



ارائه مدل ارزیابی عملکرد بانک‌های بوری کشور با رهیافت داده کاوی

الهام آدخ^۱

عارفه فدوی اصغری^۲

محمد ابراهیم محمد پورزرندی^۳

تاریخ دریافت مقاله: ۹۸/۰۷/۱۴ تاریخ پذیرش مقاله: ۹۸/۰۹/۱۱

چکیده

با رشد بانک‌های خصوصی و مؤسسات مالی و اعتباری، رقابت بین آنها به منظور ارائه خدمات بهتر افزایش یافته است. با توجه به اهمیت موضوع، تدوین یک الگوی جامع و کامل برای ارزیابی بانک‌ها ضروری به نظر می‌رسد. هر سازمان برای آگاهی از نقاط قوت و ضعف خود به ویژه در محیط‌های پویا نیازمند ارزیابی عملکرد می‌باشد. موضوع ارزیابی عملکرد آنقدر مورد توجه است که حتی در این باره صاحب‌نظران مدیریت معتقدند: "آنچه را که نتوان ارزیابی نمود، نمی‌توان مدیریت کرد". بنابراین بانک‌ها نیز همانند سایر سازمان‌ها در ایران برای ارائه خدمات متنوع‌تر و سریع‌تر و همچنین توسعه خود، نیازمند ارزیابی عملکرد می‌باشند. [6]

هدف اصلی این مقاله ارائه مدلی به منظور ارزیابی عملکرد بانک‌های بوری کشور با استفاده از روش‌های داده کاوی می‌باشد. در این پژوهش، ۴ مدل داده کاوی درخت تصمیم C5.0، درخت تصمیم C4.5، الگوریتم نایو بیز و جنگل تصادفی، به منظور ارزیابی عملکرد بانک‌های پذیرفته شده در بورس اوراق بهادار تهران پیاده‌سازی و با یکدیگر مقایسه می‌گردند. بدین منظور نمونه‌ای مشتمل بر ۲۸ نسبت مالی شامل نسبت‌های سودآوری، نقدینگی، حوزه کیفیت مدیریت، حوزه کیفیت دارایی و کفایت سرمایه در ۱۸ بانک پذیرفته شده در بورس و فرابورس اوراق بهادار تهران در فاصله بین سالهای ۱۳۹۳ تا ۱۳۹۶ به عنوان متغیرهای مستقل و عملکرد بانک‌ها در سه دسته قابل قبول، غیرقابل قبول و متوسط به عنوان متغیر وابسته انتخاب گردیدند. نتایج حاکی از آن است که در بین کلیه مدل‌ها درخت تصمیم C5.0 با صحت ۹۴٫۴٪ بهترین مدل ارائه شده در این پژوهش می‌باشد.

کلمات کلیدی

ارزیابی عملکرد، داده کاوی، نسبت‌های مالی، بانک‌های پذیرفته شده در بورس اوراق بهادار

۱- گروه مدیریت صنعتی دانشکده مدیریت، واحد تهران مرکز، دانشگاه آزاد اسلامی، تهران، ایران. eladel20081@gmail.com

۲- گروه مدیریت صنعتی دانشکده مدیریت، واحد تهران مرکز، دانشگاه آزاد اسلامی، تهران، ایران. (نویسنده مسئول) arefehfadaviasghari@gmail.com

۳- گروه مدیریت صنعتی دانشکده مدیریت، واحد تهران مرکز، دانشگاه آزاد اسلامی، تهران، ایران. pourzarandi@yahoo.com

ارائه مدل ارزیابی عملکرد بانک‌های بوری کشور.../آدخ، فدوی اصغری و محمدپورزندی

مقدمه

ارزیابی عملکرد جزء لاینفک پاسخگویی بوده و از اهمیت زیادی برخوردار است. الگوها و روش‌های متعددی جهت ارزیابی عملکرد شرکت‌ها و مؤسسات مالی وجود دارد از جمله استفاده از شاخص‌ها و نسبت‌های عملکردی که در این روش از یک سری متغیر و نسبت استفاده می‌شود. این متغیرها و نسبت‌ها باید به گونه‌ای شناسایی شوند که نتایج روش‌های ارزیابی عملکرد نیاز استفاده کنندگان را برآورد کرده و با بهبود عملکرد، باعث تداوم فعالیت سازمان‌ها و استفاده بهینه در سطح خرد و کلان گردد. پیشتر متدولوژی‌های متعدد به منظور ارزیابی عملکرد مؤسسات مالی در ارتباط با شاخص‌های مالی انجام می‌گرفت به گونه‌ای که بررسی‌های نخست اصولاً از تکنیک‌های آماری سنتی استفاده می‌کردند، اما بررسی‌های جدیدتر، روش‌های اتخاذ تصمیم پیشرفته را به کار می‌گیرند. با افزایش رقابت در عرصه تولید و خدمات، سازمان‌ها به شاخص‌ها و الگوهایی برای بررسی عملکرد خود نیاز دارند. [18]

در عصر حاضر، حوزه مدیریت مالی نیز همگام با حوزه فناوری اطلاعات در حال پیشرفت است و با وجود تحولات شگرف دانش، وجود نظام ارزیابی موضوعی اجتناب ناپذیر است. شرکت‌ها برای حضور در عرصه بازار سرمایه باید بتوانند همواره اصل بهبود مستمر را سرلوحه فعالیت‌های خود قرار دهند. ارزیابی عملکرد یکی از مباحث مهم در توانمند سازی و قابلیت پاسخگویی در چارچوب اصول و مفاهیم مدیریت است. در سال‌های اخیر تغییرات چشمگیری در زمینه‌های مدیریت مالی، سیستم‌های اطلاعاتی و آمار صورت گرفته است. افزایش تصاعدی در قدرت محاسباتی باعث شده است تا سیستم‌های اطلاعاتی هوشمند بتوانند در زمان اندک حجم بالایی از داده‌ها را پردازش کنند. تکنیک‌های داده کاوی ابزارهای شناخته شده‌ای برای تجزیه و تحلیل داده‌ها هستند. این تکنیک‌ها با استفاده از روش‌های مختلف، الگوهای موجود در مجموعه داده‌ها را استخراج می‌کنند. [15]

علم داده کاوی و کشف اطلاعات ارزشمند از حجم عظیم داده‌ها یکی از حوزه‌های تحقیقاتی جذاب و پرکاربرد است که طی دو دهه گذشته بسیار مورد توجه مدیران حوزه مدیریت و بازاریابی قرار گرفته است. در واقع هدف از داده کاوی ایجاد مدل‌هایی برای تصمیم‌گیری است. زیرا داده کاوی یک سری ابزار در اختیار دارد که به صورت نیمه خودکار و با حداقل دخالت کاربران اطلاعات سودمند و در اصطلاح الگوهای مفید (روابط منطقی بین داده‌ها) را از میان حجم انبوه داده‌ها کشف می‌کند. تکنیک‌های مختلف داده کاوی از جمله خوشه بندی، طبقه بندی، درخت‌های تصمیم، رگرسیون، قوانین

فصلنامه مهندسی مالی و مدیریت اوراق بهادار / شماره چهل و دوم / بهار ۱۳۹۹

تداعی، الگوهای توالی و شبکه‌های عصبی مصنوعی این امکان را به تحلیل‌گر می‌دهند تا به دانش پنهان در داده‌های خام دست یابد. [19]

تاکنون در صنعت بانکداری با استفاده از فنون داده کاوی، طبقه بندی مشتریان تسهیلاتی، پیاده سازی سیستم مدیریت با مشتریان، بخش بندی مشتریان و ... صورت پذیرفته است، لیکن در حوزه ارزیابی عملکرد و پیش بینی عملکرد آنها از فنون داده کاوی بهره برده نشده است. لذا پژوهشگران در این مقاله بر آن شدند که با استفاده از تکنیک‌های داده کاوی، بهترین مدل را جهت ارزیابی عملکرد و پیش بینی عملکرد بانک‌ها ارائه نمایند.

مبانی نظری پژوهش

در محیط رقابتی امروز، بقای بنگاه‌های اقتصادی در گرو بهبود مستمر عملکرد و به منظور حفظ و افزایش توان رقابتی و کسب منافع بیشتر است. این مهم از طریق تعیین اهداف و برنامه ریزی و به تبع آن ارزیابی عملکرد برای آگاهی از میزان موفقیت در دستیابی به اهداف از پیش تعیین شده محقق می‌گردد. به دلیل اهمیت و نقش اساسی بانک‌ها در اکثر فعالیت‌های اقتصادی لازم است که عملکرد آنها مورد ارزیابی قرار گرفته و در جهت بهبود نقاط قوت و رفع نقاط ضعف آنها بکوشیم.

نسبت‌های مالی ابزار سنتی ولی همچنان قوی و متداول برای تجزیه و تحلیل صورت‌های مالی است که از دیرباز مورد علاقه استفاده کنندگانی نظیر تحلیل‌گران مالی، اعتبار دهندگان، سرمایه گذاران و مدیران مالی بوده است. در این تحلیل‌ها به جای به کارگیری ارقام کلی صورت‌های مالی، از نسبت‌های ترکیبی این ارقام به منظور دستیابی به تصویر واضحی از شرکت استفاده می‌شود. [۱۳]

از سویی با افزایش مبادلات تجاری، اقتصادی و پیشرفت فناوری اطلاعات، حجم زیادی از داده‌های مالی به سرعت انباشته شده و هزینه نگهداری و پردازش اطلاعات نیز به طرز شگفت‌انگیزی کاهش داشته‌است. این موضوع به نوبه خود مشکلاتی را برای استفاده بهینه و کارا از این داده‌ها بوجود آورده‌است. از این رو به منظور استفاده بهینه و مؤثر از داده‌های مالی در تصمیم‌گیری‌ها، از فن‌های داده کاوی استفاده می‌شود. داده کاوی فرآیندی است که با نگرشی نو به مسأله استخراج اطلاعات از داده‌ها پرداخته و به کمک مجموعه‌ای از روش‌های آماری و مدل سازی می‌تواند الگوها و روابط پنهان موجود در داده‌ها را در کمترین زمان ممکن و با دقتی بالا تشخیص دهد. این فرآیند می‌تواند برای پیش بینی رویه‌ها و رفتارهای آتی در بازار مالی مورد استفاده قرار گیرد. [۳]

ارائه مدل ارزیابی عملکرد بانک‌های بوری کشور.../آدخ، فدوی اصغری و محمدپورزندی

هر پیش بینی مستلزم شناسایی متغیرهای مؤثر بر متغیر هدف و انتخاب مدلی برای به کارگیری آن متغیرهاست. اطلاعات صورتهای مالی مهم‌ترین و قابل اتکاترین اطلاعات در ارزشیابی محسوب می‌شود. [۹]

شناسایی نسبت‌های مالی مؤثر بر عملکرد شرکت‌ها، همواره مورد توجه سرمایه‌گذاران و ذی‌نفعان بوده است. تمرکز تحقیقات گذشته بیشتر بر اساس تکنیک‌های آماری مانند تحلیل عاملی، رگرسیون خطی و تجزیه و تحلیل واریانس بوده است. بر اساس ابزارهای فناوری اطلاعات نوین و بانک‌های اطلاعاتی تحقیقات اخیر به سمت مباحث داده کاوی رفته است.

درخت تصمیم یکی از ابزارهای قوی و متداول برای دسته بندی و پیش بینی می‌باشد. در ساختار درخت تصمیم پیش بینی به دست آمده در درخت در قالب یکسری قواعد بیان می‌شود. استفاده از درخت آسان بوده و دسته بندی‌هایی که توسط درخت تصمیم ایجاد می‌شوند، از روی شباهت داده‌های ذخیره شده در پارامترهای پیش بینی کننده، قابل انجام می‌باشد. [۷]

پیشینه مطالعاتی پژوهش

"حجازی و همکاران (۱۳۹۱) دقت پیش بینی مدیریت سود را با استفاده از شبکه های عصبی و درخت تصمیم گیری بررسی و همچنین با مدل های خطی مقایسه کردند. برای این منظور از یازده متغیر تأثیر گذار بر مدیریت سود به عنوان متغیرهای مستقل و اقلام تعهدی اختیاری به عنوان متغیر وابسته استفاده شده است. در این تحقیق تعداد ۵۵ شرکت از سال ۱۳۸۵ تا سال ۱۳۸۸ به صورت فصلی مورد بررسی قرار گرفته است. از روش رگرسیون پنلی جهت مدل خطی و از شبکه شبکه عصبی پیشخور تعمیم یافته و درخت تصمیم کارت جهت بررسی از طریق شبکه عصبی و درخت تصمیم استفاده شد. نتایج نشان داد که روش شبکه عصبی و درخت تصمیم در پیش بینی مدیریت سود نسبت به روش‌های خطی دقیق تر و دارای سطح خطای کمتری است [۴]."

البرزی و همکاران در مقاله ای به نام " به کارگیری ژنتیک در بهینه سازی درختان تصمیم گیری برای اعتبار سنجی مشتریان بانک‌ها" به ارائه مدلی برای اعتبار سنجی مشتریان بانک‌ها، به منظور اعطای تسهیلات اعتباری پرداختند. مدل طبقه بندی در بانک ملت مورد بررسی قرار گرفته و نتایج پژوهش نشان داد امکان استفاده از مدل طبقه بندی پیشنهادی برای ساخت و آزمون درختان تصمیم گیری به منظور اعتبار سنجی مشتریان وجود دارد [۲]."

فصلنامه مهندسی مالی و مدیریت اوراق بهادار / شماره چهل و دوم / بهار ۱۳۹۹

مهدوی و قربانی (۱۳۹۱) در پژوهشی با عنوان "بررسی مقایسه‌ای نقش شاخص‌های نوین و سنتی نقدینگی در ارزیابی عملکرد مالی شرکت‌های پذیرفته شده در بورس اوراق بهادار تهران"، ارتباط میان نسبت جاری، نسبت آنی (به عنوان شاخص های سنتی نقدینگی) و همچنین شاخص فراگیر نقدینگی، دوره تبدیل وجه نقد، مانده نقدی خالص (به عنوان شاخص نوین نقدینگی) با بازده دارایی، بازده حقوق صاحبان سهام و کیوتوین (به عنوان شاخص عملکرد) بررسی کردند. نتایج نشان داد که شاخص‌های نوین نقدینگی در ارزیابی عملکرد مالی شرکت‌ها تصویر دقیق‌تری نسبت به شاخص‌های سنتی نقدینگی به استفاده کنندگان از اطلاعات مالی در جهت اخذ تصمیمات بهینه ارائه می‌دهند [۸]. صادقی و غنی ورزنه (۱۳۹۴) به پیش بینی عملکرد مالی شرکت‌های بورسی با استفاده از الگوریتم درخت تصمیم پرداختند. آنان از یک روش تجزیه و تحلیل دومارحله‌ای استفاده کردند. در ابتدا بعدهای اساسی نسبت‌های مالی را مشخص نموده و با استفاده از تکنیک آماری PCA نسبت‌های مالی که دارای بیشترین با عاملی بودند انتخاب گردیده و در مرحله دوم با استفاده از تکنیک درخت تصمیم به کشف روابط بالقوه بین عملکرد شرکت و نسبت‌های مالی پرداخته می‌شود و نشان دادند که این رویکرد می‌تواند روش مناسبی جهت پیش بینی عملکرد شرکت‌ها به کارگرفته شود [۵]. میرغفوری و همکاران (۱۳۹۱) در پژوهشی تحت عنوان ارزیابی عملکرد مالی با رویکرد تحلیل پوششی داده‌ها و تئوری خاکستری برای ارزیابی عملکرد مخابرات استانی استفاده نمودند. نتایج اجرای این تکنیک‌ها به علاوه روش کارایی متقاطع برای واحدهای کارا امکان رتبه‌بندی استانی را فراهم کرده است [۱۰]. کوینسگی و ازگلباس (۲۰۱۲) با استفاده از درخت تصمیم CHAID شرکت‌های کوچک و متوسط را از لحاظ بحران مالی بررسی کردند و با استفاده از ۳۱ نسبت مالی یک سیستم هشدار به هنگام بحران طراحی کردند [۱۶]. دین و همکاران (۲۰۱۳)^۱

در پژوهشی با عنوان ارزیابی عملکرد واحد تجاری با استفاده از نسبت‌های مالی با رویکرد درخت تصمیم به مقایسه الگوریتم‌های درخت تصمیم پرداخته و با استفاده از نسبت‌های مالی، عملکرد شرکت‌های بورس استانبول را بررسی کردند. نتایج حاصل نشان می‌دهد که نسبت سود قبل از مالیات به حقوق صاحبان سهام و حاشیه سود خالص دو متغیر اساسی در پیش بینی عملکرد شرکت‌ها هستند [۱۳].

روش تحقیق

این تحقیق بر اساس هدف از نوع کاربردی است و از لحاظ گردآوری اطلاعات از نوع پیمایشی می‌باشد.

ارائه مدل ارزیابی عملکرد بانک‌های بورسی کشور.../آدخ، فدوی اصغری و محمدپورزند

سؤالات تحقیق

➤ آیا می‌توان الگوریتم مناسبی از الگوریتم‌های داده کاوی را برای ارزیابی عملکرد بانک‌ها بیان نمود؟

➤ مدل مناسب برای ارزیابی عملکرد بانک‌ها کدام است؟

ابزار گردآوری داده‌ها در این تحقیق با بهره‌گیری از روش‌های کتابخانه ای و پایگاه‌های داده‌ای مرتبط با موضوع می‌باشد. این داده‌ها مشتمل بر چندین فیلد و رکورد می‌باشند که نیازمند پردازش‌های اولیه است تا بتوانیم آنها را مورد بهره برداری قرار دهیم. در این تحقیق ۲۸ نسبت مالی بر اساس بهترین نسبت‌های پژوهش‌های پیشین، انتخاب گردیدند.

قلمرو زمانی تحقیق در سال‌های ۱۳۹۳-۱۳۹۶ می‌باشد.

جامعه آماری در این تحقیق، کلیه بانک‌های پذیرفته شده در بورس و فرابورس اوراق بهادار تهران می‌باشند که صورتهای مالی خود را در فاصله سالهای ۱۳۹۳ الی ۱۳۹۶ به بورس اوراق بهادار ارائه نموده‌اند. ۱۸ بانک، شرایط مذکور را داشتند و بانک شهر، بانک قوامین و قرض الحسنه رسالت به دلیل عدم ارائه صورتهای مالی حسابرسی شده خود در سال ۹۶ از جامعه آماری حذف گردیدند.

برای توسعه پروژه‌های داده کاوی رویه‌های مختلفی توسعه داده شده است. در این پژوهش روش استاندارد داده کاوی کریسپ دی ام استفاده گردیده است

مراحل کریسپ دی ام:

۱- درک تجارت: هدف اصلی در این تحقیق شناسایی شاخص‌های مؤثر در ارزیابی عملکرد بانک‌های بورس اوراق بهادار تهران و ارائه مدلی جهت پیش بینی عملکرد آنها می‌باشد.

۲- شناخت داده: اطلاعات، مجموعه‌ای از نسبت‌های بدست آمده از صورتهای مالی مربوط به بانک‌های پذیرفته شده در بورس اوراق بهادار تهران است و از سایت کدال جمع آوری می‌گردند.

۳- آماده‌سازی داده‌ها: در این مرحله پیش پردازش داده‌ها صورت می‌پذیرد که مراحل آن به شرح ذیل می‌باشد:

الف) پاکسازی داده^۲

مهمترین فعالیت‌های این بخش عبارت است از تخمین مقادیر ناموجود در پایگاه داده‌ها، از بین بردن اختلال^۳ در داده‌ها، حذف کردن داده‌های پرت و نامربوط و از بین بردن ناسازگاری در داده‌ها نکته مهم این است که هر چه این گام از داده کاوی بهتر انجام شود، خروجی الگوریتم‌ها و تکنیک‌های داده

فصلنامه مهندسی مالی و مدیریت اوراق بهادار / شماره چهل و دوم / بهار ۱۳۹۹

کاوی کیفیت بالاتری خواهد داشت.

(ب) یکپارچه سازی داده^۴

در بسیاری از موارد ممکن است داده‌ها در فایل‌ها و منابع مختلف نگهداری شوند و در این صورت نیاز است تا داده‌ها پیش از اجرای تکنیک‌های داده کاوی با یکدیگر یکپارچه شوند. در این تحقیق کلیه اطلاعات در نرم افزار اکسل جمع آوری می‌گردند.

(ج) تبدیل داده^۵

فعالیت‌هایی مانند نرمال سازی داده‌ها و گسسته سازی داده‌ها در این حوزه جای می‌گیرند.

۴- مدل سازی: در این پژوهش روشهای مدل‌سازی ذیل را بر روی داده‌های پیش پردازش شده

اجرا می‌گردد:

الف) الگوریتم درخت تصمیم C4.5

(ب) الگوریتم بیز

(ج) الگوریتم درخت تصمیم C5.0

(د) جنگل تصادفی

۵- ارزیابی: ارزیابی نتایج حاصل از مدل‌سازی صورت می‌پذیرد. به منظور اعتبار سنجی مدل، داده‌ها به دو بخش داده‌های آموزشی و آزمون تقسیم می‌شوند. مدل اصلی با استفاده از داده‌های آموزش ساخته شده و مدل ساخته‌شده بر روی داده‌های آزمون مورد تست قرار می‌گیرد. درصد نمونه‌هایی از داده‌های آزمون که ویژگی هدف آنها توسط مدل، درست تشخیص داده شده باشند دقت مدل را بیان می‌کند [۱۴].

۶- طرح ریزی برای توسعه: در توسعه مدل از بهترین مدل ارائه شده برای پیش بینی‌های آتی

استفاده می‌گردد.

در کلیه مدل‌های اجرایی در این تحقیق، عملکرد بانک‌ها در هر سال به سه دسته قابل قبول، متوسط و غیر قابل قبول تقسیم و به عنوان فیلد هدف در نظر گرفته شده است. این تقسیم بندی بر اساس استانداردهای شاخص‌های مالی بانک مرکزی و نظر شش نفر از خبرگان بانکی که شرایط ذیل را داشته باشند تقسیم گردیده است:

الف) اشتغال در مدیریت مالی یکی از بانک‌های پذیرفته شده در بورس اوراق بهادار تهران

(ب) سابقه کاری، حداقل ۲۰ سال

ارائه مدل ارزیابی عملکرد بانک‌های بورسی کشور.../آدخ، فدوی اصغری و محمدپورزرنندی

(ج) دارابودن حداقل مدرک فوق لیسانس در زمینه مدیریت مالی

(د) آگاهی کامل از صورتهای مالی و نحوه تجزیه و تحلیل آنها

با توجه به اینکه در روش دلفی تعداد شرکت کنندگان می بایست بین ۵ تا ۲۰ نفر باشند، در این تحقیق ۶ نفر خبره انتخاب گردیدند.

سپس به هرنسبت در سال مربوطه براساس قرار گیری در بازه غیرقابل قبول، متوسط و قابل قبول، اعداد ۱، ۲ و ۳ تخصیص داده می شود در نهایت بر اساس فاصله بین طبقات در هر سال، عملکرد بانکها به سه دسته مجزا در هر سال تقسیم می گردند. می بایست بدین موضوع اشاره نمود که اگر تعداد سالهای قلمرو تحقیق و تعداد بانکهای به کارگرفته در تحقیق تغییر یابد ممکن است بازه متوسط صنعت در نسبت‌های مالی تغییر یابد.

جدول تقسیم بندی نسبت‌های مالی در جدول ۱ نمایش داده شده است:

جدول ۱: تقسیم بندی نسبت های مالی

ردیف	شاخص/عملکرد بانک	غیرقابل قبول	متوسط صنعت	قابل قبول
۱	نسبت سود خالص به میانگین دارایی‌ها (ROA)	کمتر یا مساوی ۰	بین ۰ تا ۰,۰۱	بزرگتر یا مساوی ۰,۰۱
۲	نسبت سود خالص به سپرده ها	کمتر یا مساوی ۰	بین ۰ تا ۰,۰۱	بزرگتر یا مساوی ۰,۰۱
۳	نسبت سود بر میانگین حقوق صاحبان سهام (ROE)	کمتر یا مساوی ۰	بین ۰ تا ۰,۱۲	بزرگتر یا مساوی ۰,۱۲
۴	سُرانه جذب سود پرسنل (نسبت سود به پرسنل)	کمتر یا مساوی ۰	بین ۰ تا ۹۵۰	بزرگتر یا مساوی ۹۵۰
۵	نسبت سود به شعب	کمتر یا مساوی ۰	بین ۰ تا ۱۵۹۰۰	بزرگتر یا مساوی ۱۵۹۰۰
۶	نسبت سود تسهیلات به درآمدهای مشاع	کمتر یا مساوی ۰,۷۷	بین ۰,۷۷ تا ۰,۸۸	بزرگتر یا مساوی ۰,۸۸
۷	نسبت سود حاصل از سپرده گذاریها و سرمایه گذاریها به درآمد مشاع	کمتر یا مساوی ۰,۱۱	بین ۰,۱۱ تا ۰,۲	بزرگتر یا مساوی ۰,۲
۸	نسبت سود خالص به کل درآمدها(حاشیه سود)	کمتر یا مساوی ۰	بین ۰ تا ۰,۲	بزرگتر یا مساوی ۰,۲
۹	نسبت EPS (نسبت سود خالص قابل تقسیم بر تعداد سهام)	کمتر یا مساوی ۰	بین ۰ تا ۱۷۴	بزرگتر یا مساوی ۱۷۴
۱۰	نسبت سود تسهیلات به تسهیلات اعطایی	کمتر یا مساوی ۰,۲	بین ۰,۲ تا ۰,۲۵	بزرگتر یا مساوی ۰,۲۵
۱۱	نسبت تسهیلات به سپرده‌ها	کمتر از ۰/۴ و بزرگتر از ۰/۸۵	بزرگتر یا مساوی ۰,۴ و کوچکتر یا مساوی ۰,۶	بزرگتر از ۰,۶ و کوچکتر یا مساوی ۰,۸۵

فصلنامه مهندسی مالی و مدیریت اوراق بهادار / شماره چهل و دوم / بهار ۱۳۹۹

۱۲	نسبت سپرده به تعداد شعب	کمتر یا مساوی ۵۰۰,۰۰۰	بین ۵۰۰,۰۰۰ تا ۱,۲۰۰,۰۰۰	بزرگتر یا مساوی ۱,۲۰۰,۰۰۰
۱۳	نسبت سپرده به پرسنل	کمتر یا مساوی ۵۰,۰۰۰	بین ۵۰,۰۰۰ تا ۹۰,۰۰۰	بزرگتر یا مساوی ۹۰,۰۰۰
۱۴	نسبت وجه نقد به جمع دارایی‌ها	کمتر یا مساوی ۰,۰۰۵	بین ۰,۰۰۵ تا ۰,۰۲	بزرگتر یا مساوی ۰,۰۲
۱۵	نسبت سپرده های مدت دار به کل سپرده ها	بزرگتر یا مساوی ۰,۹	بین ۰,۷۵ تا ۰,۹	کمتر یا مساوی ۰,۷۵
۱۶	نسبت سپرده به دارایی ها	کمتر از ۰/۳ و بزرگتر از ۰,۸	بزرگتر یا مساوی ۰,۶ و کوچکتر یا مساوی ۰,۸	بزرگتر یا مساوی ۰,۳ و کوچکتر از ۰,۶
۱۷	نسبت تسهیلات به دارایی ها	کمتر یا مساوی ۰,۴ و بزرگتر از ۰,۷	بین ۰,۴ تا ۰,۶	بزرگتر یا مساوی ۰,۶ و کوچکتر یا مساوی ۰,۷
۱۸	نسبت (مجموع مطالبات لاوصول و مشکوک الوصول و معوق) به مانده تسهیلات و وام ها	بزرگتر یا مساوی ۰,۱۵	بین ۰/۱۵ و ۰/۰۵	کمتر یا مساوی ۰,۰۵
۱۹	نسبت مطالبات مشکوک الوصول به کل مطالبات	بزرگتر یا مساوی ۰,۵	بین ۰/۲ و ۰/۵	کمتر یا مساوی ۰,۲
۲۰	نسبت بدهی به حقوق صاحبان سهام	کمتر یا مساوی ۰ و بزرگتر یا مساوی ۲۴	بین ۱۵ تا ۲۴	بزرگتر از ۰ و کوچکتر یا مساوی ۱۵
۲۱	جمع بدهی ها به دارایی ها	بزرگتر یا مساوی ۰/۹۵	بین ۰/۸۵ و ۰/۹۵	کمتر یا مساوی ۰/۸۵
۲۲	میانگین دارایی ها به میانگین حقوق صاحبان سهام	کمتر یا مساوی ۱۴	بین ۱۴ تا ۲۴	بزرگتر یا مساوی ۲۴
۲۳	نسبت درآمد به متوسط کل دارایی ها	کمتر یا مساوی ۰,۰۳	بین ۰,۰۳ تا ۰,۰۶	بزرگتر یا مساوی ۰,۰۶
۲۴	نسبت درآمدهای غیرمشاع بر کل درآمدها	کمتر یا مساوی ۰,۳ و بزرگتر یا مساوی ۱	بین ۰,۳ تا ۰,۶	بزرگتر یا مساوی ۰,۶ و کمتر از ۱
۲۵	هزینه مطالبات مشکوک الوصول به کل هزینه ها	بزرگتر یا مساوی ۰,۲	بین ۰,۱ تا ۰,۲	کمتر یا مساوی ۰,۱
۲۶	هزینه اداری و عمومی به کل هزینه ها	بزرگتر یا مساوی ۰,۶	بین ۰,۵ تا ۰,۶	کمتر یا مساوی ۰,۵
۲۷	نسبت کفایت سرمایه	کمتر یا مساوی ۰,۰۸	بین ۰,۰۸ تا ۰,۱۲	بزرگتر یا مساوی ۰,۱۲
۲۸	نسبت حقوق صاحبان سهام به کل دارایی ها	کمتر یا مساوی ۰,۰۴	بین ۰,۰۴ تا ۰,۰۶	بزرگتر یا مساوی ۰,۰۶

ارائه مدل ارزیابی عملکرد بانک‌های بورسی کشور.../آدخ، فدوی اصغری و محمدپورزندی

فرآیند ساخت مدل: شناسایی و آماده سازی داده ها در این تحقیق به روش مشاهده‌ای جمع‌آوری گردیده است. اطلاعات از صورت‌های مالی حسابرسی شده در سایت کدال استخراج گردیده است. نسبت‌های مالی داده‌ها همگی از نوع داده‌های نسبتی و این متغیرهای کمی با عدد قابل سنجش می‌باشند. معمولاً این گام از فرآیند داده کاوی و کشف دانش طولانی‌تر و مهم‌تر محسوب می‌شود به طوری که هر قدر کیفیت آماده سازی داده‌ها بهتر انجام شود و زمان بیشتری برای آن صرف شود مدل سازی و نتیجه گیری بهتر و آسان تر خواهد بود و نتایج اعتبار بیشتری دارد. در این تحقیق از ۴ مدل الگوریتم درخت تصمیم C4.5، الگوریتم بیز، الگوریتم درخت تصمیم C5.0 و جنگل تصادفی استفاده گردیده و مورد ارزیابی قرار گرفته و نتایج با یکدیگر مقایسه می‌شوند. در کلیه مدل‌ها، داده‌های سال‌های ۹۳، ۹۴ و ۹۵ به عنوان داده‌های آموزشی و داده‌های سال ۹۶ به عنوان داده‌های تست انتخاب گردیده و براساس ماتریس اغتشاش هرمدل، شاخصه صحت بدست می‌آید. دیتاست شامل ۷۲ رکورد و ۲۹ فیلد است. که اطلاعات آماری فیلد هدف به شرح ذیل می‌باشد:

جدول ۲: دیتاست فیلد هدف

معتبر	تعداد	درصد
قابل قبول	۲۰	۲۷,۷۸
متوسط	۳۴	۴۷,۲۲
غیرقابل قبول	۱۸	۲۵

۱- الگوریتم درخت تصمیم C4.5: این الگوریتم در سال ۱۹۹۳ توسط کوئینلن^۷ تهیه شد. این الگوریتم متغیرهای پیوسته و گسسته را در محاسبات خود لحاظ کرده و مقادیر مفقود را نیز در الگوریتم خود در نظر می‌گیرد. [۱۱]

الگوریتم، لزوماً دودویی نیست و برای انتخاب یک جداکننده بهینه در طول مسیر، درخت تصمیم‌گیری از شاخص کسب اطلاعات یا کاهش آنتروپی استفاده می‌کند. [۱۷]

این الگوریتم یکی از تعمیم‌های الگوریتم ID3 است که از معیار Gain Ratio جهت انتخاب صفت خاصه استفاده می‌کند. الگوریتم هنگامی متوقف می‌شود که تعداد نمونه‌ها کمتر از مقدار مشخص شده‌ای باشد. این الگوریتم از تکنیک پس هرس استفاده می‌کند و همانند الگوریتم‌های قبلی داده‌های عددی را نیز می‌پذیرد. با کمی تغییر هم می‌توان برای داده‌های ناقص از آن استفاده کرد. [۱]

نتایج حاصل از اجرای الگوریتم در نرم افزار رپیدماینر به شرح ذیل می‌باشد:

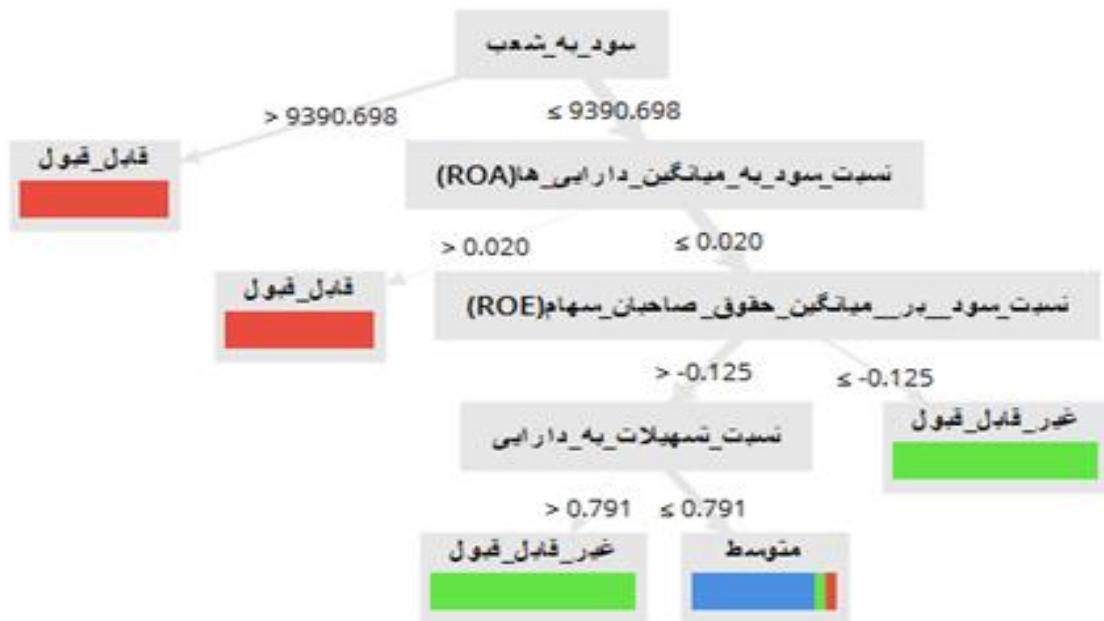
فصلنامه مهندسی مالی و مدیریت اوراق بهادار / شماره چهل و دوم / بهار ۱۳۹۹

جدول ۳: درخت تصمیم c4.5

نوع بانک	قابل قبول	متوسط	غیر قابل قبول	دقت کلاس
قابل قبول	۱	۰	۰	٪۱۰۰
متوسط	۱	۱۱	۱	٪۸۴,۶
غیر قابل قبول	۰	۰	۴	٪۱۰۰

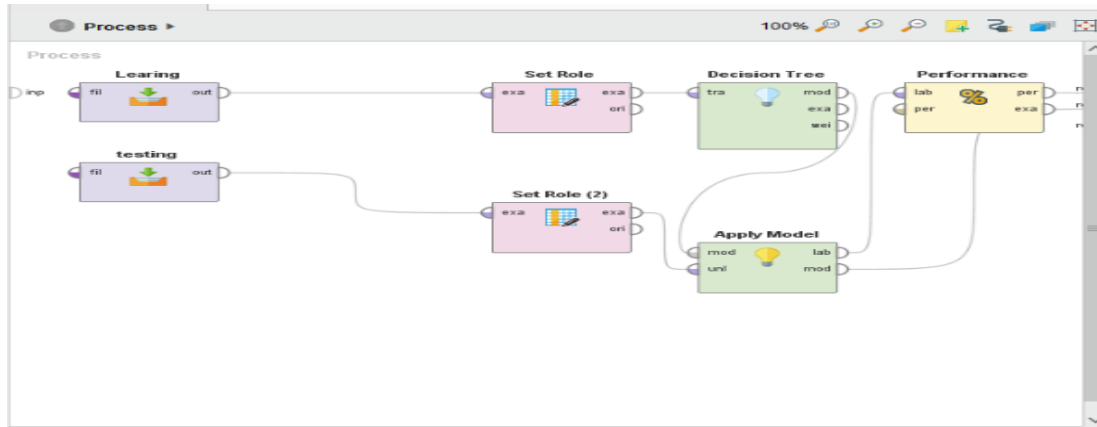
جدول ۴: صحت پیش بینی درخت تصمیم c4.5

پیش بینی	تعداد	درصد
درست	۱۶	٪۸۸,۹
غلط	۲	٪۱۱,۱
مجموع	۱۸	٪۱۰۰



شکل ۱: درخت تصمیم c4.5

ارائه مدل ارزیابی عملکرد بانک‌های بورسی کشور.../آدخ، فدوی اصغری و محمدپورزندگی



شکل ۲: مدل درخت تصمیم c4.5

همانگونه که مشاهده می‌گردد درخت تصمیم با کاهش شاخه‌های درخت و انتخاب بهترین فیله‌های ورودی تشکیل گردیده است. شکست اصلی در درخت، سود به شعب بوده، شکست بعدی در سود به میانگین دارایی‌ها و شکست آخر مدل در شاخص نسبت تسهیلات به دارایی‌ها انجام پذیرفته است.

تنظیم پارامترهای مدل به شرح ذیل می‌باشد:

Criterion: gain ratio

Maximal depth:20

Confidence: 0.5

Minimal gain: ۰,۱

Minimal leaf size:2

Minimal size for Split:4

Number of prepruning alternatives:3

۲- الگوریتم نایویز: یکی از فرمول‌های مهم احتمال، فرمول احتمال بیز است که می‌توانیم به کمک آن احتمال برچسب کلاس یک نمونه از داده‌ها را تخمین بزنیم. استفاده از این قانون برای طبقه بندی، دقت و سرعت خوبی را در پایگاه داده‌های بزرگ به همراه دارد. در این روش از قانون احتمال بیز استفاده گردیده است. در اجرا فرض بر این است که تأثیر مقدار یک صفت خاصه بر روی برچسب کلاس، مستقل از مقادیر دیگر صفت خاصه است و این موضوع استقلال شرطی^۱ کلاس نامیده می‌شود. [۱]

نتایج حاصل از اجرای مدل در نرم افزار رپیدماینر به شکل ذیل می‌باشد:

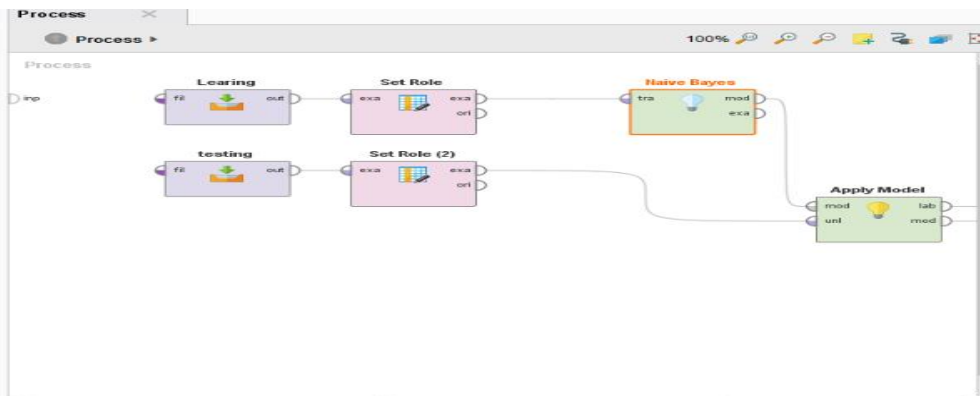
فصلنامه مهندسی مالی و مدیریت اوراق بهادار / شماره چهل و دوم / بهار ۱۳۹۹

جدول ۵: الگوریتم نایو بیس

نوع بانک	قابل قبول	متوسط	غیرقابل قبول	دقت کلاس
قابل قبول	۱	۰	۰	%۱۰۰
متوسط	۱	۹	۱	%۸۱,۸۲
غیرقابل قبول	۰	۲	۴	%۶۶,۶۷

جدول ۶: صحت پیش بینی مدل نایو بیس

پیش بینی	تعداد	درصد
درست	۱۴	%۷۷,۸
غلط	۴	%۲۲,۲
مجموع	۱۸	%۱۰۰



شکل ۳: مدل الگوریتم نایو بیس

با اجرای مدل نایو بیس، پیش بینی مدل با تعداد ۱۴ مورد صحیح و ۴ مورد نادرست به %۷۷,۸ می‌رسد.

۳- الگوریتم C5 :

یک نوع درخت تصمیم گیری تک متغیره و بهبود یافته C4.5 است. این الگوریتم ابتدا درختی تقریباً بر ایجاد می‌کند ولی استراتژی هرس آن کاملاً متفاوت است. این الگوریتم دسته بندی را با تقسیم کردن داده‌ها به زیر مجموعه‌هایی که شامل رکوردهای همگن‌تر از والد خود هستند انجام

ارائه مدل ارزیابی عملکرد بانک‌های بورسی کشور.../آدخ، فدوی اصغری و محمدپورزرنندی

می‌دهد. در C5 تقسیم کردن نمونه‌ها بر اساس فیلدی که بیشترین بهره اطلاعات را دارد صورت می‌پذیرد. این الگوریتم روشی افزایشی از هرس کردن درخت را به کار می‌گیرد تا خطای طبقه بندی کردن ناشی از نویز یا جزئیات خیلی زیاد را در داده‌های آموزشی کاهش دهد. هرس کردن با جایگزینی گره داخلی با گره برگ رخ می‌دهد که بدان وسیله درصد یا میزان خطا کاهش می‌یابد [۱۲].

نتایج حاصل از اجرای مدل در نرم افزار کلماتین به شرح ذیل می‌باشد:

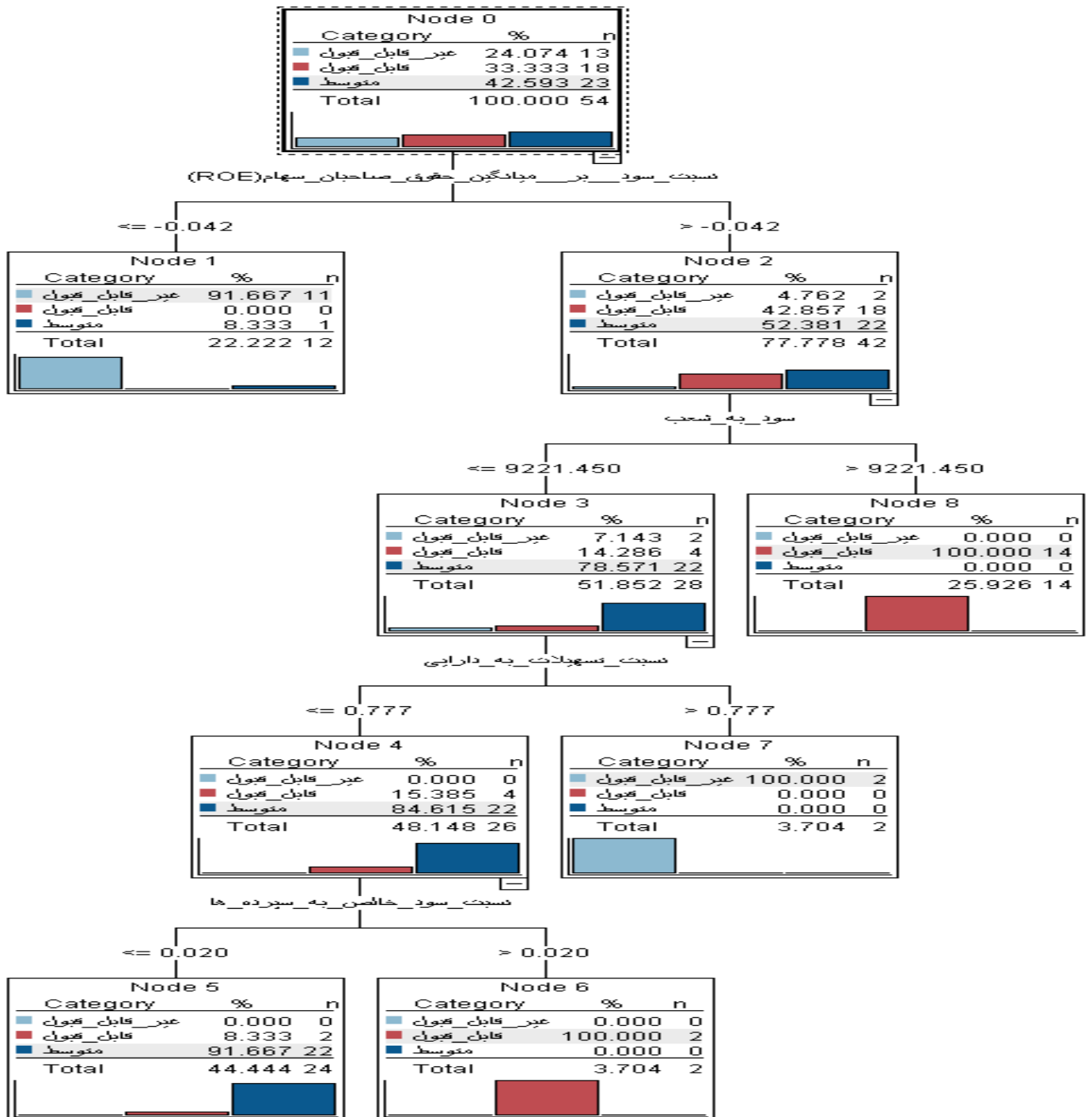
جدول ۷: الگوریتم C5

نام بانک	قابل قبول	متوسط	غیرقابل قبول	دقت کلاس
قابل قبول	۱	۰	۰	٪۱۰۰
متوسط	۱	۱۱	۰	٪۹۱,۶۷
غیرقابل قبول	۰	۰	۵	٪۱۰۰

جدول ۸: صحت پیش بینی مدل C5

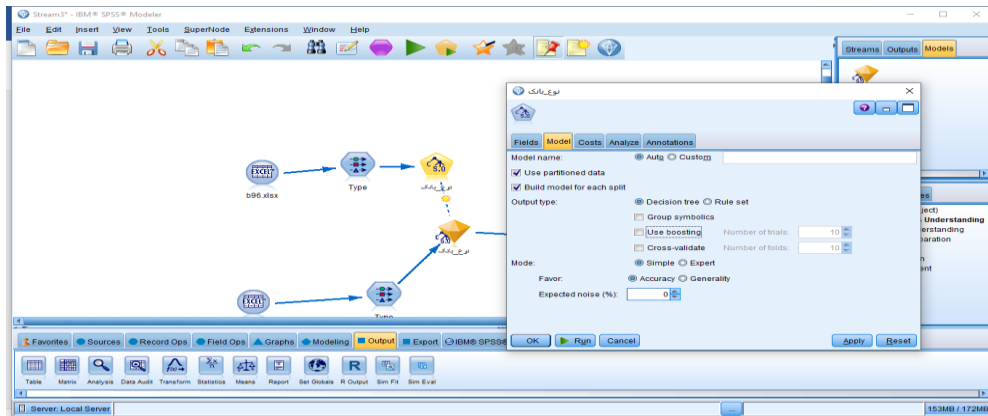
پیش بینی	تعداد	درصد
درست	۱۷	٪۹۴,۴
غلط	۱	٪۵,۶
مجموع	۱۸	٪۱۰۰

نوع بانک



شکل ۴: درخت تصمیم c5

ارائه مدل ارزیابی عملکرد بانک‌های بورسی کشور.../آذخ، فدوی اصغری و محمدپورزندی



شکل ۵: مدل الگوریتم درخت تصمیم c5

در انتخاب پارامترهای مدل، Decision tree، Simple و Accuracy انتخاب می‌گردند. همانگونه که مشاهده می‌گردد فیلدهای بااهمیت درخت به ترتیب سود به میانگین حقوق صاحبان سهام، سود به شعب، تسهیلات به دارایی و سود خالص به دارایی‌ها می‌باشد. و مدل با ۱۷ پیش بینی درست و دقت ۹۴٫۴٪ بهترین مدل را ارائه نموده است.

۴- الگوریتم جنگل تصادفی

الگوریتم جنگل‌های تصادفی از ترکیبی از درخت‌های تصمیم مستقل برای مدل‌سازی داده‌ها و ارزیابی اهمیت متغیرها استفاده می‌کند. هر درخت تصمیم در یک جنگل با استفاده از نمونه‌ای خودسازمانده^۹ از داده‌ها تشکیل می‌شود. در درخت استاندارد، هر گره از طریق بهترین انشعاب^{۱۰} (تجزیه) از بین تمام متغیرها تفکیک می‌شود. در جنگل‌های تصادفی، هر گره از طریق بهترین زیرمجموعه از متغیرهایی که به صورت تصادفی در هر گره انتخاب شده است، تفکیک می‌شود. این راهبرد در مقایسه با راهبرد بسیاری از طبقه‌بندی‌کننده‌ها از قبیل تحلیل ممیزی، ماشین بردار پشتیبان و شبکه‌های عصبی بهتر و در مقابل بیش‌تر برآزش مقاوم‌تر است. برای تشکیل هر درخت، دسته متفاوتی از الگوهای موجود، با در نظر گرفتن جایگزینی دوباره هر الگوی انتخاب شده، انتخاب می‌گردد. اندازه این دسته نمونه برداری شده برابر تعداد کل الگوهای موجود خواهد بود. این طریقه نمونه برداری معمولاً در حدود یک سوم از الگوهای موجود را بیرون دسته قرار می‌دهد. هر درخت بر اساس دسته الگوی انتخاب شده، تا ماکزیمم عمق از پیش تعیین شده رشد داده می‌شود. این عمق بر اساس حداقل تعداد الگوها در هر گره انتهایی، تعیین می‌شود. بر اساس الگوریتم جنگل‌های تصادفی، در مرحله رشد هر درخت،

فصلنامه مهندسی مالی و مدیریت اوراق بهادار / شماره چهل و دوم / بهار ۱۳۹۹

در هر گره، دسته ای از ویژگی‌ها (متغیرها) به صورت تصادفی انتخاب و بهترین انشعاب در میان دسته ویژگی انتخاب شده برای تشکیل گره‌های جدید بعدی در نظر گرفته می‌شود [۲۰].

جنگل‌های تصادفی دارای چندین مزیت نسبت به سایر روش‌های مدل‌سازی است. متغیرهای مورد استفاده می‌تواند پیوسته یا طبقه‌ای باشد. به دلیل ایجاد تعداد زیاد درخت و میانگین‌گیری در اجرای جنگل‌های تصادفی، این طبقه‌بندی کننده به نتایج با تعصب اندک و تغییر پذیری کم، ولی پیش‌بینی‌های دقیق منجر می‌شود. این راهبرد به طور معناداری بهتر از طبقه بندی کننده های تحلیل ممیزی، ماشین بردار پشتیبان و شبکه‌های عصبی است و در برابر بیش برزش^{۱۱} مقاوم است. افزون بر این جنگل‌های تصادفی، فقط دو شاخص اصلی (تعداد متغیرها در هر گره و تعداد درختان در جنگل) دارد و معمولاً به ارزش‌های آنها چندان حساس نیست. با توجه به این مزایا کاربرد الگوریتم جنگل‌های تصادفی در حال افزایش است [۲۰].

نتایج حاصل از اجرای مدل به شرح ذیل می‌باشد:

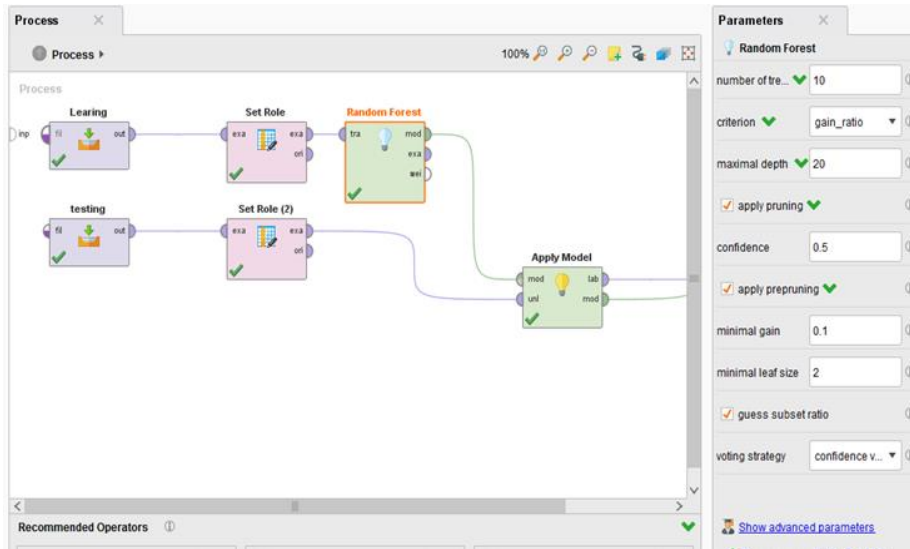
جدول ۹: مدل جنگل تصادفی

نوع بانک	قابل قبول	متوسط	غیر قابل قبول	دقت کلاس
قابل قبول	۱	۰	۰	٪۱۰۰
متوسط	۱	۱۰	۰	٪۹۰,۹
غیر قابل قبول	۰	۱	۵	٪۸۳,۳۳

جدول ۱۰: صحت پیش‌بینی مدل جنگل تصادفی

پیش‌بینی	تعداد	درصد
درست	۱۶	٪۸۸,۹
غلط	۲	٪۱۱,۱
مجموع	۱۸	٪۱۰۰

ارائه مدل ارزیابی عملکرد بانک‌های بورسی کشور.../آدخ، فدوی اصغری و محمدپورزندی



شکل ۶: مدل الگوریتم جنگل تصادفی

تنظیم پارامترهای مدل به شرح ذیل می باشد:

Numer of trees:10	Crित्रion: gain ratio	Maximal depth:20
Minimal gain: ۰,۱	Confidence: 0.5	Minimal leaf size:2
Minimal size for Split:4	Number of prepruning alternatives:3	

نتیجه گیری

پژوهش حاضر با هدف ارائه مدل ارزیابی عملکرد بانک‌های پذیرفته شده در بورس اوراق بهادار تهران انجام پذیرفت. بدین منظور ابتدا مقدار ۲۸ نسبت مالی مشتمل بر نسبت‌های سودآوری، نقدینگی، حوزه کیفیت مدیریت، حوزه کیفیت دارایی و کفایت سرمایه از ۱۸ بانک پذیرفته شده در بورس اوراق بهادار تهران در فاصله زمانی سال‌های ۱۳۹۳ الی ۱۳۹۶ از سامانه کدال استخراج گردیدند. بانک‌های مورد مطالعه شامل ۱۰ بانک پذیرفته شده در بورس و ۸ بانک لیست شده در فرابورس می‌باشد که اسامی آنها در جدول ۱۱ لیست گردیده است.

فصلنامه مهندسی مالی و مدیریت اوراق بهادار / شماره چهل و دوم / بهار ۱۳۹۹

جدول ۱۱: اسامی بانکهای بورسی و فرابورسی

بانک های بورسی	بانک های فرابورسی
بانک ملت	بانک خاورمیانه
بانک انصار	بانک دی
بانک اقتصاد نوین	بانک سرمایه
بانک کارآفرین	بانک گردشگری
پست بانک	بانک حکمت ایرانیان
بانک سینا	بانک سامان
بانک تجارت	بانک آینده
بانک پارسیان	بانک ایران زمین
بانک پاسارگاد	
بانک صادرات	

سپس با اجرای روش داده کاوی کریسپ دی ام ابتدا به جمع آوری داده‌ها پرداخته، پس از آن پاکسازی و یکپارچه سازی داده ها صورت پذیرفت. در ادامه مدل‌های داده کاوی درخت تصمیم C4.5، الگوریتم نایو بیس، درخت تصمیم C5.0 و جنگل تصادفی بر روی داده ها پیاده سازی گردیدند. به نحوی که ۲۸ نسبت به عنوان متغیرهای مستقل و عملکرد بانکها در سه دسته قابل قبول، متوسط و غیر قابل قبول به عنوان متغیر وابسته انتخاب گردیدند.

در تمامی مدل ها نسبت های سالهای ۹۳ الی ۹۵ به عنوان داده های آموزشی و نسبت های سال ۹۶، به عنوان داده های تست قرار گرفتند. پس از اجرای ۴ مدل ، نتایج جدول شماره ۱۲ حاکی از آن است که کلیه مدلها دقتی بالای ۷۰٪ داشته و با تقریب خوبی از قابلیت پیش بینی برخوردارند. درخت تصمیم C5.0 با قدرت پیش بینی ۹۴٫۴٪ بهترین مدل را برای ارزیابی عملکرد بانکها ارائه نموده است. پس از آن، درخت تصمیم C4.5 و الگوریتم جنگل تصادفی در رده دوم و مدل نایو بیس در رده سوم قرار دارد.

ارائه مدل ارزیابی عملکرد بانک‌های بورسی کشور.../آدخ، فدوی اصغری و محمدپورزند

دقت کلاس قابل قبول در تمام مدل‌ها بالا و برابر با ۱۰۰٪ می‌باشد. کمترین دقت کلاس مربوط به کلاس غیرقابل قبول در الگوریتم نایوبیز با دقت ۶۶,۶۷٪ می‌باشد.

جدول ۱۲: نتایج حاصل از اجرای مدل‌های داده کاوی

دقت الگوریتم	دقت کلاس قابل قبول	دقت کلاس متوسط	دقت کلاس غیرقابل قبول	دقت پیش‌بینی مدل
درخت تصمیم	٪۱۰۰	٪۸۴,۶	٪۱۰۰	٪۸۸,۸۹
الگوریتم نایوبیز	٪۱۰۰	٪۸۱,۸۲	٪۶۶,۶۷	٪۷۷,۸
جنگل تصادفی	٪۱۰۰	٪۹۰,۹	٪۸۳,۳۳	٪۸۸,۹
درخت تصمیم	٪۱۰۰	٪۹۱,۶۷	٪۱۰۰	٪۹۴,۴

با توجه به مطالب فوق پیشنهادات اجرایی زیر ارائه می‌گردد:

۱- تحقیق حاضر با توجه به نتایج حاصله دارای اطلاعات مفیدی بوده و از دقت بالایی برخوردار است، از این رو توصیه می‌شود محققین از این روش‌ها در پیش‌بینی عملکرد بانک‌ها استفاده نموده و متناسب با سال مورد استفاده و شاخص‌های انتخابی پارامترها بهینه گردند.

۲- این پژوهش به کارشناسان حوزه بورس، مدیران بانکی و سایر نهادهای مالی فعال در زمینه بورس و دانشجویان جهت بهره‌برداری در تحقیقات آتی توصیه می‌گردد.

توصیه برای تحقیقات آتی به شرح ذیل می‌باشد:

۱- به محققین توصیه می‌شود بازه زمانی داده‌های مورد استفاده را افزایش داده و نتایج را با تحقیق حاضر مقایسه نمایند.

۲- شاخص‌های مورد استفاده تحقیق را در روش‌های دیگر داده کاوی از جمله شبکه عصبی یا بردار پشتیبان به کارگرفته و نتایج را مقایسه نمایند.

فصلنامه مهندسی مالی و مدیریت اوراق بهادار / شماره چهل و دوم / بهار ۱۳۹۹

منابع

- ۱) اسماعیلی مهدی. داده کاوی مفاهیم و تکنیک‌ها. ویراست سوم. انتشارات نیاز دانش، ۱۳۹۳.
- ۲) البرزی محمود، محمد پورزندی محمد ابراهیم، خان بابایی محمد. به کارگیری الگوریتم ژنتیک در بهینه سازی درختان تصمیم‌گیری برای اعتبار سنجی مشتریان بانک‌ها. نشریه مدیریت فناوری اطلاعات، دوره ۲، شماره ۴، بهار و تابستان ۱۳۸۹: ۳۸-۲۳.
- ۳) باقرپور و لاشانی محمد علی، ساعی محمد جواد، مشکانی علی، باقری مصطفی. پیش بینی گزارش حسابرس مستقل در ایران، رویکرد داده کاوی. مجله تحقیقات حسابداری دوره ۱۹، سال ۱۳۴۱: ۱۳۹۱-۱۵۰.
- ۴) حجازی رضوان، محمدی شاپور، اصلانی زهرا، آقاجانی مجید. پیش بینی مدیریت سود با استفاده از شبکه عصبی و درخت تصمیم در شرکت‌های پذیرفته شده در بورس اوراق بهادار تهران. مجله بررسی‌های حسابداری و حسابرسی دانشکده مدیریت دانشگاه تهران، دوره ۱۹، شماره ۲، تابستان ۱۳۹۱: ۳۱-۴۶.
- ۵) صادقی حجت اله، غنی ورزنده فریبا. پیش بینی عملکرد مالی شرکت‌های بورسی اوراق بهادار با رویکرد الگوریتم درخت تصمیم با استفاده از ROE به عنوان متغیر پیش بین. رشت، اولین کنفرانس بین المللی حسابداران و مدیریت هزاره سوم، دی ماه ۱۳۹۴.
- ۶) صالحی سیدمرتضی، نیکوکار غلامحسین، محمدی ابوالفضل، تقی نتاج غلامحسین. طراحی الگوی ارزیابی عملکرد شعب بانک‌ها و مؤسسات مالی و اعتباری (مورد مطالعه: بانک قوامین). مجله دانشکده مدیریت دانشگاه تهران، دوره ۳، شماره ۷، بهار ۹۰: ۱۴۲-۱۲۷.
- ۷) غضنفری مهدی، علیزاده سمیه، تیمورپور بابک. داده کاوی و کشف دانش. چاپ پنجم. انتشارات دانشگاه علم و صنعت، ۱۳۹۵.
- ۸) مهدوی غلامحسین، قربانی اصغر. بررسی مقایسه ای نقش شاخص‌های نوین و سنتی نقدینگی در ارزیابی عملکرد مالی شرکت‌های پذیرفته شده در بورس اوراق بهادار تهران. مجله پژوهش‌های حسابداری مالی دانشگاه اصفهان، دوره ۴، شماره ۱، بهار ۱۳۹۱: ۸۸-۶۷.
- ۹) مهران‌سازان، مهران‌کاوه، کرمی غلامرضا. استفاده از اطلاعات تاریخی مالی و غیرمالی جهت تفکیک شرکت‌های موفق از ناموفق. مجله بررسی‌های حسابداری و حسابرسی، دوره ۳۸، سال ۱۳۸۳: ۷۷-۹۲.

ارائه مدل ارزیابی عملکرد بانک‌های بورسی کشور.../آدخ، فدوی اصغری و محمدپورزرنندی

۱۰) میرغفوری سید حبیب اله، شفیعی رودپشتی میثم، ندافی غزاله. ارزیابی عملکرد مالی با رویکرد تحلیل خاکستری (مورد: شرکت‌های مخابرات استانی). فصلنامه علمی پژوهشی دانش مالی تحلیل اوارق بهادار، شماره ۱۶، زمستان ۱۳۹۱: ۶۱-۷۶.

11) Aitkenhead, M. J., A co-evolving decision tree classification method, Journal of Expert Systems with Applications, 2008, 34(1), P.18-25.

12) Chattamvelli, R., Data mining Algorithm, Alpha science, 2011. (Book)

13) Delen, D., Kuzey, C., and Uyar, A., Measuring firm performance using financial ratios: A Decision tree approach, Journal of Expert System with Application 2013, 40(10), P.3970-3983.

14) Gupta, G.K., Introduction to Data Mining with Case Studies, Prentice - Hall of India, second edition, 2011. (Book)

15) Hand, D. J., Mannila, H., and Smyth, P., Principles of data mining. MIT press Cambridge, Massachusetts London England, 2001. (Book)

16) Koyuncugil, A.S., and ozgulbas, N. Financial early Warning system model and data mining application for risk detection, Journal of Expert Systems with Applications, 2012, 39(6), P.6238-6253.

17) Larose, D. T., Discovering Knowledge in Data, an Introduction to Data Mining, New Jersey, Wiley, 2005. (Book)

18) Porzan, M., Cristina, A. and Danescu, T., The role of the risk management and of the activities of internal control in supplying useful information through the accounting and fiscal reports, Journal of Procedia and Finance, 2012, 3, P.1099-1106.

19) Witten, I., Frank, E., Practical Machine Learning Tools and Techniques, Morgan Kaufmann Series in Data Management Systems, United Kingdom, third edition, 2011. (Book)

20) Yeh, C., Chi, D. j., and Lin, Y. R., Going-Concern prediction Using Hybrid Random Forests and Rough Set Approach, Journal of Information Sciences, 2014, 254, P.98-110.

- ۱- Delen &etal
- ۲-Data Cleaning
- ۳-Noise
- ۴-Data Integration
- ۵-Data transformation

۶ - میزان سودی که هریک از پرسنل در جذب آن نقش داشته اند.

- ۷-Queen Len
- ۸-Conditional Independence
- 9 -Bootstrap
- ۱۰- Split
- 11 -ver fitting