



## سودمندی رگرسیون‌های تجمیعی و روش‌های انتخاب متغیرهای پیش‌بین بهینه در پیش‌بینی بازده سهام

محمدحسین ستایش<sup>۱</sup>  
مصطفی کاظم‌نژاد<sup>۲</sup>

تاریخ دریافت: ۱۳۹۵/۰۷/۰۳ تاریخ پذیرش: ۱۳۹۵/۰۹/۰۸

### چکیده

مقاله حاضر به بررسی سودمندی رگرسیون‌های تجمیعی و روش‌های انتخاب متغیرهای پیش‌بین بهینه (شامل روش مبتنی بر همبستگی و ریلیف) برای پیش‌بینی بازده سهام شرکت‌های پذیرفته شده در بورس اوراق بهادار تهران می‌پردازد. به منظور ارزیابی عملکرد رگرسیون تجمیعی، معیارهای ارزیابی (شامل میانگین قدرمطلق درصد خطا، مجذور مربع میانگین خطا و ضریب تعیین) مربوط به پیش‌بینی این روش، با رگرسیون خطی و شبکه‌های عصبی مصنوعی مقایسه شده است. همچنین به منظور ارزیابی عملکرد روش‌های انتخاب متغیرهای بهینه پیش‌بین، معیارهای ارزیابی حاصل از پیش‌بینی با استفاده از متغیرهای انتخاب شده توسط این روش‌ها با معیارهای حاصل از پیش‌بینی با استفاده از کلیه متغیرها مقایسه شده است. یافته‌های تجربی مربوط به بررسی ۱۰۱ شرکت پذیرفته شده در بورس اوراق بهادار تهران در سال‌های ۱۳۸۳ الی ۱۳۹۲ حاکی از عملکرد بهتر روش تجمیعی نسبت به رگرسیون خطی و شبکه‌های عصبی مصنوعی است. افزون بر این، یافته‌ها حاکی از آن بود که پیش‌بینی با استفاده از متغیرهای انتخاب شده در روش‌های مبتنی بر همبستگی و ریلیف، به طور معناداری عملکرد پیش‌بینی را نسبت به استفاده از کلیه متغیرها افزایش می‌دهد.

**واژه‌های کلیدی:** پیش‌بینی بازده سهام، رگرسیون تجمیعی، روش‌های انتخاب متغیرهای پیش‌بین بهینه.

۱- استاد حسابداری، دانشگاه شیراز، شیراز، ایران، نویسنده اصلی.

۲- دانشجوی دکتری حسابداری، دانشگاه شیراز، شیراز، ایران، مسئول مکاتبات. Mkazemi5166@gmail.com

## ۱- مقدمه

در مفاهیم نظری گزارشگری مالی، به نقش اطلاعات مالی و سودمندی آن در تصمیم‌گیری‌های استفاده‌کنندگان از صورت‌های مالی اشاره شده است. بر اساس مفاهیم نظری، اطلاعات حسابداری برای مفید بودن در تصمیم‌گیری باید حائز ویژگی‌های مربوط بودن و قابلیت اتکا باشد (کمیت‌تدوین استانداردهای حسابداری ایران، ۱۳۸۹، ۷۶۲). از طرفی، سودمندی اطلاعات صورت‌های مالی، مشروط به توان تبیین و پیش‌بینی ارزش شرکت بوده و ارزش شرکت نیز به صورت مستقیم تحت تأثیر بازده فعلی و آتی آن است. در صورتی که اطلاعات حسابداری توان پیش‌بینی نداشته و در ارزیابی وقایع گذشته، حال و نسبت به تصمیم‌گیری آینده مؤثر نباشد، سودمندی آن نیز مورد سؤال قرار می‌گیرد (ثقفی و شعری، ۱۳۸۳، ۹۶). با توجه به اهمیت قیمت و بازده سهام در سرمایه‌گذاری، تاکنون مدل‌ها و نظریه‌های مختلفی شامل مدل قیمت‌گذاری دارایی‌های سرمایه‌ای، مدل‌های عاملی، مدل آربیتراژ، تحلیل‌های فنی و بنیادی برای تبیین و پیش‌بینی آن ارائه شده است. یافته‌های پژوهش‌های تجربی نیز حاکی از آن است که پیش‌بینی قیمت و بازده سهام، بخش عمده‌ای از پژوهش‌های انجام شده را به خود اختصاص داده است (کوئاری<sup>۱</sup>، ۲۰۰۱، ۲۰۸ و نمازی و ناظمی، ۱۳۸۴، ۱۴۴).

در اغلب پژوهش‌های انجام شده در بورس اوراق بهادار تهران، به‌رغم معایب و مشکلات روش‌های خطی، از این روش‌ها برای پیش‌بینی بازده سهام استفاده شده است و فقط در چند پژوهش (به‌عنوان نمونه، آذر و کریمی، ۱۳۸۸ و همت‌فر و همکاران، ۱۳۹۰) روش غیرخطی شبکه‌های عصبی مصنوعی<sup>۲</sup> مورد استفاده قرار گرفته است. یافته‌های این پژوهش‌ها حاکی از عملکرد بهتر این روش نسبت به روش خطی است. با وجود عملکرد مناسب شبکه‌های عصبی مصنوعی، یافته‌های برخی مطالعات نظری و تجربی انجام شده در حسابداری، حاکی از عملکرد بهتر سایر روش‌های غیرخطی از قبیل رگرسیون بردارهای پشتیبان<sup>۳</sup> و رگرسیون‌های تجمیعی<sup>۴</sup> نسبت به شبکه‌های عصبی مصنوعی است (به‌عنوان نمونه، مین<sup>۵</sup> و لی<sup>۶</sup>، ۲۰۰۵؛ موکامالا و همکاران<sup>۷</sup>، ۲۰۰۶؛ آلفارو و همکاران<sup>۸</sup>، ۲۰۰۸؛ لی و تاو<sup>۹</sup>، ۲۰۱۰؛ سان و همکاران<sup>۱۰</sup>، ۲۰۱۱). با این وجود، تاکنون پژوهشی که با استفاده از رگرسیون تجمیعی به پیش‌بینی بازده سهام در بورس اوراق بهادار تهران پرداخته باشد، مشاهده نشد. افزون بر این، یافته‌های پژوهش لو (۲۰۱۰) حاکی از آن است که انتخاب متغیرهای پیش‌بین و روش‌های آن، نسبت به انتخاب مدل پیش‌بینی، تأثیر بیشتری بر میانگین دقت پیش‌بینی دارد. علی‌رغم اهمیت متغیرهای پیش‌بین در پیش‌بینی، در پژوهش‌های انجام شده در زمینه پیش‌بینی بازده، هدف و تأکید اصلی، ارائه مدل‌های دقیق برای پیش‌بینی بوده و پژوهشی که به انتخاب متغیرهای پیش‌بین<sup>۱۱</sup> و روش‌های مناسب آن در داخل و خارج از کشور بپردازد، مشاهده نگردید.

پیش‌بینی بازده سهام در تصمیم‌گیری‌های مالی از اهمیت بسزایی برخوردار است و به‌طور وسیعی در ادبیات حسابداری و مالی مورد مطالعه قرار گرفته است. در اغلب پژوهش‌های انجام شده در بورس اوراق بهادار تهران، از روش‌های خطی و در پژوهش‌های اندکی نیز از شبکه‌های عصبی مصنوعی برای پیش‌بینی بازده سهام استفاده شده است. افزون بر این، در پژوهش‌های انجام شده در این زمینه، هدف و تأکید اصلی،

ارائه مدل‌های دقیق برای پیش‌بینی بازده بوده و کمتر به انتخاب متغیرهای پیش‌بین و روش‌های مناسب آن پرداخته شده است. با توجه به اهمیت پیش‌بینی بازده سهام، این مقاله، به بررسی سودمندی روش‌های انتخاب متغیرهای پیش‌بین بهینه و روش غیرخطی رگرسیون تجمیعی در پیش‌بینی بازده سهام شرکت‌های پذیرفته شده در بورس اوراق بهادار تهران می‌پردازد.

## ۲- مبانی نظری و پیشینه پژوهش

شارپ، لینتر و ... در دهه ۱۹۶۰ مدل قیمت‌گذاری دارایی‌های سرمایه‌ای را ارائه کردند. علی‌رغم تأیید نسبی مدل قیمت‌گذاری دارایی‌های سرمایه‌ای در پژوهش‌های اولیه، یافته‌های برخی از پژوهش‌های بعدی، با مدل مزبور متناقض بود. به طور کلی، شواهد تجربی حاکی از آن بود که بتای بازار به طور کامل بازده مورد انتظار سهام را تبیین نمی‌کند. در این راستا، تلاش‌های زیادی در زمینه توسعه مدل مزبور انجام شد. یکی از مهم‌ترین راهکارهای رفع مشکلات مدل مزبور مطرح کردن عوامل دیگر در فرآیند ایجاد بازده است. در این راستا، رأس در دهه ۱۹۷۰، نظریه قیمت‌گذاری آربیتراژ را پایه‌گذاری کرد. پژوهش برای انتخاب طرح‌های سرمایه‌گذاری و سبد اوراق بهادار در مسیر متفاوتی ادامه یافت و سرانجام نتایج فاما و فرنچ در دهه ۱۹۹۰ در قالب مدل سه‌عاملی (عامل صرف‌ریسک‌بازار، عامل‌اندازه‌ها عامل‌ارزش) برای پیش‌بینی بازده سهام ارائه شد. اگرچه مدل سه‌عاملی بهتر از مدل قیمت‌گذاری دارایی‌های سرمایه‌ای، بازده مورد انتظار را تبیین می‌کرد، ولیکن پژوهش‌های زیادی در زمینه یافتن عوامل مؤثر دیگری در تبیین و پیش‌بینی بازده سهام جهت دستیابی به مدل جامع‌تر انجام شده است (به‌عنوان نمونه، الهورانی و همکاران، ۲۰۰۳؛ براونوهمکاران، ۲۰۰۸؛ اسلامی بیدگلی و هنردوست، ۱۳۹۱؛ احمدپورو عظیمیان‌معز، ۱۳۹۱ و ایزدی‌نیا و همکاران، ۱۳۹۳).

گروهی از دانشمندان مالی که در جستجوی روش‌های علمی جدیدتری هستند، معتقدند نظریه بی‌نظمی و پویایی‌های غیرخطی می‌تواند مقدمه ارائه نظریه‌های علمی تازه‌ای در مدیریت مالی باشد. تناقض‌های آشکار میان نظریه‌های موجود و دنیای واقعی موجب شده تا دانش‌گامیان و صاحب‌نظران نسبت به اعتبار نظریه‌های مدرن مالی موجود، فرضیه بازار کارا و به‌ویژه مدل یک متغیره شارپ تردید کنند. (به نقل از عبده‌تبریزی و گنابادی، ۱۳۷۵، ۱۷). گروهی دیگر از صاحب‌نظران به پدیده آشفتگی توجه می‌کنند. آن‌ها با استفاده از فنون جدید ریاضی، بازار را سیستمی، پیچیده و در حال تحول می‌دانند. این گروه، مانند ریاضی‌دانان و فیزیک‌دانان اعتقاد دارند چنانچه هر پدیده‌ای به‌درستی مشاهده و بررسی شود، وقایع به ظاهر تصادفی آن، تبیین‌پذیر است و اگر در خور پیش‌بینی نباشد، حداقل قابل درک خواهد بود. طرفداران الگوی آشفتگی معتقدند که رفتار قیمت‌ها به صورت غیرخطی است. حرکات قیمت سهام ناشی از عوامل بسیار زیادی است که در هر لحظه از زمان بر آن‌ها تأثیر می‌گذارد. آن‌ها در واقع معتقدند که به تعداد سرمایه‌گذاران، عوامل مؤثر بر قیمت سهام وجود دارد (همان مأخذ).

برای اولین بار در ادبیات، پژوهش‌های مرامور<sup>۱۲</sup> و مرامور-کوستا<sup>۱۳</sup> (۱۹۹۷) و مرامور و پاهور<sup>۱۴</sup> (۱۹۹۸) نشان داد که رابطه نسبت‌های مالی و بازده سهام، غیرخطی است. مرامور و مرامور-کوستا (۱۹۹۷) به بررسی

رابطه بین بازده سهام و نسبت‌های مالی شرکت‌های اسلوونیایی پرداختند و برخلاف یافته‌های اغلب پژوهش‌های قبلی که به بررسی خطی رابطه بین نسبت‌های مالی و بازده سهام می‌پرداختند، وجود رابطه غیرخطی بین نسبت‌های مالی و بازده سهام را تأیید کردند. یافته‌های پژوهش مرامور و پاهور (۱۹۹۸) در آمریکا و ژاپن نیز وجود رابطه غیرخطی را تأیید کرد. پس از انجام این پژوهش‌ها، مطالعات زیادی به بررسی رابطه خطی و غیرخطی بین متغیرهای حسابداری، بازار، اقتصادی و سایر متغیرها با بازده سهام پرداختند. یافته‌های پژوهش کاناس<sup>۱۵</sup> و یانوپولس<sup>۱۶</sup> (۲۰۰۱)، اولسون و موسمن (۲۰۰۳) در شرکت‌های کانادایی، عمران<sup>۱۷</sup> و رجب<sup>۱۸</sup> (۲۰۰۴) در شرکت‌های مصری، مک‌میلان<sup>۱۹</sup> (۲۰۰۷) در کشورهای انگلیس، آمریکا، فرانسه و ژاپن، بویاسیگلو<sup>۲۰</sup> و آوسی<sup>۲۱</sup> (۲۰۱۰) در ترکیه، جیسن و همکاران<sup>۲۲</sup> (۲۰۱۳) در بورس ژوهانسبورگ نیز حاکی از عملکرد بهتر روش‌های غیرخطی نسبت به روش‌های خطی بود. در بورس اوراق بهادار تهران نیز یافته‌های پژوهش‌هایخالوزاده و خاکی‌صدیق (۱۳۷۷)، راعی و چاوشی (۱۳۸۲)، آذر و کریمی (۱۳۸۸)، همت‌فر و همکاران (۱۳۹۰) و عباسی و باقری (۱۳۹۰) این نتایج را تأیید می‌کند.

تاکنون پژوهشی که به بررسی انتخاب متغیرهای پیش‌بین و روش‌های آن و هم‌چنین پیش‌بینی بازده سهام شرکت‌های پذیرفته شده در بورس اوراق بهادار تهران با استفاده از رگرسیون غیرخطی تجمیعی پرداخته باشد، انجام نشده است. در اغلب پژوهش‌های انجام شده در بورس اوراق بهادار تهران، به پیش‌بینی خطی بازده سهام بر اساس متغیرهایی (عمدتاً حسابداری و در برخی موارد متغیرهای بازار و اقتصادی) پرداخته شده است (به‌عنوان نمونه، دستگیر و خدابنده، ۱۳۸۲؛ ثقفی و شعری، ۱۳۸۳؛ ثقفی و سلیمی، ۱۳۸۴؛ نمازی و محمدتبار کاسگری، ۱۳۸۶؛ و دستگیر و همکاران، ۱۳۹۱). پژوهش‌های اندکی (به‌عنوان نمونه، آذر و کریمی، ۱۳۸۸ و همت‌فر و همکاران، ۱۳۹۰) نیز با استفاده از روش غیرخطی شبکه‌های عصبی به پیش‌بینی بازده سهام می‌پردازد. یافته‌های این پژوهش‌ها حاکی از عملکرد بهتر روش غیرخطی نسبت به روش خطی است. با این وجود پژوهشی که در آن از سایر روش‌های غیرخطی استفاده شده باشد یا به مقایسه روش‌های غیرخطی پرداخته شده باشد، مشاهده نگردید. افزون بر این، پژوهشی که در بورس اوراق بهادار تهران به بررسی روش‌های انتخاب متغیرهای پیش‌بین و مقایسه آن پرداخته باشد انجام نشده است. در این راستا، پژوهش حاضر به پیش‌بینی بازده سهام با استفاده از رگرسیون غیرخطی تجمیعی و مقایسه روش‌های انتخاب متغیر مبتنی بر همبستگی و ریلیف و تأثیر آن بر عملکرد پیش‌بینی می‌پردازد.

مرحله انتخاب متغیرهای پیش‌بین، عموماً قبل از آموزش مدل‌های پیش‌بینی انجام می‌شود. با این وجود، در اغلب پژوهش‌های داخلی و خارجی انجام شده در حسابداری، این مرحله نادیده گرفته شده و متغیرهای پیش‌بین به‌صورت نظام‌مند انتخاب نشده است. این امر به انتخاب متغیرهای پیش‌بین غیربهینه و در برخی موارد، متغیرهای پیش‌بین نامناسب منجر می‌شود. در این پژوهش‌ها، متغیرهای پیش‌بین، بدون ضابطه و صرفاً با توجه به پژوهش‌های گذشته انتخاب شده است (تسای، ۲۰۰۹، ۱۲۱ و ستایش و همکاران، ۲۰۱۲، ۲۶۹).

انتخاب متغیرهای مناسب جهت رسیدن به بهترین نتیجه در پیش‌بینی، از مباحث چالش برانگیز در دو دهه اخیر بوده است. از دیدگاه تئوری، یادگیری براساس تعداد متغیرهای پیش‌بین بیشتر باعث می‌شود تا دقت پیش‌بینی بالا رود. با این وجود، شواهد تجربی بیانگر آن است که این امر همواره صادق نیست؛ زیرا تمام متغیرها، برای تشخیص و پیش‌بینی مهم نیستند و یا برخی از آن‌ها به‌طور کلی در پیش‌بینی نامربوط هستند (لیندن‌بام و همکاران ۲۳، ۲۰۰۴، ۱۳۰). با توجه به این‌که عامل‌های بسیاری از جمله کیفیت داده‌ها در موفقیت یک الگوریتم یادگیری مؤثر است، اگر داده‌ها حاوی متغیرها و یا اطلاعات تکراری و نامربوط<sup>۲۴</sup> باشند و یا حاوی اطلاعات نویزی<sup>۲۵</sup> و نامطمئن باشند، اخذ دانش از آن داده‌ها مشکل می‌شود (هل<sup>۲۶</sup>، ۲۰۰۰، ۳۶۰). افزون بر این، کاهش تعداد متغیرهای پیش‌بین نامربوط یا اضافی، علاوه بر کاهش زمان اجرای الگوریتم‌های آموزشی، به مفهومی عمومی‌تر منجر می‌شود. سایر مزایای بالقوه انتخاب متغیرهای پیش‌بین شامل تسهیل درک و تجسم داده‌ها، کاهش الزامات اندازه‌گیری و ذخیره اطلاعات، کاهش اضافه‌بار ابعاد<sup>۲۷</sup> و بهبود عملکرد پیش‌بینی و فراهم کردن بینش بهتر در مورد مفهوم زیربنایی از پیش‌بینی دنیای واقعی است (تسای، ۲۰۰۹، ۱۲۲).

دو جنبه مهم در روش‌های مختلف کاهش ابعاد (متغیرها) وجود دارد:

- استخراج متغیرهای پیش‌بین<sup>۲۸</sup>:

استخراج متغیرهای پیش‌بین یا به عبارت دیگر تبدیل متغیرهای پیش‌بین<sup>۲۹</sup> فرآیندی است که  $K$  متغیر جدید حاصل می‌شود که حاصل ترکیب  $N$  متغیر پیش‌بین اولیه هستند. شناخته شده ترین الگوریتم‌های استخراج متغیرهای پیش‌بین شامل تحلیل مولفه‌های اصلی، تحلیل عاملی و تحلیل ممیزی<sup>۳۰</sup> است.

- انتخاب متغیرهای پیش‌بین:

در مقابل الگوریتم‌های استخراج متغیرهای پیش‌بین، الگوریتم‌های انتخاب متغیر پیش‌بین، به انتخاب بهترین  $K$  متغیر از بین  $N$  متغیر اصلی می‌پردازد و بقیه متغیرهای کم اهمیت‌تر حذف می‌شود. شایان ذکر است که در انتخاب متغیرها، متغیرهای اصلی بدون تغییر، انتخاب می‌شود اما در استخراج متغیرها، متغیرها به صورت تغییر یافته استفاده می‌شوند. روش‌های مبتنی بر همبستگی و ریلیف از مهم‌ترین روش‌های انتخاب متغیر در پیش‌بینی متغیرهای پیوسته محسوب می‌شود که در این پژوهش نیز استفاده شده است. دلایل انتخاب این روش‌ها به شرح زیر است:

۱- یافته‌های پژوهش‌های انجام شده (از قبیل روبنیک-سیکونجا<sup>۳۱</sup> و کوننکو<sup>۳۲</sup> (۱۹۹۷)، هال، ۱۹۹۹؛ کارگودا و همکاران، ۲۰۱۰؛ ستایش و همکاران، ۲۰۱۲؛ کوپرینسکا و همکاران، ۲۰۱۵) حاکی از عملکرد بهتر روش‌های مزبور نسبت به سایر روش‌های انتخاب متغیر است.

۲- روش‌های مزبور از روش‌های انتخاب متغیر در مسائل پیش‌بینی است که متغیرهای اولیه بدون تغییر، انتخاب می‌شود و لیکن در روش‌های استخراج متغیرها، متغیرها به صورت تغییر یافته استفاده خواهد شد. همچنین، روش‌های مزبور در مسائل پیش‌بینی (با متغیر وابسته پیوسته) کاربرد دارد، در حالی که برخی از روش‌های انتخاب متغیر، صرفاً در مسائل طبقه‌بندی (با متغیر وابسته اسمی از قبیل ورشکستگی)

کاربرد دارد. به‌عنوان نمونه، علی‌رغم عملکرد بهتر روش rough-sets Fuzzy نسبت به روش ریلیف (ستایش و همکاران، ۲۰۱۲) روش مزبور صرفاً در مسائل طبقه‌بندی کاربرد دارد. در ادامه به تشریح روش‌های انتخاب متغیر مبتنی بر همبستگی و ریلیف پرداخته می‌شود.

ایده اصلی روش مبتنی بر همبستگی این است که متغیرهایی بهینه محسوب می‌شوند که با متغیر وابسته همبستگی بالا و با سایر متغیرهای پیش‌بین همبستگی نداشته یا همبستگی ناچیزی داشته باشد. این روش یک الگوریتم کاملاً خودکار است، یعنی محدودیتی برای شناسایی آستانه<sup>۳۳</sup> یا تعداد متغیرهایی که باید انتخاب شود، اعمال نمی‌کند. با این وجود، در صورت تمایل، اعمال آستانه یا تعداد متغیرها به راحتی امکان پذیر است (هال، ۱۹۹۹، ۶۹).

روش انتخاب متغیر ریلیف از جمله روش‌های انتخاب متغیرهای پیش‌بین مبتنی بر معیار فاصله است. در ریلیف، وزنی که میزان ارتباط هر ویژگی را با طبقه نشان می‌دهد، به وسیله فاصله اقلیدسی بین نمونه‌ها تعیین می‌شود و وزن هر ویژگی نشان‌دهنده قابلیت جداسازی طبقه‌ها به وسیله آن متغیر پیش‌بین می‌باشد. در این روش، اگر یک ویژگی به ازای نمونه‌های درون یک طبقه، مقدار یکسان و به ازای نمونه‌های دیگر طبقه‌ها مقادیر مختلفی داشته باشد، وزن بالاتری می‌گیرد. ریلیف از بین داده‌های آموزشی یک نمونه را به صورت تصادفی انتخاب می‌کند و سپس فاصله اقلیدسی آن نمونه تا نزدیک‌ترین نمونه از طبقه مشابه و نزدیک‌ترین نمونه از طبقه متفاوت را به دست می‌آورد و سپس این فاصله‌ها را برای به‌روز کردن وزن هر متغیر به کار می‌برد. در نهایت، الگوریتم آن دسته از متغیرهایی را انتخاب می‌کند که وزن آن‌ها از یک حد آستانه از پیش تعریف شده به وسیله کاربر، بیشتر است (آتیا<sup>۳۴</sup>، ۲۰۰۱، ۹۳۰).

ریلیف از جمله روش‌های انتخاب ویژگی فیلتر<sup>۳۵</sup> است که اساساً با رتبه‌بندی نزولی متغیرها عمل می‌کند و به علت سادگی و مؤثر بودن در افزایش دقت طبقه‌بندی، در بسیاری از موارد استفاده می‌شود. رتبه‌ای که ریلیف به هر متغیر می‌دهد بر اساس میزان نقش آن متغیر در جداسازی نمونه‌های متفاوت همسایه است. این الگوریتم برای هر نمونه آموزشی به دنبال نزدیک‌ترین همسایه که با آن، هم طبقه (کلاس) است می‌شود که به این نزدیک‌ترین همسایه، نزدیک‌ترین برخورد<sup>۳۶</sup> گفته می‌شود. سپس نزدیک‌ترین همسایه که کلاس آن با کلاس نمونه آموزشی، مخالف است را پیدا می‌کند که به این همسایه، نزدیک‌ترین خطا<sup>۳۷</sup> گفته می‌شود. رتبه‌ای که به هر متغیر داده می‌شود بر اساس نسبت مجموع فاصله نمونه‌های آموزشی تصویر شده روی هر متغیر از نزدیک‌ترین برخورد همسایه برای هر نمونه آموزشی به مجموع فاصله نزدیک‌ترین همسایه خطا برای هر نمونه آموزشی است (آتیا، ۲۰۰۱، ۹۳۰).

روش ریلیف که در ابتدا به وسیله کرا<sup>۳۸</sup> و رندل<sup>۳۹</sup> (۱۹۹۲) ارائه شد، را می‌توان فقط برای مسائل طبقه‌بندی با دو گروه (مثلاً ورشکسته در مقابل غیرورشکسته) استفاده کرد. روبنیک - سیکونجاو کونکو (۱۹۹۷) روش ریلیف را برای استفاده در داده‌های خروجی پیوسته بسط دادند. در پژوهش حاضر نیز با توجه به پیوسته بودن متغیر وابسته (بازده سهام) از مدل بسط یافته ریلیف برای انتخاب متغیرهای پیش‌بین (مستقل) استفاده خواهد شد.

روش‌های غیرخطی از قبیل شبکه‌های عصبی، چندین مزیت مهم در مقایسه با مدل‌های آماری از قبیل رگرسیون خطی دارند. معایب رگرسیون خطی نسبت به روش‌های غیرخطی (از قبیل شبکه‌های عصبی) به شرح زیر است (دی‌تاین و همکاران<sup>۴۰</sup>، ۲۰۰۳، ۲۳۷-۲۳۸):

- ماهیت خطی رگرسیون: یک عیب مهم رگرسیون‌های خطی این است که رگرسیون هیچ شاخص مستقیمی مبنی بر این که آیا داده‌ها در حالت خطی به بهترین صورت نشان داده می‌شود را ارائه نمی‌کند. با توجه به ماهیت علوم اجتماعی، در بسیاری از حالت‌ها تحلیل آماری خطی نامناسب است.
- از پیش مشخص کردن مدل: استفاده از مدل‌های رگرسیون، مستلزم از پیش مشخص کردن مدل پایه است. این کار باعث حل آسان‌تر مسأله می‌شود ولیکن، نیازمند حدس‌های زیاد است.
- مفروضات رگرسیون: عملکرد مدل‌های رگرسیون خطی وابسته به مفروضات گوناگونی از قبیل عدم وجود روابط خطی چندگانه و توزیع نرمال باقیمانده‌ها است.
- عدم انطباق پذیری: رگرسیون چند متغیره در حالتی که اجزای مدل را نتوان به‌وسیله حدس مشخص کرد، دارای خاصیت انطباق‌پذیری با داده‌ها نیست.

شبکه‌های عصبی مصنوعی رگرسیون تجمیعی از مهم‌ترین روش‌های غیرخطی محسوب می‌شوند که در ادامه به تشریح آن‌ها پرداخته می‌شود.

شبکه‌های عصبی یک تکنیک پردازش اطلاعات مبتنی بر روش سیستم‌های عصبی بیولوژیکی مانند مغز است. طی سال‌های گذشته، شبکه‌های عصبی به‌عنوان یک فناوری ظاهر شده است که می‌تواند الگوی داده‌ها را شناسایی و مدل‌سازی کند، کاری که با روش‌های آماری به سهولت امکان‌پذیر نیست. مفهوم بنیادی شبکه‌های عصبی، ساختار سیستم پردازش اطلاعات است که از تعداد زیادی واحدهای پردازشی (نورون<sup>۴۱</sup>) مرتبط با شبکه‌ها تشکیل شده‌اند. سلول عصبی بیولوژیکی یا نورون، واحد سازنده سیستم عصبی در انسان است. یک نورون از بخش‌های اصلی زیر تشکیل شده است (جعفریه و همکاران، ۱۳۸۵، ۶۳):

- (۱) بدنه سلولی<sup>۴۲</sup> که هسته در آن است و سایر قسمت‌های سلولی از آن منشأ گرفته است.
- (۲) هسته.

(۳) آکسون<sup>۴۳</sup> که وظیفه آن انتقال اطلاعات از سلول عصبی است.

(۴) دندریت<sup>۴۴</sup> که وظیفه آن انتقال اطلاعات از سلول‌های دیگر به سلول عصبی است.

نگاه مدرن به شبکه‌های عصبی، در دهه ۱۹۴۰ و با شروع به کار وارن مک کیولاج<sup>۴۵</sup> و والتر پیترز<sup>۴۶</sup> آغاز شد. آن‌ها نشان دادند که شبکه‌ای از نرون‌های عصبی دارای قابلیت محاسبه هر تابع ریاضی و یا منطقی می‌باشند. فعالیت آن‌ها در این زمینه را می‌توان به‌عنوان مبدأ علم شبکه‌های عصبی مصنوعی در نظر گرفت. اولین کاربرد عملی شبکه‌های عصبی در اواخر دهه ۱۹۵۰ شکل گرفت. در این سال‌ها شبکه‌های پرسپترون و قواعد یادگیری آن به‌وسیله فرانک روزنبلات<sup>۴۷</sup> (۱۹۵۸) ابداع شد. روزنبلات و همکاران (۱۹۵۸) ضمن ساختن یک شبکه پرسپترون نشان دادند که این شبکه‌ها دارای توانایی تشخیص الگو می‌باشند. این

موفقیت‌ها باعث به وجود آمدن علاقه به پژوهش در این زمینه در بین پژوهش‌گران شد. با این وجود، شبکه‌های عصبی پرسپترون، تنها قابلیت حل مجموعه محدودی از مسائل را دارا بودند (کیا، ۱۳۸۸، ۷۴). یکی از مهم‌ترین شبکه‌های عصبی، شبکه‌های عصبی پرسپترون چند لایه است که بر مبنای یادگیری با سرپرست است. ساختار پایه شبکه عصبی بر مبنای پرسپترون چند لایه، یک لایه ورودی، یک یا چند لایه مخفی و یک لایه خروجی است. هرکدام از این لایه‌ها از یک یا چندین گره تشکیل شده است. لایه ورودی به تعداد متغیرهای مستقل دارای گره است و به همین ترتیب نیز لایه خروجی به تعداد متغیرهای وابسته دارای گره می‌باشد. اما مشخص کردن ساختار لایه پنهان، دشوار است (هوگلاند<sup>۴۸</sup>، ۲۰۱۲، ۹۶۵۷). برای آموزش این شبکه عصبی از قانون یادگیری پس‌انتشار خطا استفاده می‌شود. الگوریتم پس‌انتشار خطا، نوع خاصی از الگوریتم حداقل میانگین مربعات<sup>۴۹</sup> به شمار می‌رود. مانند قاعده یادگیری حداقل میانگین مربعات، الگوریتم پس‌انتشار نیز بر مبنای تخمین بیشترین کاهش گرادیان عمل می‌کند. شاخص کارایی در این الگوریتم، میانگین مربعات خطاهاست. تنها تفاوت حداقل میانگین مربعات و پس‌انتشار در نحوه محاسبه مشتقات است. در مورد یک شبکه تک لایه خطی، خطا یک تابع خطی صریح از وزن‌های شبکه است و مشتقات بر حسب وزن‌های آن به راحتی قابل محاسبه است. در شبکه‌های چند لایه با تابع انتقال غیرخطی، رابطه بین وزن‌های شبکه و خطا بسیار پیچیده‌تر است و در راستای محاسبه مشتقات نیاز به استفاده از قاعده زنجیری<sup>۵۰</sup> وجود دارد (کیا، ۱۳۸۸، ۲۷۱). با توجه به این که روش پرسپترون چند لایه پس‌انتشار خطا بیشترین کاربرد را در حوزه مالی دارد، در پژوهش حاضر نیز از این روش استفاده شده است.

هدف رگرسیون‌های تجمیعی، ترکیب چندین مدل برای افزایش دقت پیش‌بینی در مسائل یادگیری با وجود یک متغیر هدف عددی است. این رویکرد، بخش قابل توجهی از پژوهش‌های اخیر را به خود اختصاص داده است. فرآیند یادگیری تجمیعی دارای سه مرحله ایجاد<sup>۵۱</sup>، هرس کردن<sup>۵۲</sup> و تجمیع است. در مرحله ایجاد، مجموعه‌ای از مدل‌های پیش‌بینی ایجاد می‌شود. تعدادی مدل اضافی نیز اغلب در این مرحله به وجود می‌آید. در مرحله دوم، از طریق حذف برخی از مدل‌های ایجاد شده قبلی، تجمیع هرس می‌شود و در نهایت، یک راهبرد برای ترکیب مدل‌های پایه، تعریف می‌شود. سپس از این راهبرد برای به دست آوردن پیش‌بینی از تجمیع برای موارد جدید براساس پیش‌بینی‌های مدل‌های پایه استفاده می‌شود. مزیت رویکرد ترکیبی نسبت به مدل‌های انفرادی، افزایش استحکام (ثبات<sup>۵۳</sup>) نتایج و دقت است (مندس-موریا و همکاران<sup>۵۴</sup>، ۲۰۱۲، ۱۰).

یکی از انواع مشهور رگرسیون تجمیعی، رگرسیون بگینگ<sup>۵۵</sup> است که بر پایه تجمیع چند مرتبه (تکرار) یک روش پیش‌بینی و استفاده از این پیش‌بین‌ها به عنوان یک پیش‌بین تجمیعی عمل می‌کند. به هر یک از پیش‌بینی‌کننده‌ها داده‌های آموزشی متفاوتی ارائه می‌شود که این داده‌ها به صورت تصادفی از مجموعه داده‌های آموزشی انتخاب می‌شوند. بنابراین، امکان تکرار انتخاب یک داده در یکی از مجموعه‌های ایجاد شده وجود دارد. هر یک از پیش‌بینی‌کننده‌ها با یکی از مجموعه‌های ایجاد شده آموزش می‌بینند. نتیجه



نهایی در تصمیم‌گیری برای داده جدید بر مبنای اکثریت در رأی‌گیری مشخص می‌شود (بریمن<sup>۵۶</sup>، ۱۹۹۶، ۱۲۳).

از مزایای طبقه‌بندی‌کننده‌های تجمیعی می‌توان به سادگی در پیاده‌سازی و ترکیب چند روش پیش‌بینی که ترکیب آن‌ها باعث ایجاد یک روش غیرخطی می‌شود اشاره کرد. از دیگر مزایای طبقه‌بندی‌کننده‌های تجمیعی، استفاده از هر روش پیش‌بینی دلخواه است. افزون بر این، روش مزبور، کاهش میزان واریانس خطا را در پی خواهد داشت و بر اساس پژوهش‌های انجام شده، افزون بر واریانس خطا، میزان سوگیری (بایاس) را نیز کاهش می‌دهد (هانگ و همکاران، ۲۰۰۹، ۵۲۹۹).

در این پژوهش، طبق پیش‌فرض نرم‌افزار مورد استفاده (وکا) از ۱۰ بار تکرار درخت تصمیم برای تجمیع روش پیش‌بینی استفاده شده است. درخت تصمیم، در اغلب کاربردهای بگینگ استفاده و کاربرد آن به نتایج مطلوبی منجر شده است و به راحتی برای متغیرهای کمی استفاده می‌شود (آلغارو و همکاران، ۲۰۰۸، ۱۱۸ و لئو و هانگ، ۲۰۱۰، ۳۶). افزون بر این، درخت تصمیم، پیش‌فرض نرم‌افزار وکا می‌باشد.

### ۳- فرضیه‌های پژوهش

براساس مبانی نظری و پیشینه پژوهش، فرضیه‌هایی به شرح زیر تدوین شده است:

- ۱- تفاوت معناداری بین عملکرد روش‌های پیش‌بینی، در پیش‌بینی بازده سهام شرکت‌ها وجود دارد.
  - ۱-۱- تفاوت معناداری بین عملکرد روش‌های غیرخطی رگرسیون تجمیعی و شبکه‌های عصبی مصنوعی، در پیش‌بینی بازده سهام شرکت‌ها وجود دارد.
  - ۲-۱- تفاوت معناداری بین عملکرد روش غیرخطی رگرسیون تجمیعی و رگرسیون خطی، در پیش‌بینی بازده سهام شرکت‌ها وجود دارد.
  - ۳-۱- تفاوت معناداری بین عملکرد روش غیرخطی شبکه‌های عصبی مصنوعی و رگرسیون خطی، در پیش‌بینی بازده سهام شرکت‌ها وجود دارد.
- ۲- انتخاب متغیرهای پیش‌بین بهینه، تأثیر معناداری بر عملکرد روش‌های پیش‌بینی دارد.
  - ۱-۲- روش انتخاب متغیر مبتنی بر همبستگی، تأثیر معناداری بر عملکرد پیش‌بینی رگرسیون تجمیعی دارد.
  - ۲-۲- روش انتخاب متغیر ریلیف، تأثیر معناداری بر عملکرد پیش‌بینی رگرسیون تجمیعی دارد.
  - ۳-۲- روش انتخاب متغیر مبتنی بر همبستگی، تأثیر معناداری بر عملکرد پیش‌بینی رگرسیون خطی دارد.
  - ۴-۲- روش انتخاب متغیر ریلیف، تأثیر معناداری بر عملکرد پیش‌بینی رگرسیون خطی دارد.
  - ۵-۲- روش انتخاب متغیر مبتنی بر همبستگی، تأثیر معناداری بر عملکرد پیش‌بینی شبکه‌های عصبی مصنوعی دارد.
  - ۶-۲- روش انتخاب متغیر ریلیف، تأثیر معناداری بر عملکرد پیش‌بینی شبکه‌های عصبی مصنوعی دارد.

#### ۴- روش پژوهش

این پژوهش کاربردی است. طرح پژوهش آن از نوع شبه‌تجربی و با استفاده از رویکرد پس‌رویدادی (از طریق اطلاعات گذشته) است. از روش پس‌رویدادی زمانی استفاده می‌شود که محقق پس از وقوع رویدادها به بررسی موضوع می‌پردازد. افزون بر این، امکان دستکاری متغیرهای مستقل وجود ندارد (نمازی، ۱۳۸۹، ۳۵).

در این پژوهش برای جمع‌آوری داده‌ها و اطلاعات از روش‌های کتابخانه‌ای و میدانی استفاده شده است. مبانی نظری پژوهش از کتب، مجلات و سایت‌های تخصصی فارسی و لاتین گردآوری شده و داده‌های مالی مورد نیاز با مراجعه به سایت سازمان بورس اوراق بهادار تهران، صورت‌های مالی شرکت‌ها و همچنین با استفاده از نرم‌افزارهای تدبیرپرداز و ره‌آورد نوین گردآوری شده است. در مرحله اول با بررسی ادبیات و پیشینه پژوهش (شامل حدود ۲۵۰ مقاله فارسی و انگلیسی)، ۲۰۰ متغیر پیش‌بین اولیه شناسایی شد. از بین متغیرهای شناسایی شده، ۵۲ متغیری که بیشتر در ادبیات مربوط به تبیین یا پیش‌بینی بازده سهام استفاده شده و داده‌های مورد نیاز برای سنجش آن‌ها از طریق پایگاه‌های اطلاعاتی سازمان بورس و اوراق بهادار و همچنین نرم‌افزارهای تدبیرپرداز و ره‌آورد نوین در دسترس بود، انتخاب شد. در این راستا، سعی شد که هم متغیرهای حسابداری (اقلام مبتنی بر ترازنامه، صورت سود و زیان و صورت جریان وجوه نقد) و هم متغیرهای بازار استفاده و محتوای اطلاعاتی آن‌ها مدنظر قرار گیرد. بنابراین، ابعاد سودآوری، کارایی، اهرم مالی، نقدینگی، نسبت‌های مبتنی بر هر سهم، نسبت‌های مبتنی بر جریان وجوه نقد و نسبت‌های بازار در نظر گرفته شده است. جدول ۱، میانگین این متغیرها را در شرکت‌های مورد بررسی نشان می‌دهد. در این جدول، متغیرهایی که به صورت تغییرات ( $\Delta$ ) نشان داده شده است، بیانگر درصد رشد سالانه متغیر مزبور می‌باشد. بتای هر سهم نیز از تقسیم کوواریانس بازده سهام و بازده بازار به واریانس بازده بازار به‌دست آمده است. برای محاسبه بتا در سال  $t$  از اطلاعات ماهانه بازده سهام و بازده بازار برای پنج ساله منتهی به سال  $t$  استفاده شده است. ارقام تعهدی نیز تفاوت بین سود عملیاتی و جریان نقد حاصل از فعالیت‌های عملیاتی است. در ادامه، با استفاده از روش‌های انتخاب متغیرهای پیش‌بین در نرم‌افزار وکا، از بین ۵۲ متغیر ذکر شده، متغیرهای بهینه، انتخاب شده است. در جدول ۲، هفت متغیر بهینه انتخاب شده در روش مبتنی بر همبستگی ارائه شده است. روش انتخاب متغیر ریلیف نیز کلیه متغیرهای پیش‌بین را رتبه‌بندی می‌کند. در این پژوهش، به منظور قابلیت مقایسه بهتر با روش مبتنی بر همبستگی از هفت متغیر اول (بهتر) رتبه‌بندی شده در این روش استفاده می‌شود که این متغیرها به ترتیب رتبه در جدول ۲ ارائه شده است. کلیه متغیرهای مستقل این پژوهش که نحوه انتخاب آن در بخش قبل توضیح داده شد در جدول ۱ ارائه شده است. متغیرهای مستقل بهینه انتخاب شده توسط روش‌های مبتنی بر همبستگی و ریلیف نیز در جدول ۲ ارائه شده است.

متغیر وابسته در این پژوهش، بازده سهام است که مشابه پژوهش‌های پیشین، طبق رابطه ۱ (به‌عنوان نمونه، دستگیر و خدابنده، ۱۳۸۲، ۱۰۶ و ثقفی و شعری، ۱۳۸۳، ۹۸) از تقسیم مجموع عایدات یک سهم در طول دوره مالی بر قیمت سهم در ابتدای دوره، محاسبه می‌شود. عایدات شامل افزایش قیمت سهم و سود نقدی پرداخت شده است و با توجه به تأثیر افزایش سرمایه، بازده سهام تعدیل خواهد شد.

(۱) بازده سهام

$$\text{مزایای سود سهمی} + \text{مزایای حق تقدم} + \text{سود نقدی خالص} + \text{تفاوت قیمت سهام آخر و اول دوره} = \text{بازده سهام} \times \text{قیمت سهام اول دوره}$$

به منظور افزایش دقت محاسبه، بازده سهام به صورت ماهانه محاسبه و میانگین هندسی بازده ماهانه برای یک سال، به‌عنوان بازده سالانه در نظر گرفته شده است. بازده میانگین هندسی، بازده‌های تجمعی و ترکیبی مربوط به گذشته را اندازه‌گیری می‌کند. در طول چند دوره، میانگین هندسی، میانگین نرخ رشد صحیح‌تری نسبت به میانگین حسابی را نشان می‌دهد (فیوزی و مارکوویتز، ۲۰۰۲، ۱۱۴ و ۱۱۵؛ پینتو و همکاران، ۲۰۱۰، ۴۹ و تهرانی و نوربخش، ۱۳۸۶، ۱۲۶ و ۱۲۸).

جامعه آماری این پژوهش، کلیه شرکت‌های پذیرفته شده در بورس اوراق بهادار تهران طی دوره زمانی ۱۳۸۳ الی ۱۳۹۲ می‌باشد. از این جامعه، بدون انجام نمونه‌گیری، شرکت‌های حائز شرایط زیر به‌عنوان نمونه انتخاب شده است:

- ۱) تا پایان اسفندماه سال ۱۳۸۲ در بورس اوراق بهادار تهران پذیرفته شده باشند.
- ۲) سال مالی آن‌ها منتهی به پایان اسفندماه بوده و در دوره زمانی مورد بررسی، تغییری در آن ایجاد نشده باشد.
- ۳) ارزش دفتری حقوق صاحبان سهام آن‌ها در دوره مورد بررسی، مثبت باشد.
- ۴) شرکت‌ها طی دوره مورد بررسی، فعالیت مستمر داشته و سهام آن‌ها بدون وقفه با اهمیت مورد معامله قرار گرفته باشد (حداقل ۴۰ مبادله در سال).
- ۵) جزء شرکت‌های تولیدی باشد.
- ۶) اطلاعات مالی مورد نیاز برای انجام این پژوهش را در دوره زمانی ۱۳۸۳ الی ۱۳۹۲ به‌طور کامل ارائه کرده باشند.

تعداد کل شرکت‌های پذیرفته شده تا پایان سال ۱۳۸۲، بالغ بر ۳۷۱ شرکت بود. تعداد ۱۲۴ شرکت پذیرفته شده، به دلیل غیرتولیدی بودن یا پایان دوره مالی غیر از اسفندماه، از مجموع شرکت‌ها حذف شد. از ۲۴۷ شرکت باقی‌مانده، ۵۱ شرکت دارای حقوق صاحبان سهام منفی، مبادلات ۳۶ شرکت، کمتر از ۴۰ مرتبه در سال و اطلاعات ۵۹ شرکت در دسترس نبود. با توجه به بررسی‌های انجام شده، تعداد ۱۰۱ شرکت در دوره زمانی ۱۳۸۳ الی ۱۳۹۲ حائز شرایط فوق بوده و مورد بررسی قرار گرفته است.

### ۵- روش آزمون فرضیه‌ها

پس از مشخص شدن متغیرهای بهینه به‌وسیله هر یک از روش‌های انتخاب متغیر، از روش رگرسیون خطی و روش‌های غیرخطی رگرسیون تجمیعی و شبکه‌های عصبی مصنوعی برای پیش‌بینی بازده سهام استفاده شده است. در این راستا، از پیش‌فرض‌های نرم‌افزار وکا استفاده شده است. شایان ذکر است که در این پژوهش، از داده‌های یک سال قبل شرکت‌ها برای پیش‌بینی بازده سهام استفاده شده است. به‌منظور ارزیابی عملکرد رگرسیون تجمیعی در پیش‌بینی بازده سهام، معیارهای ارزیابی (شامل میانگین قدرمطلق درصد خطا<sup>۵۷</sup>، مجذور مربع میانگین خطا<sup>۵۸</sup> و ضریب تعیین) مربوط به پیش‌بینی این روش با روش‌های متداول شبکه‌های عصبی مصنوعی و رگرسیون خطی مقایسه می‌شود. معیارهای مزبور، متداول‌ترین معیارهای ارزیابی عملکرد در مسائل پیش‌بینی محسوب می‌شود که نحوه محاسبه آن در جدول ۳ ارائه شده است.

بالا‌تر بودن ضریب تعیین و پایین‌تر بودن دو معیار دیگر بیانگر عملکرد بهتر پیش‌بینی است. علی‌رغم وجود معیارهای ارزیابی عملکرد متداول دیگر در این زمینه، به‌دلیل این‌که به‌راحتی از سه معیار بالا قابل محاسبه است، از ارائه آن خودداری شده است. به‌عنوان نمونه، مربع‌مجذور میانگین‌خطا (RMSE)، ریشه دوم مربع میانگین خطا (MSE) است و مربع میانگین خطای استاندارد نرمال شده (NMSE) و ضریب تعیین مکمل یکدیگرند (اسمیت و گوپتا، ۲۰۰۲، ۹).

جدول ۱- متغیرهای مستقل (پیش‌بین) و میانگین آن

ردیف	نام متغیر	نحوه سنجش	میانگین	ردیف	نام متغیر	نحوه سنجش	میانگین
۱	Size	Log MV <sub>(t-1)</sub>	۱۲/۴۶	۲۷	Δ OPM	Δ (OI/S)	۰/۰۵۴
۲	E/P <sub>t</sub>	NI/P <sub>t</sub>	۰/۲۴۱	۲۸	NPM	NI/S	۰/۱۷
۳	BV/MV	BV/MV	۰/۱۸۴	۲۹	Δ NPM	Δ (NI/S)	۰/۰۰۵
۴	Lev	TL/TA	۰/۶۷۳	۳۰	R <sub>(t-1)</sub>	R <sub>(t-1)</sub>	۰/۲۶۱
۵	TL/E	TL/E	۲/۰۵۸	۳۱	Cov	OI/Int	۳۹۵/۹۵۹
۶	ROA	NI/TA	۰/۱۴۵	۳۲	OCF	OCF/MV <sub>(t-1)</sub>	۰/۱۱۲
۷	Δ ROA	Δ (NI/TA)	-۱۲/۸۰۱	۳۳	Δ OCF	Δ OCF/MV <sub>(t-1)</sub>	۰/۰۳۱
۸	ROE	NI/E	۰/۳۱۵	۳۴	IRFEF	IRFEF/MV <sub>(t-1)</sub>	۰/۰۰۲
۹	Δ ROE	Δ (NI/E)	-۷/۴۴۵	۳۵	ICF	ICF/MV <sub>(t-1)</sub>	-۰/۰۶۹
۱۰	Beta	Beta	۰/۳۴۵	۳۶	FCF	FCF/MV <sub>(t-1)</sub>	-۰/۳۵۴
۱۱	EPS	EPS	۷۹۹/۵۷۹	۳۷	Δ S	Δ S	۲۴/۲۹۴
۱۲	DPS	DPS	۴۷۹/۱۱۴	۳۸	Δ A	Δ A	۲۱/۹۶۰
۱۳	EPS/P <sub>(t-1)</sub>	EPS/p <sub>(t-1)</sub>	۰/۳۱۶	۳۹	Δ E	Δ E	۲۴/۸۷۲
۱۴	DPS/P <sub>(t-1)</sub>	DPS/p <sub>(t-1)</sub>	۰/۱۹۱	۴۰	OI/TA	OI/TA	۰/۱۷۱
۱۵	Δ DPS	Δ DPS	۰/۲۹۶	۴۱	Δ OI/TA	Δ (OI/TA)	۰/۱۲۵
۱۶	CR	CA/CL	۱/۶۷۴	۴۲	ACC	ACC/MV <sub>(t-1)</sub>	۰/۰۸۲
۱۷	QR	(CA-Inv)/CL	۰/۸۷۵	۴۳	Turn	TV/OS	۰/۱۴۲
۱۸	Inv Turn	COGS/Inv <sub>(ave)</sub>	۴/۱۰۸	۴۴	NT	NT	۸۶۷۲
۱۹	Δ Inv Turn	Δ COGS/Inv <sub>(ave)</sub>	۰/۱۵۴	۴۵	DPR	DPS/EPS	۰/۶۲۱
۲۰	FA Turn	S/FA	۵/۱۴۳	۴۶	OI	OI/MV <sub>(t-1)</sub>	۰/۱۴۳
۲۱	Δ FA Turn	Δ S/FA	۰/۰۹۴	۴۷	Δ OI	Δ OI	۱۱/۶۹۹
۲۲	TA turn	S/TA	۰/۸۰۸	۴۸	NI	NI/MV <sub>(t-1)</sub>	۰/۱۲۷
۲۳	Δ TA turn	Δ S/TA	۰/۱۴۷	۴۹	Δ NI	Δ NI	۲۲/۳۲۱
۲۴	GPM	GP/S	۰/۲۶۴	۵۰	CI	CI/MV <sub>(t-1)</sub>	۰/۱۵۶
۲۵	Δ GPM	Δ GP/S	۰/۰۶۴	۵۱	Δ CI	Δ CI	۲۶/۲۵۴
۲۶	OPM	OI/S	۰/۲۱۱	۵۲	WC/TA	WC/TA	۰/۰۶۲

MV: ارزش بازار شرکت P: قیمت BV: ارزش دفتری TL: کل بدهی‌ها TA: کل دارایی‌ها E: حقوق صاحبان سهام NI: سود خالص Beta: ریسک سیستماتیک EPS: سود هر سهم DPS: سود تقسیمی هر سهم CA: دارایی‌های جاری CL: بدهی‌های جاری Inv: موجودی کالا COGS: بهای تمام شده کالای فروش رفته S: فروش FA: دارایی‌های ثابت GP: سود ناخالص OI: سود عملیاتی R: بازده سهام Int: هزینه‌های مالی OCF: جریان نقد عملیاتی IRFEF: بازده سرمایه‌گذاری‌ها و سود پرداختی بابت تأمین مالی ICF: جریان نقد فعالیت‌های سرمایه‌گذاری FCF: جریان نقد فعالیت‌های تأمین مالی ACC: اقلام تعهدی TV: حجم مبادلات OS: تعداد سهام منتشره NT: تعداد دفعات مبادله CI: سود جامع WC: سرمایه در گردش

منبع: یافته‌های پژوهشگر

جدول ۲- متغیرهای بهینه انتخاب شده در روش‌های مبتنی بر همبستگی و ریلیف

روش ریلیف				روش مبتنی بر همبستگی			
نام متغیر	ردیف	نام متغیر	ردیف	نام متغیر	ردیف	نام متغیر	ردیف
$NI\Delta$	۵	E/P	۱	$S\Delta$	۵	Beta	۱
$\Delta OI$	۶	GPM	۲	$OI\Delta$	۶	$DPS\Delta$	۲
$\Delta ROA$	۷	$S\Delta$	۳	$R_{(t-1)}$	۷	Lev	۳
		$R_{(t-1)}$	۴			GPM	۴

منبع: یافته‌های پژوهشگر

جدول ۳- معیارهای ارزیابی عملکرد پیش‌بینی مورد استفاده

نحوه سنجش	نام معیار
$\sqrt{\frac{\sum_{p=1}^P (dp - zp)^2}{P}}$	مربع‌مجدور میانگین خطا (RMSE)
$1 - \frac{\sum_{p=1}^P (dp - zp)^2}{\sum_{p=1}^P (dp - \bar{dp})^2}$	ضریب تعیین ( $R^2$ )
$\frac{100}{P} \times \sum_{p=1}^P \left  \frac{dp - zp}{dp} \right $	میانگین قدرمطلق درصد خطا (MAPE)

$Z_p$ : مقدار پیش‌بینی شده،  $dp$ : مقدار واقعی،  $\bar{dp}$ : میانگین مقادیر

منبع: اسمیت و گوپتا، ۲۰۰۲، ۹ و آذر و کریمی، ۱۳۸۸، ۸

افزون بر این، به منظور ارزیابی عملکرد روش‌های مختلف انتخاب متغیر بهینه، معیارهای ارزیابی (میانگین قدرمطلق درصد خطا، مجدور مربع میانگین خطا و ضریب تعیین) حاصل از هر یک از روش‌های انتخاب متغیر با یکدیگر و همچنین با معیارهای ارزیابی حاصل از عدم انجام مرحله انتخاب متغیرهای پیش‌بین در هر یک از روش‌های خطی و غیرخطی مقایسه می‌شود. منظور از عدم انجام مرحله انتخاب متغیرهای پیش‌بین، پیش‌بینی با استفاده از کلیه متغیرهای پیش‌بین (قبل از کاهش تعداد متغیرها) است. انتخاب متغیرهای بهینه پیش‌بین و پیش‌بینی بازده سهام با استفاده از روش‌های خطی و غیرخطی مختلف به وسیله نرم‌افزارهای Weka نسخه ۳-۷ و MATLAB نسخه ۸، انجام شده است. به منظور آزمون فرضیه‌های پژوهش نیز از آزمون  $t$  زوجی (براساس صد دقت حاصل از اجرای روایی متقابل ۱۰ بخشی<sup>۵۹</sup> با ۱۰ بار تکرار در هر روش پیش‌بینی) در نرم‌افزار SPSS نسخه ۲۱ استفاده شده است.

در این پژوهش به منظور ارزیابی عملکرد رگرسیون تجمیعی و کسب نتایج تجربی قابل مقایسه، معیارهای عملکرد (شامل میانگین قدرمطلق درصد خطا، مجدور مربع میانگین خطا و ضریب تعیین) این

روش غیرخطی با روش غیرخطی شبکه‌های عصبی مصنوعی و رگرسیون خطی مقایسه شده است. دلایل استفاده از شبکه‌های عصبی پرسپترون چندلایه با الگوریتم آموزشی پس انتشار خطا به عنوان مبنای مقایسه به شرح زیر است:

(۱) در دهه‌های اخیر، شبکه‌های عصبی مصنوعی به‌طور وسیعی برای پیش‌بینی‌های مالی استفاده شده است (آل‌فارو و همکاران، ۲۰۰۸، ۱۱۰، راوی کومار و راوی، ۲۰۰۷)، به‌طوری‌که، حدود ۹۵٪ از مطالعات تجاری کاربردی از پرسپترون چندلایه استفاده کرده‌اند و مشهورترین روش آموزشی نیز پس انتشار خطا است (تسای، ۲۰۰۹، ۱۲۳).

(۲) یافته‌های پژوهش‌های تجربی بیانگر دقت بالای این مدل هوش مصنوعی است. این مدل‌ها قادر به شناسایی روابط غیرخطی هستند و عملکرد خوبی در صورت وجود اطلاعات دارای پارازیت که اغلب در پیش‌بینی‌های مالی و حسابداری رخ می‌دهد از خود نشان می‌دهند (راوی کومار و راوی، ۲۰۰۷؛ آل‌فارو و همکاران، ۲۰۰۸، ۱۱۰؛ و تسای، ۲۰۰۹، ۱۲۳).

دلیل اصلی مقایسه عملکرد رگرسیون تجمیعی با رگرسیون خطی نیز مقایسه عملکرد روش غیرخطی مزبور با روش خطی است که در اغلب پژوهش‌های حسابداری و مالی استفاده می‌شود.

در روش holdout که در اغلب پژوهش‌های حسابداری و مالی (به‌ویژه در ایران) استفاده شده است، داده‌ها به دو دسته به نام مجموعه آموزشی و مجموعه آزمایشی تقسیم می‌شود. این روش دارای محدودیت‌های بارزی است. اول آن که تعداد داده‌های کمتری برای آموزش در اختیار روش پیش‌بینی خواهد بود، زیرا تعدادی از آن برای مرحله آزمایش کنار گذاشته می‌شود. بنابراین، ممکن است دقت مدل حاصل، کمتر از زمانی باشد که تمام داده‌ها برای آموزش استفاده می‌شوند. دوم آن که ممکن است مدل پیش‌بینی تا حد زیادی وابسته به ترکیب مجموعه‌های داده‌های آموزشی و آزمایشی باشد. هرچه اندازه مجموعه داده‌های آموزش کوچک‌تر باشد، واریانس مدل بیشتر خواهد بود. از طرف دیگر اگر اندازه مجموعه داده‌های آموزشی خیلی بزرگ باشد، دقت تخمین با توجه به اندازه کوچک‌تر مجموعه داده‌های آزمون از قابلیت اطمینان کمتری برخوردار خواهد بود. چنین تخمینی دارای بازه اطمینان زیادی خواهد بود. در نهایت، مجموعه داده‌های آموزشی و آزمایشی از یکدیگر مستقل نیستند، زیرا هر دو، زیرمجموعه‌ای از مجموعه اولیه داده‌ها هستند. طبقه‌بندی داده‌های بیشتر در یکی از مجموعه داده‌های آموزش یا آزمایشی سبب کمبود داده‌های مجموعه دیگر خواهد شد. بنابراین، روش مزبور، روش مناسبی نخواهد بود (کوهاوی، ۱۹۹۵: ۱۱۳۸).

در این پژوهش، به منظور بررسی تعمیم‌پذیری پیش‌بینی‌های انجام شده از روایی متقابل ۱۰ بخشی استفاده می‌شود. روایی متقابل ۱۰ بخشی برای برآورد نرخ خطای واقعی کاملاً قابل‌اتکا و کافی است (هو، ۲۰۱۰، ۲۵۳۵). در این روش، نمونه اصلی به ۱۰ دسته نمونه فرعی مختلف تقسیم می‌شود. ۹ نمونه فرعی به‌عنوان نمونه‌های آموزشی استفاده می‌شود و نمونه فرعی باقی‌مانده به‌عنوان نمونه آزمایشی، مورد آزمون قرار می‌گیرد. این شیوه تا حدی تکرار می‌شود که هر یک از ۱۰ نمونه فرعی به‌عنوان نمونه آزمایشی مورد آزمون قرار گیرد. مزیت روش مزبور، این است که تمام نمونه‌ها در نهایت هم به‌عنوان داده‌های آموزشی و

هم به‌عنوان آزمایشی استفاده خواهد شد. افزون بر این، استفاده از روایی متقابل، از بروز مشکل بیش‌برازش و مشکلات مربوط به نتایج برون‌نمونه‌ای نتایج جلوگیری می‌کند. به‌طور کلی، در این پژوهش، با استفاده از روش‌های پیش‌بینی رگرسیون تجمیعی، شبکه‌های عصبی و رگرسیون خطی، بازده‌آتی سهام پیش‌بینی شده است. در این راستا، از روایی متقابل ۱۰ بخشی با ۱۰ بار تکرار استفاده شد که منجر به ایجاد ۱۰۰ معیار ارزیابی در مورد هر روش پیش‌بینی می‌شود و معیارهای ارزیابی ارائه شده در این پژوهش، میانگین این ۱۰۰ عدد است. شایان ذکر است که بهتر بودن معیارهای ارزیابی (مثلاً بالاتر بودن ضریب تعیین) یک روش پیش‌بینی نسبت به روش دیگر، دال بر برتری روش مزبور نمی‌باشد و ممکن است تفاوت بین دو روش از نظر آماری معنادار نباشد. بنابراین، از آماره  $t$  زوجی به منظور بررسی معنادار بودن تفاوت بین دو روش از نظر آماری استفاده شده است.

#### ۶- یافته‌های پژوهش

جدول‌های ۴ الی ۶ معیارهای ارزیابی (شامل میانگین قدرمطلق درصد خطا، مجذور مربع میانگین خطا و ضریب تعیین) مربوط به پیش‌بینی بازده سهام را بر اساس روش‌های رگرسیون تجمیعی (ER)، شبکه‌های عصبی مصنوعی (ANN) و رگرسیون خطی (LR) در سه حالت، با استفاده از ۵۲ متغیر پیش‌بین (All)، با متغیرهای انتخابی در روش‌های مبتنی بر همبستگی (Cort) و ریلیف (R) نشان می‌دهد. با توجه به نتایج این جدول‌ها، در هر سه حالت، عملکرد رگرسیون تجمیعی بهتر از شبکه‌های عصبی مصنوعی و رگرسیون خطی است. همچنین عملکرد شبکه‌های عصبی مصنوعی بهتر از رگرسیون خطی است.

جدول ۴- عملکرد روش‌های مختلف پیش‌بینی با استفاده از کلیه متغیرهای پژوهش

روش پیش‌بینی معیار عملکرد	ER	ANN	LR
مجذور مربع میانگین خطا	۴۹/۲۴۳	۶۳/۸۵۷	۸۹/۲۴۷
میانگین قدرمطلق درصد خطا	۰/۱۹۱	۰/۲۴۵	۰/۵۱۲
ضریب تعیین	۰/۳۸۴	۰/۲۳۸	۰/۰۴۱

منبع: یافته‌های پژوهشگر

جدول ۵- عملکرد روش‌های مختلف پیش‌بینی با استفاده از متغیرهای انتخاب شده در روش

#### مبتنی بر همبستگی

روش پیش‌بینی معیار عملکرد	ER	ANN	LR
مجذور مربع میانگین خطا	۴۰/۱۹۵	۵۳/۸۵۷	۸۴/۸۹۵
میانگین قدرمطلق درصد خطا	۰/۱۲۱	۰/۲۱۴	۰/۴۶۷
ضریب تعیین	۰/۵۸۶	۰/۳۸۷	۰/۱۱۴

منبع: یافته‌های پژوهشگر



جدول ۶- عملکرد روش‌های مختلف پیش‌بینی با استفاده از متغیرهای انتخاب شده در روش ریلیف

LR	ANN	ER	روش پیش‌بینی معیار عملکرد
۷۵/۶۱۷	۴۱/۵۶۱	۳۷/۵۱۷	مجدور مربع میانگین خطا
۰/۳۴۲	۰/۱۵۴	۰/۰۹۷	میانگین قدرمطلق درصد خطا
۰/۱۶۷	۰/۴۷۲	۰/۶۹۴	ضریب تعیین

منبع: یافته‌های پژوهشگر

در ادامه، برای آزمون فرضیه اول و بررسی وجود تفاوت معنادار بین عملکرد پیش‌بینی میان هر جفت روش پیش‌بینی، از آزمون  $t$  زوجی استفاده شده است. در این راستا، از دقت‌های حاصل از ۱۰ بار تکرار روایی مقطع ۱۰ بخشی (روایی مقطع ۱۰ بخشی با ۱۰ بار تکرار) استفاده شد که منجر به ایجاد ۱۰۰ دقت در مورد هر روش پیش‌بینی می‌شود. آماره آزمون و مقدار احتمال مربوطه حاکی از وجود تفاوت معنادار در مورد کلیه معیارهای عملکرد بین هر جفت روش پیش‌بینی مورد مقایسه در هر سه حالت بود. جدول‌های ۷ الی ۹ نتایج این آزمون و مقدار احتمال مربوطه را برای معیار ضریب تعیین نشان می‌دهد. با توجه به آماره  $t$  و مقدار احتمال مربوطه ( $p < 0/05$ )، تفاوت معناداری بین ضریب تعیین هر جفت روش پیش‌بینی مورد مقایسه در هر سه حالت مورد بررسی وجود دارد. بنابراین، فرضیه اول پژوهش (و فرضیه‌های فرعی آن)، تأیید می‌شود. بنابراین، با توجه به بهتر بودن معیارهای عملکرد رگرسیون تجمیعی نسبت به شبکه‌های عصبی مصنوعی و رگرسیون خطی و معنادار بودن آن، می‌توان نتیجه گرفت که روش غیرخطی مزبور، به طور معناداری بهتر از رگرسیون خطی و شبکه‌های عصبی مصنوعی، بازده سهام را پیش‌بینی می‌کند. همچنین، با توجه به بهتر بودن معیارهای عملکرد شبکه‌های عصبی مصنوعی نسبت به رگرسیون خطی و معنادار بودن تفاوت آن، چنین استنباط می‌شود که این روش غیرخطی به طور معناداری، بهتر از روش خطی بازده سهام را پیش‌بینی می‌کند.

جدول ۷- نتایج آزمون  $t$  و مقدار احتمال مربوطه در حالت استفاده از کلیه متغیرها

ER	ANN	LR	
-	-	-	ER
-	-	۳/۲۸۵ (۰/۰۰۰)	ANN
-	۷/۵۴۸ (۰/۰۰۰)	۸/۷۲۱ (۰/۰۰۰)	LR

منبع: یافته‌های پژوهشگر

جدول ۸- نتایج آزمون t و مقدار احتمال مربوطه در حالت استفاده از روش انتخاب متغیر مبتنی

بر همبستگی

ER	ANN	LR	
-	-	-	ER
-	-	۵/۷۴۲ (۰/۰۰۰)	ANN
-	۷/۸۹۴ (۰/۰۰۰)	۹/۸۴۸ (۰/۰۰۰)	LR

منبع: یافته‌های پژوهشگر

جدول ۹- نتایج آزمون t و مقدار احتمال مربوطه در حالت استفاده از روش انتخاب متغیر ریلیف

ER	ANN	LR	
-	-	-	ER
-	-	۵/۹۸۱ (۰/۰۰۰)	ANN
-	۸/۳۵۷ (۰/۰۰۰)	۱۱/۴۲۹ (۰/۰۰۰)	LR

منبع: یافته‌های پژوهشگر

به‌منظور آزمون فرضیه دوم، معیارهای عملکرد هر روش پیش‌بینی در حالت استفاده از ۵۲ متغیر پیش‌بین با حالت استفاده از متغیرهای انتخابی توسط روش‌های مبتنی بر همبستگی و ریلیف مقایسه می‌شود. اگر عملکرد هر روش پیش‌بینی (بر اساس جدول‌های ۴ الی ۶) در حالت استفاده از روش‌های انتخاب متغیر نسبت به استفاده از ۵۲ متغیر، بهتر و تفاوت مربوط از نظر آماری (بر اساس جدول‌های ۱۰ الی ۱۲) معنادار باشد، می‌توان استنباط کرد که روش‌های انتخاب متغیر تأثیر مثبت و معناداری بر عملکرد روش پیش‌بینی دارد. نتایج مربوط به مقایسه معیار ضریب تعیین هر روش پیش‌بینی در حالت استفاده از ۵۲ متغیر و استفاده از دو روش انتخاب متغیر در جدول‌های ۱۰ الی ۱۲ ارائه شده است. با توجه به آماره t و مقدار احتمال مربوطه ( $p < 0.05$ )، هر دو روش انتخاب متغیر تأثیر مثبت و معناداری بر ضریب تعیین پیش‌بینی دارد. بدین معنا که پیش‌بینی با استفاده از متغیرهای انتخابی هر دو روش نسبت به استفاده از کلیه متغیرها (۵۲ متغیر) منجر به عملکرد بهتر (ضریب تعیین بالاتر) در پیش‌بینی می‌شود. همچنین، روش انتخاب متغیر ریلیف، تأثیر بیشتری بر عملکرد روش‌های پیش‌بینی دارد. بدین معنا که عملکرد (ضریب تعیین) پیش‌بینی با استفاده از متغیرهای انتخابی روش ریلیف نسبت به پیش‌بینی با استفاده از متغیرهای انتخابی روش مبتنی بر همبستگی، بهتر (ضریب تعیین، بالاتر) است. بنابراین، فرضیه دوم پژوهش نیز تأیید می‌شود.

جدول ۱۰- نتایج آزمون t و مقدار احتمال مربوطه با استفاده از رگرسیون تجمیعی

R	Corr	All	
-	-	-	All
-	-	۳/۳۹۳ (۰/۰۰۰)	Corr
-	۵/۱۲۷ (۰/۰۰۰)	۶/۳۵۷ (۰/۰۰۰)	R

منبع: یافته‌های پژوهشگر

جدول ۱۱- نتایج آزمون t و مقدار احتمال مربوطه با استفاده از شبکه‌های عصبی مصنوعی

R	Corr	All	
-	-	-	All
-	-	۵/۵۲۱ (۰/۰۰۰)	Corr
-	۳/۲۱۷ (۰/۰۰۰)	۵/۴۵۲ (۰/۰۰۰)	R

منبع: یافته‌های پژوهشگر

جدول ۱۲- نتایج آزمون t و مقدار احتمال مربوطه با استفاده از رگرسیون خطی

R	Corr	All	
-	-	-	All
-	-	۳/۱۶۳ (۰/۰۰۰)	Corr
-	۳/۱۵۹ (۰/۰۰۰)	۳/۸۷۹ (۰/۰۰۰)	R

منبع: یافته‌های پژوهشگر

به منظور ارزیابی نقش صنعت در پیش‌بینی بازده سهام، آزمون‌های انجام شده، به تفکیک گروه‌های صنعت نیز انجام شد. شایان ذکر است که با توجه به محدودیت‌های اعمال شده در نمونه آماری، تعداد شرکت‌های انتخابی برخی از صنایع اندک و برخی صنایع در نمونه آماری فاقد نماینده بوده و بنابراین با صنایع تقریباً مشابه ترکیب گردیده است. در نهایت، شرکت‌های مورد بررسی در ۸ گروه صنعت، طبقه‌بندی

شد. جدول ۱۳ گروه‌های صنایع مربوط به شرکت‌های مورد مطالعه و فراوانی آن‌ها در پژوهش حاضر را نشان می‌دهد.

جدول ۱۳- گروه‌بندی صنایع و فراوانی آن‌ها در نمونه آماری

ردیف	گروه صنایع	صنایع زیرمجموعه	فراوانی
۱	ماشین‌آلات و تجهیزات	ماشین‌آلات و تجهیزات، ماشین‌آلات دستگاه‌های برقی	۸
۲	خودرو و ساخت قطعات	خودرو و ساخت قطعات	۱۴
۳	غذایی و آشامیدنی	محصولات غذایی و آشامیدنی، قند و شکر	۷
۴	سیمانی	سیمان، آهک و گچ	۹
۵	دارویی	مواد و محصولات دارویی	۲۱
۶	فلزات	فلزات اساسی، ساخت محصولات فلزی، استخراج کانه‌های فلزی	۱۳
۷	شیمیایی، نفتی، لاستیک و پلاستیک	محصولات شیمیایی، فرآورده‌های نفتی، لاستیک و پلاستیک	۱۷
۸	کاشی و سرامیک، منسوجات، محصولات چوبی و کاغذی و سایر محصولات کانی غیرفلزی	کاشی و سرامیک، سایر محصولات کانی غیرفلزی، منسوجات، محصولات چوبی، محصولات کاغذی و استخراج سایر معادن	۱۲
<b>مجموع</b>			۱۰۱

منبع: یافته‌های پژوهشگر

جدول ۱۴، نتایج ضریب تعیین پیش‌بینی در روش‌های مختلف پیش‌بینی با استفاده از متغیرها یا عامل‌های متفاوت را برای شرکت‌های مورد بررسی در هر گروه صنعت نشان می‌دهد. دلیل استفاده از ضریب تعیین، این است که معیار مزبور، معروف‌ترین و پرکاربردترین معیار ارزیابی مدل‌های مختلف پیش‌بینی است (اسمیت و گوپتا، ۲۰۰۲: ۸).

با توجه به نتایج ارائه شده در جدول ۱۴، ضریب تعیین پیش‌بینی نسبت به پیش‌بینی تجمعی کل شرکت‌ها، در برخی موارد، افزایش و برخی موارد، کاهش یافته است. به نظر می‌رسد که کاهش تعداد شرکت‌های مورد بررسی منجر به کاهش ضریب تعیین و همگن‌تر شدن شرکت‌های مورد بررسی منجر به بهبود عملکرد پیش‌بینی می‌شود. دلیل دیگر کاهش ضریب تعیین در برخی موارد نسبت به پیش‌بینی با استفاده از کل شرکت‌ها (جدول ۴ الی ۶)، این است که پیش‌بینی با تعداد نمونه‌های آموزشی اندک در دوره زمانی ده ساله (با شرایط سیاسی، اقتصادی و اجتماعی مختلف) انجام شده است. افزون بر این، با توجه به نتایج ارائه شده در جدول ۱۴، ضریب تعیین پیش‌بینی در گروه صنعت دارویی، اغلب موارد، افزایش یافته است. به نظر می‌رسد با توجه به این که گروه مورد بررسی صرفاً از شرکت‌های دارویی و از تعداد نسبتاً قابل

توجهی شرکت تشکیل شده است و احتمالاً نوسان‌پذیری کمتری نسبت به شرایط سیاسی، اقتصادی و اجتماعی دارد، همگن‌تر شدن شرکت‌های مورد بررسی، (نسبت به بررسی کل شرکت‌های نمونه)، نقش بیشتری در مقایسه با کاهش تعداد شرکت‌های مورد بررسی داشته و منجر به بهبود عملکرد پیش‌بینی در اغلب موارد می‌شود. دلیل احتمالی کاهش در ضریب تعیین پیش‌بینی با استفاده از روش خطی، عدم احراز مفروضات رگرسیون خطی است.

جدول ۱۴- نتایج ضریب تعیین در گروه صنایع متفاوت

روش پیش‌بینی متغیرها			متغیرها	روش پیش‌بینی متغیرها			متغیرها		
LR	ANN	ER		LR	ANN	ER			
۰/۰۲۲	۰/۲۴۶	۰/۳۷۵	All	خودرو و ساخت قطعات	۰/۰۲۳	۰/۲۴۶	۰/۳۹۱	All	ماشین‌آلات و تجهیزات
۰/۱۲۸	۰/۳۹۴	۰/۵۹۲	Corr		۰/۱۳۶	۰/۳۵۷	۰/۵۸۲	Corr	
۰/۱۲۶	۰/۴۹۳	۰/۷۰۱	R		۰/۱۸۲	۰/۴۹۱	۰/۶۵۳	R	
۰/۰۳۵	۰/۲۴۱	۰/۳۶۵	All	سیمانی	۰/۰۳۵	۰/۲۲۱	۰/۳۸۷	All	غذایی و آشامیدنی
۰/۱۰۳	۰/۳۹۲	۰/۵۹۲	Corr		۰/۱۰۳	۰/۳۵۷	۰/۵۷۴	Corr	
۰/۱۴۴	۰/۴۷۸	۰/۶۸۳	R		۰/۱۴۴	۰/۴۴۷	۰/۶۸۱	R	
۰/۰۳۳	۰/۲۲۳	۰/۳۶۵	All	فلزات	۰/۰۳۲	۰/۲۴۱	۰/۳۹۴	All	دارویی
۰/۰۹۷	۰/۳۸۱	۰/۵۹۳	Corr		۰/۱۱۲	۰/۴۱۲	۰/۶۰۱	Corr	
۰/۱۷۳	۰/۴۷۹	۰/۶۸۷	R		۰/۱۷۴	۰/۴۸۴	۰/۷۰۳	R	
۰/۰۱۴	۰/۲۱۷	۰/۳۶۱	All	کاشی و سرامیک	۰/۰۲۴	۰/۲۴۱	۰/۳۵۷	All	شیمیایی، نفتی
۰/۰۹۱	۰/۳۶۳	۰/۵۷۴	Corr		۰/۰۸۱	۰/۳۴۶	۰/۵۹۲	Corr	
۰/۱۱۵	۰/۴۶۹	۰/۷۰۱	R		۰/۱۸۲	۰/۴۶۳	۰/۶۸۴	R	

#### ۷- بحث و نتیجه‌گیری

پیش‌بینی بازده سهام در تصمیم‌گیری‌های مالی از اهمیت بسزایی برخوردار است و همواره به عنوان موضوعی حیاتی مدنظر بوده و به طور وسیعی در ادبیات حسابداری و مالی مورد مطالعه قرار گرفته است. مرحله انتخاب متغیرهای پیش‌بین به عنوان یکی از مراحل است که باید قبل از پیش‌بینی استفاده شود و گامی مؤثر در انتخاب اطلاعات ارزشمندتر در بین اطلاعات وسیع است. به عبارت دیگر، هدف این مرحله، فیلتر کردن اطلاعات نامربوط یا اضافی است و در نتیجه می‌تواند عملکرد مدل را با کاهش تلاش برای آموزش، بهبود بخشد. علی‌رغم اهمیت مرحله انتخاب متغیرهای پیش‌بین قبل از پیش‌بینی، این مرحله کمتر در ادبیات پیش‌بینی بازده سهام بررسی شده است و بیشتر مطالعات بر ایجاد مدل‌های پیش‌بینی اثربخش‌تر با قابلیت‌های پیش‌بینی بهتر، (بدون توجه به انتخاب متغیرهای پیش‌بین)، تأکید داشته‌اند. افزون

بر این، در اغلب پژوهش‌های انجام شده در بورس اوراق بهادار تهران، از روش‌های خطی و در پژوهش‌های اندکی نیز از شبکه‌های عصبی مصنوعی برای پیش‌بینی بازده سهام استفاده شده است و علی‌رغم برتری نظری و تجربی سایر روش‌های غیرخطی از قبیل رگرسیون تجمیعی، تاکنون از این روش‌ها استفاده نشده است. در این راستا، پژوهش حاضر بر پیش‌بینی بازده سهام با استفاده از متغیرهای منتخب توسط روش‌های انتخاب متغیرهای پیش‌بین بهینه و روش غیرخطی رگرسیون تجمیعی تأکید دارد.

یافته‌های تجربی بررسی ۱۰۱ شرکت پذیرفته شده در بورس اوراق بهادار تهران در دوره زمانی ۱۳۸۳ الی ۱۳۹۲ حاکی از آن بود که عملکرد رگرسیون تجمیعی (شامل میانگین قدرمطلق درصد خطا، مجدور مربع میانگین خطا و ضریب تعیین) به طور معناداری بهتر از رگرسیون خطی و شبکه‌های عصبی مصنوعی است. این موضوع حتی بدون استفاده از روش‌های انتخاب متغیر (و در صورت استفاده از ۵۲ متغیر پیش‌بین) نیز صادق است. عملکرد شبکه‌های عصبی مصنوعی نیز به طور معناداری بهتر از رگرسیون خطی بوده است. بنابراین، روش‌های غیرخطی به طور معناداری بازده سهام را بهتر از رگرسیون خطی پیش‌بینی می‌کند. یافته‌های بررسی عملکرد و سودمندی روش‌های انتخاب متغیرهای پیش‌بین بهینه نیز حاکی از آن است که عملکرد پیش‌بینی بازده با متغیرهای انتخاب شده توسط روش‌های مبتنی بر همبستگی و ریلیف به طور معناداری بهتر از پیش‌بینی با استفاده از کلیه متغیرها (و عدم استفاده از روش‌های انتخاب متغیرهای بهینه) است. در این راستا، متغیرهای انتخاب شده در روش ریلیف عملکرد پیش‌بینی بازده را بیشتر از روش مبتنی بر همبستگی بهبود می‌دهد. بنابراین، استفاده از روش غیرخطی رگرسیون تجمیعی و بدون استفاده از روش‌های انتخاب متغیرهای پیش‌بین، عملکرد پیش‌بینی به طور معناداری بهبود می‌یابد، که این موضوع حاکی از سودمندی روش غیرخطی مزبور است. همچنین، در صورت استفاده از متغیرهای پیش‌بین بهینه انتخاب شده نسبت به استفاده از ۵۲ متغیر اولیه، عملکرد پیش‌بینی به طور معناداری افزایش می‌یابد که این موضوع بیانگر سودمندی این روش‌هاست. دلیل این امر، مسأله اضافه‌بار ابعاد می‌باشد. به نظر می‌رسد اضافه کردن متغیرهای بیشتر، پارازیت (نویز) و در نتیجه خطا را افزایش می‌دهد و اضافه کردن متغیرها تنها تا یک حد معین می‌تواند به بهبود پیش‌بینی کمک کند و اضافه کردن بیشتر متغیرها منجر به مسأله اضافه‌بار ابعاد می‌شود.

به طور کلی، نتایج پژوهش حاکی از برتری معنادار روش‌های غیرخطی رگرسیون تجمیعی و شبکه‌های عصبی نسبت به رگرسیون خطی در پیش‌بینی بازده سهام است. این نتیجه، با یافته‌های پژوهش‌های خارجه‌ای جیمرامور و مرامور-کوستا (۱۹۹۷)، مرامور و پاهور (۱۹۹۸)، کاناسویانو پولس (۲۰۰۱)، اولسونو موسمن (۲۰۰۳)، عمران و رجب (۲۰۰۴)، مک‌میلان (۲۰۰۷)، بوایسیگلو و آوسی (۲۰۱۰)، جیسن و همکاران (۲۰۱۳) و با پژوهش‌های داخلی خالوزاده و خاکی صدیق (۱۳۷۷)، راعی و چاوشی (۱۳۸۲)، آذر و کریمی (۱۳۸۸)، همت‌فر و همکاران (۱۳۹۰)، عباسی و باقری (۱۳۹۰) هماهنگ است. نتایج این پژوهش مبنی بر سودمندی روش‌های انتخاب متغیرها و وجود تفاوت معنادار بین میزان سودمندی روش‌های مختلف انتخاب متغیر، با یافته‌های پژوهش‌های تسای (۲۰۰۹) و ستایش و همکاران (۲۰۱۲) هماهنگ است.

با توجه به یافته‌های این پژوهش، پیشنهادهای زیر ارائه می‌شود:

- با توجه به نتایج آزمون فرضیه‌های فرعی مربوط به فرضیه اصلی اول، دال بر برتری روش غیرخطی تجمیعی نسبت به روش غیرخطی شبکه‌های عصبی و همچنین برتری این دو روش غیرخطی نسبت به رگرسیون خطی، به سرمایه‌گذاران و سایر استفاده‌کنندگان پیشنهاد می‌شود در تصمیم‌گیری‌های سرمایه‌گذاری مبنی بر خرید و فروش سهام، کاهش خطر سبب سرمایه‌گذاری و ارزیابی ریسک شرکت و سایر تصمیم‌گیری‌ها از رگرسیون تجمیعی استفاده کنند.
- با توجه به نتایج حاصل از آزمون فرضیه‌های فرعی مربوط به فرضیه اصلی دوم، مبنی بر این که متغیرهای انتخابی روش ریلیف بهتر از متغیرهای انتخابی روش همبستگی و همچنین این متغیرهای انتخابی بهتر از کلیه متغیرها (۵۲ متغیر) بازده سهام را پیش‌بینی می‌کند، به سرمایه‌گذاران و سایر استفاده‌کنندگان پیشنهاد می‌شود که در پیش‌بینی بازده سهام، مرحله انتخاب متغیرهای پیش‌بین را انجام دهند و صرفاً بر اساس پژوهش‌های گذشته متغیرهای پیش‌بین را انتخاب نکنند. استفاده از روش ریلیف، برای انتخاب متغیرهای پیش‌بین توصیه می‌شود.

#### فهرست منابع

- ۱) آذر، عادل و سیروس کریمی، (۱۳۸۸)، "پیش‌بینی بازده سهام با استفاده از نسبت‌های حسابداری با رویکرد شبکه‌های عصبی"، تحقیقات مالی، دوره ۱۱، شماره ۲۸، صص ۳-۲۰.
- ۲) احمدپور، احمد و امیرحسین عظیمیان معز، (۱۳۹۱)، "بررسی ارتباط رشد دارایی‌ها با بازده سهام در بورس اوراق بهادار تهران"، پژوهشنامه اقتصادی، دوره ۱۲، شماره ۳ - صص ۲۷-۴۲.
- ۳) اسلامی بیدگلی، غلامرضا، شهاب‌الدین شمس و هستی چیت‌سازان، (۱۳۸۶)، "نظریه‌های مالی نوین، (رابرت هاگن)"، تهران، انتشارات دانشگاه تهران.
- ۴) اسلامی بیدگلی، غلامرضا و اعظم هنردوست، (۱۳۹۱)، "مدل سه عاملی فاما و فرنچ و ریسک نقدشوندگی: شواهدی از بازار بورس اوراق بهادار تهران"، دانش سرمایه‌گذاری، سال ۱، شماره ۲، صص ۹۷-۱۱۶.
- ۵) ایزدی نیا، ناصر، منیژه رامشه و سعید یادگاری، (۱۳۹۱)، "پیش‌بینی بازده سهام براساس حجم معاملات سهام"، فصلنامه حسابداری مالی، شماره ۱۶، صص ۱۶۰-۱۷۴.
- ۶) ایزدی نیا، ناصر، محمد ابراهیمی و امین حاجیان نژاد، (۱۳۹۳)، "مقایسه مدل اصلی سه عاملی فاما و فرنچ با مدل اصلی چهار عاملی کارهارت در تبیین بازده سهام شرکت‌های پذیرفته‌شده در بورس اوراق بهادار تهران"، مجله مدیریت دارایی و تأمین مالی، جلد ۲، شماره ۳، صص ۱۷-۲۸.
- ۷) تهرانی، رضا و عسکر نوربخش، (۱۳۸۶)، "مدیریت سرمایه‌گذاری"، (چارلز پی جونز)، تهران، انتشارات نگاه دانش، چاپ ۳،

- ۸) تهرانی، رضا و عسکر نوربخش، (۱۳۸۷)، "مدیریت مالی پیشرفته"، (توماس کاپلند؛ ای، وستون، جان فرد و کولدپ شاستری)، تهران، انتشارات نگاه دانش، چاپ اول.
- ۹) ثقفی، علی و محمدجواد سلیمی، (۱۳۸۴)، "متغیرهای بنیادی حسابداری و بازده سهام"، مجله علوم انسانی و اجتماعی دانشگاه شیراز، دوره ۲۲، شماره ۲، پیاپی ۴۳، صص ۶۱-۷۴.
- ۱۰) ثقفی، علی و صابر شعری، (۱۳۸۳)، "نقش اطلاعات بنیادی حسابداری در پیش‌بینی بازده سهام"، فصلنامه مطالعات حسابداری، شماره ۸، صص ۸۷-۱۲۰.
- ۱۱) جعفریه، حمیدرضا، نگار معتمدی و الهه ملایی، (۱۳۸۵)، "شبکه‌های عصبی و الگوریتم‌های ژنتیک در تجارت"، ماهنامه تدبیر، سال ۱۸، شماره ۱۷۷، صص ۶۲-۶۸.
- ۱۲) خالوزاده، حمید و علی خاکی صدیق (۱۳۷۷)، "پیش‌بینی و الگوسازی فرآیند رفتار قیمت‌ها در بورس اوراق بهادار تهران در هم‌سنجی مدل‌های خطی و غیرخطی"، مدرس علوم انسانی، شماره ۶، صص ۴۵-۵۹.
- ۱۳) دستگیر، محسن، ندا تاجی و رحمان ساعدی، (۱۳۹۱)، "رابطه بین متغیرهای حسابداری با بازده سهام با استفاده از مدل بازده ژانگ"، پژوهشنامه حسابداری مالی و حسابرسی، سال ۴، شماره ۱۳، صص ۴۳-۶۴.
- ۱۴) دستگیر، محسن و رامین خدابنده، (۱۳۸۲)، "بررسی ارتباط بین محتوای اطلاعاتی اجزای اصلی صورت گردش وجه نقد با بازده سهام"، مجله علوم انسانی و اجتماعی دانشگاه شیراز، دوره ۱۹، شماره ۲، پیاپی ۳۸، صص ۱۰۰-۱۱۲.
- ۱۵) راعی، رضا و کاظم چاوشی، (۱۳۸۲)، "پیش‌بینی بازده سهام در بورس اوراق بهادار تهران: مدل شبکه‌های عصبی مصنوعی و مدل عاملی"، تحقیقات مالی، سال پنجم، شماره ۱۵، صص ۹۷-۱۲۰.
- ۱۶) عباسی، ابراهیم و سحر باقری، (۱۳۹۱)، "پیش‌بینی بازده سهام با استفاده از مدل‌های غیرخطی آستانه ای و بررسی نقش حجم معاملات در بهبود عملکرد اینمدل‌ها"، تحقیقات مالی، شماره ۳۲، صص ۹۱-۱۰۸.
- ۱۷) عبده تبریزی، حسین و محمود گنابادی، (۱۳۷۵)، "تردید در اعتبار مدل‌های مالی"، حسابدار، شماره ۱۱۵، صص ۱۳-۲۰.
- ۱۸) کمیته تدوین استانداردهای حسابداری ایران، (۱۳۸۹)، "استانداردهای حسابداری"، تهران، انتشارات سازمان حسابرسی.
- ۱۹) کیا، مصطفی، (۱۳۸۸)، "طراحی شبکه‌های عصبی، (مارتینتی هاگان، هاوارد بی دیموث، مارک بیل)"، تهران، کیان رایانه سبز، چاپ ۱.
- ۲۰) نمازی، محمد، (۱۳۸۹)، "پژوهش‌های تجربی در حسابداری: دیدگاه روش‌شناختی، (رشاد عبدالخلیق، بیپین ب. آجین کیا)"، شیراز، انتشارات دانشگاه شیراز، چاپ ۲.



- (۲۱) نمازی، محمد و امین ناظمی، (۱۳۸۴)، "بررسی تحلیلی تحقیقات انجام شده در بورس اوراق بهادار تهران"، تحقیقات مالی، شماره ۱۹، صص ۱۳۵-۱۶۶.
- (۲۲) نمازی، محمد و حسن محمدتبار کاسگری، (۱۳۸۶)، "به‌کارگیری مدل چندعاملی برای توضیح بازده سهام شرکت‌های پذیرفته شده در بورس اوراق بهادار تهران"، مجله علوم انسانی و اجتماعی دانشگاه شیراز، دوره ۲۶، شماره ۱، پیاپی ۵۰، صص ۱۵۷-۱۸۰.
- (۲۳) همت‌فر، محمود، سید علی‌اکبر حسینی، فرهاد شاه ویسی و یوسف نجفی، (۱۳۹۰)، "روابط خطی و غیرخطی بین متغیرهای حسابداری و بازده سهام شرکت‌های صنعت خودرو و ساخت قطعات"، پژوهشنامه حسابداری مالی و حسابرسی، سال ۳، شماره ۱۲، صص ۱۳۷-۱۵۴.
- 24) Alfaro, E., Garcia, N., Gamez, M., and D. Elizondo., (2008), "Bankruptcy Forecasting: An Empirical Comparison of AdaBoost and Neural Networks". *Decision Support Systems*, Vol. 45, PP. 110-122.
- 25) Al-Horani, A., Pope, P. F., and A. W. Stark., (2003), "Research and Development Activity and Expected Returns in The United Kingdom". *European Finance Review*, Vol. 7, No. 1, PP. 27-46.
- 26) Atiya, A. F., (2001), "Bankruptcy Prediction for Credit Risk Using Neural Networks: A Survey and New Results". *IEEE Transactions on Neural Networks*, Vol. 12, No. 4, PP. 929-935.
- 27) Boyacioglu, M. A., and D. Avci., (2010) , "An Adaptive Network-Based Fuzzy Inference System (ANFIS) for The Prediction of Stock Market Return: The Case of The Istanbul Stock Exchange". *Expert Systems with Applications*, Vol. 37, PP. 7908 – 7912.
- 28) Breiman, L., (1996), "Bagging Predictors". *Machine Learning*, Vol. 24, PP. 123-140.
- 29) DeTienne, K. B., DeTienne, D. H., and S. A. Joshi., (2003), "Neural Networks as Statistical Tools for Business Researchers". *Organizational Research methods*, Vol. 6, No. 2, PP. 236-265.
- 30) Fabozzi, F. J., and H. M. Markowitz., (2002), "The Theory and Practice of Investment Management". John Wiley & Sons.
- 31) Fama, E. F., (1991), "Efficient Capital Markets: II". *The Journal of Finance*, Vol. 46, No. 5, PP. 1575-1617.
- 32) Gysen, M., Huang, C. S., and R. Kruger., (2013), "The Performance of Linear Versus Non-Linear Models in Forecasting Returns on The Johannesburg Stock Exchange". *International Business & Economics Research Journal*, Vol. 12, No. 8, PP.985-994.
- 33) Hall, M. A., (2000), "Correlation-Based Feature Selection for Discrete and Numeric Class Machine Learning". In *Proceedings of The Seventeenth International Conference on Machine Learning (June 29 - July 02)*. P. Langley, Ed. Morgan Kaufmann Publishers, San Francisco, CA, PP. 359-366.
- 34) Hall, M. A., (1999), "Correlation-Based Feature Selection for Machine Learning". Ph.D Thesis, University of Waikato.
- 35) Hoglund, H., (2012), "Detecting Earnings Management Using Neural Networks". *Expert Systems with Applications*, Vol. 39, No. 10, PP. 9564-9570.
- 36) Hu, Y.C., (2010), "Analytic Network Process for Pattern Classification Problems using Genetic Algorithms". *Information Sciences*, Vol. 180, PP. 2528-2539.

- 37) Hung, C., and J. Chen., (2009), "A Selective Ensemble Based on Expected Probabilities for Bankruptcy Prediction". *Expert Systems with Applications*, Vol. 36, pp. 5297-5303.
- 38) Kanas, A., and A. Yannopoulos., (2001), "Comparing Linear and Nonlinear Forecasts for Stock Returns". *International Review of Economics and Finance*, Vol. 10, PP. 383-398.
- 39) Kira, K., and L. A. Rendell., (1992), "A Practical Approach to Feature Selection". *Proceedings of International Conference on Machine Learning*, PP. 249 -256.
- 40) Kohavi, R. (1995). "A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection", *IJCAI'95 Proceedings of The 14th International Joint Conference on Artificial Intelligence*, PP. 1137-1143.
- 41) Kothari, S. P., (2001), "Capital Markets Research in Accounting". *Journal of Accounting and Economics*, Vol. 31, PP. 105-231.
- 42) Koprinskaa, I.; Ranaa, M.; and V. G. Agelidisb., (2015), "Correlation and instance based feature selection for electricity load forecasting. *Knowledge-Based Systems*", Vol. 82, PP. 29-40.
- 43) Lee, M. C., and C. To., (2010), "Comparison of Support Vector Machine and Back Propagation Neural Network in Evaluating the Enterprise Financial Distress". *International Journal of Artificial Intelligence & Applications*, Vol. 1, No.3, PP. 31-43.
- 44) Lindenbaum, M., Markovitch, S., and D. Rusakov., (2004), "Selective Sampling for Nearest Neighbor Classifiers". *Machine Learning*, Vol. 2, PP. 125-152.
- 45) Liu, H., and S.Huang., (2010), "Integrating GA with Boosting Methods for Financial Distress Predictions". *Journal of Quality*, Vol. 17, No. 2, PP.131-157.
- 46) Lo, S. C., (2010), "The Effects of Feature Selection and Model Selection on The Correctness of Classification". *Proceedings of the 2010 IEEE IEEM*, PP. 989-993.
- 47) McMillan, D. G., (2007), "Non-Linear Forecasting of Stock Returns: Does Volume Help?". *International Journal of Forecasting*, Vol. 23, PP. 115-126.
- 48) Mendes-Moreia, J., Soares, C., Jorge, A. M., and J. F. D. Sousa., (2012), "Ensemble Approaches for Regression: A Survey". *ACM Computing Surveys*, Vol. 45, No. 1, PP. 10-40.
- 49) Min, J. H., and Y. Lee., (2005), "Bankruptcy Prediction Using Support Vector Machine with Optimal Choice of Kernel Function Parameters". *Expert Systems with Applications*, Vol. 28, PP. 603-614.
- 50) Mramor, D., and M. Pahor., (1998), "Testing Nonlinear Relationship between Excess Rate of Return on Equity and Financial Ratios". *23rd Meeting of the EURO Working Group on Financial Modelling*, PP. 119-134.
- 51) Mramor, D., and N. Mramor-Kosta., (1997), "Accounting Ratios as Factors of Rate of Return on Equity". *New Operational Approaches for Financial Modelling*, Physica -Verlag Heidelberg, PP. 335-348.
- 52) Mukkamala, S., Sung, A. H., Ribeiro, B., and A. Vieira., (2006), "Computational Intelligent Techniques for Financial Distress Detection". *Journal of Computational Intelligence Research*, Vol. 2, No. 1, PP. 60-65.
- 53) Olson, D., and C. Mossman., (2003), "Neural Network Forecasts of Canadian Stock Returns Using Accounting ratios". *International Journal of Forecasting*, Vol. 19, PP. 453-465
- 54) Omran, M., and A. Ragab., (2004), "Linear Versus Non-Linear Relationships between Financial Ratios and Stock Returns: Empirical Evidence from Egyptian Firms". *Review of Accounting and Finance*, Vol. 3, No. 2, PP.84-102.
- 55) Pinto, J., Henry, E., Robinson, T. R., and J. D. Stowe., (2010), "Equity Asset Valuation". 2 th Edition. John Wiley & Sons,.

- 56) Ravi Kumar, P., and V. Ravi., (2007), "Bankruptcy Prediction in Banks and Firms Via Statistical and Intelligent Techniques - A Review". *European Journal of Operational Research*, Vol. 180, PP. 1-28.
- 57) Robnik-Sikonja, M., and I. Kononenko., (1997), "An Adaptation of Relief for Attribute Estimation in Regression". *Machine Learning, Proceedings of 14th International Conference on Machine Learning (ICML'97)*, PP. 296-304.
- 58) Roll. R., (1997), "A Critique of The Asset Pricing Theory's Tests Part I: On Past and Potential Testability of The Theory". *Journal of Financial Economics*. Vol. 4, No. 2, PP. 129-176.
- 59) Setayesh, M. H., Kazemnezhad, M., Nikouei, M. A., and S. Azadi., (2012), "The Effectiveness of Fuzzy-Rough Set Feature Selection in The Prediction of Financial Distress: A Case of Iranian Context". *Wulfenia Journal*, Vol. 19, No. 10, PP. 268-287.
- 60) Smith, K., and J. Gupta., (2002), "Neural Networks in Business: Techniques and Applications". Idea Group Publishing.
- 61) Stambaugh, R., (1982), "On the Exclusion of Assets from Tests of The Two-Parameter Model: A Sensitivity Analysis". *Journal of Financial Economics*, Vol. 10, No. 3, PP. 237-268.
- 62) Sun, Z., (2004), "PEG Ratios and Stock Returns". Ph.D Thesis, University of Toronto.
- 63) Tsai, C., (2009), "Feature Selection in Bankruptcy Prediction". *Knowledge-Based Systems*, Vol. 22, PP. 120-127.

## یادداشت‌ها

- <sup>1</sup>. Kothari
- <sup>2</sup>. Artificial Neural Networks (ANN)
- <sup>3</sup>. Support Vector Regression (SVR)
- <sup>4</sup>. Ensemble Regression
- <sup>5</sup>. Min
- <sup>6</sup>. Lee
- <sup>7</sup>. Mukkamala et al.
- <sup>8</sup>. Alfaro et al.
- <sup>9</sup>. To
- <sup>10</sup>. Sun et al.
- <sup>11</sup>. Feature (Variable) Selection
- <sup>12</sup>. Mramor
- <sup>13</sup>. Mramor-Kosta
- <sup>14</sup>. Pahor
- <sup>15</sup>. Kanas
- <sup>16</sup>. Yannopoulos
- <sup>17</sup>. Omran
- <sup>18</sup>. Ragab
- <sup>19</sup>. McMillan
- <sup>20</sup>. Boyacioglu
- <sup>21</sup>. Avci
- <sup>22</sup>. Gysen et al.
- <sup>23</sup>. Lindenbaum
- <sup>24</sup>. Redundant
- <sup>25</sup>. Noisy
- <sup>26</sup>. Hall

27. Curse of Dimensionality
28. Feature Extraction
29. Feature Transformation
30. Discriminant Analysis
31. Robnik-Sikonja
32. Kononenko
33. Threshold
34. Atiya
35. Filter
36. Nearest Hit
37. Nearest Miss
38. Kira
39. Rendell
40. DeTienne et al.
41. Neuron
42. Cell Body
43. Axon
44. Dendrite
45. Waren McCulloch
46. Walter Pitts
47. Frank Rosenblatt
48. Hoglund
49. Least Mean Square (LMS)
50. Chain Rule
51. Generation
52. Pruning
53. Robustness
54. Mendes-Moreia et al.
55. Bagging
56. Breiman
57. Mean Absolute Percentage Error
58. Root Mean Squared Error
59. 10-Fold Cross Validation
60. Hu