

**پیش‌بینی بازده سهام شرکت‌های پذیرفته شده در بورس و اوراق بهادار تهران
با استفاده از تکنیک‌های داده‌کاوی (شبکه عصبی (MLP) و مدل افزودنی کلی (GAM)**

راضیه رجبی^۱، دکتر محمد رضا عسگری^۲، دکتر عبدالمجید دهقان^۳

۱. کارشناسی ارشد، مدیریت مالی، دانشکده مدیریت و حسابداری، واحد یادگار امام خمینی (ره) شهرری،

دانشگاه آزاد اسلامی، تهران، ایران r.rajabi2550@gmail.com

۲. دانشیار و عضو هیأت علمی، مدیریت مالی، دانشکده مدیریت و حسابداری، واحد یادگار امام خمینی (ره)

شهرری، دانشگاه آزاد اسلامی، تهران، ایران Asgari@iausr.ac.ir

۳. استادیار و عضو هیأت علمی، مدیریت مالی، دانشکده مدیریت و حسابداری، واحد یادگار امام خمینی (ره)

شهرری، دانشگاه آزاد اسلامی، تهران، ایران mjd.dehghan@gmail.com

چکیده

در این پژوهش به پیش‌بینی بازده سهام شرکت‌های پذیرفته شده در بورس و اوراق بهادار تهران با استفاده از تکنیک‌های داده‌کاوی اعم از شبکه عصبی پرسپترون چندلایه (MLP) و مدل افزودنی کلی (GAM)، طی سال‌های ۱۳۹۱ الی ۱۳۹۶ می‌پردازیم. مسئله اصلی در پژوهش حاضر این است که از بین تکنیک‌های داده‌کاوی مطرح، کدامیک در پیش‌بینی بازده سهام شرکت‌های پذیرفته شده در بورس و اوراق بهادار تهران کارا تر می‌باشد. لذا این پژوهش در قالب دو فرضیه مطرح شده و برای آزمون این مسئله ۶۷ شرکت پذیرفته شده در بورس - بازار اول - تابلوی اصلی و تابلوی فرعی به عنوان جامعه هدف انتخاب شده‌اند. همچنین از بین عوامل متعددی که بر بازده سهام اثر می‌گذارند، چهار عامل به عنوان عوامل مؤثر بر بازده سهام از قبیل سود (زیان) عملیاتی، سود هر سهم، نسبت جاری و نسبت آبی به عنوان ورودی‌های پژوهش انتخاب شده‌اند. برازش هر یک از مدل‌های مذکور بر مبنای فرم خطی تابع کاب - داگلاس می‌باشد. یافته‌های بدست آمده حاکی از توانایی مدل شبکه عصبی و عدم توانایی مدل افزودنی کلی در پیش‌بینی بازده سهام بوده و همچنین نشان‌دهنده برتری عملکرد و کارایی شبکه عصبی پرسپترون چند لایه نسبت به مدل افزودنی کلی می‌باشد.

واژه‌های کلیدی: بازده سهام، پیش‌بینی، داده‌کاوی، شبکه عصبی، مدل افزودنی کلی.

مقدمه

سال‌های متمادی است که پژوهشگران و تحلیلگران حوزه مالی در جستجوی یافتن سنجه عملکردی هستند که از سوی آن بتوانند با نزدیکترین تقریب، بازده و تغییرات آن را برای یک شرکت پیش‌بینی کنند، چرا که معمولاً بازده سهام تأثیر مستقیمی بر ثروت سهامداران دارد و اگر تصمیم گیرندگان بازار اوراق بهادار به طریقی بازده سهام را پیش‌بینی کرده و سهام با بازده آتی پایین را فروخته و سهام با بازده آتی بالا را خریداری کنند خواهند توانست ثروت خود را حداکثر کنند (۱۷/۱۳۸۹:ص ۱۲۲). در این بین به نظر می‌رسد که سرمایه‌گذاران می‌توانند به اطلاعات سیستم‌های حسابداری اتکا کنند، زیرا مدت‌ها است که اطلاعات خروجی سیستم حسابداری، وظیفه ارائه اطلاعات صحیح و مربوط به ارزیابی عملکرد شرکت‌ها را بر عهده گرفته و از مهمترین پارامترها در تصمیم‌گیری‌های اقتصادی بازیگران بازارهای مالی و سرمایه محسوب می‌شود (۱۳/۱۳۸۹:ص ۱۴۷). از سوی دیگر با توجه به پیشرفت‌های اخیر در علم داده‌کاوی، می‌توان ابزارهای مفیدی را در پیش‌بینی محیط‌های شلوغ مانند بازارهای بورس در تسخیر رفتارهای غیرخطی مورد استفاده قرار داد (۱۸/۱۳۹۱:ص ۱۵). داده‌کاوی کمک می‌کند تا سازمان‌ها با کاوش بر روی داده‌های یک سیستم، الگوها، روندها و رفتارهای آینده را کشف و پیش‌بینی کرده و بهتر تصمیم بگیرند (۱۸/۱۳۹۱:ص ۱۵). از این رو هدف از انجام این پژوهش علاوه بر پیش‌بینی بازده سهام شرکت‌های پذیرفته شده در بورس و اوراق بهادار تهران با استفاده از تکنیک‌های داده‌کاوی اعم از شبکه عصبی پرسپترون چندلایه (MLP) و مدل افزودنی کلی (GAM)، ارزیابی قدرت و تعیین کارایی هر یک از مدل‌های مذکور می‌باشد. بر اساس ماهیت پژوهش، دو سؤال به شرح زیر مطرح می‌شود که با بهره‌گیری و بررسی هر یک از تکنیک‌های مذکور، قابلیت پاسخ به این سؤالات فراهم می‌گردد:

الف) چگونه هر یک از مدل‌های شبکه عصبی و مدل افزودنی کلی (GAM) قدرت پیش‌بینی بازده سهام شرکت‌های مورد مطالعه را دارا هستند؟

ب) از بین مدل‌های مذکور کدام یک در پیش‌بینی بازده سهام شرکت‌های مورد مطالعه کارا تر است؟

مبانی نظری و پیشینه پژوهش

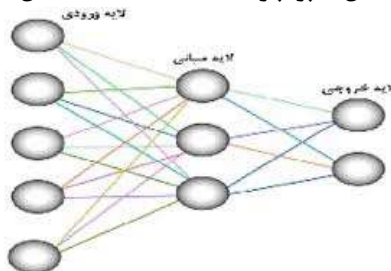
در ادبیات نظری این پژوهش به تشریح تکنیک‌های داده‌کاوی مطرح می‌پردازیم.

شبکه عصبی پرسپترون چندلایه (MLP)^۱

شبکه‌های عصبی نوعی مدل‌سازی ساده‌انگارانه از سیستم عصبی واقعی هستند که کاربرد فراوانی در حل مسائل مختلف در علوم دارند. حوزه کاربرد این شبکه‌ها آنچنان گسترده است که از کاربردهای طبقه‌بندی گرفته تا کاربردهایی نظیر درون‌یابی، تخمین، آشکارسازی و... را شامل می‌شوند. (۱۵/۱۳۸۷:ص ۳) شبکه‌ها انواع مختلفی دارند، اما همگی آنها از دو مؤلفه تشکیل می‌شوند، مجموعه‌ای از نرون‌ها و اتصالات

بین نرون ها. هر نرون در حقیقت واحد محاسباتی شبکه است که ورودی ها را گرفته و بر روی آن پردازش انجام می دهند تا خروجی بدست آید (۱۳۸۷/۱۵:ص ۴). یکی از ساده ترین و در عین حال کارآمدترین چیدمان های پیشنهادی برای استفاده در مدل سازی عصب های واقعی، مدل پرسپترون چندلایه یا به اختصار MLP می باشد که از یک لایه ورودی، یک یا چند لایه پنهان و یک لایه خروجی تشکیل شده است. در این ساختار، تمام نرون های یک لایه به تمام نرون های لایه بعد متصلند. این چیدمان اصطلاحاً یک شبکه با اتصالات کامل را تشکیل می دهد. شکل ۱ شمای یک پرسپترون دولایه است. به سادگی می توان استنباط نمود که تعداد نرون های هر لایه، مستقل از تعداد نرون های دیگر لایه ها می باشد. توجه به این نکته حائز اهمیت است که در شکل ۱، هر دایره جمع شده عمل جمع و آستانه گذاری (عبور از تابع غیرخطی سیگموئید) است (۱۳۸۷/۱۵:ص ۷).

شکل ۱. پرسپترون دو لایه با اتصالات کامل



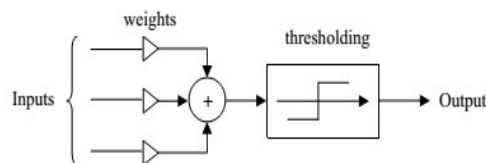
در شکل ۲ که به منظور سهولت نمایش به این فرم نشان داده شده است، با توجه به شکل، خروجی عصب i ام (در لایه آخر) را می توان به صورت زیر نشان داد (۱۳۸۷/۱۵:ص ۸):

$$O_i = \text{sgm}(\sum \text{sgm}(\sum x_i w_{lm}^h) w_{mi}^o) \quad \text{رابطه ۱}$$

که در آن h و o به ترتیب نشان دهنده لایه نهان و لایه خروجی بوده و منظور از w همان وزن لایه ها می باشد. sgm نیز تابع سیگموئید است که به صورت زیر تعریف می شود (۱۳۸۷/۱۵:ص ۸):

$$\text{Sgm}(x) = \frac{1}{1+e^{-x}} \quad \text{رابطه ۲}$$

شکل ۲. مدل ریاضی ساده شده عصب واقعی



وزن ها در محاسبات لایه های میانی استفاده می شوند و طرز استفاده آنها به این صورت است که هر نرون در لایه میانی (لایه هایی غیر از لایه اول) دارای چند ورودی از یال مختلف می باشد که هر کدام یک وزن خاص دارند. هر نرون لایه میانی، میزان هر ورودی را در وزن یال مربوطه آن ضرب می کند و حاصل این ضرب ها را با هم جمع می کند و سپس یک تابع از پیش تعیین شده (تابع فعال سازی) روی این حاصل اعمال می کند و نتیجه را به عنوان خروجی به نودهای لایه بعد می دهد. تعداد نرون ها و تعداد لایه های نهان و نحوه وصل شدن گره ها به یکدیگر معماری (توپولوژی) شبکه عصبی را مشخص می کند (۱۳۹۱/ ۱۸:ص ۱۱ و ۱۲).

مدل افزودنی کلی (GAM)^۱

مدل افزودنی کلی (GAM) که با عناوینی چون مدل افزایشی تعمیم یافته و مدل جمعی تعمیم یافته مورد استفاده قرار می گیرد، در واقع بسطی بر روش های رگرسیون خطی و رگرسیون منطقی می باشد. به این دلیل به این روش افزودنی می گویند که فرض می کنیم می توانیم مدل را به صورت مجموع چند تابع غیرخطی (هر تابع برای یک متغیر پیش بینی کننده) بنویسیم. این مدل می تواند هم به منظور رگرسیون و هم به منظور کلاس بندی داده ها استفاده گردد. این ویژگی غیرخطی بودن توابع باعث می شود که این روش نسبت به روشهای رگرسیون خطی بهتر باشد. ارتباط متدولوژیک بین رگرسیون منطقی و نگرش GAM با تشریح چارچوب مدل خطی تعمیم یافته^۲ (GLM) یافته آغاز شده است. در حقیقت مدل افزودنی کلی، یک مدل خطی تعمیم یافته یا یک پیشگویی خطی بوده که شامل مجموعه ای از توابع هموار برای متغیرها است. در این مدل ها فرض می شود که میانگین متغیر وابسته از طریق یک تابع پیوند به متغیر مستقل وابسته است. بنابراین متغیر وابسته می تواند هر توزیعی از خانواده نمایی (مثل توزیع نرمال، پواسون، دوجمله ای، گاما) را داشته باشد. در مدل خطی تعمیم یافته مقدار مورد انتظار متغیر وابسته به یک تابع یکنوای هموار از پیشگو خطی وابسته می باشد. مدل خطی تعمیم یافته ساختار اصلی زیر را دارد (۱۳۹۳/۱۹:ص ۱۶۴):

$$g(\mu_i) = X_i^T \beta$$

رابطه ۳)

که در آن $\mu_i \equiv E(Y_i)$ ، g یک تابع پیوند یکنوا هموار، بردار X_i' ، i امین سطر از ماتریس طرح X و β بردار پارامترهای نامعلوم است. چون مدل‌های خطی تعمیم یافته از طریق پیشگویی خطی $X'\beta$ تعیین می‌شود، بیشتر مفاهیم و نظریه‌های کلی مدل‌سازی خطی با مقداری تغییر، به مدل‌سازی خطی تعمیم یافته انتقال می‌یابد. انتخاب‌های مختلف تابع توزیع و معادله در نتیجه تکنیک‌های مختلف آماری بوجود می‌آیند. برای مثال، توزیع نرمال در ترکیب با معادله اتحاد منتج به مدل رگرسیون چندگانه نرمال می‌شود. وقتی پاسخ‌های Y دودویی (باینری) هستند، تابع توزیع عبارت است از توزیع دوجمله‌ای و تفسیر μ_i عبارت است از احتمال (p_i) ، Y_i با گرفتن مقدار یک. بنابراین مدل رگرسیون منطقی با اجرای معادله لوجیت عبارت است از:

$$\log_{it}\{P(X)\} \equiv \log\left\{\frac{P(X)}{1-P(X)}\right\} = \alpha + \sum_{j=1}^p \beta_j^1 X_j \quad \text{رابطه ۴}$$

مدل افزودنی کلی در حالت کلی ساختاری به شکل زیر دارد:

$$g(\mu_i) = X_i' \beta + \sum_{j=1}^p p_j = 1 f_j(X_j) \quad \text{رابطه ۵}$$

حال فرض کنید که Y پاسخ متغیر با برجسب‌های هدف دوتایی و $X_1, X_2, \dots, X_p$ یک سری متغیرهای مستقل هستند، مدل افزودنی کلی قانون رگرسیون استدلالی را به وسیله جایگزینی پیش‌بین خطی $\sum_{j=1}^p \beta_j X_j$ در رابطه ۴ با دیگر مؤلفه جمعی در رابطه ۵ که در اینجا به قرار زیر است:

$$\log_i \log_{it}\{P(X)\} \equiv \log\left\{\frac{P(X)}{1-P(X)}\right\} = \alpha + \sum_{j=1}^p p_j = 1 f_j(X_j) \quad \text{رابطه ۶}$$

این تحقیق از خط پارامتری برای $f_1(x_1), f_2(x_2), \dots, f_p(x_p)$ جهت برآورد روند غیرپارامتری برای تبعیت از آزمونگر استنباطی استدلالی در $X_1, X_2, \dots, X_p$ استفاده می‌کند. به عبارت دیگر (۱۳۹۳/۱۹:ص ۱۶۷):

$$Y = \alpha + f_1(x_1) + f_2(x_2) + \dots + f_p(x_p) \quad \text{رابطه ۷}$$

با توجه به رابطه ۷، مدل مورد استفاده در پژوهش به قرار زیر می‌باشد:

$$R_{it} = \alpha + f_1(OP) + f_2(EPS) + f_3(CR) + f_4(QR) \quad \text{رابطه ۸}$$

این پژوهش یک نمای عمومی از اصول مدل افزودنی کلی را ارائه می‌کند.

در ادامه به برخی از تحقیقات انجام شده در خصوص پیش‌بینی با استفاده از تکنیک‌های داده کاوی پرداخته می‌شود که از جمله آنها می‌توان به تحقیقات زیر اشاره کرد:

پیشینه پژوهش‌های خارجی

هنریک، سابریو و کیمارا (۲۰۱۸/۲۱:ص ۱۸۳)، با استفاده از تکنیک رگرسیون بردار پشتیبان و قیمت‌های روزانه در سه بازار مختلف به پیش‌بینی قیمت سهام پرداخته‌اند. نتایج بدست آمده نشان دهنده این است که رگرسیون بردار پشتیبان دارای قدرت پیش‌بینی است. کراناپولاس، پاپادیمیتریو و ولز تاکیز (۲۰۱۸/۲۲:ص ۵۹)، فرضیه خود مبنی بر قابل پیش‌بینی بودن بازده سهام را با استفاده از درخواست اطلاعات تخمینی با توجه به مقدار عددی شاخص جستجوی روزانه اینترنتی (SVI) از طریق گوگل مورد بررسی قرار دادند. نتایج بدست آمده نشان دهنده این است که ترکیب متغیر SVI با مدل‌های مختلف GARCH به طور معناداری عوامل ناپایداری را بهبود می‌بخشد. ناهیل و لیهایوی (۲۰۱۸/۲۰:ص ۱۶۱)، تلاش کردند تا برای پیش‌بینی بهتر قیمت سهام از مدل جامع پیش‌بینی مبتنی بر تجزیه و تحلیل مؤلفه اصلی کرنل (KPCA) و رگرسیون ماشین بردار پشتیبان (SVR) استفاده کنند. نتایج همانندسازی نشان می‌دهند که، ساده‌سازی مشخصه‌های کرنل در تمام ساختار سیستم تصمیم‌گیری سرمایه‌گذاری می‌تواند به طور مطلوب با بهسازی عملکرد مدل، ساده شده باشد. ژانگ و انک (۲۰۱۷/۲۴:ص ۱۵۲)، در پژوهش خود فرآیند داده کاوی را برای پیش‌بینی جهت‌گیری روزانه شاخص S&P500 بر اساس ۶۰ ویژگی مالی و اقتصادی را ارائه کرده‌اند. شبکه عصبی مصنوعی (ANNs) و رگرسیون لجستیک مدل‌هایی هستند که برای طبقه‌بندی استفاده می‌کنند. نتایج این طور نشان می‌دهند که شبکه‌های عصبی به طور معنادار دقت بیشتری در طبقه‌بندی نسبت به رگرسیون لجستیک دارند. همچنین آنها (۲۰۱۷/۲۵:ص ۱۲۶) مجدداً پژوهش خود را با اعمال تکنیک‌های کاهش سه بعدی کامل، از قبیل تجزیه و تحلیل مؤلفه اصلی (PCA)، مؤلفه اصلی قدرت فازی (FRPCA) و مؤلفه اصلی الگوریتم kernel-based (KPCA)، به مجموعه داده‌های کامل برای ساده‌سازی و تنظیم مجدد ساختار اصلی داده‌ها انجام دادند. نتایج حاصل از تلفیق ANNs با PCA پایین بودن دقت طبقه‌بندی نسبت به دو ترکیب دیگر را نشان می‌دهد. بارالیس، کاگلیر و گارزا (۲۰۱۷/۲۳:ص ۱۵ و ۱۶)، در پژوهش خود رویکردی مبتنی بر مجموعه اقلام برای حمایت از سرمایه‌گذاران به هنگام خرید و نگهداری سهام به وسیله تجزیه و تحلیل تکنیکال از طریق ابزارهای خودکار جهت شناسایی مجموعه سهام پربازده و متنوع، را ارائه کردند. بررسی‌های انجام شده بر روی شاخص‌های مختلف بازار سهام، تحت شرایط مختلف بازار نشان دهنده این است که روش پیشنهادی قابلیت برنامه‌ریزی سبد سهام پربازده و متنوع را دارد.

1- Henrique, Sobreiro, Kimura

2- Chronopoulos, Papadimitriou, Vlastakis

3- Nahil, Lyhyaoui

4- Zhong, Enke

5- Baralis, Cagliero, Garza

پیشینه پژوهش های داخلی

در ایران نیز، زندیه و مردانلو (۱۳۹۶/۱۰:ص ۱۰۹)، سعی کردند تا با ارائه یک مدل تکاملی چند هدفه جدید با استفاده از دو الگوریتم ژنتیک چند هدفه NSGA II و NPGA، یک سیستم سیگنال دهی خودکار به منظور پشتیبانی از تصمیمات سرمایه گذاران ایجاد کنند. نتایج بدست آمده نشان می دهند که با در نظر گرفتن دقت بالای پیش بینی، مدل می تواند به عنوان ابزاری برای تصمیم گیری در معاملات بازار سرمایه جهت افزایش میزان سوددهی در بازار سرمایه، در چارچوب های زمانی مورد استفاده قرار گیرد. سروش یار و اخلاقی (۱۳۹۶/۱۲:ص ۵۷ و ۵۸)، با استفاده از چهار الگوریتم تحلیل جداساز خطی، تحلیل جداساز غیرخطی، نزدیکترین K همسایگی و درخت تصمیم و به کمک ۱۶ متغیر مستقل به پیش بینی بازده و ریسک سیستماتیک سهام پرداخته اند. نتایج حاکی از این است که استفاده از متغیرهای مستقل منتخب (به جای استفاده از کل متغیرهای مستقل) توان الگوریتم ها را در پیش بینی بازده و ریسک سیستماتیک بهبود می بخشد. ستایش و کاظم نژاد (۱۳۹۵/۱۱:ص ۱)، برای پیش بینی بازده سهام شرکت های پذیرفته شده در بورس اوراق بهادار تهران به مقایسه عملکرد رگرسیون های تجمعی و روش مبتنی بر همبستگی و ریلیف با رگرسیون خطی و شبکه عصبی مصنوعی پرداختند. نتایج حاکی از عملکرد بهتر روش رگرسیون تجمعی نسبت به رگرسیون خطی و شبکه های عصبی مصنوعی است. افزون بر این، یافته ها حاکی از آن بود که پیش بینی با استفاده از متغیرهای انتخاب شده در روش های مبتنی بر همبستگی و ریلیف، به طور معناداری عملکرد پیش بینی را نسبت به استفاده از کلیه متغیرها افزایش می دهد. برزگری خانقاه و جمالی (۱۳۹۵/۹:ص ۷۱)، به بررسی مطالعات صورت گرفته در تعیین نقش نسبت های مالی در پیش بینی بازده سهام به عنوان یکی از اهداف عمده سهامداران پرداخته اند. در این پژوهش که به تحقیقات داخلی مراجعه شده است، نتایج حاکی از آن است که نسبت های سودآوری در بین سایر گروه های نسبت های مالی از سهم بالاتری در پیش بینی بازده سهام برخوردارند و در این بین نسبت بازده دارایی و بازده حقوق صاحبان سهام بیشترین توانایی در توضیح تغییرات بازده سهام را از خود نشان داده اند. علیمحمدی، عباسی مهر و جواهری (۱۳۹۴/۱۴:ص ۱۲۹)، به منظور بررسی توانایی نسبت های مالی در تبیین بازده معاصر و پیش بینی بازده آتی سهام، از چهار الگوریتم درخت تصمیم (شامل CHAID، ECHAID، QUEST و CRT) استفاده کردند. نتایج بدست آمده نشان می دهد که الگوریتم CRT و ECHAID در تبیین بازده معاصر و الگوریتم CHAID در پیش بینی بازده آتی بهترین عملکرد را دارند. همچنین قدرت مدل ها در تبیین بازده معاصر بیشتر از پیش بینی بازده آتی است. هرچند که در هر دو حالت، توانایی مدل ها از نظر آماری قابل اتکا نبوده و از این رو فرضیه برقراری ارتباط تبیینی بین نسبت های مالی مطرح شده در پژوهش و تغییرات بازده معاصر و آتی سهام رد می شود.

فرضیه‌های پژوهش

با توجه به اهداف پژوهش و سؤالات مطرح شده، فرضیه‌های این تحقیق به قرار زیر می‌باشد:

– شبکه عصبی قدرت پیش‌بینی بازده سهام شرکت‌های پذیرفته شده در بورس و اوراق بهادار تهران را دارد.

– مدل افزودنی کلی (GAM) قدرت پیش‌بینی بازده سهام شرکت‌های پذیرفته شده در بورس و اوراق بهادار تهران را دارد.

متغیرهای پژوهش

در این پژوهش متغیر وابسته بازده سهام و از بین عوامل متعددی که بر بازده سهام اثر می‌گذارند، چهار عامل به عنوان عوامل مؤثر بر بازده سهام (متغیرهای مستقل) انتخاب شده‌اند که به قرار زیر می‌باشند (۱۳۹۲: ۱۶ ص ۵ و ۶):

جدول ۱. متغیرهای پژوهش

نام متغیر	نوع متغیر	علامت اختصاری	نحوه محاسبه
بازده سهام	وابسته	R_{it}	بازده سهام $= \frac{\text{سهام جایزه} + \text{حق تقدم} + \text{سود نقدی سهام} + (\text{قیمت روز} - \text{قیمت پایه})}{\text{سهام جایزه} + \text{حق تقدم} + \text{سود نقدی سهام} + (\text{قیمت روز} - \text{قیمت پایه})}$ (درصد افزایش سرمایه از محل آورده) + قیمت پایه
سود (زیان) عملیاتی	مستقل	OP	هزینه‌های عملیاتی - درآمدهای عملیاتی = سود (زیان) عملیاتی
سود هر سهم	مستقل	EPS	سود خالص $= \frac{\text{سود هر سهم}}{\text{سرمایه شرکت (میانگین موزون تعداد سهام منتشره)}}$
نسبت جاری	مستقل	CR	داراییهای جاری $= \frac{\text{نسبت جاری}}{\text{بدهیهای جاری}}$
نسبت آنی	مستقل	QR	موجودی‌ها - داراییهای جاری $= \frac{\text{نسبت آنی}}{\text{بدهیهای جاری}}$

روش‌شناسی پژوهش

پژوهش حاضر از لحاظ هدف کاربردی و از نوع تحقیقات توصیفی-همبستگی می‌باشد. جهت گردآوری اطلاعات در زمینه ادبیات نظری، از منابع کتابخانه‌ای، مقالات، کتاب‌های مورد نیاز و پایان‌نامه، و جهت گردآوری داده‌ها، از طریق مراجعه به صورت‌های مالی حسابرسی شده، داده‌ها و اطلاعات تاریخی

شرکت‌های پذیرفته شده در بورس و اوراق بهادار تهران در سایت رسمی بورس و اوراق بهادار تهران، سایت کدال، سایت شرکت پردازش اطلاعات مالی ایران^۳ و همچنین با بهره‌گیری از نرم افزار رهاورد نوین^۳، صورت گرفته است. جامعه آماری این پژوهش را شرکت‌های پذیرفته شده در بورس و اوراق بهادار تهران در بازه زمانی ۱۳۹۱ الی ۱۳۹۶ تشکیل می‌دهند، لذا برای تعیین جامعه هدف از روش حذف سیستماتیک استفاده شده است. از این رو شرکت‌هایی که شرایط زیر را دارا باشند به عنوان جامعه هدف انتخاب و مابقی شرکت‌ها حذف می‌شوند که با توجه به شرایط ذکر شده، تعداد ۶۷ شرکت به عنوان جامعه هدف انتخاب شده‌اند.

شرکت‌ها قبل از سال ۱۳۹۱ در بازار سرمایه پذیرفته شده و جزء بازار اول (تابلوی اصلی و فرعی) باشند؛ دوره مالی آنها منتهی به اسفند ماه هر سال (۱۳۹۱ الی ۱۳۹۶) باشد؛ صورت‌های مالی آنها برای هر دوره در دسترس باشد.

یافته‌های پژوهش

تخمین روابط بین متغیرها بر اساس مدل آماری پژوهش

روش به کار گرفته شده فرم خطی تابع کاب – داگلاس می‌باشد. هدف ما از به کارگیری این مدل بررسی درصد تغییر پذیری و به نوعی پیش‌بینی بازده سهام شرکت‌های پذیرفته شده در بورس و اوراق بهادار تهران نسبت به تغییر متغیرهای مستقل است. آنچه ما باید در این معادله تخمین بزنیم، ضرایب معادله هستند. با توجه به پژوهش حاضر، فرم کلی و فرم خطی تابع کاب – داگلاس برای معادله‌ای با چهار متغیر مستقل به صورت زیر می‌باشد:

$$y = \alpha x_1^{\beta_1} x_2^{\beta_2} x_3^{\beta_3} x_4^{\beta_4} \quad \text{رابطه ۹}$$

$$\log(y) = \log(\alpha) + \beta_1 \log(x_1) + \beta_2 \log(x_2) + \beta_3 \log(x_3) + \beta_4 \log(x_4) \quad \text{رابطه ۱۰}$$

در معادله فوق y همان متغیر وابسته یعنی بازده سهام؛ x_1 ، x_2 ، x_3 و x_4 متغیرهای مستقل هستند که به ترتیب بیانگر سود (زیان) عملیاتی، سود هر سهم، نسبت جاری و نسبت آئی می‌باشند. به عبارت دیگر:

$$\log(R_{it}) = \log(\alpha) + \beta_1 \log(OP) + \beta_2 \log(EPS) + \beta_3 \log(CR) + \beta_4 \log(QR) \quad \text{رابطه ۱۱}$$

جهت سهولت در انجام فرآیند، از نرم افزار EViews9 برای تخمین ضرایب و تحلیل نتایج حاصل از مدل آماری استفاده شده است.

1- <http://new.tse.ir>
2- <http://www.codal.ir>
3- <http://www.fipiran.com>

جدول ۲. نتایج بدست آمده از تخمین ضرایب مدل آماری

متغیرها	ضرایب تخمین	انحراف معیار	آماره t
عرض از مبدأ	۰/۶۶۰۷۶۰	۰/۰۲۷۹۰۰	۲۳/۶۸۲۷۹
سود (زیان) عملیاتی	۰/۰۴۶۱۷۳	۰/۰۲۸۷۲۰	۱/۶۰۷۷۱۹
سود هر سهم	۰/۰۲۰۴۸۵	۰/۰۲۸۹۶۵	۰/۷۰۷۲۲۰
نسبت جاری	-۰/۰۱۸۸۲۱	۰/۰۴۷۶۱۲	-۰/۳۹۵۳۰۷
نسبت آتی	۰/۰۳۸۴۴۷	۰/۰۶۲۲۷۴	۰/۶۱۷۳۸۷
معیار آکائیک (AIC)	۱/۳۵۹۹۰۳		

با توجه به فرم خطی تابع کاب - داگلاس و جدول ۲ معادله اصلی به صورت زیر می باشد:

$$y = (0/660760) + (0/046173)OP + (0/020485)EPS + (-0/018821)CR + (0/038447)QR \quad (\text{رابطه ۱۲})$$

در معادله فوق y همان $\log(R_{it})$ می باشد.

از آنجا که هدف ما تخمین ضرایب معادله می باشد، لذا با توجه به ضرایب بدست آمده؛

- به ازای ۱ واحد تغییر در سود (زیان) عملیاتی، $0/046173$ واحد تغییر در متغیر وابسته (بازده سهام) ایجاد می شود. به عبارت دیگر ۱ واحد افزایش سود (زیان) عملیاتی، بازده سهام را $0/046173$ واحد افزایش می دهد.

- به ازای ۱ واحد تغییر در سود هر سهم، $0/020485$ واحد تغییر در متغیر وابسته (بازده سهام) ایجاد می شود. به عبارت دیگر ۱ واحد افزایش سود هر سهم، بازده سهام را $0/020485$ واحد افزایش می دهد.

- به ازای ۱ واحد تغییر در نسبت جاری، $0/018821$ واحد تغییر در متغیر وابسته (بازده سهام) ایجاد می شود. به عبارت دیگر ۱ واحد افزایش نسبت جاری، بازده سهام را $0/018821$ واحد کاهش می دهد.

- به ازای ۱ واحد تغییر در نسبت آتی، $0/038447$ واحد تغییر در متغیر وابسته (بازده سهام) ایجاد می شود. به عبارت دیگر ۱ واحد افزایش نسبت آتی، بازده سهام را $0/038447$ واحد افزایش می دهد.

آزمون فرضیه اول

در طراحی شبکه جهت پیش بینی بازده سهام از نرم افزار SPSS20 استفاده شده است. به منظور شبیه سازی شبکه عصبی، داده های ورودی را به دو قسمت داده های آموزش و تست دسته بندی کرده ایم که $67/7$ درصد داده ها معادل ۲۷۲ داده به عنوان داده های آموزش و $32/3$ درصد داده ها معادل ۱۳۰ داده به عنوان داده های تست مورد استفاده قرار گرفته است.

جدول ۳. خلاصه فرآیند آماده سازی داده‌ها

		تعداد	درصد
نمونه	داده‌های آموزش	۲۷۲	۶۷.۷٪
	داده‌های تست	۱۳۰	۳۲.۳٪
کل		۴۰۲	۱۰۰.۰٪

در طراحی شبکه باید اندازه مجموعه‌های آموزش و تست، تعداد لایه‌های پنهان شبکه، تعداد واحدهای (نرون‌های) هر لایه، تابع تبدیل و تابع عملکرد مشخص شود. در این پژوهش ۴ معیار به منظور پیش‌بینی بازده سهام مطرح است. بنابراین در این تحقیق تعداد نرون‌های لایه ورودی ۴ و با توجه به اینکه شبکه دو خروجی دارد تعداد لایه‌های خروجی ۲ می باشد.

جدول ۴. اطلاعات شبکه

لایه ورودی	متغیرهای مستقل	۱ ۲ ۳ ۴	سود (زیان) عملیاتی سود هر سهم نسبت جاری نسبت آتی
لایه‌های پنهان (میانی)	تعداد واحدها مقیاس بندی متغیرهای وابسته تعداد لایه‌های پنهان تعداد واحدها در لایه پنهان اول تعداد واحدها در لایه پنهان دوم	۴ ۲ ۳ ۲	استاندارد شده سیگموئید
لایه خروجی	متغیر وابسته تعداد واحدها تابع عملکرد تابع خطا	۲ سیگموئید مجموع مربعات	بازده سهام

جدول ۴، اطلاعات مربوط به شبکه عصبی را نشان داده و برای حصول از اینکه موارد اختصاص یافته صحیح می باشند، کاربرد دارد. به موارد زیر توجه نمایید:

– تعداد واحدهای موجود در لایه ورودی برابر با تعداد متغیرهای مستقل می باشد.

– تعداد واحدهای خروجی در کل ۲ واحد لایه خروجی است.

– دو لایه پنهان در خواست شده و فرآیند ۳ واحد در لایه اول و ۲ واحد را در لایه دوم پنهان، قرار داده است.

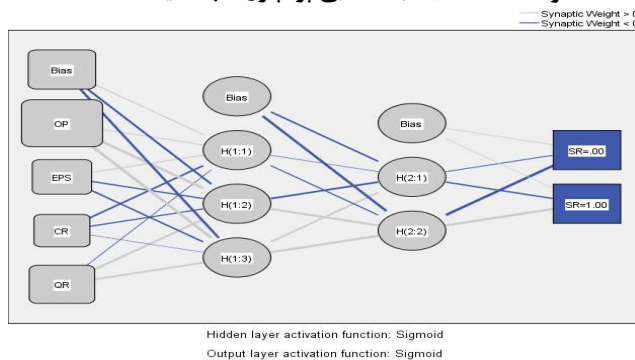
– برای هر یک از متغیرهای وابسته به مقیاس یک واحد خروجی جدا و مستقل ایجاد گردیده است که در آن برای لایه خروجی نیاز به تابع فعال‌سازی سیگموئید داریم.

– در جدول، میزان مجموع مربعات خطاها نیز آورده شده است.

ساختار شبکه عصبی و توپولوژی

ساختار شبکه عصبی و توپولوژی یا به اصطلاح معماری شبکه نشان دهنده ساختار شبکه در این پژوهش می‌باشد که این ساختار شامل ۴ متغیر مستقل به عنوان لایه ورودی، فرایند ۳ واحد در لایه پنهان اول، فرایند ۲ واحد در لایه پنهان دوم و متغیر وابسته که شامل دو طبقه افزایش و کاهش یعنی ۲ خروجی می‌باشد.

نمودار ۱. معماری شبکه عصبی پرسپترون چندلایه (MLP)



خلاصه مدل

جدول ۵. خلاصه مدل (متغیر وابسته : بازده سهام)

آموزش	مجموع مربعات خطا درصد خطای پیش‌بینی عدم کاهش الگوریتم تخمین	۶۱/۰۹۹ ۳۵/۰۷ توقف الگوریتم
تست	مجموع مربعات خطا درصد خطای پیش‌بینی	۲۷/۷۲۹ ۳۳/۰۱

در این بخش اطلاعات مربوط به نتایج آموزش و تست انجام شده است. در این جدول مجموعه مربعات خطاها نشان داده شده است. این همان تابع خطایی است که شبکه تلاش دارد در طی عملیات آموزش آن را به کمترین میزان خود برساند. باید توجه داشت که مقادیر مربوط به مجموع مربعات و مابقی خطاها که در جدول آورده شده است، برای مقادیر از نو مقیاس شده متغیر وابسته، محاسبه گردیده اند.

میزان پیش بینی ها اشتباه که به صورت درصد نشان داده شده است. از آنجایی که میزان خطا پس از گذشت یک مرحله از الگوریتم تخمین کاهش نیافته، این الگوریتم متوقف گردیده است.

با توجه به جدول ۵ مجموع مربعات خطای بدست آمده از برازش این مدل برابر ۶۱/۰۹۹ در نمونه آموزش و ۲۷/۷۲۹ در نمونه تست می باشد و از آنجایی که مقدار درصد خطای پیش بینی بدست آمده در نمونه آموزش بیشتر از نمونه تست است بنابراین برازش مدل قابل قبول است.

طبقه بندی

جدول طبقه بندی نشان می دهد که شبکه به شکل قابل ملاحظه ای در پیش بینی افزایش بازده سهام نسبت به کاهش بازده سهام بهتر عمل می نماید.

جدول ۶. طبقه بندی

پیش بینی			
	کاهش بازده سهام (کد ۰)	افزایش بازده سهام (کد ۱)	درصد محاسبه شده
نمونه آموزش:			
کاهش بازده سهام (کد ۰)	۰	۹۷	۰/۰
افزایش بازده سهام (کد ۱)	۰	۱۷۵	۱۰۰/۰
کل بر حسب درصد	۰/۰	۱۰۰/۰	۶۴/۳
نمونه تست:			
کاهش بازده سهام (کد ۰)	۰	۴۳	۰/۰
افزایش بازده سهام (کد ۱)	۰	۸۷	۱۰۰/۰
کل بر حسب درصد	۰/۰	۱۰۰/۰	۶۶/۹

بر اساس جدول ۶، برای ۴۰۲ داده مورد استفاده ۹۷ مورد خطای طبقه بندی در نمونه آموزش و ۴۳ مورد خطای طبقه بندی در نمونه تست وجود دارد (در نمونه آموزش ۹۷ مورد از داده ها که مربوط به تعیین میزان

کاهش بازده سهام هستند در خصوص افزایش بازده پیش بینی شده‌اند و خطایی در خصوص داده‌ها که مربوط به تعیین میزان افزایش بازده سهام هستند رخ نداده است؛ در نمونه تست ۴۳ مورد از داده‌ها که مربوط به تعیین میزان کاهش بازده سهام هستند در خصوص افزایش بازده پیش‌بینی شده‌اند و خطایی در خصوص داده‌ها که مربوط به تعیین میزان افزایش بازده سهام هستند رخ نداده است) بر اساس جدول فوق، شبکه در نمونه آموزش و نمونه تست به ترتیب ۶۴/۳ و ۶۶/۹ درصد بازده سهام شرکت‌ها را بر اساس متغیرهای مستقل به درستی طبقه‌بندی کرده است.

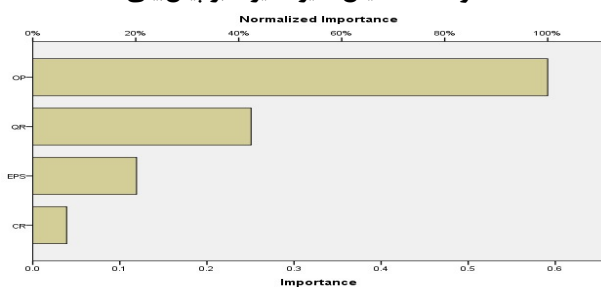
اهمیت متغیرهای مستقل

اهمیت متغیرهای مستقل در تشخیص این نکته است که به چه میزان مقادیر پیش‌بینی شده توسط شبکه، با مقادیر متغیر مستقل، تغییر می‌نماید. نرمال سازی این اهمیت بسیار ساده است و با تقسیم مقادیر اهمیت به بزرگترین مقدار آن حاصل می‌شود و به صورت درصد بیان می‌گردد. در این پژوهش آرایش تأثیر متغیرها بر پیش‌بینی بر اساس فرم خطی تابع کاب - داگلاس، مدل آماری پژوهش و با استفاده از شبکه عصبی (MLP) مشخص شده‌اند که سود (زیان) عملیاتی، سود هر سهم، نسبت جاری و نسبت آتی به ترتیب با ضرایب نفوذ ۰/۵۹۱، ۰/۱۱۹، ۰/۰۳۹ و ۰/۲۵۱ بر بازده سهام اثرگذار هستند. در نمودار ۲، اهمیت هر یک از متغیرهای مستقل و تأثیر آنها بر بازده سهام که به صورت نزولی مرتب شده‌اند، به روشنی قابل مشاهده است.

جدول ۷. اهمیت متغیرهای مستقل

متغیرها	اهمیت	اهمیت‌های نرمال شده
سود (زیان) عملیاتی	۰/۵۹۱	٪ ۱۰۰/۰
سود هر سهم	۰/۱۱۹	٪ ۲۰/۱
نسبت جاری	۰/۰۳۹	٪ ۶/۶
نسبت آتی	۰/۲۵۱	۴۲٪/۴

نمودار ۲. آرایش تأثیر متغیرها بر پیش‌بینی



آزمون فرضیه دوم

برای برآورد مدل افزودنی کلی جهت پیش بینی بازده سهام از نرم افزار SPSS20 و 9 EViews استفاده شده است. حجم نمونه معتبر ۴۰۲ است و داده ناقص یا گمشده ای وجود ندارد. به بیان دیگر ۱۰۰ درصد داده ها که برابر با ۴۰۲ می باشد در پردازش، شرکت داده شده اند. جدول ۸ بیانگر این مطلب است و جدول ۹، بیانگر میانگین، خطای استاندارد و ضریب تغییرات می باشد.

جدول ۸. خلاصه فرآیند انتخاب داده ها برای برآورد مدل افزودنی کلی

درصد نهایی	تعداد	
۱۰۰٪/۰	۴۰۲	حجم نمونه معتبر
	۰	داده های ناقص یا گمشده
	۴۰۲	کل

جدول ۹. آمار توصیفی برای مدل افزودنی کلی

ضریب تغییرات	خطای استاندارد	میانگین	متغیر وابسته: بازده سهام متغیرهای مستقل: سود (زیان) عملیاتی، سود هر سهم، نسبت جاری، نسبت آبی
۱۴۱٪/۴	۰/۷۰۷۱۱	۰/۵۰۰۰	کل

جدول ۱۰، اطلاعات لازم برای قبول یا رد شدن برازش مدل را در اختیار ما می گذارد. با توجه به مدل آماری پژوهش و مقادیر معیارهای ذکر شده در جدول زیر، مقدار معیار آکائیک مورد قبول در این مدل باید کمتر یا برابر با مقدار معیار آکائیک مدل آماری پژوهش باشد در حالیکه این مقدار در مدل افزودنی کلی بیش از مدل آماری پژوهش است و تفاوت زیادی را بین دو مدل نشان می دهد، لذا بر اساس مدل آماری پژوهش و مقادیر بدست آمده برازش مدل قابل قبول نیست.

جدول ۱۰. معیارهای تأیید مدل

مدل	انحراف معیار	آماره F	معیار آکائیک (AIC)	معیار بیزی (BIC)
مدل افزودنی کلی (GAM)	۰/۰۱۶۰۶۹	۷۵۰/۴۴۳	۵۵۲/۳۳۴	۵۵۶/۳۲۸
مدل آماری پژوهش	۰/۴۷۷۰۱۲	۱/۹۹۸۱۸۶	۱/۳۵۹۹۰۳	۱/۴۰۹۶۱۰

نتیجه گیری و بحث

در پژوهش حاضر با استفاده از تکنیک‌های داده کاوی از قبیل مدل‌های شبکه عصبی (پرسپترون چند لایه (MLP)) و مدل افزودنی کلی (GAM) و با بهره‌گیری از متغیرهای سود (زیان) عملیاتی، سود هر سهم، نسبت جاری و نسبت آنی مقادیر کاهش و افزایش بازده سهام را پیش‌بینی کرده‌ایم. با توجه به جدول ۱۱، براساس مدل آماری پژوهش، شبکه عصبی (پرسپترون چند لایه (MLP))، قادر است تا به ترتیب اثرگذاری متغیرهای مستقل بر متغیر وابسته از آنها استفاده و به پیش‌بینی دقیق با ۹۷ مورد خطای طبقه‌بندی بپردازد، در حالی که روش مدل افزودنی کلی قدرت پیش‌بینی بازده سهام را ندارد، زیرا اساس پذیرش برازش این مدل بر مبنای معیار آکائیک می‌باشد و از آنجایی که معیار آکائیک مربوط به این روش (۵۵۲/۳۳۴) از معیار آکائیک مربوط به مدل آماری پژوهش (۱/۳۵۹۹۰۳) خیلی بیشتر است لذا برازش مدل قابل قبول نبوده و این روش قادر به پیش‌بینی بازده سهام نیست. با توجه به این مطلب ملاحظه می‌شود که شبکه عصبی (پرسپترون چند لایه (MLP)) عملکرد مناسبی داشته است که نتیجه بدست آمده با نتایج مطالعات گذشته چون مطالعات محمد حسین ستایش و مصطفی کاظم نژاد (۱۳۹۵) در تحقیقی با عنوان سودمندی رگرسیون‌های تجمیعی و روش‌های انتخاب متغیرهای پیش‌بین بهینه در پیش‌بینی بازده سهام، بیان کردند که نتایج پژوهش حاکی از برتری معنادار روش‌های غیرخطی رگرسیون تجمیعی (مدل افزودن کلی) نسبت به رگرسیون خطی و شبکه‌های عصبی در پیش‌بینی بازده سهام است، مطابقت ندارد.

جدول ۱۱. خلاصه تکنیک‌های داده کاوی مورد استفاده در پژوهش

روش	شبکه عصبی (پرسپترون چند لایه (MLP))	مدل افزودنی کلی (GAM)
متغیرهای مستقل پذیرفته شده	سود (زیان) عملیاتی سود هر سهم نسبت جاری نسبت آنی	سود (زیان) عملیاتی سود هر سهم نسبت جاری نسبت آنی
مشاهدات	کاهش بازده سهام (کد ۰) افزایش بازده سهام (کد ۱)	کاهش بازده سهام (کد ۰) افزایش بازده سهام (کد ۱)
مقدار پیش‌بینی شده	۰ ۱۷۵	-
مقدار خطای پیش‌بینی شده	۹۷	-
معیار اطلاعات آکائیک (AIC)	-	۵۵۲/۳۳۴
تحلیل	برازش مدل بر اساس مدل آماری پژوهش قابل قبول است	برازش مدل بر اساس مدل آماری پژوهش قابل قبول نیست

با توجه به گستره نظری موجود و یافته ها در این پژوهش، از لحاظ کاربردی پیشنهاد می شود که کارگاه های آموزشی در زمینه افزایش و کیفیت بهره گیری از تکنیک های داده کاوی در تمام حوزه های مربوط به تحلیل داده های مالی، به ویژه در خصوص پیش بینی بازده سهام برگزار گردد. همچنین در خصوص تحقیقات آتی به محققین پیشنهاد می شود تا:

– پژوهش با حجم نمونه بزرگتر انجام گیرد تا ادبیات متراکم و منسجمی در خصوص نحوه به کارگیری متغیرهای ذکر شده در این تحقیق فراهم آید.

– با توجه به نظر کارشناسان و صاحب نظران خبره نسبت های مالی را رتبه بندی و با توجه به اولویت آنها، نسبت به انتخاب نسبت های مالی به عنوان متغیرهای مستقل، اقدام نموده و از نسبت های منتخب در تکرار پژوهش حاضر استفاده شود.

– برای پژوهش های آتی پیشنهاد می شود تا محققین، پژوهش حاضر را بر کل شرکت های پذیرفته شده در بورس و شرکت های فرابورسی تکرار نمایند. این کار هم قابلیت تعمیم پذیری را افزایش می دهد و هم امکان مقایسه را فراهم می کند.

– با توجه به مطالعات انجام شده در این پژوهش، پیشنهاد می شود تا پژوهش حاضر را با استفاده از شبکه های عصبی آدالاین و یادگیری LMS انجام داده و نتایج بدست آمده را با نتایج پژوهش حاضر مورد قیاس قرار دهند.

– الگوریتم های تجمعی Bagging و Boosting به جای مدل افزودنی کلی استفاده شود و نتایج مورد بررسی قرار گیرند.

فهرست منابع و مآخذ

- آذر، عادل؛ مؤمنی، منصور (۱۳۹۲). آمار و کاربرد آن در مدیریت (جلد دوم). ویراست سوم. تهران، انتشارات سمت، چاپ نوزدهم: ۱۳۹۴.
- ژیاوی هان - میشلین کمبر - ژان پی (۱۳۹۳)؛ داده کاوی مفاهیم و تکنیک ها. ترجمه: مهدی اسماعیلی. تهران، انتشارات نیاز دانش، چاپ سوم - ۱۳۹۶.
- صنیعی آباده، محمد؛ محمودی، سینا؛ طاهر پور، محدثه (۱۳۹۴). داده کاوی کاربردی. تهران، انتشارات نیاز دانش، چاپ سوم - ۱۳۹۶.
- فرانک کی - رایلی کیت سی - براون (۱۳۹۳). تجزیه و تحلیل سرمایه گذاری و مدیریت سبد اوراق بهادار و مهندسی مالی ترجمه: فریدون رهنمای رودپشتی؛ فرشاد هییتی؛ غلامرضا اسلامی بیدگلی، تهران، انتشارات ترمه، چاپ چهارم (چاپ دوم ترمه ۱۳۹۶).
- مؤمنی، منصور؛ فعال قیومی، علی (۱۳۹۴). تحلیل های آماری با استفاده از SPSS. تهران، انتشارات گنج شایگان، چاپ هشتم - ۱۳۹۴.
- نیکومرام، هاشم؛ شاهوردیانی، شادی (۱۳۹۰). مدیریت مالی راهبردی (ارزش آفرینی) مبتنی بر مدیریت ریسک، تهران، انتشارات حکیم باشی، چاپ چهارم - ۱۳۹۵.
- آذر، عادل؛ کریمی، سیروس (۱۳۸۸). «پیش بینی بازده سهام با استفاده از نسبت های حسابداری با رویکرد شبکه های عصبی»، تحقیقات مالی، دوره ۱۱، شماره ۲۸، صفحات مقاله (از ص ۳ تا ص ۲۰).
- انواری رستمی، علی اصغر؛ آذر، عادل؛ نوروزی، محمد (۱۳۹۲). «الگوسازی و پیش بینی EPS شرکت های پذیرفته شده در بورس اوراق بهادار تهران با رویکرد شبکه عصبی GMDH». بررسی های حسابداری و حسابرسی، دوره ۲۰، شماره ۱، صفحات مقاله (از ص ۱ تا ص ۱۸).
- برزگری خانقاه، جمال؛ جمالی، زهرا (۱۳۹۵). «پیش بینی بازده سهام با استفاده از نسبت های مالی، کنکاشی در پژوهش های اخیر». پژوهش حسابداری، دوره ۶، شماره ۲، صفحات مقاله (از ص ۷۱ تا ص ۹۲).
- زنده، مصطفی؛ مردانلو، سیما (۱۳۹۶). «کاوش قوانین همبستگی کمی در بازار سهام با استفاده از الگوریتم های فراابتکاری چند هدفه مبتنی بر الگوریتم ژنتیک». مجله مهندسی مالی و مدیریت اوراق بهادار، دوره ۸، شماره ۳۱، صفحات مقاله (از ص ۹۵ تا ص ۱۱۲).
- ستایش، محمدحسین؛ کاظم نژاد، مصطفی (۱۳۹۵). «سودمندی رگرسیون های تجمیعی و روش های انتخاب متغیرهای پیش بینی بهینه در پیش بینی بازده سهام». پژوهش های حسابداری مالی و حسابرسی، دوره ۸، شماره ۳۲، صفحات مقاله (از ص ۱ تا ص ۲۸).
- سروش یار، افسانه؛ اخلاقی، محمد (۱۳۹۶). «ارزیابی مقایسه ای اثربخشی تکنیک های داده کاوی در پیش بینی ریسک و بازده سهام شرکت های پذیرفته شده در بورس اوراق بهادار تهران». پژوهش های حسابداری مالی، دوره ۹، شماره ۱، صفحات مقاله (از ص ۵۷ تا ص ۷۵).

- سعیدی، علی؛ اکبری، نوردین (۱۳۸۹). «رابطه بین ارزش افزوده اقتصادی و ارزش افزوده تعدیل شده بر اساس تورم با بازده و قیمت سهام». پژوهشنامه اقتصادی، شماره ۷، صفحات مقاله (از ص ۱۴۵ تا ص ۱۷۲).
- علیمحمدی، علی محمد؛ عباسی مهر، محمدحسین؛ جواهری، احمد (۱۳۹۴). «پیش بینی بازده سهام شرکت ها با استفاده از نسبت های مالی تحت رویکرد درخت تصمیم». دانشگاه الزهرا (س)، دانشکده علوم اجتماعی و اقتصادی، راهبرد مدیریت مالی، دوره ۳، شماره ۴، صفحات مقاله (از ص ۱۲۵ تا ص ۱۴۶).
- صیادی، امید (۱۳۸۷). «آشنایی مقدماتی با شبکه های عصبی مصنوعی». دانشکده مهندسی برق، دانشگاه صنعتی شریف، صفحات (از ص ۳ تا ۱۶).
- مروتی شریف آبادی، علی؛ گلشن، مریم (۱۳۹۲). «پیش بینی بازده سهام با استفاده از شبکه های عصبی مصنوعی و رگرسیون چندمتغیره (مطالعه موردی: شرکت های پذیرفته شده در بورس اوراق بهادار تهران)». دومین کنفرانس ملی حسابداری، مدیریت مالی و سرمایه گذاری، استان گلستان، گرگان، کد COI مقاله: CAFM02-044.
- معین الدین، محمود؛ نایب زاده، شهناز و منصوری، ابراهیم (۱۳۸۹). «مقایسه توان تبیین معیارهای سنتی و نوین در ارزیابی عملکرد». فصلنامه حسابداری مالی، شماره ۶، صفحات مقاله (از ص ۱۲۱ تا ص ۱۴۰).
- مهرآرا، اسداله؛ عطف، زهرا؛ عسکری، زهرا (۱۳۹۱). «تکنیک های داده کاوی در پیش بینی قیمت سهام بورس اوراق بهادار». کنفرانس ملی حسابداری، مدیریت مالی و سرمایه گذاری، دانشگاه جامع علمی — کاربردی استان گلستان، کد COI مقاله: CAFM01-179.
- محمدیان مصمم، علی؛ عباسی، ملیحه (۱۳۹۳). «تحلیل قیمت مسکن شهر تهران با استفاده از مدل های جمعی تعمیم یافته». مجله بررسی های آمار رسمی ایران، دوره ۲۵، شماره ۲، صفحات مقاله (از ص ۱۶۱ تا ص ۱۷۴).

— Anass Nahil, & Abdelouahid Lyhyaoui (2018). « Short-term stock price forecasting using kernel principal component analysis and support vector machines: the case of Casablanca stock exchange», Procedia Computer Science, Volume 127/ 2018/, Pages from 161 to 169.

— Bruno Miranda Henrique, Vinicius Amorim Sobreiro, & Herbert Kimura (2018). «Stock Price Prediction Using Support Vector Regression on Daily and Up to the Minute Prices». The Journal of Finance and Data Science, 4(3), Pages from 183 to 201.

— Dimitris K. Chronopoulos, Fotios I. Papadimitriou, & Nikolaos Vlastakis (2018). «Information demand and stock return predictability», Journal of International Money and Finance, Volume 80, February 2018, Pages from 59 to 74.

- ___ Elena Baralis, Luca Cagliero, & Paolo Garza (2017). « Planning stock portfolios by means of weighted frequent itemsets», Expert Systems With Applications, Volume 86, 15 November 2017, Pages from 1 to17.
- ___ Xiao Zhong, & David Enke (2017). «A Comprehensive Cluster and Classification Mining Procedure for Daily Stock Market Return Forecasting», Neurocomputing, Volume 267, 6 December 2017, Pages from 152 to 168.
- ___ Xiao Zhong, & David Enke (2017). «Forecasting daily stock market return using dimensionality reduction», Expert Systems With Applications, Volume 67, January 2017, Pages from 126 to 139.