

ارزیابی ریسک اعتباری مشتریان بانکها

سید حبیب اله میرغفوری^{۱*}، زهره امین آشوری^۲

^۱ دانشیار مدیریت و حسابداری، دانشکده اقتصاد، دانشگاه یزد

^۲ دانشجوی کارشناسی ارشد مدیریت صنعتی - گرایش تولید، دانشگاه یزد

چکیده

بانک‌های تجاری به منظور مدیریت ریسک اعتباری، از روش‌های امتیازدهی اعتباری متفاوتی برای ارزیابی عملکرد مالی شرکت‌های متقاضی تسهیلات اعتباری استفاده می‌کنند. در این تحقیق از یک روش پارامتریک (رگرسیون لجستیک) و یک روش ناپارامتریک (درخت تقسیم و رگرسیون) برای ایجاد مدل امتیازدهی اعتباری استفاده شده است. برای ساخت مدل امتیازدهی اعتباری داده‌های مربوط به ۲۸۲ شرکت کوچک و متوسط وام‌گیرنده از یکی از شعب بانک تجارت استان تهران مورد استفاده قرار گرفت. ۱۳ نسبت مالی به عنوان شاخص‌های تعیین کننده وضعیت مالی شرکت‌های انتخاب شده به کار گرفته شدند. با استفاده از این دو روش، نسبت‌های مؤثر و همچنین دقت روش‌های مذکور در طبقه‌بندی مشتریان مشخص شد. با مشاهده نتایج حاصل از ارزیابی این روش‌ها می‌توان فهمید که روش‌های ناپارامتریک دارای دقت قابل رقابتی با روش‌های پارامتریک می‌باشند.

کلمات کلیدی: ریسک اعتباری، مدل امتیازدهی اعتباری، رگرسیون لجستیک، درخت تقسیم و رگرسیون.

طبقه‌بندی JEL: G21, E51

مقدمه

حوزه فعالیت‌های بانکی به‌گونه‌ای است که ریسک‌های متنوعی را معطوف عملیات بانکی می‌سازد. همچنین بانک، اصلی‌ترین نهاد فعال در زمینه ارائه اعتبار می‌باشد که خود باعث شده است تا فعالیت بانک‌ها با ریسک اعتباری مواجه شود. به همین دلیل بانک‌ها به بحث مدیریت پویای ریسک و طراحی مدل‌های درونی مدیریت ریسک توجه ویژه‌ای داشته و ساختار و تشکیلاتی متنوعی را نیز جهت مدیریت بهینه ریسک در بانک‌ها تنظیم کنند. ریسک اعتباری یکی از اصلی‌ترین ریسک‌هایی است که بانک با آن روبه‌رو است. عدم مدیریت صحیح این ریسک و کنترل آن، بانک را دچار بحران و ورشکستگی می‌کند و از آنجایی که بانک یک نهاد تأثیرگذار در سیستم اقتصادی هر کشوری است، این بحران‌ها کل سیستم اقتصادی و اجتماعی را دچار مشکل می‌کنند (ریگو، ۲۰۰۸).

ریسک اعتباری، ریسکی است که بر اساس آن پرداخت وام از سوی مشتری یا با تأخیر صورت می‌گیرد و یا این‌که اصلاً پرداخت نمی‌شود و این امر باعث بروز مشکلات در گردش وجوه نقد بانک‌ها شده و بر روی نقدینگی و بازده سرمایه‌گذاری بانک‌ها اثر منفی دارد. برای کنترل و کاهش ریسک اعتباری، بانک نیاز دارد که متقاضیان تسهیلات اعتباری خود را به‌درستی بشناسد و بتواند بین متقاضیانی که توانایی پرداخت به‌موقع وام خود را دارند و در واقع دارای ریسک پایین هستند و متقاضیان با ریسک بالا تمایز قایل شوند که این امر از طریق مدیریت کارا و اثربخش ریسک اعتباری امکان‌پذیر است. به این منظور به سیستم رتبه‌بندی و امتیازدهی اعتباری مشتریان بانک به عنوان یک ابزار اصلی در مدیریت ریسک اعتباری نیاز است. با استفاده از سیستم رتبه‌بندی اعتباری می‌توان به دسته‌بندی متقاضیان تسهیلات اعتباری پرداخت و آن‌ها را به دو دسته مشتریان خوش‌حساب و بدحساب تقسیم کرد (چن و همکاران، ۲۰۱۲).

کمیته بالسوئیس تعریف خلاصه‌تری از ریسک اعتباری ارائه داده است که بدین شرح است: «ریسک اعتباری عبارت است از امکان بالقوه این که قرض‌گیرنده از بانک و یا از طرف حساب وی در اجرای تعهدات خود در مقابل بانک مدت مشخصی ناتوان شود» (کمیته بالسوئیس، ۲۰۰۰).

با توجه به محیط رقابتی حاکم بر نظام بانکی در کشورها و بخصوص ایران، بانکها باید تمامی تلاش خود را در راستای حفظ و توسعه موقعیت‌های رقابتی خود معطوف دارند و در این خصوص توجه ویژه‌ای به ریسک اعتباری مشتریان خود داشته باشند و بازگشت منابع سرمایه‌ای خود را تضمین کنند. بنابراین در این مقاله تلاش بر آن است تا مدلی جهت ارزیابی ریسک اعتباری مشتریان بانکها با رویکرد داده کاوی ارائه گردد. در این مدل سعی شده است با استفاده از نسبت‌های مالی که از صورت‌های مالی و ترازنامه حسابرسی شده شرکت‌های کوچک و متوسط نزد بانک تجارت استان تهران استخراج شده است و به‌کارگیری روش‌های رگرسیون لجستیک و درخت تقسیم و رگرسیون شاخص‌های مؤثر بر نکول وام (عدم پرداخت اصل و فرع وام) این شرکت‌ها و دسته‌بندی آن‌ها در دو گروه خوش‌حساب و بدحساب مورد بررسی قرار گیرد. مدیران بانکها می‌توانند با استفاده از مدل این تحقیق به صورت دقیق‌تر و علمی سطح ریسک اعتباری مشتریان خود را تعیین کرده و تصمیمات مربوط به اعطای وام را بر اساس آن اتخاذ کنند.

ادبیات تحقیق

ریسک، امری ذاتی در بسیاری از فعالیت‌های انسانی است. در دنیای امروز نیز با پیچیده‌تر شدن محیط فعالیت سازمان‌ها و تغییرات ساختارهای سازمانی در جهت سازگاری هر چه بیشتر با این تحولات و همین‌طور گسترش رقابت‌ها در سطوح جهانی عملاً سازمان‌ها در معرض ریسک‌های مختلفی قرار گرفته‌اند. بانکها و مؤسسات مالی و اعتباری نیز از این قاعده مستثنی نیستند؛ چون با توجه به جهانی شدن و ادغام بازارهای مالی و رشد چشمگیر دانش مالی، تنوع و پیچیدگی فعالیت‌های بانکی بیشتر شده است و در عین حالی که فرصت‌های جدیدی را برای بانکها فراهم کرده است، این روند آن‌ها را در معرض ریسک‌ها و مخاطرات مختلفی همچون ریسک‌های مالی و اعتباری، ریسک‌های عملیاتی، ریسک‌های محیطی قرار داده است. علی‌رغم ابداعات و نوآوری‌های موجود در نظام بانکی هنوز هم ریسک اعتباری به عنوان دلیل عمده عدم موفقیت بانکها محسوب می‌شود که بنا به گفته کارشناسان علت آن این است که معمولاً ۸۰ درصد ترازنامه یک بانک تسهیلات

اعطایی به مشتریان است (گرانینگ و همکاران، ۲۰۰۹). بانک برای بقای خود و جلوگیری از ورشکستگی نیازمند یک سیستم مدیریت ریسک جامع است تا قادر باشد زیان ناشی از عملیات بانکی را پیش‌بینی و به موقع از آن جلوگیری کند. یکی از موارد مدیریت ریسک در بانک، مدیریت ریسک اعتباری است که به دنبال مهار ریسک اعتباری موجود در سرمایه بانک است. همچنین ثبات دارایی بانک و حصول اطمینان از بازده مناسب یکی دیگر از اهداف مدیریت ریسک اعتباری می‌باشد. با مدیریت این ریسک، بانک قادر خواهد بود تا تعهدات خود را انجام دهد و برای سهامداران ارزش ایجاد کند. یکی از ابزارهای مدیریت ریسک اعتباری، سیستم رتبه‌بندی و امتیازدهی اعتباری می‌باشد که یک سیستم تحلیلی مبتنی بر آمار است و به منظور تعیین درجه ریسک متقاضیان تسهیلات اعتباری به‌کار می‌رود و به وام‌دهندگان در شناسایی مشتریان خوش‌حساب و بد حساب کمک می‌کند و از این طریق می‌توان ریسک اعتباری را مدیریت کرد. مدل‌های مختلفی برای رتبه‌بندی و امتیازدهی اعتباری ارائه شده است که مدل‌هایی همچون مدل‌های رگرسیون لجستیک، تحلیل ممیزی، شبکه‌های عصبی و غیره نمونه‌هایی از این روش‌ها هستند (سینگ، ۲۰۱۳).

در تحقیقات مختلف از شاخص‌های گوناگونی برای ارزیابی ریسک اعتباری شرکت‌های وام‌گیرنده از بانک استفاده شده است؛ که در جدول زیر چند نمونه ارائه شده‌اند.

جدول ۱- شاخص‌های مورد استفاده در تحقیقات مربوط به ارزیابی ریسک اعتباری

پژوهشگر، تاریخ	شاخص مورد استفاده	ردیف
O. Soares, P. Pina, S. Ribeiro, Catalão- Lopes, 2011.	شاخص‌های مالی شامل: نسبت‌های نقدینگی (تعادل نقدی، نسبت آنی و دوره مورد نیاز تأمین مالی)، نسبت‌های اهرمی (پوشش ارقام دارایی‌های ثابت، نسبت حقوق صاحبان سهام و نسبت تعداد دفعات تحقق بهره)، نسبت‌های سودآوری (بازده ارزش ویژه، بازگشت دارایی، حاشیه اعتبار اقتصادی، درصد هزینه‌های ثابت، سود ناخالص، جریان نقدی شرکت و ارزش فعلی خالص پروژه) شاخص‌های بازار شامل: حساسیت شرکت نسبت به اقتصاد کلان، حساسیت شرکت نسبت به محیط قانونی و سیاسی، سهم بازار، گرایش بازار، تنوع تقاضا، تکنولوژی و نوآوری، انعطاف‌پذیری تولید، محصولات مخلوط، سطح قیمت، جایابی، ارزش برند و وابستگی به مشتریان و تأمین‌کنندگان شاخص‌های مدیریتی شامل: داده‌های برنامه‌ریزی و پیش‌بینی، گزارش‌دهی به موقع و قابل اعتماد، حسابداری تصدیق‌شده، رفتار انجام شده نسبت به وام‌های بانکی، رفتار انجام شده نسبت به تعهدات مالیاتی، رفتار انجام شده نسبت به تأمین‌کنندگان، رفتار انجام شده نسبت به مشتریان، وثیقه، انگیزش و بهره‌وری نیروی انسانی، پیوستگی مدیریت و تجارت، تجربه و عملکرد گذشته مدیر، تعهد و مهارت تیم مدیریت، عملکرد مدیریت، مدیریت اخلاقی و جهت‌گیری نسبت به ارزش محیط و مشتری.	۱
Yu, Wang, Lai, 2008	شاخص‌های مالی شامل: فروش، سود قبل از مالیات به سرمایه به‌کار گرفته شده، سود قبل از کسر مالیات و استهلاک به بدهی کل، بدهی جاری به‌علاوه بدهی بلندمدت به دارایی کل، بدهی جاری به دارایی کل، دارایی جاری به بدهی جاری، دارایی جاری منهای سرمایه به بدهی جاری، دارایی جاری منهای بدهی جاری به دارایی کل تعداد سال‌هایی که شرکت مشغول به فعالیت بوده است، تعداد روزهای بین پایان سال مالی و تاریخ گزارش سالیانه، تغییر حساب‌برسان در طی سه سال گذشته	۲

با توجه به تنوع روش‌های مورد استفاده در امتیازدهی اعتباری در جدول زیر به‌طور خلاصه به این روش‌ها اشاره شده است.

جدول ۲- روش‌های امتیازدهی اعتباری مورد استفاده در پژوهش‌های انجام یافته در خارج از کشور

ردیف	مدل امتیازدهی اعتباری	پژوهشگر، تاریخ
۱	شبکه عصبی	Yu et al., 2008
۲	ترکیب تحلیل ممیزی و شبکه عصبی	Lee et al., 2002
۳	ترکیب رگرسیون لجستیک و شبکه عصبی	Mileris, 2011; Lin, 2009
۴	درخت تصمیم	Yanping et al., 2012; Lee et al., 2006
۵	ترکیب درخت تصمیم و شبکه عصبی	Kabari&Nwachukwu, 2013
۶	الگوریتم ژنتیک	Chen &Haung, 2003
۷	ماشین بردار پشتیبان	Yu et al., 2010
۸	روش‌های ترکیبی ماشین بردار پشتیبان	Yu et al., 2008
۹	تحلیل پوششی داده‌ها	Min & Lee, 2008; Emel et al., 2003

جدول ۳- روش‌های امتیازدهی اعتباری مورد استفاده در پژوهش‌های انجام یافته در داخل کشور

ردیف	مدل امتیازدهی اعتباری	منبع
۱	رگرسیون لجستیک	میرزایی و همکاران، ۱۳۹۰؛ عرب‌مازار و روئین‌تن، ۱۳۸۵؛ تهرانی و فلاح شمس، ۱۳۸۴
۲	شبکه عصبی	نظری و علی‌دادی، ۱۳۹۲؛ قدسی پور و همکاران، ۱۳۹۱؛
۳	تحلیل پوششی داده‌ها	صفری و همکاران، ۱۳۸۹
۴	تحلیل ممیزی	کریمی، ۱۳۸۹؛ شریعت‌پناهی و هاشمی، ۱۳۸۸
۵	ماشین بردار پشتیبان	راعی و فلاح‌پور، ۱۳۸۷
۶	روش‌های ترکیبی	پورداراب و همکاران، ۱۳۹۰

روش تحقیق

این پژوهش از نظر نوع کاربردی و در حوزه پژوهش‌های میدانی و توصیفی می‌باشد. در این راستا، اطلاعات و داده‌های مورد نیاز از بولتن اطلاعاتی موجود در بانک استخراج شده است.

جامعه آماری این تحقیق، شرکت‌های کوچک و متوسطی هستند که طی سال‌های ۸۲-۱۳۸۱ اقدام به اخذ وام از یکی از شعب بانک تجارت در استان تهران کرده‌اند و صورت‌های مالی این شرکت‌ها توسط حسابرسان معتبر، حسابرسی شده است. از میان اعضای جامعه آماری تعداد ۲۸۲ شرکت به صورت تصادفی انتخاب شدند که این نمونه شامل ۱۴۱ مشتری خوش حساب و ۱۴۱ مشتری بد حساب می‌باشد. با توجه به سه وضعیت مطالبات جاری (تا حداکثر ۲ ماه پس از سررسید تسهیلات

دریافتی نسبت به بازپرداخت وام خود اقدام ننموده‌اند)، سررسید گذشته (بین ۲ تا ۶ ماه پس از سررسید تسهیلات دریافتی اقدامی جهت بازپرداخت وام خود نکرده‌اند) و معوق (بین ۶ تا ۱۸ ماه پس از سررسید، بدهی خود را تسویه ننموده‌اند) که بانک مزبور براساس قوانین بانک مرکزی برای وضعیت مطالبات بانکی خود در نظر گرفته بود، در این تحقیق شرکت‌هایی که وضعیت مطالبات آن‌ها به صورت جاری بوده، جزء گروه مشتریان خوش حساب و شرکت‌هایی که وضعیت مطالبات آن‌ها به صورت معوق و سررسید گذشته بوده، جزء گروه مشتریان بد حساب به شمار آمده‌اند.

در این تحقیق یک متغیر وابسته وجود دارد که همان وضعیت بازپرداخت وام توسط شرکت‌های وام‌گیرنده بوده است که به دو دسته خوش حساب و بد حساب تقسیم شده‌اند. این متغیر از خصوصیت گسسته برخوردار است و می‌تواند دو حالت صفر و یک را به خود اختصاص دهد که مقدار "صفر" برای مشتریان خوش حساب و مقدار "یک" برای مشتریان بد حساب در نظر گرفته شده است.

پس از بررسی و مروری بر ادبیات ریسک اعتباری و همچنین با توجه به داده‌های موجود در بانک و نظر کارشناسان امور بانکی، این نتیجه حاصل شد که معیارهای مالی تأثیر بیشتری بر روی ارزیابی ریسک اعتباری و پیش بینی احتمال بازپرداخت به موقع وام دارند. از این رو با استفاده از ترانزنامه و صورت‌های مالی شرکت‌های وام‌گیرنده از بانک، ۱۳ نسبت مالی استخراج شد. که به عنوان متغیرهای ورودی (مستقل) در نظر گرفته شده‌اند. این نسبت‌های مالی عبارتند از: سرمایه در گردش به دارایی کل (X_1)، ذخیره درآمد به دارایی کل (X_2)، درآمد قبل از بهره و مالیات به دارایی کل (X_3)، فروش به دارایی کل (X_4)، دارایی جاری به بدهی جاری (X_5)، بدهی جاری به کل دارایی (X_6)، درآمد قبل از مالیات به بدهی کل (X_7)، بدهی کل به حقوق صاحبان سهام (X_8)، فروش به بدهی کل (X_9)، دارایی جاری به کل دارایی (X_{10})، فروش خالص به موجودی‌ها (X_{11})، فروش خالص به حقوق صاحبان سهام (X_{12}) و سود خالص به حقوق صاحبان سهام (X_{13}). توجه به این نکته ضروری است که استفاده از نسبت‌های مالی به جای اندازه‌های مالی موجب حذف اندازه شرکت‌ها و قابل مقایسه کردن آن‌ها می‌باشد (طالبی و همکاران، ۱۳۹۰، ص ۲۴۶).

پیش پردازش داده‌ها

در این مرحله تبدیل داده انجام گرفت تا داده‌ها برای داده‌کاوی مناسب شوند که در اینجا از روش نرمال‌سازی Min-Max برای این عمل استفاده شد. این روش یک تبدیل خطی بر روی داده‌های اصلی انجام می‌دهد. اگر Min_a و Max_a به ترتیب حداقل و حداکثر مقادیر یک ویژگی (a) باشند نرمال‌سازی Min-Max طبق فرمول زیر یک مقدار v از a را به مقدار v' تبدیل می‌کند. در این روش رابطه بین مقادیر اصلی حفظ می‌شود. از داده‌های نرمال‌ایز شده در مدل شبکه عصبی تابع شعاع مدار استفاده شده است (غضنفری و همکاران، ۱۳۸۷).

$$v' = \frac{v - Min_a}{Max_a - Min_a} \quad (1)$$

تقسیم مجموعه داده‌ها به دو دسته آموزش و آزمایش برای استفاده در مدل‌های مختلف تحقیق برای پی بردن به قدرت تفکیک‌کنندگی و پیش‌بینی مدل باید داده‌ها را به دو دسته آموزش و آزمون تقسیم کرد تا با استفاده از داده‌های آموزش مدل ساخته شود و با استفاده از داده‌های آزمون مدل اعتبارسنجی شود. این اقدام در رابطه با مدل‌های به‌کار رفته در تحقیق انجام گرفت و به صورت تصادفی، ۸۰ درصد داده‌ها در ساخت مدل استفاده شد و ۲۰ درصد بقیه برای تست مدل به عنوان نمونه شاهد مورد استفاده قرار گرفت.

مدل‌سازی

با استفاده از داده‌های آموزش و مدل پارامتریک رگرسیون لجستیک و مدل ناپارامتریک درخت تقسیم و رگرسیون به ساخت مدل در زمینه طبقه بندی مشتریان بانک مبادرت شد. به این منظور از نرم افزارهای SPSS 16.0 و CLEMENTINE 12.0 استفاده شده است.

ارزیابی مدل‌ها

ماتریس اغتشاش^۱ ابزاری مفید برای تحلیل چگونگی عملکرد روش دسته‌بندی در تشخیص داده‌ها یا مشاهدات دسته‌های مختلف است. برای این‌که یک روش

دسته‌بندی، صحت بالایی داشته باشد، حالت ایده‌آل آن است که اکثر داده‌های مرتبط به مشاهدات بر روی قطر اصلی ماتریس قرار گرفته باشند و بقیه مقادیر ماتریس صفر یا نزدیک به صفر باشند. در این‌جا به توضیح اصطلاحات به‌کار رفته در این ماتریس می‌پردازیم. عنصر "مثبت درست" به مشاهداتی از دسته C_1 دلالت دارد که توسط روش دسته‌بندی به‌درستی تشخیص داده شده است (در این‌جا شرکت‌هایی که بد حساب بودند و به‌درستی طبقه‌بندی شده‌اند). عنصر "منفی درست" به مشاهداتی از دسته C_2 دلالت دارد که توسط روش روش دسته‌بندی به‌درستی تشخیص داده شده است (شرکت‌های خوش‌حسابی که به‌درستی طبقه‌بندی شده‌اند). به‌طور مشابه "منفی غلط" مشاهداتی از دسته C_1 است که به اشتباه در دسته دیگر قرار گرفته است و عنصر "مثبت غلط" مشاهداتی از دسته C_2 که به اشتباه در دسته دیگر قرار گرفته است (غضنفری و همکاران، ۱۳۸۷).

مدل‌های تحقیق

جهت تحلیل داده‌ها در این تحقیق از مدل‌های رگرسیون لجستیک و درخت تقسیم استفاده شده است که در ادامه این مدل‌ها معرفی می‌شوند.

رگرسیون لجستیک

زمانی که متغیر وابسته تحقیق (در این‌جا وضعیت مطالبات) در مقیاس فاصله‌ای یا نسبی نباشد و مقیاس آن به صورت اسمی (دو وجهی یا چند وجهی) باشد برای شناسایی عوامل پیش‌بینی‌کننده تغییرات یک متغیر اسمی باید از روش رگرسیون لجستیک استفاده کرد. در رگرسیون لجستیک به جای حداقل کردن مجذور خطاها (کاری که رگرسیون خطی انجام می‌دهد)، احتمالی را که یک واقعه رخ می‌دهد، حداکثر می‌کند (حبیب‌پور و همکاران، ۱۳۸۸).

تابع رگرسیون لجستیک که شامل چندین متغیر مستقل و یک متغیر وابسته می‌باشد به صورت زیر تعریف می‌شود:

$$\pi = P_r = (Y = 1 | X_1 = x_1, \dots, X_p = x_p) = \frac{e^{\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p}}{1 + e^{\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p}} \quad (2)$$

به منظور ایجاد یک تابع خطی از پارامترهای $\beta_0, \beta_1, \dots, \beta_p$ باید از نسبت بخت‌ها (احتمال وقوع یک پدیده به عدم وقوع آن پدیده) لگاریتم گرفت. بنابراین تابع رگرسیونی لجستیک به صورت ضابطه زیر تعریف می‌شود:

$$\ln\left(\frac{\pi}{1-\pi}\right) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p \quad (3)$$

در رگرسیون لجستیک برازش با کارکردن با لجیت‌ها (لگاریتم نسبت بخت‌ها) انجام می‌پذیرد. تبدیل لجیت الگویی را حاصل می‌کند که نسبت به پارامترهای مدل خطی است. روشی که برای برآورد ضرایب در مدل رگرسیون لجستیک استفاده می‌شود، روش برآورد حداکثر درست‌نمایی^۱ می‌باشد (نیرومند، ۱۳۸۴). آلتمن (۲۰۰۶) بر این عقیده بود که استفاده از لگاریتم داده‌های اصلی به جای خود داده‌ها منجر به نتیجه دقیق‌تری می‌شود. بنابراین در این تحقیق از لگاریتم داده‌های اصلی در مدل رگرسیون لجستیک استفاده شد.

درخت تقسیم و رگرسیون

روش CART می‌تواند در سه مرحله خلاصه شود: مرحله اول، رشد درخت را شامل می‌شود که با استفاده از تکنیک تقسیم‌بندی بازگشتی و شاخص‌های ایجاد انشعاب، متغیرها و نقاط انشعاب انتخاب می‌شوند. برای انتخاب نقطه انشعاب روش‌های مختلفی همچون شاخص جینی، آنتروپی و غیره وجود دارد. در این روش‌ها برمبنای شاخص‌های مبتنی بر ناخالصی به انتخاب نقطه انشعاب مبادرت می‌شود. در نهایت هدف ساخت درخت، به دست آوردن گره‌های خالص، یعنی گره‌هایی که فقط شامل مشاهداتی از یک کلاس تنهاست، می‌باشد که این هدف با کاهش تابع ناخالصی $I(t)$ یا کاهش گوناگونی در هر سطح میسر است. در روش CART به طور معمول از شاخص جینی^۲ برای انتخاب نقطه شکست استفاده می‌شود. در استفاده از این روش، هدف کاهش تابع ناخالصی است که این مفهوم به عنوان تابعی از فراوانی‌های نسبی کلاس‌ها در یک گره به صورت زیر تعریف می‌شود.

1-Maximum likelihood estimation(MLE)
2-Gini Index

$$I(t) = \varphi(p_1, p_2, \dots, p_j) \quad (4)$$

$I(t)$ تابع ناخالصی است، P_j فراوانی نسبی j کلاس مختلف در یک گره است و $\varphi(p_1, p_2, \dots, p_j)$ تابعی متقارن از p_1, p_2, \dots, p_j می باشد. شاخص جینی به صورت زیر تعریف می شود؛

$$\text{Gini}(t) = 1 - \sum p_i^2 \quad (5)$$

حال بهترین انتخاب برای تقسیم مجموعه S به دو مجموعه S_1 و S_2 از معیار زیر تبعیت می کند، یعنی حداقل کردن تابع زیر:

$$I(S) = \frac{S_1}{S_2} [1 - \sum (p_j^f)^2] + \frac{S_2}{S_1} [1 - \sum (p_j^f)^2] \quad (6)$$

که p_j^f نشان دهنده مقدار فراوانی نسبی در انشعاب سمت چپ و p_j^f نشان دهنده مقدار فراوانی نسبی در انشعاب سمت راست می باشد.

مرحله دوم، شامل هرس کردن درخت می شود. هرس کردن درخت به دو دسته روش پیش هرس و پس هرس تقسیم می شود که در روش CART از روش پس هرس با معیار نرخ پیچیدگی استفاده می شود. در روش پس هرس پایین ترین گره هرس نشده به گره برگ تبدیل می شود و با کلاس با فراوانی بیشتر در بین شاخه های سابقش برچسب زده می شود. برای هر گره غیر برگ در درخت، این الگوریتم نرخ خطای مورد انتظار را محاسبه می کند که اگر این زیر درخت در این گره هرس می شد، رخ می داد. بعد نرخ خطای مورد انتظار حاصل اگر گره هرس نمی شد با استفاده از نرخ های خطا برای هر شاخه محاسبه می شود. اگر هرس کردن این گره به نرخ خطای مورد انتظار بزرگتری منجر شود، آن گاه زیر درخت حفظ می شود. در غیر این صورت، زیر درخت هرس می شود. بعد از به وجود آوردن مجموعه ای از درختان به تدریج هرس شده یک مجموعه آزمون مستقل برای ارزیابی دقت هر مدل به کار می رود. درخت تصمیمی که نرخ خطای مورد انتظار را حداقل می کند، ترجیح داده می شود.

معیار نرخ پیچیدگی برای یک درخت در روش CART از فرمول زیر به دست می آید:

$$C_{\alpha}(T) = R(T) + \alpha |\tilde{T}| \quad (7)$$

$C_{\alpha}(T)$ هزینه کل یا نرخ پیچیدگی درخت T می باشد. $R(T)$ کسری از داده های آزمون است که به وسیله درخت T اشتباه رده بندی شده است. $|\tilde{T}|$ تعداد برگ های درخت T و α مقدار جریمه برای هر گره می باشد.

در مرحله نهایی، درخت بهینه انتخاب می شود. بهترین راه جهت انتخاب درخت نهایی از دنباله ایجاد شده با هزینه پیچیدگی هرس کردن عبارت از انتخاب درختی با پایین ترین نرخ خطا در یک مجموعه آزمون است. برآورد نرخ خطای درخت T بر نمونه آزمون را با $R^{ts}(T)$ نشان می دهیم. برای برآورد نرخ خطای واقعی از خطای استاندارد R^{ts} استفاده می شود که فرمول آن به شرح زیر است:

$$SE(R^{ts}) = \sqrt{\frac{R^{ts}(1 - R^{ts})}{n_{test}}} \quad (8)$$

که در آن n_{test} تعداد عناصر در مجموعه آزمون می باشد (مشکانی و همکاران، ۱۳۸۸؛ لی و همکاران، ۲۰۰۶).

نتایج حاصل از تجزیه و تحلیل داده ها

در این بخش نتایج حاصل از تحلیل اطلاعات از روش های مختلف آورده شده است.

روش رگرسیون لجستیک

جدول (۴) نشان می دهد که کدام متغیرها بعد از اجرای رگرسیون لجستیک در مدل باقی مانده اند. در این جدول با محاسبه آماره والد^۱ معنی داری حضور هر متغیر مستقل در مدل مورد آزمون قرار می گیرد که از طریق سطح معنی داری آن می توان به این امر پیبرد. در مورد این آماره اگر سطح معنی داری کوچک تر از (۰/۰۵) باشد، در آن صورت نتیجه می گیریم که وجود آن متغیر در مدل مفید و اثر آن معنی دار است.

1-Wald

جدول ۴- ضرایب رگرسیون لجستیک

	B	S.E.	Wald	Df	Sig.
Step ۴	-۲۵/۰۵۱	۵/۴۲۱	۲۱/۳۵۹	۱	۰
X ₁	-۸/۴۳۱	۱/۱۸۷	۵۰/۴۲۰	۱	۰
X ₅	۲/۹۰۲	۱/۰۵۵	۷/۷۵۷	۱	۰/۰۰۶
X ₆	-۱/۶۲۲	۰/۶۷۲	۵/۸۱۹	۱	۰/۰۱۶
X ₇	-۱۲/۳۱۸	۱/۷۹۵	۴۹/۰۵۵	۱	۰
ثابت					

براساس نتایج این جدول می‌توانیم مدل رگرسیونی لجستیک را براساس مرحله چهارم به شرح زیر نشان دهیم:

$$\ln(p/1-p) = -۲۵/۰۵۱ (\text{عدد ثابت}) - ۱۲/۳۱۸ (\text{درآمد قبل از بهره و مالیات به بدهی کل}) + ۲/۹۰۲ (\text{دارایی جاری به بدهی جاری}) + ۸/۴۳۱ (\text{تأثیر متغیرهای سرمایه در گردش به دارایی کل، دارایی جاری به بدهی جاری و درآمد قبل از بهره و مالیات به بدهی کل بر متغیر وابسته منفی و خلاف جهت یکدیگر می‌باشد، در حالی که تأثیر متغیر بدهی جاری به کل دارایی بر متغیر وابسته مثبت و هم جهت است.})$$

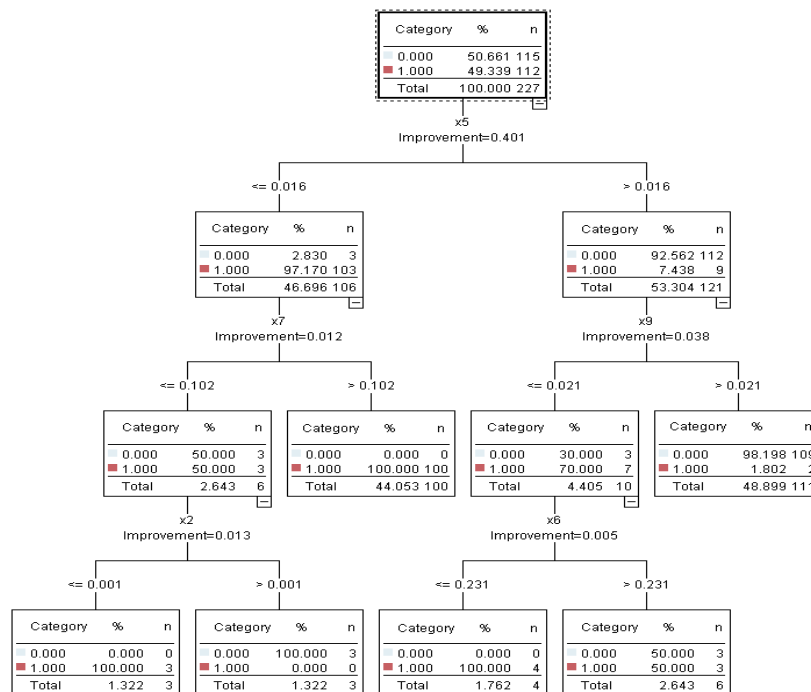
جدول (۵) نشان دهنده جدول طبقه‌بندی با استفاده از داده‌های آموزش با حد آستانه ۰/۰۵ است و نشان‌دهنده این است که کارایی مدل تحقیق ۹۲/۹ درصد است. بدین معنا که تعداد ریسک اعتباری ۲۱۰ شرکت به طور صحیح و ریسک اعتباری ۱۶ شرکت به طور اشتباه پیش‌بینی شده است.

جدول ۵- بررسی قدرت پیش‌بینی مدل در حد آستانه ۰/۵

حد آستانه ۰/۵	Y=1 (بد حساب)		Y=0 (خوش حساب)	
	ریسک اعتباری	درجه حساسیت	درجه تشخیص	ریسک تجاری
P(Y) ≤ 0.5	۹		۱۰۶	
P(Y) > 0.5	۱۰۴		۷	
جمع کل	۱۱۳		۱۱۳	
میزان نیکویی برازش (کارایی مدل)		۹۲/۰٪	۹۳/۸٪	۶/۲٪

نتایج حاصل از درخت تقسیم و رگرسیون

همان طور که می‌دانیم با استفاده از درخت تصمیم می‌توان مجموعه بزرگی از داده‌ها را به وسیله یک مجموعه قوانین ساده متوالی به مجموعه کوچکتر تقسیم کرد. بنابراین با استفاده از درخت تقسیم و رگرسیون به تحلیل مسئله پرداخته شد. در شکل (۱) درخت تصمیم ساخته شده با استفاده از داده‌های آموزش نشان داده شده است.



شکل ۱- درخت تصمیم ساخته شده با استفاده از داده‌های آموزش

قوانین حاصل از درخت تصمیم

قوانین استخراج شده در این قسمت به شرح زیر است:

قوانین برای طبقه ۰

- قانون ۱: اگر $X_5 \leq 0.016$ ، $X_7 \leq 0.102$ و $X_2 > 0.001$ آنگاه ۰
- قانون ۲: اگر $X_5 > 0.016$ ، $X_9 \leq 0.021$ و $X_6 > 0.231$ آنگاه ۰
- قانون ۳: اگر $X_5 > 0.016$ و $X_9 > 0.021$ آنگاه ۰

قوانین برای طبقه ۱

- قانون ۱: اگر $X_5 \leq 0.016$ ، $X_7 \leq 0.102$ و $X_2 \leq 0.001$ آنگاه ۱
- قانون ۲: اگر $X_5 \leq 0.016$ و $X_7 > 0.102$ آنگاه ۱
- قانون ۳: اگر $X_5 > 0.016$ ، $X_9 \leq 0.021$ و $X_6 \leq 0.231$ آنگاه ۱

در شکل زیر جدول طبقه‌بندی حاصل از روش درخت تقسیم و رگرسیون با استفاده از داده‌های آموزش نشان داده شده است. به کارگیری این روش باعث شده است تا کارایی مدل به ۹۷/۷ درصد افزایش یابد و در این مرحله ریسک اعتباری تنها ۵ شرکت به طور پیش‌بینی نشده است.

جدول ۶- جدول طبقه‌بندی با استفاده از داده‌های آموزش

پیش‌بینی شده مشاهده شده	Y=1 (بدحساب)		Y=0 (خوش‌حساب)	
Y=0	۰		۱۱۵	
Y=1	۱۰۷		۵	
جمع کل	۱۰۷		۱۲۰	
ریسک تجاری	ریسک اعتباری	درجه حساسیت	درجه تشخیص	ریسک اعتباری
%۴/۱۶	۰	%۱۰۰	%۹۵/۸	%۹۷/۷

ارزیابی و مقایسه مدل‌های ارائه شده در تحقیق

نتایج مربوط به مدل‌های رگرسیون لجستیک و درخت تقسیم و رگرسیون مربوط به متغیر وابسته در جداول (۷)، (۸) و (۹) ارائه شده است. جدول طبقه‌بندی با استفاده از داده‌های آزمون در شکل زیر نشان داده شده است. با استفاده از این جدول می‌توان قدرت مدل در تفکیک افراد در طبقات متغیر وابسته را تعیین کرد و به صحت مدل در طبقه‌بندی افراد پی برد.

به نسبتی از مشاهدات ($Y=1$) که به درستی پیش‌بینی شده‌اند، درجه حساسیت و به نسبتی از مشاهدات ($Y=0$) که به درستی پیش‌بینی شده‌اند، درجه تشخیص مدل گفته می‌شود. به عبارت دیگر خطای نوع اول (ریسک اعتباری) که عبارت است از طبقه‌بندی مشتری بد در میان مشتریان خوب برابر است با عدد یک منهای درجه حساسیت و مقدار خطای نوع دوم یعنی ریسک تجاری که عبارت است از طبقه‌بندی مشتری خوب در میان مشتریان بد برابر است با عدد یک منهای درجه تشخیص (عرب‌مازار و همکاران، ۱۳۸۵).

جدول ۷- ماتریس اغتشاش

مدل	نقش داده‌ها	منفی غلط	منفی درست	مثبت غلط	مثبت درست
درخت تقسیم و رگرسیون	تست	۴	۲۳	۳	۲۳
رگرسیون لجستیک لگاریتمی	تست	۲	۲۰	۸	۲۶

اعداد حاصل در جدول ۷ نشان‌دهنده این است که در مدل درخت تقسیم و رگرسیون تعداد ۴ مشتری بد به عنوان مشتری خوب تشخیص داده شده و تعداد ۳ مشتری خوب نیز در بین مشتریان بد قرار گرفته‌اند. بنابراین سطح ریسک اعتباری و ریسک تجاری به ترتیب $14/9$ و $11/6$ درصد است که در جدول ۸ نشان داده شده است. این اعداد در حقیقت سطح کارایی این مدل است. اعداد مربوط به مدل رگرسیون لجستیک لگاریتمی نیز به همین صورت تفسیر می‌شوند.

جدول ۸- پارامترهای کارایی مدل

مدل	نقش داده‌ها	درجه حساسیت	ریسک اعتباری	درجه تشخیص	ریسک تجاری
درخت تقسیم و رگرسیون	تست	$85/1$	$14/9$	$88/4$	$11/6$
رگرسیون لجستیک لگاریتمی	تست	$92/8$	$7/2$	$71/42$	$28/58$

با توجه به میزان ریسک اعتباری و تجاری هر کدام از مدل‌ها کارایی نهایی این مدل‌ها بر اساس میانگین این دو ریسک به شرح جدول ۹ است:

جدول ۹- کارایی مدل

درخت تقسیم و رگرسیون	رگرسیون لجستیک لگاریتمی
$86/7$	$82/1$

نتیجه‌گیری

از آن‌جا که یکی از وظایف مهم بانکها در زمینه اعطای تسهیلات اعتباری است و انجام درست این وظیفه موجبات رونق اقتصادی را فراهم می‌آورد؛ از این‌رو توجه به این امر در سیستم بانکی کشور از اهمیت ویژه‌ای برخوردار است. رتبه‌بندی و امتیازدهی اعتباری متقاضیان تسهیلات اعتباری بانکها به منظور پیش‌بینی احتمال عدم بازپرداخت وام، یکی از سازوکارهای مهم مدیریت ریسک اعتباری است که برای کنترل ریسک‌های موجود در فعالیت‌های وام‌دهی بانکها مورد استفاده قرار می‌گیرد. به همین دلیل، ساخت یک مدل امتیازدهی اعتباری دقیق برای سیستم بانکداری کشور بسیار ضروری می‌نماید. هدف از این مقاله مقایسه توانمندی مدل‌های مختلف پارامتریک و ناپارامتریک در طبقه‌بندی اعتباری است. با مروری بر ادبیات تحقیق درمی‌یابیم که بهترین مدل برای ارزیابی ریسک اعتباری وجود ندارد و عملکرد این مدل‌ها به ساختار داده‌ها، کیفیت داده‌ها و هدف طبقه‌بندی بستگی دارد. اما در این تحقیق، رویکرد مقایسه‌ای نشان می‌دهد که مدل درخت تقسیم و رگرسیون نسبت به مدل‌های پارامتریک همچون رگرسیون لجستیک از قدرت پیش‌بینی بالایی برخوردارند و به‌طور کلی از نظر دقت طبقه‌بندی با روش‌های پارامتریک قابل رقابت هستند. این امر در حالی است که این مدل‌ها نیازی به بررسی پیش فرض‌های خاص ندارد و دارای مزیت‌های سادگی و سرعت ساخت و فهم آسان هستند.

نتایج حاصل از تحلیل اطلاعات در روش درخت تقسیم و رگرسیون بیانگر این است که متغیر نسبت دارایی جاری به بدهی جاری اولین متغیر ایجادکننده انشعاب است و بیشترین تأثیر را در ایجاد شاخه‌های همگن دارد. سپس به‌ترتیب متغیرهای فروش به بدهی کل (X9)، درآمد قبل از بهره و مالیات به بدهی کل (X7)، ذخیره درآمد به دارایی کل (X2) و بدهی جاری به کل دارایی (X6) بیشترین تأثیر را در طبقه‌بندی شرکت‌های وام‌گیرنده از نظر وضعیت بازپرداخت وام دارند. بنابراین بانک هنگام تصمیم در اعطای وام به شرکت‌ها باید به این نسبت‌های مالی اهمیت بیشتری بدهد تا حجم مطالبات معوق بانک را کاهش دهد. مقایسه نتایج حاصل از این تحقیق با تحقیق عبدلی و همکاران (۱۳۹۴) نشان می‌دهد که مهم‌ترین عوامل موثر بر ریسک اعتباری این دو تحقیق با یکدیگر همخوانی دارد. همچنین نتایج این تحقیق با تحقیقات عرب مازار و همکاران (۱۳۸۵) و صفری (۱۳۸۹) تطابق دارد.

با استقرار این مدل‌ها، این امکان برای سیستم بانکی کشور فراهم می‌آید تا بتواند تضمین‌های دریافتی از مشتریان را با توجه به ریسک آن‌ها مشخص کند، به نحوی که بانک بتواند ریسک اعتباری مشتریان را پوشش داده و بازگشت منابع بانک را تضمین کند. لذا پیشنهاد می‌شود که بانک‌ها نسبت به ایجاد بانک اطلاعاتی قوی از متغیرهای موثر بر ریسک اعتباری مشتریان اقدام کرده و بر اساس مدل‌های تحقیق نسبت به رتبه‌بندی و تعیین سطح ریسک مشتریان خود اقدام کنند.

منابع و مأخذ

- ۱- تهرانی، رضا. فلاح شمس، میرفیض (۱۳۸۴). طراحی و تبیین مدل ریسک اعتباری در نظام بانکی کشور. مجله علوم اجتماعی و انسانی دانشگاه شیراز، ۴۳، ۴۵-۶۰.
- ۲- چترجی، سمپریت. اس. هادی، علی. پرایس، برترام (۱۳۸۴). تحلیل رگرسیون با مثال. حسینعلی نیرومند. مشهد: دانشگاه فردوسی مشهد.
- ۳- حبیب‌پور، کرم. صفری، رضا (۱۳۸۸). راهنمای جامع کاربرد SPSS در تحقیقات پیمایش (تحلیل داده‌های کمی). تهران: لویه.
- ۴- راعی، رضا. فلاح‌پور، سعید (۱۳۸۷). کاربرد ماشین بردار پشتیبان در پیش‌بینی درماندگی مالی شرکت‌ها با استفاده از نسبت‌های مالی. بررسی‌های حسابداری و حسابرسی، ۵۳، ۱۷-۳۴.
- ۵- شریعت‌پناهی، س. برکادهی، س. (۱۳۸۸). ارائه مدلی برای اعتبار سنجی مشتریان در بانک صنعت و معدن. فصلنامه مطالعات حسابداری، ۲۱، ۶۱-۸۲.
- ۶- صفری، سعید. ابراهیمی شقاقی، مرضیه. شیخ، محمدجواد (۱۳۸۹). مدیریت ریسک اعتباری مشتریان حقوقی در بانک‌های تجاری با رویکرد تحلیل پوششی داده‌ها (رتبه بندی اعتباری). پژوهش‌های مدیریت در ایران (مدرس علوم انسانی)، ۱۴، ۱۳۷-۱۶۴.
- ۷- طالبی، محمدصادق. شیرزادی، نازنی (۱۳۹۰). ریسک اعتباری: اندازه‌گیری و مدیریت. انتشارات سمت.
- ۸- عبدلی، قهرمان. فرد حریری، علیرضا (۱۳۹۴). الگوسازی سنجش ریسک اعتباری مشتریان حقوقی بانک رفاه. نظریه‌های کاربردی اقتصاد، ۲(۱)، ۱-۲۴.
- ۹- عرب‌مازار، عباس. روئین‌تن، پونه (۱۳۸۵). عوامل مؤثر بر ریسک اعتباری مشتریان بانکی؛ مطالعه موردی بانک کشاورزی. جستارهای اقتصادی، ۶، ۴۵-۸۰.

- ۱۰- غضنفری، مهدی. علیزاده، سمیه. تیمورپور، بابک (۱۳۸۷). داده‌کاوی و کشف دانش. تهران: دانشگاه علم و صنعت ایران.
- ۱۱- قدسی پور، حسن. سالاری، میثم. دلاوری، وحید (۱۳۹۱). ارزیابی ریسک اعتباری شرکت های وام گیرنده از بانک با استفاده از تحلیل سلسله مراتبی فازی و شبکه عصبی ترکیبی درجه بالا، نشریه بین المللی مهندسی صنایع و مدیریت تولید، شماره ۱، جلد ۲۳.
- ۱۲- کرمی، مصطفی (۱۳۸۹). ارائه مدلی برای ارزیابی وضعیت ریسک اعتباری مشتریان قبل از اعطای تسهیلات بانکی. بانک، ۱۴۲-۱۴۶.
- ۱۳- مشکانی، علی،، ناظمی، عبدالرضا (۱۳۸۸). مقدمه‌ای بر داده‌کاوی. چاپ و انتشارات دانشگاه فردوسی مشهد، چاپ اول.
- ۱۴- میرزایی، حسین. نظریان، رافیک. باقری، رعنا (۱۳۹۰). بررسی عوامل مؤثر بر ریسک اعتباری اشخاص حقوقی بانکها. فصل‌نامه روند پژوهش‌های اقتصادی، شماره ۵۸، ۶۷-۹۸.
- 15-Altman, Edward I. Sabato, Gabriele. (2006). Modeling Credit Risk for SMEs: Evidence from the US Market. In Proceedings of the international conference C.R.E.D.I. 25-26 September, Venice, Italy.
- 16-Basel Committee on Banking Supervision.; “Credit rating and complementary sources of credit quality information”; Working Paper, No.3, 2000.
- 17-Chen, W, Xiang, G. Liu, Y. Wang, K. (2012). Credit risk Evaluation by hybrid data mining technique. *Systems Engineering Procedia*, 3(0), 1, 20-94.
- 18-Emel, AhmetBurak. Oral, Muhittin. Reisman, Arnold. Yolalan, Reha. (2003). A credit scoring approach for the commercial banking sector. *Socio-Economic Planning Sciences*, 37, 103-123.
- 19-Greuning, Hennie van. Bratanovic, Sonja Brajovic. (2009). Analyzing Banking Risk: A Framework for Assessing Corporate Governance and Risk Management. World Bank Publications.
- 20-Kabari, L. G. Nwachukwu, E. O. (2013). Credit Risk Evaluating System Using Decision Tree – Neuro Based Model. *International Journal of Engineering Research & Technology*, 2, 2738-2745.
- 21-Lee, Tian-Shyug. Chiu, Chih-Chou. Lu, Chi-Jie. Chen, I-Fei. (2002). Credit scoring using the hybrid neural discriminant technique. *Expert Systems with Applications*, 23, 245-254.
- 22-Lee, T. S., Chiu, C. C., Chou, Y. C., Lu, C. J. (2006). Mining the customer credit using classification and regression tree and multivariate adaptive regression splines. *Computational Statistics & Data Analysis*, 50, 1113-1130.
- 23-Lin, S. (2009). A new two-stage hybrid approach of credit risk in banking industry. *Expert Systems with Applications*, 36, 8333-8341.
- 24-Mileris, Ricardas. Boguslauskas, Vytautas. (2011). Credit Risk Estimation Model Development Process: Main Steps and Model Improvement. *Inzinerine Ekonomika-Engineering Economics*, 22, 126-133.
- 25-Min, Jae H. Lee, Young-Chan. (2008). A practical approach to credit scoring. *Expert Systems with Applications*, 35, 1762-1770.

- 26-Nazari, Mohsen. Alidadi, Mojtaba. (2013). Measuring Credit Risk of Bank Customers Using Artificial Neural Network. *Journal of Management Research*, 5, 17-27.
- 27-Pourdarab, Sanaz. Nadali, Ahmad. EslamiNosratabadi, Hamid.(2011). A Hybrid Method for Credit Risk Assessment of Bank Customers. *International Journal of Trade, Economics and Finance*, 2, 125-131.
- 28-Singh, S., Murthi, B. P. S., & Steffes, E. (2013). Developing a measure of risk adjusted revenue (RAR) in credit cards market: Implications for customer relationship management. *European Journal of Operational Research*, 224(2), 425-434.
- 29-Soares, J., Pina, J., Ribeiro, M., Catalão-Lopes, M. (2011). Quantitative vs. Qualitative Criteria for Credit Risk Assessment. *Frontiers in Finance and Economics*, 8(1), 69-87.
- 30-Yanping, Y., Zhengming, Q., Min, Y., Rui, G., Liting, F., Penghui, G. (2012). Research on the Application of Decision Tree to the Analysis of Individual Credit Risk. *Information Technology*, 25, 209-214.
- 31-Yu, L., Wang, S., Lai, K. (2008). Credit risk assessment with a multistage neural network ensemble learning approach. *Expert Systems with Applications*. 34, 1434–1444.
- 32-Yu, L., Yue, W., Wang, S., Lai, K.K.(2010). Support vector machine based multiagent ensemble learning for credit risk evaluation. *Expert Systems with Applications*, 37, 1351-1360.