

پژوهش‌های تجربی حسابداری

سال چهارم، شماره ۱۵، بهار ۱۳۹۴، صص ۱۸۱ - ۲۰۳

مقایسه دقت پیش بینی مدیریت سود با استفاده از الگوریتم‌های مورچگان و غذایابی باکتری

عزیز گرد*، سید حسام وقفی**، سید جواد حبیب زاده بایگی***

سارا خواجه زاده****

تاریخ دریافت: ۹۴/۰۳/۰۴

تاریخ پذیرش: ۹۴/۰۷/۱۱

چکیده

هدف این تحقیق بررسی این موضوع است که آیا می‌توان مدیریت سود را براساس مدل‌های مبتنی بر یادگیری ماشین کشف کرد. در این تحقیق برای پیش‌بینی مدیریت سود از مدل‌های مبتنی بر یادگیری ماشین (الگوریتم کلونی مورچه‌ها و غذایابی باکتری) استفاده شده است. برای این منظور ۱۴۳ شرکت پذیرفته شده در بورس اوراق بهادار تهران طی دوره زمانی ۱۳۸۸ تا ۱۳۹۲ مورد مطالعه قرار گرفتند. در این تحقیق با استفاده از الگوریتم حرکات ذرات اقدام به شناسایی متغیرهای معنادار با مدیریت سود شده و در نهایت بوسیله نرم‌افزار متلب اقدام به پیش‌بینی مدیریت سود شده است. نتایج برازش الگوریتم غذایابی باکتری و کلونی مورچه‌ها نشان می‌دهد که این دو الگوریتم با دقت بالای ۹۸ درصد توانایی پیش‌بینی مدیریت سود را دارند. نتایج مبین آن است که مدل کلونی مورچه‌ها توانایی بیشتری (خطای ۰/۹۷ درصد) در پیش‌بینی مدیریت سود نسبت به مدل غذایابی باکتری (خطای ۱/۱۹ درصد) دارد.

واژه‌های کلیدی: پیش‌بینی، مدیریت سود، الگوریتم کلونی مورچه‌ها، الگوریتم غذایابی باکتری

طبقه بندی موضوعی: C12, M41

* استادیار حسابداری دانشگاه پیام نور، (afmgord@yahoo.com)

** مربی حسابداری دانشگاه پیام نور، (نویسنده مسئول)، (h.vaghfi2012@gmail.com)

*** مدرس دانشگاه پیام نور، (mfakhar1369@yahoo.com)

**** دانشجوی کارشناسی ارشد مدیریت مالی، دانشگاه آزاد واحد الکترونیک، (sarakhmt20@yahoo.com)

مقدمه

برای تکنولوژی‌های سنتی حسابرسی مشکل است که زمان، منابع انسانی، مخارج و تأثیر رفتارهای غیرعادی بر اطلاعات مالی پیچیده و بزرگ را محدود نمایند. لذا توسعه مدل پیش بینی برای مدیریت سود برای حسابرسان نیز به منظور شناسایی درجه دستکاری در صورت‌های مالی مفید است (چانگ و همکاران، ۲۰۰۵). پیش‌بینی از این نظر که عنصر بنیادین و کلیدی در تصمیم‌گیری استفاده‌کنندگان درون سازمانی و همچنین برون سازمانی محسوب می‌شود، مهم است. براین اساس، کارایی و اثربخشی نهایی هر تصمیم، به نتایج رویدادهایی بستگی دارد که به دنبال هر تصمیم روی می‌دهد. بدین ترتیب تصمیمی کارا و اثربخش خواهد بود که براساس پیش‌بینی‌هایی انجام گیرد که مبنای آن صحیح بوده باشد. یکی از این پیش‌بینی‌ها، پیش‌بینی مدیریت سود می‌باشد. از نظر رونن و سادن (۱۹۸۱) مدیریت سود تلاش عمدی مدیریت جهت انتقال اطلاعات خاص به استفاده‌کنندگان صورت‌های مالی تعریف می‌شود. شپیر (۱۹۸۹) دخالت هدفمند در فرآیند گزارش‌گری مالی برون سازمانی به منظور تحصیل منافع شخصی را مدیریت سود تعریف می‌کند. پیش‌بینی‌های دقیق و به موقع مدیریت سود، موجب بهبود تصمیم‌گیری استفاده‌کنندگان از گزارش‌های حسابداری می‌شود. چنانچه سرمایه‌گذاران گذشته را مبنایی برای آینده قرار دهند، آن‌ها با استفاده از اطلاعات گذشته می‌توانند حدس بزنند که در آینده وضعیت شرکت چگونه خواهد بود، ولی تنها این یک حدس است که بر اساس اطلاعات گذشته به دست آمده است؛ در حالیکه آینده ممکن است بسیار با گذشته تفاوت داشته باشد. در این رابطه نظریه نمایندگی بر اساس فرض تضاد منافع می‌تواند به بیان دلایل مدیریت سود پردازد (حیب زاده، ۱۳۸۹). بر اثر ایجاد تضاد منافع که از نتایج ایجاد شرکتهای سهامی بزرگ است، رابطه مالک - نماینده و نظریه نمایندگی شکل گرفته و وجود مکانیسم نظارتی موثر بر مدیریت، جهت اطمینان از اعمال مدیریت صحیح در جهت منافع سهام داران، ضرورت می‌یابد. در ساختار کسب و کار جاری، تفکیک مالکیت از مدیریت اجتناب‌ناپذیر است و عدم وجود مکانیسم نظارتی موثر بر مدیریت در چنین شرکت‌هایی احتمال تخصیص ناکارآمد منابع و گسترش مشکلات سازمانی را افزایش می‌دهد و این امر غالباً منجر به صدور گزارش‌های غیر شفاف و گمراه‌کننده برای پنهان کردن مشکلات از دید سهام‌داران می‌گردد (جوهری و همکاران، ۲۰۰۸). در مدیریت سود، مدیریت قضاوت خود را در ارتباط با

گزارش‌گری مالی یا سازماندهی رویدادها به منظور تغییر گزارش‌های مالی به کار می‌گیرد. چه این کار به منظور گمراه کردن سهامداران در مورد عملکرد شرکت، یا به منظور تاثیر بر نتایج قراردادهای متکی به ارقام گزارش شده حسابداری باشد. صورت‌های مالی نادرست تا زمانی که عدم صحت آن‌ها معلوم شود متناوباً مدیران خوش‌بین را ترغیب به سرمایه‌گذاری می‌کند. تحت چنین شرایطی تلاش جهت دستیابی به ارزش مورد انتظار بازار سرمایه و اهداف سوددهی بالا، می‌تواند سرمایه‌گذاران، کارمندان، مشتریان و غیره را تحت تأثیر قرار دهد (خانی، ۱۳۸۲). عمل مدیریت سود با توجه به مدیریت کردن ادراک استفاده‌کنندگان از گزارش‌های مالی، یک عمل غیراخلاقی به شمار می‌آید (جوهراری و هم‌کاران، ۲۰۰۸). دیچاو و اسکینر (۲۰۰۰) بیان می‌کنند که حسابداران، مدیریت سود را به عنوان مشکلی می‌دانند که نیاز به یک عمل کنترلی فوری دارد. طبق نتایج تحقیق رافیک در سال ۲۰۰۲ که با استفاده از ابزار پرسشنامه انجام داده است، اکثریت پاسخ‌دهندگان باور ندارند که دست‌کاری سود یک عمل اخلاقی است. از سوی دیگر، برخی نیز بر این عقیده‌اند که مدیریت سود عملی است که به وسیله شرکت‌ها در جهت منافع سرمایه‌گذاران صورت می‌پذیرد. هیلی و والن (۱۹۹۹) معتقدند که گزارش‌گری مالی می‌تواند ارزش شرکت را افزایش دهد. از این‌رو، استانداردهای پذیرفته شده حسابداری باید برای مدیران جایگزین‌های مورد نیاز را در روش‌ها و به کارگیری قضاوت و برآورد در جهت گردآوری اطلاعات عملکرد شرکت، از قبیل به کارگیری مفروضات براساس اصول پذیرفته شده عمومی تشریح شده در استانداردها فراهم آورد.

از این رو دو عقیده کاملاً متضاد وجود دارد و این اختلاف نظر، کنترل اختیار مدیریت در انتخاب رویه‌های متنوع حسابداری را دشوار می‌سازد. در این ارتباط بلک (۱۹۹۸) معتقد است مدیران شرکت‌هایی که مالکان متعدد دارند و از سهامداران عمده برخوردار نیستند، انگیزه بیشتری برای مدیریت سود دارند. زیرا، هزینه پردازش اطلاعات برای سهامداران جزء توجه اقتصادی ندارد. در نتیجه، آن‌ها مجبورند به اطلاعات سود و زیان گزارش شده توسط مدیریت شرکت‌ها اتکا کنند. دیچو و همکاران (۱۹۹۵) معتقدند شرکت‌هایی که تعداد اعضای هیأت مدیره آن‌ها زیاد است و این اعضا نیز عمدتاً موظف هستند، در مقایسه با دیگر شرکت‌ها، از مدیریت سود بیشتری استفاده می‌کنند. به اعتقاد پورتر (۱۹۹۲) در شرکت‌هایی که سرمایه‌گذاران نهادی بخش عمده مالکیت شرکت را به عهده دارند، به دلیل این که آنان بر

سودهای نزدیک به سودهای پیش بینی شده تاکید می کنند، احتمال انجام مدیریت سود در مقایسه با دیگر شرکت‌ها بیشتر است. ماتسوموتو (۲۰۰۲) به این نتیجه رسید شرکت‌هایی که عمده سهام آن‌ها متعلق به سرمایه‌گذاران نهادی است، با احتمال بیشتری از مدیریت سود برای دستیابی به شاخص از پیش تعیین شده (سود) استفاده می کنند. فرانکل و همکاران (۲۰۰۲) نسبت بازده دارایی‌ها را به عنوان شاخص اندازه‌گیری عملکرد شرکت در نظر گرفتند و به این نتیجه رسیدند که نسبت مزبور بر مدیریت سود موثر است. مک نیکولز (۲۰۰۰) و بارتون و سیمکو (۲۰۰۲) شاخص اندازه‌گیری عملکرد را نسبت بازده حقوق صاحبان سهام و میزان فروش شرکت انتخاب کردند و به این نتیجه رسیدند که هر دو شاخص مزبور با مدیریت سود رابطه دارند. دیچو و دیچو (۲۰۰۲) اعتقاد دارند میزان فروش بر مدیریت سود موثر است. نجاری و همکاران (۲۰۱۴) با استفاده از متغیرهای نسبت جاری، نسبت وجه نقد عملیاتی به دارایی‌ها، نسبت جاری، نسبت بدهی به حقوق صاحبان سهام، اهرم مالی، مالکان نهادی، هزینه‌های سیاسی، تغییرات سود، ساختار مالکیت، تغییرات حسابرس و نوع صنعت مدلی را ایجاد کردند که با دقت ۹۴ درصد می‌توانست به پیش‌بینی مدیریت سود بپردازد. بداغی و بزازاده (۱۳۸۷) کیفیت افشا را به عنوان متغیری موثر بر مدیریت سود معرفی کردند. کردستانی و آشتاب (۱۳۸۸) از سود هر سهم برای پیش‌بینی مدیریت سود استفاده نمودند. چالاکلی و یوسفی (۱۳۹۱) با استفاده از درصد مالکیت سهامداران نهادی، نسبت بدهی، اندازه شرکت، مالیات بر درآمد، تغییر پذیری فروش، تغییر پذیری سود، وجوه نقد حاصل از فعالیت‌های عملیاتی، نسبت کیفیت سود، گردش مجموع دارایی‌ها، بازده فروش، بازده سرمایه‌گذاری و بازده حقوق صاحبان سهام مدلی را با استفاده از درخت تصمیم ایجاد نمودند که دقت ۷۴ درصدی در پیش‌بینی مدیریت سود داشت.

هدف این پژوهش معرفی متغیرهای موثر بر مدیریت سود و همچنین ارائه الگو و کشف مدیریت سود با استفاده از الگوریتم کلونی مورچه‌ها و غذایابی باکتری می‌باشد.

فرضیه‌های تحقیق

طبق نتایج مطالعات تسی و چیو (۲۰۰۹) نه تنها می‌توان با استفاده از مدل‌های ریاضی، مدیریت سود را کشف نمود، بلکه مدل‌های مبتنی بر رگرسیون خطی در کشف مدیریت سود در مقایسه با مدل‌های مبتنی یادگیری ماشین، دارای قدرت توضیح‌دهندگی کمتر (یا خطای

بیشتر) است. در این پژوهش با استفاده از دو الگوریتم کلونی مورچه‌ها و غذایابی باکتری اقدام به پیش‌بینی مدیریت سود گردیده است. لذا فرضیه‌های این تحقیق به شرح زیر طراحی و آزمون شده است:

فرضیه ۱: کشف مدیریت سود بر اساس الگوریتم کلونی مورچه‌ها، امکان‌پذیر است.

فرضیه ۲: کشف مدیریت سود بر اساس الگوریتم غذایابی باکتری، امکان‌پذیر است.

مروری بر پیشینه

سو کوچیپ و همکاران (۲۰۱۳) به بررسی اثر متغیرهای مرتبط با هیات‌مدیره بر مدیریت سود پرداخت. نتایج تحقیق رابطه معنی‌داری را بین متغیرهای اندازه هیات‌مدیره، جلسات هیات‌مدیره و دوگانگی نقش مدیرعامل با مدیریت سود نشان نمی‌دهد. سلیمان و رجب (۲۰۱۳) به بررسی عوامل موثر بر مدیریت سود در بورس اوراق بهادار مصر پرداختند. این تحقیق به بررسی اثر استقلال هیات‌مدیره، دوگانگی نقش مدیرعامل و اندازه هیات‌مدیره بر مدیریت سود پرداخته است. نتایج تحقیق رابطه معنی‌داری بین متغیرهای مزبور با مدیریت سود نشان نداده است. چی و همکاران (۲۰۱۴) در تحقیق خود به بررسی تأثیر مالکیت بلندمدت و کوتاه‌مدت صندوق‌های سرمایه‌گذاری با توجه به انواع مختلفی از مدیریت سود در چین پرداخته است. نتایج نشان می‌دهد که مدیریت سود در صندوق‌های سرمایه‌گذاری تحت کنترل دولت کمتر از شرکت‌های غیردولتی می‌باشد. فیلیپ و رافورنیر در سال ۲۰۱۴ به بررسی بحران مالی بر رفتار مدیریت سود شرکت‌های اروپایی پرداخته‌اند. نتایج نشان می‌دهد که مدیریت سود بطور قابل توجهی در سال‌های بحران مالی کاهش یافته است. اندرینی و یانگ در سال ۲۰۱۴ ارتباط وضعیت رقابت در بازارهای فروش محصول و اقلام تعهدی مورد بررسی قرار داده‌اند. نتایج نشان می‌دهد که شرکت‌هایی که در بازارهای با رقابت کمتر فعالیت دارند، تمایل بیشتر به مدیریت سود دارند. نجاری و همکاران (۲۰۱۴) در تحقیق خود اقدام به پیش‌بینی مدیریت سود نمودند. تحقیق مزبور بین سال‌های ۲۰۰۴ تا ۲۰۱۰ در بورس اوراق بهادار تهران صورت پذیرفته است. مدیریت سود در این تحقیق با استفاده از اقلام تعهدی اختیاری اندازه‌گیری شده است و جهت پیش‌بینی مدیریت سود نیز از الگوریتم ماشین بردار پشتیبان استفاده شده است. نتایج تحقیق بیانگر آن است که الگوریتم مذکور توانایی مناسبی جهت پیش‌بینی مدیریت سود دارد.

اخگر (۲۰۱۵) به بررسی رابطه بین اظهارنظر حسابرس، مدیریت سود و مدیریت سود واقعی پرداخت. در این تحقیق جهت بررسی فرضیه‌ها از ۲۸۱۸ سال-شرکت و از اقلام تعهدی جهت کمی نمودن مدیریت سود استفاده گردید. نتایج بیانگر رابطه‌ای معنی دار بین اظهارنظر حسابرس و مدیریت سود می‌باشد؛ لکن رابطه‌ای بین اظهارنظر حسابرس و مدیریت سود واقعی یافت نشده است.

حجازی و همکاران (۱۳۹۱) نشان دادند که روش شبکه عصبی و درخت تصمیم در پیش‌بینی مدیریت سود نسبت به روش‌های خطی دقیق‌تر و دارای سطح خطای کمتری است. ضمناً مدیریت سود با متغیرهای اقلام تعهدی اختیاری و غیر اختیاری دوره قبل و عملکرد شرکت، اندازه، تداوم سود در هر دو روش دارای بیشترین ارتباط است. رحمانی و اربابی در سال ۱۳۹۳ نشان دادند که تفاوت مالیات ابرازی و تشخیصی با مدیریت سود رابطه معناداری دارد و مدل‌های مختلف مدیریت سود در تبیین این رابطه تفاوت چندانی ندارند.

روش پژوهش

در تحقیق حاضر مدیریت سود با استفاده از الگوریتم‌های کلونی مورچه‌ها و غذایابی باکتری پیش بینی می‌شود. متغیر وابسته تحقیق حاضر مدیریت سود است. در این تحقیق، اقلام تعهدی اختیاری به عنوان نماینده مدیریت سود در نظر گرفته شده است. با توجه به این که براساس نتایج تحقیقات گذشته مدل تعدیل شده جونز (۱۹۹۱) قوی‌ترین مدل اندازه‌گیری اقلام تعهدی است (دیچاو و همکاران، ۱۹۹۵)، در تحقیق حاضر جهت تعیین اقلام تعهدی از این مدل استفاده شده است. در مدل مذکور، در مرحله اول می‌بایست کل اقلام تعهدی اندازه‌گیری شود. جمع اقلام تعهدی در تحقیق حاضر با استفاده از روش ترازنامه ای و به شرح مدل (۱) محاسبه گردیده است:

$$TA_t = (\Delta CA_t - \Delta Cash_t) - (\Delta CL_t - \Delta CPL_t) - DEP_t \quad \text{مدل (۱)}$$

که در این رابطه:

TA_t : جمع اقلام تعهدی در سال t

ΔCA_t : تغییر دارایی‌های جاری در سال t

$\Delta Cash_t$:	تغییر وجه نقد در سال t
ΔCL_t	:	تغییر بدهی جاری در سال t
ΔCPL_t	:	تغییر حصة جاری بدهی بلندمدت در سال t
DEP_t	:	هزینه استهلاک دارایی‌های ثابت در سال t

سپس اقلام تعهدی غیر اختیاری به تفکیک صنعت با استفاده از مدل (۲) محاسبه می‌گردد:

$$NDA_t = \alpha \left(\frac{1}{A_{t-1}} \right) + \beta_1 \left(\frac{\Delta REV_t - \Delta REC_t}{A_{t-1}} \right) + \beta_2 \left(\frac{PPE_t}{A_{t-1}} \right) \quad \text{مدل (۲)}$$

که در این رابطه:

NDA_t	اقلام تعهدی غیر اختیاری
A_{t-1}	کل دارایی‌ها در سال $t-1$
ΔREV_t	تغییر در درآمد سالانه (مبلغ درآمد سال t منهای مبلغ درآمد سال قبل)
ΔREC_t	تغییر در حساب‌های دریافتی (حساب‌های دریافتی سال t منهای حساب‌های دریافتی سال قبل)
PPE_t	اموال و ماشین‌آلات همان سال
α و β_1 و β_2	پارامترهای خاص شرکت می‌باشد که با استفاده از مدل (۳) بدست می‌آید:

$$\frac{TA_t}{A_{t-1}} = \alpha \left(\frac{1}{A_{t-1}} \right) + \beta_1 \left(\frac{\Delta REV_t}{A_{t-1}} \right) + \beta_2 \left(\frac{PPE_t}{A_{t-1}} \right) + \varepsilon \quad \text{مدل (۳)}$$

و در نهایت اقلام تعهدی اختیاری به عنوان نماینده مدیریت سود به مدل (۴) محاسبه می‌گردد:

$$DA_t = \frac{TA_t}{A_{t-1}} - NDA_t \quad \text{مدل (۴)}$$

که در این رابطه:

$$DA_t = \text{اقلام تعهدی اختیاری (نماینده مدیریت سود)}$$

متغیرهای مستقل تحقیق شامل نسبت‌های سودآوری، نسبت‌های فعالیت، نسبت‌های نقدینگی، نسبت‌های بدهی و متغیرهای کیفی می‌باشند. تعریف عملیاتی و نحوه برآورد و اندازه‌گیری هر یک از متغیرهای مذکور به شرح زیر ارائه شده است.

نگاره (۱): متغیرهای تحقیق

نام متغیر/روش محاسبه/منبع مورد استفاده
حاشیه سود ناخالص: سود ناخالص به فروش خالص (نجاری و همکاران، ۲۰۱۴)
بازده دارایی: سود خالص به جمع دارایی‌ها (مردای و همکاران، ۲۰۱۲)
بازده حقوق صاحبان سهام: سود خالص به حقوق صاحبان سهام (بارتون و سیمکو، ۲۰۰۲)
سود هر سهم: سود خالص به تعداد سهام (کردستانی و آشتاب، ۱۳۸۸)
هزینه‌های سیاسی: لگاریتم طبیعی تعداد کارکنان شرکت (نجاری و همکاران، ۲۰۱۴)
تغییرات سود: تفاوت سود دوره جاری نسبت به دوره قبل تقسیم بر جمع دارایی‌های دوره قبل (نجاری و همکاران، ۲۰۱۴)
نسبت گردش کل دارایی: درآمد حاصل از فروش به کل دارایی‌ها (چالاکی و یوسفی، ۱۳۹۱)
تغییرات حسابرس: اگر حسابرس نسبت به دوره قبل تغییر کرده باشد از متغیر مصنوعی یک و در غیر این صورت متغیر مصنوعی صفر استفاده می‌شود (نجاری و همکاران، ۲۰۱۴)
مالیات بردرآمد: هزینه‌های مالیات تقسیم بر سود قبل از مالیات (چالاکی و یوسفی، ۱۳۹۱)
نسبت کیفیت سود: جریانات نقدی ناشی از فعالیت‌های عملیاتی بر سود عملیاتی (چالاکی و یوسفی، ۱۳۹۱)
وجه نقد عملیاتی به دارایی‌ها: نسبت وجه نقد عملیاتی به جمع دارایی‌ها (نجاری و همکاران، ۲۰۱۴)
نسبت جاری: دارایی جاری به بدهی جاری (نجاری و همکاران، ۲۰۱۴)
بدهی به دارایی: نسبت جمع بدهی‌ها به جمع دارایی‌ها (مردای و همکاران، ۲۰۱۲)
بدهی به حقوق صاحبان سهام: نسبت بدهی‌ها به حقوق صاحبان سهام (نجاری و همکاران، ۲۰۱۴)
کیفیت افشا: امتیاز افشا گزارش شده توسط بورس اوراق بهادار (بداغی و بزاززاده، ۱۳۸۷)
نقش دوگانه مدیرعامل: متغیر مصنوعی است و در صورتی که مدیرعامل و رئیس هیات‌مدیره یک نفر باشند به آن عدد یک و در غیر این صورت عدد صفر اختصاص داده می‌شود (مردای و همکاران، ۲۰۱۲)
اعضای غیرموظف هیات‌مدیره: نسبت اعضای غیرموظف به کل اعضا (مردای و همکاران، ۲۰۱۲)
مالکان نهادی: سهام متعلق به سهامداران بالای ۵ درصد (پورتر، ۱۹۹۲)
وجود حسابرس داخلی: متغیر مصنوعی است و در صورت وجود حسابرس داخلی به آن عدد یک و در غیر این صورت عدد صفر اختصاص داده می‌شود (حبیب زاده، ۱۳۸۹)
نوع مالکیت: متغیر مصنوعی است و در صورتی که عمده سهام شرکت متعلق به بخش دولتی باشد به آن عدد یک و در غیر این صورت عدد صفر اختصاص داده می‌شود (نجاری و همکاران، ۲۰۱۴)

جامعه آماری این تحقیق شامل تمامی شرکت‌های پذیرفته شده در بورس اوراق بهادار تهران است که در دوره مورد بررسی تغییر دوره مالی نداشته باشند، شرکت‌های سرمایه‌گذاری، واسطه‌گری‌های مالی، بانک و لیزینگ نباشند و داده‌های مورد نظر آن‌ها در دسترس باشد. بر این اساس ۱۴۳ شرکت طی سال‌های ۱۳۸۸ تا ۱۳۹۲ به عنوان جامعه آماری انتخاب شده است.

تجزیه و تحلیل آماری

در این تحقیق جهت پیش‌بینی مدیریت سود از روش‌های غذایابی باکتری^۲ و الگوریتم کلونی مورچه‌ها^۳ استفاده شده است. پنج فرآیند در روش پیشنهادی وجود دارد که به ترتیب عبارت است از انتخاب داده‌ها، پاکسازی داده‌ها، تقسیم داده‌ها به مجموعه‌های آموزشی و ارزیابی، فرآیند آموزش مدل و ارزیابی مدل آموزش داده شده با داده‌های ارزیابی که تاکنون توسط الگوریتم‌ها مشاهده نشده است. مرحله اول، انتخاب داده‌ها است. در این مرحله داده‌های مورد نیاز که شامل مدیریت سود و متغیرهای ورودی لیست شده در نگاره (۱) است جمع‌آوری می‌گردد.

مرحله دوم، پاکسازی داده‌ها است. در این تحقیق جهت اجرای بخش دوم از رگرسیون گام به گام در انتخاب متغیرهای ورودی استفاده می‌گردد. انتخاب این متغیرها یکی از مهم‌ترین اهداف در پیش‌پردازش داده‌ها است. این مساله شامل فرآیند تعیین ورودی‌های مرتبط و حذف صفاتی که زائد بوده و اطلاعات اندکی فراهم می‌کنند، می‌باشد. انجام فرآیند انتخاب ورودی‌ها قبل از به کارگیری یک الگوریتم یادگیری مزایای فراوانی دارد. با حذف تعداد زیادی از ورودی‌های نامربوط، روش‌های یادگیری آموزشی، هزینه محاسباتی و زمان کمتری را متحمل می‌شوند. همچنین مدل به دست آمده ساده‌تر می‌شود که غالباً تفسیر آن آسان‌تر و در عمل مفیدتر و سودمندتر می‌باشد. همچنین مدل‌های ساده هنگامی که برای پیش‌بینی به کار می‌روند، دارای کلیت و عمومیت بهتری می‌باشند. بنابراین مدلی که دارای ورودی‌های کمتری است، مزیت‌های بیشتر و دقت بالاتری دارد. یکی از روش‌هایی که برای این منظور به کار می‌رود روش PSO می‌باشد. روش PSO در سال ۱۹۹۵ توسط دکتر ابرهارت و دکتر کندی ارائه شد و ایده اصلی آن از رفتار دسته جمعی ماهی‌ها یا پرندگان به هنگام جستجوی غذا الهام گرفته شده

است (دودا و همکاران، ۲۰۰۱). برای انتخاب مؤلفه‌های تأثیرگذار بر مدیریت سود از دو بردار به شرح زیر استفاده شده است:

بردار اول (S): یک بردار با مقادیر دودویی که مشخص می‌کند که آیا متغیر مستقل بر مدیریت سود تأثیرگذار است یا خیر.

بردار دوم (W): یک بردار با مقادیر واقعی که برای محاسبه وزن‌های معادله خطی ۱ استفاده می‌شود. این دو بردار در مدل (۶) و (۷) نشان داده شده‌اند (دودا و همکاران، ۲۰۰۱).

$$Z = S_1 W_1 X_1 + S_2 W_2 X_2 + \dots + S_n W_n X_n \quad \text{مدل (۵)}$$

$$S = \{S_1, \dots, S_n\}, S_i \in \{0, 1\}, i = 1, \dots, n \quad \text{مدل (۶)}$$

$$w = \{w_1, \dots, w_n\}, -10 \leq w_i \leq 10, i = 1, \dots, n \quad \text{مدل (۷)}$$

با در نظر گرفتن این که n تعداد کل است، مدل (۵) تابعی است که الگوریتم PSO سعی در یافتن ضرایب $w_i, i = 1, \dots, n$ دارد. برای وارد کردن این کدگذاری به PSO هر ذره PSO به صورت یک بردار دو قسمتی انتخاب کردیم. به این صورت که هر ذره دارای $2 * n$ قسمت است که n قسمت اول برای بردار X که برای تأثیرگذار بودن یا نبودن متغیر در پیش‌بینی مدیریت سود است و n قسمت دوم برای Z است که هر مقدار آن متناظر با یکی از X_i ها است. برای متغیرهایی که مقدار S_i آنها برابر صفر است، w_i تأثیری در مدل (۵) ندارد؛ به این معنی که متغیر مربوطه در پیش‌بینی مدیریت سود تأثیرگذار نیست. مقدار n در این مقاله عدد ۲۰ است؛ زیرا ۲۰ متغیر مستقل وجود دارد. در نگاره زیر خطاهای مختلف در پیش‌بینی مدیریت سود برای داده‌های آموزش، اعتبارسنجی و ارزیابی نشان داده شده است.

نگاره (۲): خطای پیش بینی

Train	Validation	Test	
۰/۰۰۹	۰/۰۱۲۸	۰/۰۱۲۲	MSE
۰/۵۰۸۱	۰/۷۱۲۸	۰/۶۹۶۹	SMAPE
۰/۴۵۳۹	۰/۳۵۷۶	۰/۴۸۶۸	R2

در نگاره زیر متغیرهای انتخابی و وزن هر کدام از متغیرها با دقت ۳ رقم اعشار برای مدل (۵) نشان داده شده است. الگوریتم پیشنهادی ۶ متغیر را از بین ۲۰ متغیر انتخاب کرده است.

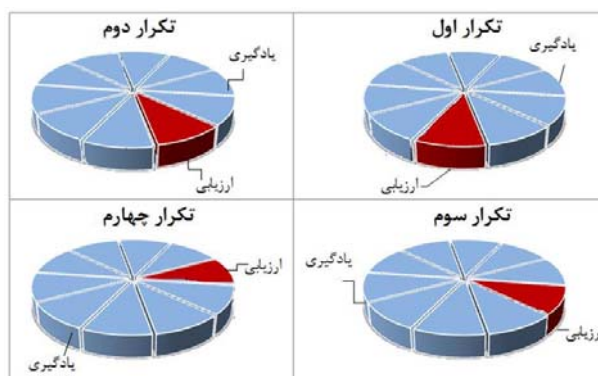
نگاره (۳): لیست متغیرهای مستقل انتخابی

وزن متغیر	نام متغیر	ردیف	وزن متغیر	نام متغیر	ردیف
۰/۰۰۰	وجه نقد عملیاتی به دارایی‌ها	۱۱	۰/۰۰۰	حاشیه سود ناخالص	۱
۰/۷۷۵	نسبت جاری	۱۲	-۰/۹۵	بازده دارایی	۲
۰/۰۰۰	بدهی به دارایی	۱۳	۰/۰۰۰	بازده حقوق صاحبان سهام	۳
۰/۰۰۰	بدهی به حقوق صاحبان سهام	۱۴	۰/۰۰۰	سود هر سهم	۴
۰/۰۰۰	کیفیت افشا	۱۵	۰/۰۰۰	هزینه‌های سیاسی	۵
۰/۰۰۰	نقش دوگانه مدیرعامل	۱۶	۱/۱۴۱	تغییرات سود	۶
۰/۰۰۰	اعضای غیر موظف هیات مدیره	۱۷	۰/۰۰۰	نسبت گردش کل دارایی	۷
-۰/۲۱۴	مالکان نهادی	۱۸	۰/۰۰۰	تغییرات حسابرس	۸
۰/۰۸۸	وجود حسابرس داخلی	۱۹	۰/۰۰۰	مالیات بردرآمد	۹
۰/۰۰۰	نوع مالکیت	۲۰	-۰/۴۷۴	نسبت کیفیت سود	۱۰

طبق نتایج نگاره فوق ۶ متغیر بازده دارایی، تغییرات سود، نسبت کیفیت سود، نسبت جاری، مالکان نهادی و وجود حسابرس داخلی جهت ورود به مدل داده کاوی تایید گردیده است. سایر متغیرها نیز به دلیل آنکه وزنی در مدل ندارند، وارد مدل داده کاوی نمی گردند.

مرحله سوم تقسیم داده‌ها می‌باشد. یکی از معیارهایی که برای ارزیابی یک تخمین گر مورد استفاده قرار می‌گیرد، نرخ خطا^۴ است که انواع مختلفی دارد. بطور کلی نمی‌توان با مقایسه خطای محاسبه شده روی داده‌های یادگیری، قضاوت مناسبی در خصوص توانایی‌های الگوریتم‌ها انجام داد. معمولاً نرخ خطا روی داده‌های یادگیری کمتر از نرخ خطا روی داده‌هایی است که در فرآیند یادگیری دیده نشده‌اند. با این استدلال، نمی‌توان از خطای یادگیری برای مقایسه دو الگوریتم استفاده نمود. دلیل این است که برای مدل‌های پیچیده‌تر، تخمین‌گرهایی که معمولاً دارای پارامترهای بیشتری هستند، دارای مرز پیچیده‌تری هستند. این مرز پیچیده باعث کاهش خطا بر روی داده‌های یادگیری در مقایسه با مدل‌های ساده‌تر می‌شود. بنابراین علاوه بر مجموعه داده‌های یادگیری، مجموعه‌ای از داده‌ها برای ارزیابی^۵ مورد نیاز است. از داده‌های آموزش برای یادگیری مدل و از داده‌های ارزیابی به منظور محاسبه نرخ خطای الگوریتم روی داده‌هایی که تا کنون مشاهده نکرده است، استفاده می‌شود. البته برای اینکه ارزیابی مناسب

باشد، یک بار اجرای الگوریتم کفایت نمی کند. معمولاً الگوریتم‌ها تمایل دارند که نرخ خطای تخمینی خود را به نرخ خطای واقعی نزدیک کنند و این امر با اجرای بارها و بارها فرآیند یادگیری و ارزیابی امکان پذیر است. بنابراین زمانی که یک مجموعه داده در اختیار گذاشته می شود، بایستی بخشی از آن را برای ارزیابی نهایی کنار گذاشت و از بقیه برای یادگیری استفاده کرد و مجدداً دو مجموعه را تغییر داده و دوباره مدل را ارزیابی کرد. یکی از روش های معمول برای این منظور روش اعتبارسنجی ده گانه^۶ نام دارد (آلبایدین، ۲۰۱۰). در این روش مجموعه داده‌ها به K قسمت مساوی، به صورت تصادفی تقسیم میگردد. R_X زوج مجموعه X_1, X_2, \dots, X_d به صورت تصادفی استخراج می شود که در آن X_i متغیرهای مستقل و γ_i متغیر وابسته نمونه λ م است. در اجرای اول قسمت اول از K قسمت به منظور ارزیابی، $K-1$ قسمت باقیمانده برای یادگیری استفاده می شود. در اجرای دوم قسمت دوم از K قسمت به منظور ارزیابی، $k-1$ قسمت باقیمانده برای یادگیری استفاده می شود. Y مرتبه الگوریتم به همین روال اجرا می گردد. مجموعه داده های یادگیری و ارزیابی باید به اندازه کافی بزرگ باشند تا خطای تخمینی، به مقدار واقعی نزدیک تر باشد. در عین حال داده های یادگیری و ارزیابی با داده های یادگیری و ارزیابی سایر تکرارها، باید کمترین همپوشانی را داشته باشند تا به این وسیله تمام داده ها در فرآیند یادگیری و ارزیابی دخالت داده شوند. در این روش دو نکته وجود دارد. نکته اول اینکه نسبت مجموعه ارزیابی به یادگیری کوچک است. همچنین هر چقدر مقدار N (تعداد کل نمونه های مجموعه داده ها) افزایش یابد می توان مقدار پارامتر K را کاهش داد و اگر مقدار N کوچک باشد، باید مقدار K را آنقدر بزرگ در نظر گرفت که تعداد نمونه های لازم برای عمل یادگیری فراهم باشد. چنانچه مقدار K برابر N در نظر گرفته شود، این روش به روش خارجی تبدیل می شود. در نمودار زیر چهار تکرار اول انتخاب مجموعه داده های یادگیری و ارزیابی روش اعتبارسنجی ده گانه نشان داده شده است (آلبایدین، ۲۰۱۰).



نمودار (۱): روش اعتبارسنجی ده گانه

در هر بار تکرار یک نرخ خطا برای داده‌های یادگیری و ارزیابی محاسبه می‌گردد و در نهایت میانگین نرخ‌های خطای بدست آمده به عنوان نرخ خطا داده‌های یادگیری و داده‌های ارزیابی انتساب داده می‌شود. برای ارزیابی مدل‌های پیش‌بینی از معیار ارزیابی با نام میانگین قدرمطلق خطا^(۸) (MSE) استفاده شده است که با استفاده از مدل (۸) محاسبه می‌گردند.

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - d_i)^2 \quad \text{مدل (۸)}$$

که در آن d_i مدیریت سود شرکت i ام، y_i مدیریت سود پیش‌بینی شده توسط الگوریتم هوشمند، n تعداد داده‌های مجموعه مورد بررسی (آموزش، اعتبارسنجی، ارزیابی) است. $y_i - d_i$ میزان خطای پیش‌بینی را برای شرکت i ام نشان می‌دهد. هر چه MSE نزدیک‌تر به صفر باشد، پیش‌بینی الگوریتم‌ها به واقعیت نزدیک‌تر است (آلپایدین، ۲۰۱۰).

مرحله چهارم، فرآیند آموزش و ارزیابی مدل‌ها در الگوریتم‌های تحقیق است. پس از تقسیم نمونه‌ها به دو دسته داده‌های یادگیری و ارزیابی، با استفاده از داده‌های آموزشی، مدل آموزش ایجاد می‌شود. برای حل مسئله ابتدا به معرفی مدل آن پرداخته می‌شود. رابطه زیر تابعی است که الگوریتم تحقیق سعی در یافتن ضرایب $b_i, i=1, \dots, m$ خواهد داشت.

$$z = b_0 + b_1x_1 + b_2x_2 + \dots + b_mx_m \quad \text{مدل (۹)}$$

که در آن b_0 عرض از مبدا و $b_i, i=1, \dots, m$ وزن‌های هر کدام از متغیرها (X_i) (ضرایب تخمین‌گر) است. X_i مقدار متغیر مستقل i ام است. b_i ها توسط الگوریتم با استفاده از داده‌های آموزشی محاسبه می‌گردند که اصطلاحاً به آن آموزش مدل گفته می‌شود و سپس با استفاده از داده‌های ارزیابی، به ارزیابی مدل پرداخته می‌شود. یعنی پس از محاسبه b_i ها، داده‌های ارزیابی به رابطه بالا وارد شده و مقدار MSE محاسبه می‌گردد (آلبایدین، ۲۰۱۰).

الگوریتم کلونی مورچگان

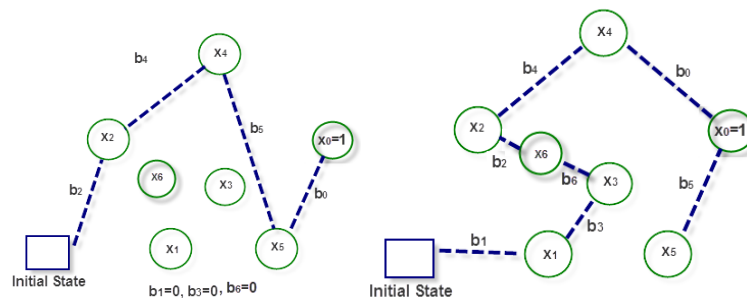
برای حل مسئله ابتدا به معرفی مدل این الگوریتم پرداخته می‌شود. مدل (۱۰) تابعی است که الگوریتم کلونی مورچگان سعی در یافتن ضرایب $b_i, i = 1, \dots, m$ خواهد داشت وقتی که $m=7$ باشد.

$$z = b_0 + b_1x_1 + b_2x_2 + \dots + b_mx_m \quad \text{مدل (۱۰)}$$

که در آن b_0 عرض از مبدا و $b_i, i = 1, \dots, m$ وزن‌های هر کدام از متغیر (X_i) (ضرایب تخمین‌گر) است. X_i مقدار متغیر مستقل i ام است. b_i ها توسط الگوریتم ACO با استفاده از داده‌های آموزشی محاسبه می‌گردند که اصطلاحاً به آن آموزش مدل گفته می‌شود و سپس با استفاده از داده‌های ارزیابی، به ارزیابی مدل پرداخته می‌شود. یعنی پس از محاسبه b_i ها، داده‌های ارزیابی به مدل (۹) وارد شده و مقدار MSE محاسبه می‌گردد (ادلستین، ۱۹۹۹).

در ادامه به روند یافتن پارامترهای مدل (۱۰) با استفاده از الگوریتم ACO و داده‌های آموزشی پرداخته می‌شود. هنگامی که مورچه‌ها بسوی منابع غذایی یا برعکس از منابع غذایی بسوی لانه حرکت می‌کنند ماده‌ای بنام فرومون روی زمین ترشح می‌کنند. مورچه‌ها می‌توانند فرومون را بچشند و وقتی می‌خواهند راه خود را انتخاب کنند، احتمالاً راهی را انتخاب می‌کنند که دارای غلظت فرومون زیادتری است. هرچه غلظت فرومون بیشتر باشد علاقه مورچه به طی این مسیر بیشتر می‌شود. فرومون در اثر گذشت زمان تبخیر می‌شود و در نتیجه در مسیرهای که زیاد طی نشده‌اند، فرومون کمتری انباشته می‌شود. با گذشت زمان کوتاهترین مسیر با حرکت مورچه‌ها بدست می‌آید، این مسیر تقویت شده و مسیرهای دیگر تضعیف می‌شوند تا همه مورچه‌ها از یک مسیر کوتاه رفت و آمد کنند (ادلستین، ۱۹۹۹).

مسئله یافتن پارامترها به صورت مناسب می‌تواند در یک مساله ACO فرمول‌بندی شود. در گراف ACO، گره‌ها و یال‌های بین آنها به ترتیب نشان‌دهنده متغیرهای مستقل و انتخاب متغیرهای مستقل توسط مورچه و پارامتر (b_i) مربوط به متغیر مستقل است. گراف توسط یک مورچه برای پیدا کردن یک زیرمجموعه بهینه از متغیرهای مستقل تا جایی پیمایش می‌شود که تعداد حداقلی از گره‌ها ملاقات شوند. پیمایش یک مورچه زمانی متوقف می‌شود که به یک معیار توقف مشخص برسد. نمودار (۲) نشان‌دهنده مسیر انتخاب شده دو مورچه است. متغیرهای مستقل توسط مورچه و پارامتر مربوط به متغیر مستقل (میزان فرمون) توسط مورچه را نشان می‌دهد. برای سازگار کردن مدل با مدل، مورچه‌ها ابتدا از یک گره مجازی حرکت می‌کنند. شش گره که پنج تای آنها بیانگر متغیرهای مستقل و یکی بیانگر عرض از مبدا که مقدار آن همواره یک است $(x_0=1)$ ، می‌باشد. حال با این نوع بیان مسئله می‌توان آن را با الگوریتم ACO حل نمود (ادلستین، ۱۹۹۹).



نمودار (۲): دو نمونه از حرکت یک مورچه.

تمایل اکتشافی حرکت مورچه‌ها و میزان فرومن یال‌ها با هم‌دیگر قانون احتمال انتقال را تشکیل می‌دهند که در مدل (۱۱) نشان داده شده است: (ادلستین، ۱۹۹۹).

$$p_{i,j}^k(t) = \begin{cases} \frac{[\tau_{i,j}(t)]^\alpha [\eta_{i,j}(t)]^\beta}{\sum_{l \in J_i^k} [\tau_{i,l}(t)]^\alpha [\eta_{i,l}(t)]^\beta} & \text{مدل (۱۱)} \\ 0 & \text{otherwise} \end{cases}$$

که در آن $\eta_{i,j}$ یک معیار اکتشافی، $\tau_{i,j}$ میزان فرومن بین گره i ام و j ام، k شماره مورچه است و t زمان (تکرار) است. α و β به ترتیب وزن فرومن و معیار اکتشافی هستند که در این

تحقیق معادل عدد ۱.۰ در نظر گرفته شده است. J_i^k همسایگان گره i ام برای مورچه k ام را نشان می دهد که در اینجا نشان دهنده متغیرهای مستقلی است که هنوز انتخاب نشده اند. ابتدا مورچه ها در حالت اولیه قرار می گیرند. در هر گام از پیمایش، مورچه k ام قانون احتمال انتقال را به کار می برد. احتمال اینکه مورچه k ام در متغیر مستقل i ام (گره i ام) باشد، متغیر مستقل j ام (گره j ام) را در تکرار t ام انتخاب کرده باشد از مدل (۱۲) محاسبه می گردد.

بروز شدن فرومن بر اساس رابطه زیر انجام می شود.

$$\tau_{i,j} = \rho \cdot \tau_{i,j} + \sum_{k=1}^n \Delta \tau_{i,j}^k(t) \quad \text{مدل (۱۲)}$$

که در آن n تعداد مورچه ها و $0 < \rho < 1$ میزان تبخیر فرومن را نشان می دهد. همچنین پارامتر ρ برای جلوگیری از تراکم بیش از حد فرومن بکار برده می شود که در این تحقیق معادل عدد ۰.۲ در نظر گرفته شد و الگوریتم را قادر می سازد که تصمیمات اشتباهی که قبلاً گرفته شده است، فراموش شوند. اگر یک یال بوسیله مورچه ها انتخاب نشود، غلظت فرومن آن بطور تدریجی کاهش پیدا میکند. $\Delta \tau_{i,j}^k(t)$ مقدار فرومنی است که مورچه k ام در زمان t بر روی یال هایی که ملاقات کرده است، اضافه می کند و مقدار آن از مدل (۱۳) بدست می آید (ادلستین، ۱۹۹۹).

$$\Delta \tau_{i,j}^k(t) = \begin{cases} Q, & \text{if the edge } (i,j) \text{ is chosen by ant}_k \\ L_k(t), & \\ 0, & \text{otherwise} \end{cases} \quad \text{مدل (۱۳)}$$

که در آن Q یک ثابت برای بروزرسانی فرومن است که معمولاً یک در نظر گرفته می شود و $L_k(t)$ میزان خطای تخمینگر MSE است. در هنگام محاسبه خطا چون همواره میزان فرمون ها مثبت است و ضرایب b_i ها می توانند منفی نیز باشند، از فرمون ها مقدار ثابت ۳۰ کم گردید. این عدد با سعی و خطا بدست آمده است. در اینجا مجموعه داده ها به دو دسته داده های آموزشی و ارزیابی تقسیم شدند. داده های آموزشی برای یادگیری (بدست آوردن پارامترهای b رابطه مدل (۱۰) با استفاده از ACO استفاده می گردد. با استفاده از $\tau_{i,j}^k(t)$ مقدار فرومن ترشح شده بین گره ها برای مورچه k ام در زمان t ، $L_k(t)$ با استفاده از رابطه ۴-۲ برای داده های آموزشی محاسبه می گردد. لازم به ذکر است که $\tau_{i,j}^k(t) = b_j$ مربوط به مورچه k ام در لحظه t ، b ها در مدل (۱۰) است.

الگوریتم غذایابی باکتری:

ایده الگوریتم غذایابی باکتری بر این واقعیت استوار است که در طبیعت، جانداران با روش غذایابی ضعیف احتمال انقراض بیشتری نسبت به جاندارانی با استراتژی غذایابی موفق دارند. پس از نسل‌های زیاد، جانداران با روش غذایابی ضعیف نابود شده و یا به حالت‌های بهتر تغییر شکل می‌دهند. باکتری E-coil که در روده انسان زندگی می‌کند، روش غذایابی دارد که بر چهار مرحله استوار است. این مرحله عبارتند از: حرکت، عملکرد گروهی، تولید مثل و حذف پراگندگی (بری و لینوف، ۱۹۹۷).

حرکت: در این مرحله باکتری‌ها شروع به جنبش و شنا می‌کنند. در واقع بسته به چرخش دم باکتری، باکتری جست و خیز کرده و شروع به حرکت می‌کند (جنبش). اگر در مسیر جدید مقدار غذا بهتر بود، باکتری شروع به حرکت در همان مسیر می‌کند (شنا). فرض کنید می‌خواهیم مقدار حداقل $J(\theta)$ ، $\theta \in R^p$ را پیدا کنیم که در اینجا تابع هدف پرتفوی است. θ مکان باکتری (یا همان x_i ها) و $J(\theta)$ نشان‌دهنده مقدار غذا (مقدار تابع هدف) در مکان θ است. $J(\theta) > 0$ ، $J(\theta) = 0$ ، $J(\theta) < 0$ به ترتیب به این معنی می‌باشد که باکتری در مکان θ دارای غذای خوب، خنثی و بد است. برای انجام جنبش، یک بردار با طول واحد به نام $\Phi(i)$ تولید می‌شود. این بردار برای تعریف جهت جدید حرکت باکتری بعد از انجام جنبش، به کار می‌رود. مکان جدید باکتری به صورت زیر تعریف می‌شود.

$$\theta^i(j, k, 1) = \theta^i(j + 1, k, 1) + c(i)\Phi(i)$$

باکتری i ام در مرحله j ام، تولید مثل k ام و نابودی و پخش i ام می‌باشد. $C(i)$ اندازه حرکت باکتری در جهت حرکت $\Phi(i)$ می‌باشد. اگر اندازه $J(i, j, k, 1)$ در $\theta^i(j + 1, k, 1)$ کمتر از اندازه آن در $\theta^i(j, k, 1)$ باشد، آنگاه یک گام حرکت دیگر به اندازه $C(i)$ در جهت $\Phi(i)$ انجام می‌شود و باکتری شروع به شنا کردن در جهت $\Phi(i)$ می‌کند. این شنا کردن تا زمانی که اندازه $J(\theta)$ کاهش می‌یابد و حداکثر تا ماکزیمم تعداد مراحل مجاز شنا کردن، N ادامه می‌یابد و نشان می‌دهد که باکتری تا زمانی که در جهت حرکت خود محیط بهتری از لحاظ غذا بیابد، به حرکت در همان جهت ادامه خواهد داد (بری و لینوف، ۱۹۹۷).

عملکرد گروهی: وقتی که یک باکتری مسیر بهتری برای غذا پیدا می کند، باکتری های دیگر را به سمت خود جذب کرده و باکتری ها سریع تر به محل غذای اصلی می رسند. عملکرد دسته جمعی سبب حرکت گروهی باکتری ها به سمت غذا می شود. اگر $P(j, k, 1) = \{\theta^i(j, k, 1) | i = 1, 2, \dots, S\}$ را مجموعه مکان های باکتری فرض کنیم، عملکرد دسته جمعی به صورت مدل (۱۴) می باشد.

$$J_{cc}(\theta, P(i, j, l)) = \sum_{i=1}^S J_{cc}^i(\theta, \theta^i(i, j, l)) = \sum_{i=1}^S [-d_{attact} \exp(-\omega_{attact} \sum_{m=l}^P (\theta_m - \theta_m^i)^2)] + \sum_{i=1}^S [-h_{repelen} \exp(-\omega_{repelen} \sum_{m=l}^P (\theta_m - \theta_m^i)^2)] \quad \text{مدل (۱۴)}$$

که $J_{cc}(\theta, P(i, j, l))$ بسته به حرکت همه باکتری ها، تابعی وابسته به زمان بوده و به مقدار تابع هزینه، $J(i, j, k, l)$ افزوده می شود. بنابراین، باکتری ها شروع به تلاش برای پیدا کردن غذا نموده، از مکان های بی غذا فرار کرده و در همان حین یکدیگر را جذب می کنند و در عین حال بیش از حد به هم نزدیک نمی شوند. "S" تعداد کل باکتری ها بوده و "P" تعداد پارامترهایی است که باید بهینه شوند و به عنوان مختصات مکان باکتری در فضای p بعدی محسوب می شوند. ω_{attact} ، d_{attact} ، $\omega_{repelen}$ ، $h_{repelen}$ ضرایبی هستند که باید مقدار مناسبی برای آنها بسته به مسئله مورد نظر انتخاب شود (بری و لینوف، ۱۹۹۷).

تولید مثل: نصف تعداد باکتری ها که غذای خوبی پیدا نکرده اند، نابود شده و نصف دیگر شامل باکتری های سالم هر یک به دو باکتری تقسیم شده که در همان مکان قبلی باکتری قرار می گیرند. این عمل، تعداد جمعیت باکتری ها را ثابت نگه می دارد. (بری و لینوف، ۱۹۹۷).

حذف و پراکندگی: زندگی جمعیت باکتری ها به مرور با مصرف غذا و یا ناگهان در اثر موارد دیگر دچار تغییر می شود. حوادث می توانند موجب کشته شدن و یا پراکنده شدن باکتری ها شوند. این عمل اگر چه در ابتدا ممکن است منجر به برهم خوردن مرحله حرکت به سمت غذا باشد، اما می تواند تأثیر مثبتی هم بر آن داشته باشد. زیرا پراکندگی باکتری ها ممکن است آن ها را در مکان هایی نزدیک به منابع غذایی خوب قرار دهد. مرحله حذف و پراکندگی از به دام افتادن باکتری ها در نقطه بهینه محلی جلوگیری می کند. در هر مرحله حذف و پراکندگی، هر باکتری موجود در جمعیت با احتمال P_{cd} در معرض حذف و پراکندگی قرار می گیرد. برای

ثابت نگه داشتن تعداد باکتری‌ها، اگر یک باکتری نابود شود، باکتری جدیدی را به صورت رندم در محدوده فضای جستجو قرار می‌دهیم. در مسأله برنامه‌ریزی تابع سعی در کاهش مقدار این حاصل جمع یا به عبارت دیگر سعی در افزایش تعداد صفرها در ضرایب دارد. پس از آموزش مدل با استفاده از داده‌های آموزشی ضرایب ذخیره می‌گردد تا در مرحله ارزیابی از آن استفاده گردد (تیبشیرانی، ۱۹۹۶). در مرحله آخر مدل با توجه به طی کردن مراحل قبل مورد ارزیابی قرار خواهد گرفت و نتایج ارائه می‌گردد. ارزیابی مدل، آخرین مرحله است. نتایج حاصل از اعمال همه داده‌ها به صورت زیر است:

تکراه (۴): نتایج تحقیق

الگوریتم کلونی مورچه‌ها		الگوریتم غذایی باکتری		Fold
خطای داده ارزیابی	خطای داده یادگیری	خطای داده ارزیابی	خطای داده یادگیری	
۰/۹۲۴	۱/۰۲۵	۱/۰۲۵	۱/۰۲۵	۱
۰/۸۹۵	۰/۹۰۵	۱/۱۲۵	۱/۱۱۵	۲
۱/۰۰۲	۱/۰۱۲	۱/۳۵۲	۱/۶۵۱	۳
۰/۹۸۱	۰/۸۱۲	۰/۹۸۶	۱/۰۵۲	۴
۱/۰۹۱	۱/۰۸۸	۱/۲۱۱	۱/۲۳۱	۵
۱/۱۲۱	۱/۰۱۴	۱/۰۲۴	۱/۰۲۹	۶
۱/۰۲۵	۰/۹۶۳	۱/۰۳۶	۱/۲۱۴	۷
۰/۷۸۲	۰/۸۹۲	۰/۹۱۲	۰/۹۸۹	۸
۰/۸۹۱	۰/۹۸۸	۱/۲۵۲	۱/۴۲۲	۹
۱/۱۴	۱/۰۱۲	۱/۰۰۴	۱/۱۹۱	۱۰
۰/۹۸۵	۰/۹۷۱	۱/۱۱۳	۱/۱۹۲	میانگین

نتیجه‌گیری

مدیریت سود زمانی رخ می‌دهد که مدیر برای گزارشگری مالی از قضاوت شخصی خود استفاده می‌کند. ارقام تعهدی سود از یک سو به مدیران اجازه می‌دهند تا سود را طوری محاسبه کنند که گویای ارزش واقعی بنگاه اقتصادی باشد و از سوی دیگر این ارقام به مدیران اختیار می‌دهند تا از انعطاف پذیری روش‌های و اصول پذیرفته شده حسابداری استفاده کرده و محتوای اطلاعاتی سود را مخدوش کنند. هدف این مطالعه ارزیابی این موضوع بود که آیا می‌توان

مدیریت سود را بر اساس روش‌های مبتنی بر یادگیری ماشین کشف کرد. این تحقیق که با استفاده از اطلاعات مالی ۱۴۳ شرکت بین سال‌های ۱۳۸۸ الی ۱۳۹۲ انجام شده است، نشان می‌دهد که بین متغیرهای توان بازپرداخت بهره، نسبت سرمایه در گردش، گردش دارایی ثابت، مالکان نهادی، نسبت آتی، وجود حسابرس داخلی، نسبت گردش موجودی‌ها و بازده دارایی با مدیریت سود، همبستگی وجود دارد. همچنین در این تحقیق با استفاده از الگوریتم غذایابی باکتری و کلونی مورچه‌ها اقدام به پیش‌بینی مدیریت سود شده است که نتیجه پژوهش نشان می‌دهد هردو الگوریتم توانایی بالایی (بیش از ۹۸٪) جهت پیش‌بینی مدیریت سود دارند. این نتایج با نتیجه پژوهش هاگلوند (۲۰۱۲) تسی و چپو (۲۰۰۹) سازگار است.

این تحقیق با محدودیتهایی نیز همراه بوده است. از جمله این که نتایج حاصل از تحقیق حاضر فقط قابل تعمیم به شرکت‌های پذیرفته شده در بورس است. لذا تعمیم نتایج به سایر شرکت‌ها باید با احتیاط صورت پذیرد. با توجه به نتایج تحقیق پیشنهادهای زیر ارائه می‌شود:

۱. با توجه به دقت بالاتر از ۹۹ درصد الگوریتم کلونی مورچه‌ها به سرمایه‌گذاران و سایر استفاده‌کنندگان از صورت‌های مالی پیشنهاد می‌گردد از این الگوریتم جهت پیش‌بینی مدیریت سود در شرکتها استفاده نمایند.

۲. با توجه به نتایج تحقیق مبنی بر این که مشابه الگوریتم کلونی مورچه‌ها، الگوریتم غذایابی باکتری نیز دقت بالایی (بالاتر از ۹۸ درصد) در پیش‌بینی مدیریت سود دارد، به سرمایه‌گذاران و استفاده‌کنندگان از صورت‌های مالی پیشنهاد می‌گردد به کاربرد الگوریتم کلونی مورچه‌ها در پیش‌بینی مدیریت سود توجه نمایند.

همچنین برای تحقیق‌های آتی پیشنهاد می‌گردد:

۱. با توجه به نتایج تحقیق حاضر، مدلی ترکیبی از الگوریتم‌های کلونی مورچه‌ها و غذایابی باکتری طراحی گردد تا در افزایش دقت پیش‌بینی موثر باشد.

۲. با استفاده از سایر تکنیک‌های گروه هوش مصنوعی از قبیل الگوریتم کلونی زنبورها و آدابوست نیز مدلی برای پیش‌بینی مدیریت سود ارائه و نتایج آن با نتایج این تحقیق مقایسه شود.

پی‌نوشت‌ها

- | | |
|-----------------------|---------------------------|
| ۱ Jones | ۲ Bacterial Foraging |
| ۳ Adabost | ۴ Error Rate |
| ۵ Test | ۶ K-Fold Cross Validation |
| ۷ Mean Absolute Error | ۸ Most Correlated |
| ۹ Convex | |

منابع

- بداغی، حمید؛ بزاززاده، حمیدرضا. (۱۳۸۷). رابطه بین مدیریت سود و کیفیت افشا. فصلنامه مطالعات حسابداری، شماره ۱۷: ۱۷۳-۱۹۸.
- پورزمانی، زهرا؛ اولی، محمدرضا و عبدالهیان، جواد. (۱۳۹۳). بررسی تاثیر اجرای آئین نامه نظام راهبری شرکتی سازمان بورس اوراق بهادار تهران بر مدیریت سود. حسابداری مدیریت، دوره ۷، شماره ۲۰: ۱۲-۱.
- چالاک، پری؛ یوسفی، مرتضی. (۱۳۹۱). پیش بینی مدیریت سود با استفاده از درخت تصمیم گیری. *مطالعات حسابداری و حسابرسی*، سال اول، شماره ۱: ۱۱۰-۱۲۳.
- حبیب زاده بایگی، سید جواد. (۱۳۸۹). بررسی ارتباط بین ویژگی‌های هیات مدیره و مدیریت سود در شرکت‌های پذیرفته شده در بورس اوراق بهادار تهران. پایان نامه کارشناسی ارشد دانشگاه آزاد مشهد.
- حجازی، رضوان؛ محمدی، شاپور؛ اصلانی، زهرا؛ آفاجانی، مجید. (۱۳۹۱). پیش بینی مدیریت سود با استفاده از شبکه عصبی و درخت تصمیم در شرکت‌های پذیرفته شده در بورس اوراق بهادار تهران. *بررسی‌های حسابداری و حسابرسی*، دوره ۱۹، شماره ۶۸: ۳۱-۴۶.
- خانی، عبدالله. (۱۳۸۲). مدیریت سود و پاداش مدیران، فصل نامه مطالعات حسابداری، شماره ۳: ۱۲۸-۱۵۳.
- رحمانی، علی؛ اربابی بهار، زهرا. (۱۳۹۳). رابطه‌ی تفاوت مالیات تشخیصی و ابرازی با مدیریت سود. *فصلنامه پژوهش‌های تجربی حسابداری*، ۳ (۴): ۶۱-۸۳.
- طغرابی، زهره. (۱۳۹۲). بررسی اثرات هم خطی در مدل‌های رگرسیونی چندگانه. نشریه ندا، سال پنجم، شماره ۱: ۳۱-۳۹.
- کردستانی، غلام رضا؛ آشتاب، علی. (۱۳۸۸). پیش بینی مدیریت سود بر مبنای تعدیل سود هر سهم. *مجله توسعه و سرمایه*، شماره ۴: ۱۴۱-۱۵۸.

- Akhgar, M. (2015). Qualified Audit Opinion, Accounting Earnings Management and Real Earnings Management. *Asian Economic and Financial Review*, 5 (1): 46- 57.
- Alpaydin, E. (2010). Introduction to machine learning, 2nd ed.: Cambridge, Mass.: MIT Press.
- Barton, J. J. and P. J. Simko. (2002). The balance sheet as an earnings management constraint. *The Accounting Review* (Supplement): 1-27.
- Berry, M and G. Linoff. (1997). Data Mining Techniques for marketing, sales, and customer support. New York.
- Black, E. L, K. F. Sellers and T. S. Manly (1998). Earning management using asses sales. *Journal of Business Finance & Accounting*, 25 (9): 25-38.
- Chi, J; Yang, J; Young, M. (2014). Mutual funds' holdings and listed firms' earnings management: Evidence from China. *Journal of Multinational Financial Management*, 6: 62- 78.
- Chung, R. , Firth, M. and Kim, J. B. (2005). Earnings Management, Surplus Free Cash Flow, and External Monitoring, *Journal of Business Research*, 58: 766– 776.
- Dechow, P. M. , R. G. Sloan and Sweeney, A. P. (1995). Detecting Earnings Management. *Accounting Review*, vol. 70, n 2: 193- 225.
- Dechow, P. M. , And Dichev, I. D. (2002). The quality of accruals and earnings: The role of accrual estimation errors. *The Accounting Review* (Supplement): 35.
- Dechow, P. M. , and Skinner, D. J. (2000). Earnings management: Reconciling the views of accounting academics, practitioners and regulators. *Accounting Horizons*, 14 (2): 235-250.
- Duda, R. O. Hart, P. E. and Stork, D. G. (2001). Pattern classification vol. 2: wiley New York.
- Edelstein, H. (1999). Introduction to data mining and knowledge discovery, 3rd ed. Potomac, MD, USA: Two Crows Corporation.
- Indrarani, L; Yang, Y (2014). Product market competition and earnings management: Evidence from discretionary accruals and real activities manipulation. *Advances in Accounting*, 6: 263- 275.
- Fayyad. U. , Uthwrsusamy. R. (1996). Data Mining and Knowledge Discovery in Databases. Communication of ACM.
- Filip, A, Raffournier, B. (2014). Financial Crisis And Earnings Management: The European Evidenc. *The International Journal of Accounting*, 6: 455-478.
- Frankel, R. M. , Johnson, M. F. and Nelson, K. K. (2002). The relation between auditor's fees for nonaudit services and earnings management. *The Accounting Review* (Supplement): 71-106.

- Johari , Nor Hashimah ; Mohd Saleh , Norman ; Jaffar , Romlah ; Sabri Hassan , Mohamat. (2008). The Influence of Board Independence, Competency and Ownership on Earnings Management in Malaysia. *Journal of Economics and Management* 2 (2) , 281– 306.
- Healy, P. M. , and Wahlen, J. M. (1999). A review of the earnings management literature and its implications for standard setting. *Accounting Horizons*, 13 (4): 365- 383.
- Matsumoto, D. A. (2002). Management’s incentives to avoid negative earnings surprises. *The Accounting Review*, 77 (July): 483– 514.
- McNichols, M. (2000). Research design issues in earnings management studies. *Journal of Accounting and Public Policy* 19 (4-5): 313- 345.
- Moradi, M. , Salehi, M. , Habibzadeh Baygi, S. J. & Najari, M. (2012). A Study of Relationship between Board Characteristics and Earning Management: Iranian Scenario, *Universal Journal of Management and Social Sciences*, 2 (3): 12- 29.
- Najari, M. , Hazrati, A. , Rezaie, P. , Habibzadeh Baygi, J. (2014). Forecasting of Earning Management by Support Vector Machine: Case Study in Tehran Exchange Stock. *Middle-East Journal of Scientific Research* 19 (7): 1007- 1017.
- Rafik, Z. A. (2002). Determinants of earnings management ethics among accountants. *Journal of Business Ethics*, 40: 33- 45.
- Ronen, A. and Sadan, S. (1981). Smoothing income numbers, Objectives, Means and Implication. reading, MA, Addition Wesley.
- Petroni, k. (1992). optimistic reporting in the property- casualty insurance industry. *Journal of Accounting and Economics*, 15 (december): 485- 509.
- Schipper, K. (1989). Earnings management. *Accounting Horizon*, 3 (4) , 91- 102.
- Soliman, M. M. , and Ragab, A. A. (2013). Board of director’s attributes and earning management: Evidence from Egypt. *Journal of Business and Social Sciences Research*, 6: 52- 63.
- Sukeecheep, S. , Yarram, S. R. , and Al Farooq, O. (2013). Earnings management and board characteristics in Thai Listed Companies. *Journal of International Conference on Business, Economics and Accounting*, 4 (6): 74- 93.
- Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society, Series B* 58: 267–288.
- Tsai, C. Chiou F. (2009). Earnings Management Prediction: A Pilot Study of Combining Neural Networks and Decision Trees. *Expert Systems with Applications*, 36: 7183-7191