط آن بیشتر شود.

مراجع

- [1] G. Schuh, C. Reuter, J.-P. Prote and F. Brambri, "Increasing data integrity for improving decision making in production planning and control," *CIRP Annals*, vol. 66, no. 1, pp. 425-428, 2017.
- [2] G. Schuh, G. Reinhart, J.-P. Prote, F. Sauermann, J. Horsthofer, F. Oppolzer and D. Knoll, "Data Mining Definitions and Applications for the Management of Production Complexity," *Procedia CIRP*, vol. 81, pp. 874-879, 2019.
- [3] ف. رهنمای رودپشتی، «داده‌کاوی و کشف تقلب‌های مالی»، دانش حسابداری و حسابرسی مدیریت، vol. 3, no. 1, pp. 17-34, 1391.
- [4] R. Jain and D. Singh, "Data Mining and Analysis of Economic Data," *International Journal of Advanced Research in Computer Science and Software Engineering*, vol. 3, no. 8, pp. 683-688, 2013.
- [5] H. Liu, "Research on measurement of digital economy based on Data Mining," in the 7th International Conference on Intelligent Information Processing, New York, NY, USA, Sep. 2022, pp. 1-5.
- [6] P. Akulwar, S. Pardeshi and A. Kamble, "Survey on Different Data Mining Techniques for Prediction," in *2nd International Conference on I-SMAC*, Palladam, India, 30-31 Aug. 2018, pp. 513-519.
- [7] D. Papakyriakou and I. Barbounakis, "Data Mining Meth-

می‌کند که با مقایسه نتایج به دست آمده از رابطه خطی استخراج شده و نتایج واقعی، معیارهای ارزیابی اختلاف بسیار کمی را نشان می‌دهند که نشان‌دهنده دقت مدل استخراج شده است. همچنین مدل استخراج شده از روش پرسپترون چندلایه نیز یک شبکه سه لایه متشکل از چهار نورون را ارائه می‌کند که با مقایسه نتایج به دست آمده از مدل استخراج شده و نتایج واقعی، معیارهای ارزیابی اختلاف بسیار کمی را نشان می‌دهند و حتی نتایج ارزیابی مدل استخراج شده از روش پرسپترون چندلایه از نتایج مدل رگرسیون نیز بهتر بود. بنابراین فرضیه اولیه مورد تایید قرار گرفت و نتایج به دست آمده نشان داد که این شش شاخص مالی ارتباط تنگاتنگی با یکدیگر دارند و با استفاده از مدل‌های استخراج شده می‌توان تأثیر تغییر مقدار هر شاخص مالی را بر سایر شاخص‌ها محاسبه و ارزیابی کرد. علاوه بر این، با استفاده از مدل‌های استخراج شده می‌توان مقدار هر شاخص مالی را در آینده پیش‌بینی کرد. به عنوان نمونه نشان داده شد که چگونه قیمت دلار می‌تواند بر اساس قیمت سایر شاخص‌های مالی محاسبه شود.

۵. نتیجه‌گیری

در این مقاله چهار گام فرآیند داده‌کاوی شامل «انتخاب شاخص‌ها و جمع‌آوری داده»، «پاک‌سازی و نرمال‌سازی داده‌ها»، «انتخاب روش داده‌کاوی و کشف الگوها» و «تفسیر و ارائه نتایج» به منظور استخراج مدل‌های پیش‌بینی انجام شد. برای این منظور ابتدا داده‌های شش شاخص مالی شامل طلای ۱۸ عیار، سکه تمام بهار، نفت (ریال)، نفت (دلار)، دلار، و یورو از تاریخ ۱۳۹۸/۰۱/۰۱ الی ۱۴۰۲/۰۲/۳۱ (در مجموع ۱۵۲۲ رکورد) جمع‌آوری شد. سپس بعد از پاک‌سازی و نرمال‌سازی، این داده‌ها با استفاده از روش‌های رگرسیون و پرسپترون چندلایه مورد تحلیل قرار گرفته و مدل‌های پیش‌بینی از آنها استخراج شد. با مقایسه نتایج مدل‌های استخراج شده با

- [۱۲] ا. تفتیان و ع. تیموری، «ارائه چارچوب داده‌کاوی مناسب جهت پیش‌بینی بازارهای مالی با تأکید بر دانش تخصصی»، in «یازدهمین کنفرانس بین‌المللی حسابداری، مدیریت و نوآوری در کسب و کار، تهران، ۱۴۰۱».
- [13] R. Tahmasbi, A. A. Anvary Roostamy, S. J. Sadeghi Sharif and A. Khorshidi, "Information Content of Intellectual Capital and Financial Performance Indicators in Financial Prediction by Data Mining Approach," *Financial Management Perspective*, vol. 8, no. 24, pp. 129-158, 2019.
- [14] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann and I. H. Witten, "The WEKA data mining software: an update," *ACM SIGKDD Explorations Newsletter*, vol. 11, no. 1, pp. 10-18, 2009.
- ods: A Review," *International Journal of Computer Applications*, vol. 183, no. 49, pp. 5-19, 2022.
- [8] Y. Bai, M. Zhao, R. Li and P. Xin, "A new data mining method for time series in visual analysis of regional economy," *Information Processing & Management*, vol. 59, no. 1, p. 102741, 2022.
- [9] T. Daglis, G. Tsironis and K. P. Tsagarakis, "Data mining techniques for the investigation of the circular economy and sustainability relationship," *Resources, Conservation & Recycling Advances*, vol. 19, p. 200151, 2023.
- [۱۰] م. خدایاری، ا. یعقوب نژاد و م. خلیلی عراقی، «مقایسه برآورد تلاطم بازارهای مالی با استفاده از مدل رگرسیون و مدل شبکه عصبی»، *اقتصاد مالی*, vol. 14, no. 52, pp. 223-240, 1399.
- [۱۱] ا. تفتیان و ا. اولیایی، «کاربرد تکنیک‌های داده‌کاوی در صنعت مالی»، in «هشتمین کنفرانس بین‌المللی پژوهش‌های نوین حسابداری، مدیریت و علوم انسانی در هزاره سوم، تهران، ۱۴۰۰».