



نخستین کنفرانس ملی توسعه مدیریت پولی و بانکی

تحول در مدیریت نظام بانکی، گامی بروسای حاشیه اقتصادی

۸ و ۹ بهمن ماه ۱۳۹۲ تهران - مرکز همایش‌های بین‌المللی صدا و سیما



رتبه‌بندی اعتباری مشتریان حقیقی بانک ملت با استفاده از روش‌های داده‌کاوی

ابراهیم فرید (بایزیدی)^۱ * محمود محمدی^۲

کارشناس ارشد آمار ریاضی، ، : Ebrahim.bayazidi@gmail.com

استادیار دانشگاه آزاد اسلامی واحد تهران مرکز، دانشکده مدیریت: mah.mohammadi@iauctb.ac.ir

چکیده

هدف از ارائه این مقاله اندازه‌گیری ریسک اعتباری مشتریان حقیقی بانک ملت با استفاده از روش‌های داده‌کاوی مانند مدل شبکه‌های عصبی و مدل ساختار درختی در مقابل روش‌های پارامتریک آماری مانند رگرسیون لجستیک و ارائه راهکارهای صحیح جهت مقابله با افزایش مطالبات معوق می‌باشد. بدین منظور از ۵۲۴ پرونده اعتباری مشتریان حقیقی که طی سال‌های ۱۳۷۶-۱۳۸۸ تسهیلات دریافت نموده‌اند؛ با نمونه‌گیری طبقه‌بندی تصادفی ساده و روش انتساب متناسب برای جمع‌آوری نمونه آماری استفاده شد. در این مطالعه پس از بررسی پرونده‌های اعتباری هر یک از نمونه‌ها، در ابتدا ۷ متغیر توضیحی شامل متغیرهای کمی و کیفی شناسایی شده و مورد بررسی قرار گرفت. نتایج این مطالعه نشان می‌دهد که در بین متغیرهای مرتبط با ویژگی‌های شخصیتی متقاضی، متغیرهایی نظیر جنسیت و شغل و در بین متغیرهای مرتبط با ویژگی‌های تسهیلات اعطایی، متغیرهایی نظیر نرخ سود، نرخ جریمه، تعداد تسهیلات دریافتی، مدت تسهیلات و مبلغ تسهیلات بیشترین اهمیت را در جداسازی مشتریان خوب و بد از هم داشته و اثر معناداری آن بر متغیر وابسته (احتمال عدم بازپرداخت) به وسیله آزمون‌های آماری تأیید شده است. ملاک مقایسه سه مدل، دقت طبقه‌بندی مشتریان خوش حساب و بد حساب می‌باشد. بنابراین دقت هر روش پس از محاسبه در نرم‌افزار Spss برای مدل رگرسیون لجستیک برابر ۶۶٫۹٪، مدل ساختار درختی برابر ۶۷٫۴٪ و برای مدل شبکه عصبی مصنوعی برابر ۷۰٫۱٪ به دست آمد. بنابراین برازش مدل شبکه عصبی مصنوعی در این زمینه کارآمدتر می‌باشد.

واژه‌های کلیدی: رتبه‌بندی اعتباری، داده‌کاوی، شبکه عصبی، ساختار درختی، رگرسیون لجستیک، ریسک



۱- مقدمه

ارتباط صحیح بین نظام‌های تولیدی و مالی در هر کشوری از مهم‌ترین عوامل در رشد و توسعه اقتصادی محسوب می‌شود. بانک‌ها به عنوان بخش مهم نظام مالی، نقش اساسی را در تأمین مالی بخش‌های تولیدی، تجاری، مصرفی و حتی دولتی بر عهده دارند. بانکداری یکی از با اهمیت‌ترین فعالیت‌های اقتصادی به شمار می‌رود و بانک‌ها با سازماندهی و هدایت دریافت‌ها و پرداخت‌ها می‌توانند مبادلات بازرگانی را تسهیل و موجب گسترش بازارها، رشد و شکوفایی اقتصادی شوند و همچنین با تجهیز پس‌اندازها و هدایت آن‌ها به سمت بنگاه‌های تولیدی، تجاری و خانوارها خدمات ارزشمندی را ارائه دهند. در واقع در بازارهای پولی، بانک‌ها به عنوان یک کانال ارتباطی بین پس‌اندازکنندگان و سرمایه‌گذاران بر مبنای استفاده بهینه و کارا از عامل سرمایه، عمل می‌کنند. اعطای تسهیلات بخش مهمی از عملیات هر بانک را تشکیل می‌دهد.

از طرف دیگر یکی از فعالیت‌های اساسی و عمده بانک‌ها تجهیز و تخصیص منابع مالی می‌باشد. بانک‌ها می‌توانند تحت عنوان سپرده‌های قرض‌الحسنه جاری و سپرده‌های سرمایه‌گذاری به جمع‌آوری منابع بپردازند. بخش دیگر موفقیت و بقای بانک به تخصیص منابع مالی (اعتبارات) مرتبط است. بانک‌ها می‌بایست این منابع مالی محدود را به صورت مطلوب به تولید کالا و خدمات تخصیص دهند. امروزه بانک‌ها در تخصیص اعتبار به متقاضیان تسهیلات به منظور ارتقاء بهره‌وری بانکی علاوه بر عوامل فوق، به عوامل دیگری نیز توجه دارند. برخی از مهم‌ترین این عوامل عبارتند از:

۱. رضایت متقاضیان تسهیلات
۲. کاهش مطالبات معوق
۳. افزایش سود حاصل از اعطای اعتبار
۴. بهبود وضعیت اقتصادی کشور

با توجه به مطالب فوق، نکته‌ای که در آن برجسته می‌باشد توجه به کاهش مطالبات معوق است که یکی از اهداف اصلی بانک‌ها می‌باشد. رسیدن به این هدف یعنی حداقل نمودن مطالبات معوق به طور خودکار اهداف دیگر بانک را نیز عملی می‌کند. کاهش مطالبات معوق باعث می‌شود جریان پول خارج شده از بانک سریع‌تر در اختیار قرار گیرد تا با آن به افراد بیشتری، تسهیلات اعطا کنند که این خود موجب رضایت‌مندی مشتریان می‌شود. همچنین کمک می‌کند تا بانک‌ها با در اختیار قرار دادن تسهیلات در اختیار سرمایه‌گذاران اقتصادی چرخه صنعت به گردش درآید و به توسعه و تحول اقتصادی کشور کمک کند. بنابراین در این تحقیق عوامل موثر در جهت کاهش مطالبات معوق توسط روش‌های آماری و داده‌کاوی مدل‌بندی شده و دقت آنها با هم مقایسه می‌شود.

۲- مبانی و ادبیات نظری

بانک‌ها به منظور آگاهی از نیازمندی‌های مشتریان خود، در اعطای تسهیلات اعتباری باید به شناسایی ویژگی‌های آنها بپردازند. این امر از طریق اعتبارسنجی^۱ منجر به کاهش ریسک‌های بانکی از جمله ریسک اعتباری^۲ می‌شود (توماس^۳ ۲۰۰۶). اعتبارسنجی به عملی اطلاق می‌شود که در آن اعتبار مشتریان حقیقی و حقوقی مؤسسات مالی اعتباری و بانک‌ها با توجه به اطلاعات دریافتی از آن‌ها اندازه‌گیری شده و امکان شناخت بیشتر را نسبت به وضعیت و توان مالی افراد جهت بازپرداخت تسهیلات دریافتی و دریافت خدمات بیشتر فراهم می‌کند. بر اساس این روش، ریسک اعتباری افراد اندازه‌گیری شده و افراد و مشتریان بر اساس ریسک اعتباری خود طبقه‌بندی و امتیازدهی می‌شوند (خداوردی، ۱۳۸۸).

رتبه‌بندی اعتباری ابزاری برای مدیریت ریسک است که با استفاده از آمار و اطلاعات کمی متقاضیان تسهیلات و نیز تکنیک‌های آماری به رتبه‌بندی مشتریان می‌پردازد. رتبه‌بندی اعتباری اثرات ویژگی‌های مختلف متقاضیان تسهیلات را بر نکول شدن آن به طور مجزا بررسی کرده و با محاسبه احتمال نکول شدن تسهیلات متقاضی، به رتبه‌بندی اعتباری آن‌ها در واحدهای ریسک می‌پردازد. مدل‌های رتبه‌بندی اعتباری متقاضیان اعتبار را به دو گروه اعتباری خوب و بد تقسیم می‌کند. گروه اعتباری خوب گروهی است که دیون خود را به موقع بازپرداخت می‌کنند و گروه اعتباری بد گروهی است که به احتمال بالا دیون‌شان نکول خواهد شد. با توجه به روند توسعه و

^۱. Accreditation

^۲. Credit Risk

^۳. Thomas



پویایی صنعت اعتبار، امروزه این صنعت نقش مهمی در اقتصاد کشورها یافته است. ۵٪ اگر چه افزایش تقاضای اعتبار، افزایش رقابت و به وجود آمدن کانال‌های جدید در فضای اقتصاد نوین، فرصت‌های جدیدی برای مؤسسات اعتباردهنده به وجود آورده است، اما از طرفی آن‌ها را نیازمند ابزارها و روش‌های جدیدی نیز نموده است. این مسئله، مؤسسات مزبور را به سمت تجدید نظر، توانمند سازی و ورود فن آوری‌های جدید در فرایند مدیریت اعتبار سوق داده است.

مدل‌های رتبه‌بندی اعتباری، یکی از مهم‌ترین و اساسی‌ترین سیستم‌های تصمیم‌گیری هستند که بخش عمده‌ای از اطلاعات مورد نیاز مؤسسات اعتباردهنده در مدیریت اعتبار را فراهم می‌کنند. هدف مدل‌های رتبه‌بندی اعتباری، پیش‌بینی احتمال عدم پرداخت اعتبار از سوی مشتری و یا طبقه‌بندی متقاضیان اعتبار به دو گروه خوب و بد است. به عبارت دیگر، رتبه‌بندی اعتباری مجموعه‌ای از مدل‌های تصمیم‌گیر و روش‌های مرتبط با آن‌هاست که به اعتبار دهندگان در اعطای اعتبار به مشتریان کمک می‌کند. یک فرض دهنده می‌بایست دو نوع تصمیم بگیرد: اول اینکه آیا به یک متقاضی جدید اعتبار بدهد یا نه، و دوم اینکه با مشتریان موجود چگونه رفتار کند. روش‌هایی که در مورد اول تصمیم‌گیری می‌کنند، رتبه‌بندی اعتباری نامیده می‌شوند، در حالی که روش‌هایی که در مورد مشتریان موجود تصمیم‌گیری می‌کنند رتبه‌بندی رفتاری نامیده می‌شود.

مؤسسات اعتباری و بانک‌ها به دو دلیل به وجود سیستمی برای رتبه‌بندی اعتباری مشتریان خود نیازمندند. سیستم رتبه‌بندی اعتباری مشتریان این امکان را برای بانک‌ها و مؤسسات اعتباری فراهم می‌کند که با اتکا به چنین سیستمی و بر اساس نرخ‌های تکلیفی موجود، ریسک پرتفوی اعتباری خود را تا حد ممکن کاهش داده و از بین متقاضیان دریافت تسهیلات، معتبرترین و کم ریسک‌ترین مشتریان را گزینش نمایند. در مؤسسات اعتباری که امکان تعیین نرخ تسهیلات بر اساس ریسک و درجه اعتباری مشتریان می‌باشد، سیستم رتبه‌بندی اعتباری می‌تواند این گونه سازمان‌ها را در طراحی پرتفوی اعتباری خود بر اساس رعایت اصل تنوع یاری دهد.

۳- شناسایی ریسک‌ها

تقریباً هر فعالیتی که توسط بانک صورت می‌پذیرد، در بطن خود با ریسک‌هایی مواجه است. ریسک در بانکداری، بنا به دلایلی مانند تعداد و تنوع عملیات بانکی، ماهیت‌های متفاوت این عملیات، وضعیت سرمایه بانک و محدودیت میزان آن، حفظ منافع سهامداران بانک، وضعیت منابع سپرده‌گذاران و فراوانی آن، حفظ منافع سپرده‌گذاران، متفاوت بودن منافع بانک و سپرده‌گذاران، نوع وظیفه بانک‌ها که آن‌ها را متمایز از بنگاه‌های دیگر اقتصادی می‌کند، تعداد گیرندگان تسهیلات اعتباری، متفاوت بودن وضعیت مالی هر یک از گیرندگان تسهیلات، سر و کار داشتن گروه زیادی از کارکنان بانک با منابع مالی، نبود تخصص کافی در زمینه‌های مدیریت منابع، ثبت عملیات زیاد مالی، نقل و انتقالات زیاد پولی داخلی، خارجی و... ماهیتی متفاوت از ریسک در دیگر بخش‌های اقتصادی دارد. به عبارتی مدیریت ریسک در بانک‌ها حساس‌تر، پیچیده‌تر و مشکل‌تر از مدیریت ریسک در بخش‌های دیگر اقتصادی می‌باشد.

مدیریت ریسک در وهله اول مستلزم شناسایی ریسک‌های فعالیت بانک می‌باشد. البته دسته‌بندی ریسک در دیدگاه‌های مختلف متمایز از هم هستند. به عنوان مثال طبق استانداردهای کمیته بال، ریسک‌های مختلفی که در فعالیت بانکی باید مورد توجه قرار گیرد عبارتند از: ۱- ریسک اعتباری^۱ ۲- ریسک بازار^۲ شامل الف: ریسک نوسانات نرخ ارز^۳ ب: ریسک نوسانات قیمت بازار^۴ ج: ریسک نوسانات نرخ بهره یا تورم^۵ ۳- ریسک نقدینگی^۶ ۴- ریسک عملیاتی و تسویه حساب‌ها^۷ ۵- ریسک قانون^۸

همچنین طریق ا... خان و حبیب احمد^۹ (۲۰۰۱) نیز ریسک‌هایی که بانک‌ها با آن مواجه می‌شوند را به دو دسته ریسک مالی و ریسک غیرمالی تقسیم می‌کنند.

ریسک مالی، ریسک اعتباری، ریسک نقدینگی، ریسک بازار و ریسک سرمایه را شامل می‌شود.

^۱. Credit risk

^۲. market risk

^۳-Currency Risk

^۴-Price Risk

^۵-Interest Rate or Inflation Risk

^۶-Liquidity Risk

^۷-Operational Risk

^۸-Legal Risk

^۹ - Tariqullah khan & Ahmed,H. ۲۰۰۱



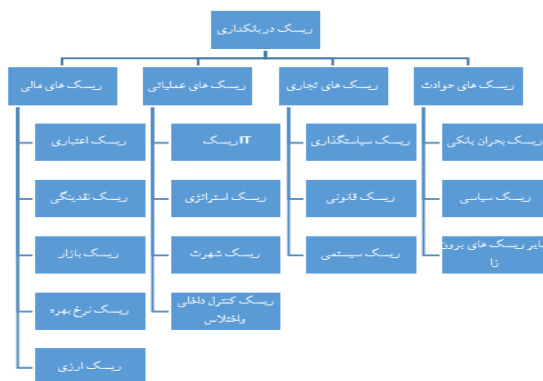
نخستین کنفرانس ملی توسعه مدیریت پولی و بانکی

تحول در مدیریت نظام بانکی، گامی بزرگ برای توسعه اقتصادی

۸ و ۹ بهمن ماه ۱۳۹۲ تهران - مرکز همایش‌های بین‌المللی صدا و سیما



ریسک غیر مالی، ریسک عملیاتی و تسویه حساب‌ها و ریسک حقوقی را دربر می‌گیرد. همچنین سینکی^۱ ریسک بانکی و حوزه مدیریت آن را در سه بخش ریسک اعتباری، ریسک نرخ بهره و ریسک نقدینگی می‌داند که به ترتیب از ناتوانی مشتریان در انجام تعهدات و نوسانات ناشی از نرخ بهره و عدم کفایت نقدینگی به وجود می‌آیند. بانکداری اسلامی نیز یک تقسیم‌بندی از انواع ریسک دارد که عبارتند از: ریسک اعتباری، ریسک سرمایه‌گذاری‌های مشارکتی، ریسک بازار، ریسک نقدینگی، ریسک نرخ بازدهی، ریسک عملیاتی. ریسک سرمایه‌گذاری مشارکتی: ریسک ناشی از ورود بانک‌های اسلامی در انواع مشارکت‌ها (به صورت ضامن یا تأمین کننده بخشی از سرمایه مشارکت) می‌باشد. بدیهی است در این حالت، سرمایه مشارکت در معرض ریسک فعالیت تجاری مربوطه قرار می‌گیرد. ریسک نرخ بازدهی: ریسک ناشی از اثرات عوامل بازاری بر نرخ بازدهی دارایی‌ها بوده و زمانی حاصل می‌شود که نرخ بازدهی کمتر از نرخ مورد انتظار دارندگان حساب‌های سرمایه‌گذاری گردد. اما در نگاه کلی ریسک‌هایی که بانک‌ها با آن مواجه می‌باشند، از نظر منشاء ایجاد آن به چهار دسته کلی ریسک‌های مالی، عملیاتی، تجاری، حوادث تقسیم می‌شوند. ریسک‌های مالی، ریسک‌هایی هستند که به شدت به یکدیگر وابسته بوده و در مجموع ریسک کلی بانک را افزایش می‌دهند. به عنوان مثال، در بیشتر وام‌های پرداختی، ریسک اعتباری وجود دارد. این ریسک نیز به دنبال خود، ریسک نقدینگی را در پی دارد. ریسک‌های عملیاتی به کنترل و اداره نامناسب کلی فعالیت‌های بانک مانند ضعف در مدیریت و یا اخلال در سیستم کنترل‌های داخلی مربوط می‌شوند. ریسک‌های تجاری، ریسک‌هایی هستند که به محیط فعالیت بانک از قبیل وضعیت کلان اقتصادی، سیاست‌گذاری و قانونی مربوط می‌شوند. این ریسک‌ها همچنین به عنوان ریسک کشوری شناخته می‌شوند. ریسک‌های حوادث، شامل کلیه ریسک‌های برون‌زا مانند ریسک سیاسی و ریسک بحران‌های بانکی می‌باشند که می‌توانند در عملیات بانک اخلال ایجاد کنند و یا این که وضعیت مالی آن را تضعیف نمایند. نمودار، ریسک‌های یاد شده را به تفکیک زیر گروه‌های مربوطه نشان می‌دهد. شایان ذکر است که ریسک‌ها همچنین از نظر منشاء ایجاد آنها به دو دسته درونی و بیرونی تقسیم می‌شوند. ریسک‌های گروه‌های تجاری و حوادث در زمره ریسک‌های خارج از کنترل بانک می‌باشند، لیکن مدیریت بانک می‌بایست شرایطی را فراهم آورد تا حتی الامکان آثار نامطلوب آن‌ها کاهش یابد. با توجه به این که در بیشتر اوقات عامل اساسی زیان‌های وارده به بانک‌ها ریسک‌های مالی و عملیاتی به بیان دیگر، ریسک‌های درونی می‌باشند، از این رو، در این قسمت به بررسی ریسک‌های مالی می‌پردازیم. (شربت اوغلی و همکاران، ۱۳۸۵)



شکل ۱: گستره ریسک در صنعت بانکداری

ریسک‌های مالی

ریسک نقدینگی

^۱ - Sinkey



نخستین کنفرانس ملی توسعه مدیریت پولی و بانکی

تحول در مدیریت نظام بانکی، گامی بزرگ برای توسعه اقتصادی

۸ و ۹ بهمن ماه ۱۳۹۲ - تهران - مرکز همایش‌های بین‌المللی صدا و سیما



به عدم توانایی بانک در ایفای تعهداتش به دیگران در زمان سررسید، ریسک نقدینگی گویند. این ریسک عمدتاً از ساختار دارایی‌ها و بدهی‌های بانک‌ها ناشی می‌شود. لازم به توضیح است که به طور سنتی، تعهدات بانک کوتاه مدت و در مقابل، دارایی‌های آن بلندمدت است. پس قصور بانک در پوشش شکاف زمانی بین این دو، موجب بروز ریسک نقدینگی می‌شود. این ریسک از ریسک‌های اساسی در بانکداری، به ویژه در صورت وارد شدن هر نوع شوک به نظام مالی است. همانطور که گفته شد، منشاء اصلی ریسک نقدینگی، عدم تطابق زمانی بین جریان‌های ورودی و خروجی به بانک می‌باشد. از این رو، ریسک نقدینگی را می‌توان به دو دسته ریسک نقدینگی دارایی‌ها و ریسک نقدینگی منابع تقسیم نمود. ریسک نقدینگی دارایی‌ها به عدم توانایی بانک در فروش دارایی‌های خود جهت برآوردن نیازهای نقدی غیر منتظره اشاره دارد. ریسک نقدینگی منابع نیز به عدم جذب منابع توسط بانک به طریق عادی (مانند جذب سپرده‌ها و یا استفاده از تسهیلات و خطوط اعتباری بانک‌ها) بر می‌گردد. بدیهی است که در چنین شرایطی بانک مجبور به فروش دارایی‌های خود در قیمت‌های بسیار پایین می‌باشد و نهایتاً این موضوع به زیان‌های بیشتر و در حالت افراطی به ورشکستگی بانک منجر می‌شود. در مجموع احتمال عدم توان اجرا به تعهدات مالی کوتاه مدت، ریسک نقدینگی تعریف می‌شود.

این ریسک به حالت‌های زیر خود را نشان می‌دهد:

- عدم توان اجرا به تعهدات کوتاه مدت
- عدم توان تأمین منابع مالی کوتاه مدت در زمان لازم
- عدم توان تأمین منابع مالی کوتاه مدت با هزینه مقرون به صرفه
- برای مدیریت کردن این ریسک از روش‌های زیر استفاده می‌شود:
- سرمایه‌گذاری در اوراق مالی قابل فروش با قدرت به نقد کردن بالا
- ماهنگی و طراحی حجم و سررسید منابع و مصارف در سرسیدهای مختلف
- هماهنگی با نهادهای مالی مختلف برای دریافت خطوط اعتباری کوتاه مدت
- استفاده از روش تبدیل به اوراق بهادار کردن دریافتی‌های مالی
- استفاده از اوراق مالی مشتقه در دل یا همراه قرارداد تسهیلات و سپرده‌ها

ریسک بازار

این ریسک عبارتست از احتمال تغییر در قیمت و نرخ بازارهای مالی که ارزش دارایی‌ها یا پرتفوی بانک را کاهش می‌دهد. به بیانی دیگر، ریسک بازار را می‌توان به صورت احتمال زیان ناشی از حرکت نامطلوب در ارزش دارایی‌ها (شامل اوراق بهادار، سهام، وام، ارز، کالا و یا اوراق مشتقه مرتبط با این دارایی‌ها) به علت تغییر قیمت و نرخ بازار تعریف کرد. بر این اساس شکل زیر گروه‌های زیر برای ریسک بازار معرفی می‌شوند. (شربت اوغلی و همکاران، ۱۳۸۵)



شکل ۲: انواع ریسک بازار

در کل خطا در پیش‌بینی‌هایی که حاصل از نوسانات نرخ‌ها در بازار باشد، ریسک بازار تعریف می‌شود. این ریسک به حالت‌های زیر خود را نشان می‌دهد:



- ریسک حاصل از نوسانات نرخ بهره و تورم و تاثیر آن بر حاشیه خالص^۱ و ارزش دارائی‌ها و بدهی‌ها
 - ریسک حاصل از نوسانات نرخ ارز و تاثیر آن بر حاشیه خالص و ارزش دارایی‌ها و بدهی‌ها
 - ریسک حاصل از نوسانات قیمت دارائی‌ها و تاثیر آن بر حاشیه خالص و ارزش دارائی‌ها و بدهی‌ها
- برای مدیریت کردن این ریسک از روش‌های زیر استفاده می‌شود:
- کوتاه مدت کردن سر رسید منابع و مصارف و یا متغیر و هماهنگ کردن نرخ سود آن‌ها
 - متنوع سازی سبد و هماهنگ نمودن مانده درآمدها و هزینه‌های دارائی‌ها و بدهی‌های ارزی
 - متغیر و وابسته کردن نرخ سود تسهیلات به شاخص کالاها و قیمت‌ها در بازار
 - هماهنگ کردن حساسیت ارزش دارائی‌ها و بدهی‌ها^۲
 - استفاده از اوراق مالی مشتقه در دل یا همراه قرارداد تسهیلات و سپرده‌ها
 - استفاده از روش پیشرفته "ارزش در خطر"^۳ (VAR) یا محاسبه احتمال کاهش ارزش دارائی‌ها تا حد تعریف شده برای یک دوره مشخص

ریسک نرخ بهره

این ریسک، ریسکی است که بانک در قبال تغییرات نامطلوب در نرخ‌های بهره با آن مواجه می‌شود. تمامی مؤسسات مالی و بانک‌ها در معرض ریسک نرخ بهره قرار دارند. شایان ذکر است که در بررسی خزانه‌داری آمریکا مشخص شد که بیش از ۹۰ درصد مؤسسات با ریسک نرخ بهره مواجه می‌شوند. نوسان در نرخ‌های بهره، بر درآمدها، هزینه‌ها و همچنین بر ارزش اقتصادی دارائی‌ها، بدهی‌ها و وضعیت اقلام زیر ترانزنامه تأثیر می‌گذارد و نهایتاً اثر خالص آن بر درآمد و سرمایه بانک منعکس می‌شود. در واقع، آنچه در اینجا دارای اهمیت است، همانا حاشیه خالص نرخ بهره و به عبارتی، تفاوت نرخ بهره دریافتی و پرداختی است.

ریسک نرخ بهره به چهار گروه فرعی ریسک نرخ‌گذاری مجدد، ریسک منحنی نرخ بازده، ریسک مبنا و ریسک حق اختیار تقسیم می‌شود. ریسک نرخ بهره در حالت کلی، عبارتست از احتمال کافی نبودن درآمد بهره، یعنی بهره‌های پرداختی بیشتر از بهره‌های دریافتی است، به طوری که حاشیه خالص نرخ بهره، تکافوی هزینه‌های بانک (شامل دستمزدها، اجاره‌ها و غیره) را ننماید.

ریسک ارزی

این ریسک احتمال زبانی است که از نوسانات نامطلوب نرخ ارز به وجود می‌آید و ناشی از عدم تطابق بین دریافتی‌ها و پرداختی‌های ارزی است. ریسک ارزی، ماهیت سوداگرانه دارد و احتمال وقوع این ریسک در بازارهای نقدی و سلف ارز وجود دارد. بانک‌های بین‌المللی به شدت در معرض این ریسک قرار دارند.

در بانک‌هایی که فعالیت ارزی دارند، نوسانات نرخ ارز، پیش‌بینی ارزش دارائی‌ها، بدهی‌ها، درآمدهای ناشی از دارائی‌ها (درآمد تسهیلات ارزی) و هزینه‌های بدهی‌ها (هزینه تامین منابع ارزی) را تحت تأثیر قرار می‌دهد. نوسانات نرخ ارز هم خطا در پیش‌بینی حاشیه سود ناخالص بانک‌ها را افزایش می‌دهد و هم موجب افزایش خطا در پیش‌بینی ارزش دارائی‌ها و بدهی‌های ارزی می‌شود.

ریسک اعتباری

ریسک اعتباری، به احتمال عدم ایفای تعهدات توسط مشتریان (وام گیرندگان) به عمد یا غیر عمد در بازپرداخت اصل و فرع بدهی اشاره دارد. این ریسک، ریسک کلاسیک بانکداری بوده که به نوبه خود می‌تواند مشکل نقدینگی را نیز برای بانک در پی داشته باشد. علیرغم نوآوری و ابداعات در بازارهای مالی، ریسک اعتباری کماکان عامل اصلی ورشکستگی بانک‌ها می‌باشد. سه دسته اصلی ریسک‌های اعتباری عبارتند از:

ریسک مشتریان (اشخاص)، ریسک شرکت‌ها، ریسک کشورها

ذکر این نکته هم ضروری است که برای ریسک اعتباری زیر گروه‌هایی از قبیل ریسک تسویه و ریسک اعتباری احتمالی نیز تعریف می‌شود. ریسک تسویه عموماً در بازار سهام و بازار ارز رخ می‌دهد و به زیان ناشی از قصور در پرداخت توسط طرف مقابل بعد از

^۱ - Net Margin

^۲ - Asset/Liability Duration Matching

^۳ - Value At Risk



خرید ارز یا سهام اطلاق می‌شود. ریسک اعتباری احتمالی هم زیان ناشی از امتناع یا قصور طرف مقابل در پرداخت یا تمدید اعتبار در آینده است.

در مجموع ریسک اعتباری، احتمال عدم برگشت اصل و فرع تسهیلات اعطا شده تعریف می‌شود.

این ریسک به حالت‌های زیر خود را نشان می‌دهد:

- احتمال کاهش توان باز پرداخت اصل و فرع تسهیلات دریافتی
 - احتمال عدم باز پرداخت اصل و فرع تسهیلات دریافتی
 - احتمال معوق شدن بازپرداخت اصل و فرع تسهیلات دریافتی
- برای مدیریت کردن این ریسک از روش‌های زیر استفاده می‌شود:
- طراحی نظام تعیین درجه اعتباری برای دریافت کنندگان تسهیلات
 - تعیین سقف تسهیلات قابل اعطا و وثیقه‌های لازم با در نظر گرفتن درجه اعتباری تعیین شده
 - ایجاد تنوع در طیف تسهیلات اعطاء شده با درجات اعتباری مختلف در سطوح مختلف
 - تعیین هزینه یا جرائم متعدد در صورت معوق شدن باز پرداخت‌های اصل و فرع (صدوقی، ۱۴۶)

۴- داده‌کاوی

در دو دهه قبل توانایی‌های فنی بشر برای تولید و جمع‌آوری داده‌ها به سرعت افزایش یافته است. عواملی نظیر استفاده گسترده از بارکد برای تولیدات تجاری، به خدمت گرفتن کامپیوتر در کسب و کار، علوم، خدمات دولتی و پیشرفت در وسائل جمع‌آوری داده، از اسکن کردن متون و تصاویر تا سیستم‌های سنجش از دور ماهواره‌ای، در این تغییرات نقش مهمی دارند.

به طور کلی استفاده همگانی از وب و اینترنت به عنوان یک سیستم اطلاع‌رسانی جهانی ما را مواجه با حجم زیادی از داده و اطلاعات می‌کند. این رشد انفجاری در داده‌های ذخیره شده، نیاز مبرم وجود تکنولوژی‌های جدید و ابزارهای خودکاری را ایجاد کرده که به صورت هوشمند به انسان یاری رسانند تا این حجم زیاد داده را به اطلاعات و دانش تبدیل کند، داده‌کاوی به عنوان یک راه حل برای این مسائل مطرح می‌باشد. در یک تعریف غیر رسمی داده‌کاوی فرآیندی است، خودکار برای استخراج الگوهایی که دانش را بازنمایی می‌کنند، که این دانش به صورت ضمنی در پایگاه داده‌های عظیم، انبار داده^۱ و دیگر مخازن بزرگ اطلاعات، ذخیره شده است. داده‌کاوی به طور همزمان از چندین رشته علمی بهره می‌برد نظیر: تکنولوژی پایگاه داده، هوش مصنوعی، یادگیری ماشین، شبکه‌های عصبی، آمار، شناسایی الگو، سیستم‌های مبتنی بر دانش^۲، حصول دانش^۳، بازیابی اطلاعات^۴، محاسبات سرعت بالا^۵ و بازنمایی بصری داده^۶. داده‌کاوی در اواخر دهه ۱۹۸۰ پدیدار گشته، در دهه ۱۹۹۰ گام‌های بلندی در این شاخه از علم برداشته شده و انتظار می‌رود در این قرن به رشد و پیشرفت خود ادامه دهد.

واژه‌های «داده‌کاوی» و «کشف دانش در پایگاه داده^۷» اغلب به صورت مترادف یکدیگر مورد استفاده قرار می‌گیرند.

کشف دانش در پایگاه داده فرایند شناسایی درست، ساده، مفید، و در نهایت الگوها و مدل‌های قابل فهم در داده‌ها می‌باشد. داده‌کاوی، مرحله‌ای از فرایند کشف دانش می‌باشد و شامل الگوریتم‌های مخصوص داده‌کاوی است، بطوریکه، تحت محدودیت‌های مؤثر محاسباتی قابل قبول، الگوها و یا مدل‌ها را در داده کشف می‌کند، به بیان ساده‌تر، داده‌کاوی به فرایند استخراج دانش ناشناخته، درست، و بالقوه مفید از داده اطلاق می‌شود. تعریف دیگر این است که، داده‌کاوی گونه‌ای از تکنیک‌ها برای شناسایی اطلاعات و یا دانش تصمیم‌گیری از قطعات داده می‌باشد، به نحوی که با استخراج آن‌ها، در حوزه‌های تصمیم‌گیری، پیش‌بینی، پیشگویی، و تخمین مورد استفاده قرار گیرند. داده‌ها

^۱ Data warehouses

^۲ Knowledge-based system

^۳ Knowledge-acquisition

^۴ Information retrieval

^۵ High-performance computing

^۶ Data visualization

^۷ Knowledge Discovery in Database



اغلب حجیم، اما بدون ارزش می‌باشند، داده به تنهایی قابل استفاده نیست، بلکه دانش نهفته در داده‌ها قابل استفاده می‌باشد. به این دلیل اغلب به داده‌کاوی، تحلیل داده‌ای ثانویه^۱ گفته می‌شود.

۵- شبکه عصبی

شبکه‌های عصبی مصنوعی^۲ روشی است که بر پایه اتصال چندین واحد پردازش ساخته می‌شود. به زبان ساده‌تر شبکه‌های عصبی سیستم‌ها و روش‌های محاسباتی نوینی هستند که برای یادگیری ماشینی، نمایش دانش، و در انتها اعمال دانش به دست آمده در جهت پیش‌بینی پاسخ‌های خروجی از سامانه‌های پیچیده استفاده می‌شود. ایده اصلی این گونه شبکه‌ها (تا حدودی) الهام گرفته از شیوه کارکرد سیستم عصبی زیستی، برای پردازش داده‌ها، و اطلاعات به منظور یادگیری و ایجاد دانش می‌باشد. عنصر کلیدی این ایده، ایجاد ساختارهایی جدید برای سامانه پردازش اطلاعات است. شبکه از تعداد دلخواهی سلول یا گره یا واحد یا نرون تشکیل می‌شود که مجموعه ورودی را به خروجی ربط می‌دهند.

۵-۱-۱- شبکه‌های عصبی پرسپترون چند لایه

نوعی از شبکه عصبی بر مبنای یک واحد محاسباتی به نام پرسپترون^۳ ساخته می‌شود. پرسپترون یک نوع شبکه عصبی مصنوعی است که در سال ۱۹۵۷ در لابراتوار کرنل آرونوتیکال به وسیله فرانک روزنبلت ابداع شد. می‌توان آن را ساده‌ترین نوع شبکه عصبی‌هایی که در آن‌ها خروجی فقط به سمت جلو حرکت می‌کنند^۴ در نظر گرفت. یک شبکه عصبی پرسپترون، برداری از ورودی‌های با مقادیر حقیقی را گرفته و یک ترکیب خطی از این ورودی‌ها را محاسبه می‌کند. اگر حاصل از یک مقدار آستانه بیشتر بود خروجی پرسپترون برابر با ۱ و در غیر این صورت معادل ۰ خواهد بود.

۵-۱-۲- آموزش پرسپترون

چگونه وزن‌های یک پرسپترون واحد را یاد بگیریم به نحوی که پرسپترون برای مثال‌های آموزشی مقادیر صحیح را ایجاد نماید؟ یکی از راه‌های آموزش پرسپترون استفاده از قانون پرسپترون است که در زیر به آن اشاره می‌شود.

۵-۱-۳- الگوریتم یادگیری پرسپترون

مقادیری تصادفی به وزن‌ها نسبت می‌دهیم

۱. پرسپترون را به تک تک مثال‌های آموزشی اعمال می‌کنیم. اگر مثال غلط ارزیابی شود مقادیر وزن‌های پرسپترون را تصحیح می‌کنیم.
۲. آیا تمامی مثال‌های آموزشی درست ارزیابی می‌شوند:
 - بله (پایان الگوریتم)
 - خیر (به مرحله ۲ برمی‌گردیم)

۵-۱-۴- قدرت نمایش توابع

گرچه قدرت نمایش توابع توسط یک شبکه رو به جلو بسته به عمق و گستردگی شبکه دارد، با این وجود موارد زیر را می‌توان به صورت قوانین کلی بیان نمود:

توابع بولی: هر تابع بولی را می‌توان توسط یک شبکه دو لایه پیاده‌سازی نمود.
توابع پیوسته: هر تابع پیوسته محدود را می‌توان توسط یک شبکه دو لایه تقریب زد. تئوری مربوطه در مورد شبکه‌هایی که از تابع سیگموئید در لایه پنهان و لایه خطی در شبکه خروجی استفاده می‌کنند صادق است.
توابع دلخواه: هر تابع دلخواه را می‌توان با یک شبکه سه لایه تا حد قابل قبولی تقریب زد.

۱. Secondary Data Analysis

۲. Artificial Neural Network-ANN

۳. Perceptron

۴. Feedforward



۶- درخت تصمیم^۱

درخت تصمیم یک ساختار درختی شبیه نمودار گردش کار^۲ است که برگ‌های آن نشان دهنده‌ی رده‌ها یا توزیع‌های رده‌ای می‌باشند و با شروع از ریشه‌ی درخت و عبور از گره‌های میانی، با رسیدن به برگ‌ها تخمینی از رده‌ی مورد جستجو به دست می‌آید. درخت تصمیم یک ساختار درختی شبیه نمودار گردش کار است که هر گره آن معرف یک آزمون بر روی یک مقدار صفت است، هر شاخه نمایانگر یک برآمد از آزمون و برگ‌های درخت نشان دهنده‌ی رده‌ها یا توزیع‌های رده‌ای می‌باشند.

سوالی که در اینجا مطرح می‌شود این است که چگونه می‌توان از درخت‌های تصمیم برای رده‌بندی استفاده کرد. فرض می‌کنیم که می‌خواهیم برچسب رده‌بندی یک تاپل داده شده را تعیین کنیم. بدین منظور مقادیر صفات آن تاپل در درخت تصمیم مورد آزمون قرار می‌گیرند؛ مسیری از ریشه‌ی درخت تا یک گره برگ دنبال می‌شود و بدین ترتیب پیش‌بینی رده‌ی آن تاپل انجام می‌شود. درخت‌های تصمیم را می‌توان به سادگی به قوانین رده‌بندی تبدیل کرد.

درختان تصمیم، نمونه‌ها را با مرتب کردن آن‌ها در درخت از گره ریشه به سمت گره‌های برگ دسته‌بندی می‌کنند. هر گره داخلی در درخت، صفتی از نمونه را آزمایش می‌کند و هر شاخه‌ای که از آن گره خارج می‌شود متناظر یک مقدار ممکن برای آن صفت می‌باشد. همچنین به هر گره برگ، یک دسته‌بندی منتسب می‌شود. هر نمونه، با شروع از گره ریشه درخت و آزمایش صفت مشخص شده توسط این گره و حرکت در شاخه متناظر با مقدار صفت داده شده در نمونه، دسته‌بندی می‌شود. این فرآیند برای هر زیردرختی که گره‌ی جدید ریشه آن می‌باشد تکرار می‌شود.

در حالت کلی، درختان تصمیم یک ترکیب فصلی از ترکیبات عطفی قیود روی مقادیر صفات نمونه‌ها را بازنمایی می‌کنند. هر مسیر از ریشه به درخت به یک برگ متناظر با یک ترکیب عطفی صفات تست موجود در آن مسیر بوده و خود درخت نیز متناظر با ترکیب فصلی همه این ترکیبات عطفی می‌باشد.

درخت‌های تصمیم روشی برای نمایش یک سری از قوانین هستند که منتهی به یک رده یا مقدار می‌شوند. برای مثال، می‌خواهیم متقاضیان وام را به دارندگان ریسک اعتبار خوب و بد تقسیم کنیم. شکل زیر یک درخت تصمیم را که این مسئله را حل می‌کند نشان می‌دهد و همه مؤلفه‌های اساسی یک درخت تصمیم در آن نشان داده شده است: نود تصمیم، شاخه‌ها و برگ‌ها.

بر اساس الگوریتم، ممکن است دو یا تعداد بیشتری شاخه داشته باشد. برای مثال، CART درختانی با تنها دو شاخه در هر نود ایجاد می‌کند. هر شاخه منجر به نود تصمیم دیگر یا یک نود برگ می‌شود. با پیمایش یک درخت تصمیم از ریشه به پایین به یک مورد، یک رده یا مقدار نسبت می‌دهیم. هر نود از داده‌های یک مورد برای تصمیم‌گیری درباره آن انشعاب استفاده می‌کند.

درخت‌های تصمیم از طریق جداسازی متوالی داده‌ها به گروه‌های مجزا ساخته می‌شوند و هدف در این فرآیند افزایش فاصله بین گروه‌ها در هر جداسازی است.

یکی از تفاوت‌ها بین متدهای ساخت درخت تصمیم این است که این فاصله چگونه اندازه‌گیری می‌شود. درخت‌های تصمیمی که برای پیش‌بینی متغیرهای دسته‌ای استفاده می‌شوند، درخت‌های طبقه بندی^۳ نامیده می‌شوند زیرا نمونه‌ها را در دسته‌ها یا رده‌ها قرار می‌دهند. درخت‌های تصمیمی که برای پیش‌بینی متغیرهای پیوسته استفاده می‌شوند درخت‌های رگرسیون^۴ نامیده می‌شوند.

۷- مدل لجستیک^۵

رگرسیون در مواردی استفاده می‌شود که متغیر وابسته قابل مشاهده نباشد. متغیر وابسته در این موارد به صورت انتخاب دوگانه^۶ ظاهر می‌شود، به عبارت دیگر دارای دو یا بیش از دو حالت است. اساس این که پدیده مورد نظر اتفاق افتاده یا اتفاق نیفتاده باشد دو حالت صفر و یک را به خود اختصاص می‌دهند و لذا در بالا و پایین نمودار مزبور قرار می‌گیرند. وقوع یا عدم وقوع پدیده مزبور با توجه به سطوح مختلف

۱ Decision Trees

۲ flowchart

۳ classification

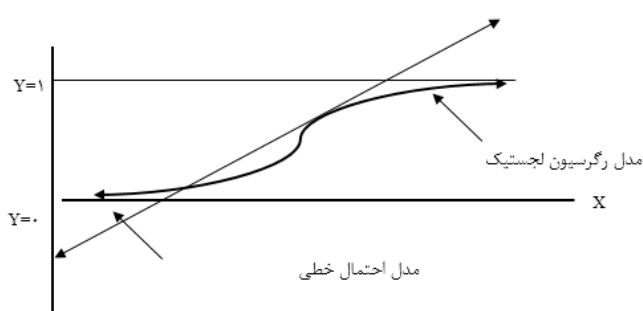
۴ regression

۵ logit model

۶ Binary Choice



از ترکیبات خطی متغیرهای مستقل تعیین می‌شود. یکی از مزایای رگرسیون لجستیک این است که می‌خواهیم بدانیم پدیده مورد نظر (به طور مثال خرید یک کالا، ریسک اعتباری، موفقیت یا عدم موفقیت یک شرکت) اتفاق افتاده است یا خیر. در این صورت می‌توان از یک متغیر وابسته گسسته مانند یک یا صفر برای نشان دادن پدیده مذکور استفاده نمود. بدین ترتیب از این متغیر وابسته می‌توان به منظور تخمین وقوع و یا عدم وقوع اتفاق مورد نظر سود جست. مقادیر احتمال بین صفر و یک می‌باشند و لذا مقادیر مورد پیش‌بینی نیز باید بین صفر و یک قرار گیرند. در صورتی که احتمال وقوع بیشتر از حد آستانه (نرخ توقف)^۱ که معمولاً در بیشتر مطالعات ۰.۵ در نظر گرفته می‌شود پیش‌بینی شود، وقوع پدیده مورد نظر محتمل تلقی شده و در غیر این صورت غیرمحتمل خواهد بود. در این ارتباط رگرسیون لجستیک رابطه‌ای S شکل را همان‌گونه که در شکل زیر آورده شد. بین احتمال وقوع یک پدیده (متغیر وابسته) و ترکیب خطی از متغیرهای مستقل فرض می‌نماید. (لاری، ۱۳۸۸)



شکل ۳: مقایسه مدل خطی و لجستیک

همان‌گونه که در شکل بالا ملاحظه می‌شود در سطوح پایین متغیر مستقل، احتمال وقوع به سمت صفر میل می‌نماید و با افزایش سطوح متغیر مستقل احتمال افزایش یافته و در سطوح بالاتر متغیر مستقل این مقدار به سمت یک میل می‌کند. رابطه‌ی بین متغیر مستقل و متغیر وابسته غیرخطی است لذا عملاً نمی‌توان از رگرسیون خطی معمولی برای برآورد آن استفاده کرد. جدای رابطه غیرخطی، ارتباط مزبور را نمی‌توان در قالب رگرسیون معمولی در نظر گرفت چرا که رگرسیون معمولی مفروضاتی دارد که در این حالت برقرار نیست. اول این که خطاهای مربوط به مقادیر گسسته از توزیع دوجمله‌ای و نه نرمال پیروی می‌کند لذا تمام آزمون‌های آماری مربوط به آن بی‌اعتبار می‌شود. دوم این که واریانس جمله خطا ثابت نمی‌باشد. این امر موجب به وجود آمدن پدیده‌ای به نام واریانس ناهمسانی^۲ می‌شود و دلیل سوم این که احتمالات پیش‌بینی شده در روش مدل رگرسیون خطی ممکن است بیشتر از یک و یا کمتر از صفر شود. رگرسیون لجستیک به منظور رفع مشکلات فوق ایجاد شده است. بدست آوردن رابطه‌ی منحصر به فرد متغیر وابسته با متغیرهای مستقل در این رگرسیون نیازمند روش‌های برآورد، تعیین نیکویی برازش و تفسیر ضرایب می‌باشد.

۷-۱-۱- برآورد مدل رگرسیون لجستیک

با وجود این که اطلاعات متقاضی‌ها و اطلاعاتی از عدم پرداخت‌های قبلی ولی متغیرهای غیر قابل مشاهده وجود دارد باید از رگرسیون لجستیک استفاده نمود. برای توضیح رگرسیون لجستیک می‌توان از تابع توزیع تجمعی استفاده نمود، توابع توزیع تجمعی، مجموعه‌ای از تغییراتی که مقدار p در معادله (۱) را در فاصله صفر و یک قرار می‌دهد، به وجود می‌آورند، در حالیکه همچنان خواص یکنواختی را دارا می‌باشند. (بدین معنا که آن‌ها توابع یکنواخت افزایشی یا کاهش هستند). فرض می‌کنیم که یک توزیع نرمال استاندارد برای بیان احتمال انتخاب شده باشد

$$p(y/x) = \Phi(b'x) = \int_{-\infty}^{b'x} \phi(Z) dz \quad (1)$$

^۱ . Cut-off Rate

^۲ -Heteroscedasticity



λ : بردار متغیرهای توضیحی

P : احتمال تجمعی وقوع پیشامد

b' : بردار ضرایب

که در آن تابع چگالی نرمال استاندارد است.

این تابع معرف مدل پروبیت می‌باشد. اگر تابع توزیع لجستیک برای بیان احتمال تجمعی وقوع پیشامد (P) مورد استفاده قرار گیرد، به مدل لجستیک منجر خواهد شد. در این موارد:

$$P(Y / X) = \phi(B'X) = \int_{-\infty}^{b'x} \eta(Z) dz = \frac{1}{1 + e^{-\beta'x}} = \frac{e^{\beta'x}}{1 + e^{\beta'x}} \quad (2)$$

$$1 - P_i = \frac{1}{1 + e^{\beta'x}}$$

که در آن تابع چگالی لجستیک می‌باشد. در مدل لجستیک متغیر وابسته به صورت صفر و یک تعریف می‌شود و دارای توزیع دو جمله ای می‌باشند

در این صورت نسبت $\frac{P}{1-P}$ یعنی نسبت احتمال وقوع پیشامد مورد نظر به احتمال عدم وقوع آن به صورت رابطه زیر می‌باشد:

$$\frac{P}{1-P} = e^{\beta_0 + \beta_1 x_1 + \dots + \beta_n x_n} \quad (3)$$

در این رابطه داریم:

P = احتمال وقوع پدیده مورد نظر

$1-P$ = احتمال عدم وقوع پدیده مورد نظر

$B_0 + B_1 X_1 + \dots + B_n X_n$ = ترکیب خطی متغیرهای مستقل

e = پایه لگاریتم طبیعی یا عدد نپر

نسبت $\frac{P}{1-P}$ نسبتی است که نشان دهنده شانس احتمال وقوع حادثه مورد نظر بر عدم وقوع آن می‌باشد، و نسبت بخت^۱ یا نسبت شانس نام دارد.

به دلیل غیر خطی بودن معادله فوق، از تبدیل خطی این تابع به صورت زیر استفاده می‌شود. با گرفتن لگاریتم، رابطه (۲) به صورت زیر در

می‌آید:

$$L = \ln\left(\frac{P}{1-P}\right) = \beta_0 + \beta_1 X_1 + \dots + \beta_n X_n \quad (4)$$

بدین ترتیب رابطه شماره (۳) که لگاریتم نسبت بخت نامیده می‌شود نسبت به متغیرهای مستقل خطی بوده و لذا ضرایب این متغیرها به ترتیبی که در قسمت بعد توضیح داده خواهد شد، قابل برآورد خواهند بود.

در رابطه شماره (۳) ضریب عرض از مبدا β_0 میزان تغییر مقدار لگاریتم نسبت شانس به نفع وقوع حادثه مورد نظر را در زمانی که مقادیر متغیرهای مستقل صفر است نشان می‌دهد و به طور مثال ضریب زاوی β_1 میزان تغییر در L را به ازای یک واحد تغییر در X نشان می‌دهد.

نهایتاً در مدل لجستیک اثر یک واحد تغییر در X_1 بر احتمال وقوع پیش آمد مورد نظر ($Y=1$) که به نام اثر نهایی خوانده می‌شود، از رابطه زیر بدست می‌آید:

(۵)

$$ME = \frac{\partial P}{\partial X_1} = f(\beta'X) \beta_1 = \frac{\exp(\beta'X)}{(1 + \exp(\beta'X))^2} \times \beta_1$$

^۱ -Odd ratio



بنابراین در این الگو مقدار تغییر در احتمال، بستگی به احتمال اولیه و بنابراین بستگی به ارزش‌های ابتدایی همه متغیرهای مستقل و ضرایب دارد.

۷-۱-۲- محاسبه ضرایب متغیرهای مستقل در رگرسیون لجستیک

هرچند در رگرسیون لجستیک ضرایب متغیرهای مستقل مانند رگرسیون چند متغیره برآورد می‌شود، لیکن نحوه برآورد آن کاملاً متفاوت است. در رگرسیون چند متغیره از روش حداقل مربعات استفاده می‌شود. در رگرسیون لجستیک بدلیل طبیعت غیرخطی تبدیل لجستیک، از روش حداکثر درست‌نمایی استفاده می‌شود. فرض نمائید: (x_i, y_i) به ازاء $i = 1, 2, \dots, n$ شامل n جفت مشاهده مستقل از همدیگر باشند در اینجا با استفاده از روش حداکثر درست‌نمایی^۲ می‌بایست مقادیر بردار β برآورد گردد. فرض نمائید $X = (x_1, x_2, \dots, x_n)$ برداری شامل n متغیر مستقل و متغیرها پیوسته و توزیع شرطی متغیر پاسخ به شرط x برابر $P(Y = \frac{1}{x}) = \pi(x)$ باشد بنابراین تابع لجستیک چند متغیره به صورت زیر خواهد بود.

$$g(x) = \beta_0 + \beta_1 x_1 + \dots + \beta_n x_n \quad (6)$$

با توجه به نمایش $\pi(n)$ به صورت تابع لجستیک از $g(x)$ خواهیم داشت:

$$\pi(x) = \exp\left\{\frac{g(x)}{1 + \exp(g(x))}\right\} \quad (7)$$

تابع درست‌نمایی در این صورت به شکل زیر تعریف خواهد شد:

$$\xi(x) = \prod_{i=1}^n \pi_i^{y_i} (1 - \pi_i)^{1-y_i} \quad (8)$$

با توجه به اینکه π_i همان رابطه مربوط به مدل رگرسیون لجستیک یک متغیره می‌باشد. لگاریتم تابع درست‌نمایی به صورت زیر در می‌آید:

$$L(\beta) = \sum_{i=1}^n \{Y_i \ln \pi_i - (1 - Y_i) \ln(1 - \pi_i)\} \quad (9)$$

در صورتی که نسبت به ضرایب β مشتق‌گیری شود خواهیم داشت:

$$\frac{\partial L(\beta)}{\partial \beta_0} = \sum y_i - \pi(x_i) = 0 \quad (10)$$

$$\frac{\partial L(\beta)}{\partial \beta_i} = \sum_{i=1}^n x_{ik} (y_i - \pi(x_i)) = 0 \quad (11)$$

با حل معادلات (۹) و (۱۰) برداره β برآورد می‌شود. ماتریس‌های واریانس و کواریانس با استفاده از نظریه برآورد حداکثر درست‌نمایی برآورد می‌شود. این نظریه بیان می‌کند که واریانس برآوردکننده‌ها به وسیله مشتق جزئی دوم از لگاریتم تابع درست‌نمایی بدست می‌آید.

$$\begin{aligned} \frac{\partial^2 L\beta}{\partial^2 \beta_i} &= -\sum x_{ik} \pi_i (1 - \pi_i) = -\sum_{i=1}^n x_{ik}^2 \left[\frac{\pi_i}{(1 - \pi_i)} \right] [1 - \pi_i]^2 \\ &= \sum_{i=1}^n x_{ik}^2 g(x) [1 + \exp\{g(x)\}]^2 \end{aligned} \quad (12)$$

۸- تحقیقات انجام شده در داخل و خارج از کشور

پژوهش‌ها و کاربردهای متعددی در حوزه اعتبارسنجی، برای شناسایی مشتریان خوب و بد بانک‌ها صورت گرفته است. روش قضاوتی در اعتبارسنجی به دلیل خطا و زمان زیاد به تدریج جای خود را به روش‌های پارامتریک و ناپارامتریک داده است. (توماس، ۲۰۰۶). روش‌های

^۱ - Ordinary Least Square, (OLS)

^۲ - Maximum Likelihood



پارامتریک^۱ مثل لاجیت و پروبیت^۲، تحلیل ممیزی^۳ و رگرسیون لجستیکی^۴ در ابتدای ظهور اعتبارسنجی مورد استفاده قرار گرفتند و سپس استفاده از روش‌های ناپارامتریک^۵ و داده‌کاوی^۶ مثل: درختان تصمیم‌گیری^۷، شبکه‌های عصبی^۸ و سیستم‌های خبره^۹ مورد توجه محققین گرفت (سبزواری و همکاران، ۲۰۰۶). به طور کلی مطالعات بسیاری در این زمینه صورت پذیرفته است که از میان آن‌ها می‌توان به مطالعات خارجی: سالچین برگر^{۱۰} (۱۹۹۲)، کوانتس و فانت^{۱۱} (۱۹۹۳)، آلتمن^{۱۲} و همکاران (۱۹۹۴)، لاچر^{۱۳} و همکاران (۱۹۹۵)، دسای^{۱۴} (۱۹۹۶)، پیراموتا^{۱۵} (۱۹۹۸)، دیوید وست^{۱۶} (۲۰۰۰)، تیان شیونگ لی^{۱۷} و همکاران (۲۰۰۲)، مو چین چین و شی هسین هوانگ^{۱۸} (۲۰۰۳)، حسین عبده^{۱۹} و همکاران (۲۰۰۷)، جی جنگ هوانگ^{۲۰} و همکاران (۲۰۰۹) و مطالعات داخلی شامل علی منصوری (۱۳۸۲)، پونه و داده‌کاوی از جمله شبکه‌های عصبی به مراتب بالاتر از روش‌های سنتی آماری می‌باشد.

سلطانی در پایان‌نامه ارشد خود تحت عنوان "بررسی عوامل مؤثر بر افزایش مطالبات معوق بانک‌های ملی استان مازندران" (سلطانی، ۱۳۷۹) از چهار فرضیه استفاده نمود. عواملی همچون نوع تسهیلات ارائه شده، مشخصات گیرندگان تسهیلات و میزان تورم و میزان نظارت را با استفاده از یک مدل نیمه لگاریتمی به عنوان عوامل مؤثر بر مطالبات معوق بررسی نمود.

$$\text{LN}Y : o + o\text{ln} + c\text{T} + a\text{PP} + e\text{CR} + o\text{iDi} + \text{Vi} \quad (13)$$

N: تعداد اخطارها، PP: نرخ سود T: مدت زمان بازپرداخت CR: میزان تسهیلات بانکی

LN Y: میزان تغییر در مطالبات معوق D_i : i امین متغیر موهومی

جامعه آماری وی برای برآورد مدل رگرسیونی نیمه لگاریتمی ۱۵۵ پرونده مطالبات معوق بوده است که اطلاعات مربوط به سال‌های ۱۳۶۰-۱۳۷۸ را شامل می‌شود. نتایج او از برازش مدل، بیان کننده وجود رابطه معنادار بین نوع تسهیلات و مشخصات فردی با افزایش مطالبات معوق است. علاوه بر آن، در مورد اثر نرخ تورم بر افزایش مطالبات معوق نتیجه گرفت به دلیل اینکه رابطه‌ی معناداری میان میزان تورم و مطالبات معوق وجود ندارد، نمی‌توان برای آن مدل مناسبی بدست آورد.

مدرس و ذکاوت در تحقیق خود تحت عنوان «مدل‌های ریسک اعتباری مشتریان بانک» (مدرس، ۱۳۸۶) با تعاریف از ریسک و انواع آن و ارائه روش‌هایی برای تعیین میزان ریسک اعتباری مشتریان حکومتی به تحلیل‌های آماری برای آن پرداخته و برای این منظور از مدل رگرسیون لجستیک استفاده کرده که نیازی به مفروضات محدود کننده و آماری در رابطه با متغیرها نیست. نویسندگان در مقاله خود نمونه را به دو دسته خوش حساب و بدحساب تفکیک نموده و بر اساس پنج نسبت مالی به تعیین میزان ریسک اعتباری پرداخته‌اند.

۱. Parametric
۲. Logit and Probit
۳. Discriminant Analysis
۴. Logistic regression
۵. Nonparametric
۶. Data mining
۷. Decision trees
۸. Nero Network
۹. Expert systems
۱۰. Salchenberger
۱۱. Coats and Fant
۱۲. Altman
۱۳. Lacher
۱۴. Desai
۱۵. Piramutha
۱۶. David West
۱۷. Tian-Shyug Lee
۱۸. Mu-Chen Chen and Shih-Hsien Huang
۱۹. Hussein Abdou
۲۰. Jih-Jeng Huang



$$Z = -2,836 + 3,977R_1 - 1,264R_2 + 1,339R_3 + 1,094R_4 \quad (14)$$

$$Z = -2,922 + 4,494R_1 - 1,505R_2 + 1,258R_4 \quad (15)$$

R_1 : نسبت جاری

R_2 : نسبت بدهی‌های جاری به مجموع دارایی‌ها

R_3 : نسبت حقوق صاحبان سهام به مجموع دارایی‌ها

R_4 : نسبت سود قبل از کسر مالیات به حقوق صاحبان سهام

در مدل اول (۱۴) تابع تشخیص بر اساس تمام نمونه ۱۲۰ تایی استخراج شده است. اگر مقدار این تابع بر اساس متغیرهای توضیح‌دهنده مثبت باشد، مشاهده مورد نظر (مشتری) به گروه مشتریان خوش حساب و در صورتی که این مقدار منفی باشد، مشتری به گروه مشتریان بد حساب تعلق خواهد داشت.

طی روش دوم فرمول (۱۵) تابع ممیزی بر اساس نمونه ۸۰ تایی برآورد و سپس اعتبار آن بر روی نمونه ۴۰ تایی به عنوان شاهد آزمون نمود.

رگرسیون لجستیک با ۴ متغیر توضیح‌دهنده R_1, R_2, R_3, R_4 بر روی تمام نمونه تصادفی ۱۲۰ تایی انجام و طی آن تابع زیر برآورد گردید که P در آن احتمال بد حساب بودن شرکت است.

$$p = 1 / (1 + e^{10,382 + 14,208R_1 + 3,715R_2 + 4,160R_3 + 5,025R_4}) \quad (16)$$

از نمونه‌های دیگر تحقیق منصوری تحت عنوان «طراحی و تبیین مدل ریاضی تخصیص تسهیلات بانکی، رویکرد مدل‌های کلاسیک و شبکه‌های عصبی» (منصوری، ۱۳۸۲) می‌باشد که در آن تحقیق با استفاده از یک سری متغیرهای مستقل و بهره‌گیری از شبکه‌های عصبی پرسپترون چندلایه، ریسک اعتباری و ظرفیت اعتباری شرکت‌ها و سازمان‌های در خواست کننده اعتبار به طور همزمان مورد تحلیل قرار گرفته است. به منظور سنجش کارایی مدل‌های شبکه‌های عصبی در مقایسه با مدل‌های کلاسیک نتایج حاصل از این مدل با نتایج مدل‌های رگرسیون خطی و لجستیک مقایسه شده است. بررسی نتایج نشان داد که مدل‌های شبکه‌های عصبی و رگرسیون لجستیک در برآورد ریسک اعتباری از قابلیت مشابهی برخوردار هستند لیکن مدل شبکه‌های عصبی در برآورد ظرفیت اعتباری مشتریان از توان بالاتری در برآورد ریسک و ظرفیت اعتباری برخوردار است.

از نمونه‌های دیگر تحقیق محمدپور با عنوان «اندازه‌گیری ریسک اعتباری در یک بانک تجاری» (محمد پور، ۱۳۸۷) می‌باشد که در مطالعه خود از ۴۴۲ پرونده اعتباری استفاده نمود. وی برای تعیین مدل مناسب جهت اندازه‌گیری ریسک اعتباری بانک، با استفاده از رگرسیون لاجیت، متغیر وابسته احتمال عدم بازپرداخت را بر روی ۱۸ متغیر مستقل نظیر ویژگی‌های شخصی متقاضی، نوع تسهیلات، مبلغ تسهیلات، تعداد تسهیلات دریافتی، نوع و مبلغ وثیقه و ... برآزش نمود که در نهایت ۷ متغیر توضیحی که بالاترین رابطه معناداری را با متغیر وابسته داشته که انتخاب شده و در مدل وارد گردید. در میان متغیرهای توضیحی نهایی نوع تسهیلات، نرخ سود بانکی، مدت تسهیلات و جنسیت رابطه مستقیم و مثبت با افزایش مطالبات معوق دارند و متغیرهای سن، تحصیلات و نوع وثیقه رابطه غیرمستقیم و منفی با افزایش مطالبات معوق دارند.»

هوین و کلیمیر (اندسون هوین، ۱۹۸۴) در مقاله خود تحت عنوان «رتبه‌بندی اعتبار برای بازار بانکداری خرد ویتنام» به بررسی مدل‌های درجه‌بندی اعتبار برای تسهیلات‌های جزئی ویتنام به عنوان کشور در حال توسعه در مقابل سایر کشورهای توسعه یافته پرداختند. در این راستا از مدل رگرسیونی لاجیت استفاده شده است. جامعه آماری آن‌ها برای برآورد مدل، ۵۶۰۳۷ پرونده از پرونده‌های اعطای تسهیلات بانک‌های تجاری در سال ۲۰۰۵ بوده است و از ۲۲ متغیر مستقل نظیر درآمد سالانه متقاضی، سطح تحصیلات، معدل یک ساله حساب بانکی، نرخ سود تسهیلات، مدت و میزان تسهیلات و ... استفاده شد. در برآزش مدل، آن‌ها نشان دادند که چگونگی تشخیص پارامترها و مشخصات متقاضیان باید بخشی از مدل رتبه‌بندی اعتبار باشد. همچنین نشان داده شد که بین استفاده از مدل‌های رتبه‌بندی اعتبار در زمینه تجاری و نسبت تسهیلات‌دهی رابطه معناداری وجود دارد. مدل برآزش شده به شرح زیر می‌باشد:



$$Z = -3,176 - 1,774X_1 - 1,557X_2 - 0,938X_3 - 0,845X_4 - 0,755X_5 - 0,652X_6 - 0,551X_7 - 0,492X_8 - 0,402X_9 - 0,356X_{10} - 0,285X_{11} - 0,233X_{12} - 0,19X_{13} - 0,181X_{14} - 0,156X_{15} - 0,125X_{16}$$

در مدل فوق X ها به ترتیب: مدت زمان همکاری با بانک، جنسیت، تعداد تسهیلات، وضعیت حساب سپرده، منطقه، وضعیت سکونت، حساب جاری، ارزش وثیقه، تعداد افراد تحت تکفل، مدت زمان سکونت در آدرس فعلی، وضعیت تأهل، نوع وثیقه، داشتن تلفن شهری (منزل)، سطح تحصیلات، درآمد و هدف از گرفتن تسهیلات می‌باشند.

۹- فرضیه‌های پژوهش

طراحی یک مدل ریاضی برای کنترل ریسک به عنوان هدف اصلی پژوهش قلمداد می‌شود. از این رو فرضیه‌های تحقیق نیز در راستای طراحی مدل ریاضی تحقیق به صورت زیر تدوین گردید.

فرضیه اصلی: کارایی مدل‌های داده‌کاوی در رتبه‌بندی مشتریان حقیقی بانک ملت در مقایسه با مدل‌های آماری پارامتریک بیشتر می‌باشند.

فرضیه‌های فرعی:

- ۱- بین جنسیت متقاضی و درجه ریسک اعتباری متقاضیان رابطه معنی‌داری وجود دارد.
- ۲- بین شغل متقاضی و درجه ریسک اعتباری متقاضیان رابطه معنی‌داری وجود دارد.

۱۰- متغیرهای پژوهش

در این پژوهش از ۷ متغیر استفاده شده است. این متغیرها شامل ویژگی‌های فردی متقاضی تسهیلات و متغیرهای مرتبط با بانکداری متقاضی می‌باشند که در جدول زیر خلاصه شده‌اند:

جدول ۱- تعریف متغیرها

متغیرها	ماهیت متغیر	تعریف عملیاتی (شاخص)	نام‌گذاری متغیر	مقیاس سنجش متغیر
تسهیلات اعطایی	کمی	مبلغ تسهیلات، مدت تسهیلات، تعداد تسهیلات، نرخ سود تسهیلات، نرخ جریمه	مستقل	فاصله‌ای / نسبی
درجه ریسک اعتباری متقاضیان	کیفی	مشتریان خوش حساب یا بد حساب (مطالبات معوقه)	وابسته	نسبی
جنسیت	کیفی	مرد / زن	مستقل	اسمی
شغل	کیفی	افراد شاغل در یک ارگان	مستقل	اسمی

۱۱- روش تحقیق

روش تحقیق بر حسب هدف کاربردی است به طوری که تحقیق حاضر به دنبال شیوه و روشی جدید در رتبه‌بندی اعتباری مشتریان بانک بر اساس روش‌های داده‌کاوی است، روش تحقیق برحسب نوع داده‌ها؛ کمی بوده که با استفاده از تحلیل‌های همبستگی به دنبال رابطه بین متغیرها می‌باشد. این تحقیق میدانی محسوب می‌گردد و محقق با حضور خود اقدام به جمع‌آوری اطلاعات می‌کند.

۱۲- جامعه و حجم نمونه

جامعه آماری این پژوهش تمامی افراد حقیقی هستند که از شعب مختلف بانک ملت شهرستان کرج تسهیلات دریافت نموده‌اند. محدوده زمانی مورد نظر بین سال‌های ۱۳۸۸-۱۳۷۶ می‌باشد.



علت انتخاب این حوزه برای مطالعه، نیاز به همکاری بالایی بود که با توجه به حساسیت مدیران شعب در حفظ اطلاعات حوزه مدیریتی خود و حفظ اطلاعات محرمانه مشتریان و همچنین حجم بالای کار به علت تعدد زیاد پرونده‌های اعتباری مورد نیاز، از سوی شعب انتظار می‌رفت. این حوزه از ۱۷ شعبه با درجه‌های ۱ و ۲ و ۳ و ۴ و تشکیل شده است. در بین انواع تسهیلات اعطایی بانک، چهار نوع تسهیلات مشارکت مدنی، مضاربه، فروش اقساطی و جعاله که عمده تسهیلات اعطایی شعبه‌ها را تشکیل می‌دهند، انتخاب گردیدند. جامعه آماری فوق از نظر وضعیت اعتباری به دو گروه تقسیم شده‌اند:

مشتریان خوش حساب: افراد حقیقی هستند که پس از دریافت تسهیلات در سررسید مربوطه نسبت به پرداخت و یا تسویه تعهدات خود (اصل و سود تسهیلات) اقدام می‌نمایند.

مشتریان برحساب: افراد حقیقی که پس از دریافت تسهیلات، در سررسید مربوطه نسبت به پرداخت و یا تسویه تعهدات خود (اصل و سود تسهیلات) اقدام نمی‌نمایند. افرادی که در بازپرداخت تسهیلات اعطایی بانک با مشکل مواجه گردیده و توانایی بازپرداخت تسهیلات را ندارند.

۱۳- حجم نمونه

یکی از مسائل عمده که در طرح‌های تحقیقاتی قابل اهمیت است دانستن حجم نمونه لازم می‌باشد زیرا چنانچه تعداد نمونه را زیاد از حد بزرگ انتخاب کنیم متضمن کار زیاد و در نتیجه هزینه زیاد خواهد بود و اگر تعداد نمونه را کمتر از حد معین بگیریم نتیجه‌های حاصل نادرست بوده و اطلاعات لازم و دقیقی به دست نخواهد آمد.

به منظر تعیین حجم نمونه ابتدا پیش آزمونی^۱ انجام شد. در این پیش‌آزمون ۹۰ مورد از پرونده‌های اعتباری مربوط به استفاده‌کنندگان تسهیلات مورد بررسی قرار گرفت. با توجه

به احتمال عدم بازپرداخت بدست آمده طبق

$$n = \frac{N.t^2.S^2}{t^2.S^2 + N.D^2} = \frac{12925 * (1.96)^2 * (0.2275)^2}{(1.96)^2 * (0.2275)^2 + 12925 * (0.04)^2} = 524$$

فرمول زیر حجم نمونه محاسبه شد:

(۱۸)

که در آن: n حجم نمونه و N کل جامعه آماری و D دقت مطلوب احتمالی (نصف فاصله اطمینان) که معمولاً بین ۰،۱-۰،۰۱ است (در اینجا D=۰/۰۴ در نظر گرفته شد).

$S^2 = p.q$ واریانس، p میانگین احتمال عدم بازپرداخت (۰/۰۳۵) و q=۰/۶۵، آماره‌ی t که در نتیجه آن ۵۲۴ مورد از پرونده‌ها مورد بررسی و داده‌های بدست‌آمده مورد تجزیه و تحلیل قرار گرفت.

برای تعیین نمونه در هر یک از شعبات دارای درجه‌های مختلف بانکی (شعبات درجه ۱ و ۲ و ۳ و ۴) و بین چهار نوع تسهیلات داده شده از روش انتساب متناسب^۲ استفاده شده است.

۱۴- تجزیه و تحلیل اطلاعات

اشخاص، به دو گروه حقیقی و حقوقی تقسیم‌بندی می‌شوند. اشخاص حقیقی، اشخاصی هستند که قائم بر وجود خودشان هستند. اشخاص حقیقی خود به دو دسته مذکر و مؤنث تقسیم می‌شوند

در این مطالعه از پرونده‌های اعتباری اشخاص حقیقی استفاده شده است. با توجه به نتایج زن‌ها در حدود ۱۵،۶ درصد از حجم نمونه آماری و مردها نیز ۸۴،۴ درصد از حجم نمونه آماری ما را تشکیل می‌دهند. در تجزیه و تحلیل نمونه گردآوری شده مشخص شده است که از میان پرونده‌های تسهیلات اعتباری که ۳۶،۶۴٪ آن‌ها معوق می‌باشد، ۳۱،۴۹٪ آن مربوط به آقایان و ۵،۱۵٪ آن مربوط به خانم‌ها می‌باشد. نوع شغل اثر مستقیم بر ثروت می‌گذارد. در این پژوهش بیش از ۴۰ شغل در داده‌های جمع‌آوری شده وجود دارد. این داده‌ها به ۴ دسته: (۱) صاحبان کارخانجات، شرکت‌ها و کارگاه‌ها (اعم از تولیدی، توزیعی و خدماتی) (۲) پزشک، فرهنگی، کارمند و ... (۳) کشاورز (زراعت، باغداری، دامداری، پرورش گل و گیاه، پرورش ماهی و ...)، خانه‌دار و دانشجو (۴) سایر (شامل شغل‌های آزاد مانند: جوشکاری، فروشندگی لوازم خانگی، خیاطی، آرایشگری، خبازی، قنادی، خواروبار فروشی، تعمیرگاه‌ها و ...) طبقه‌بندی شده‌اند. برآش اولیه نشان می‌دهد بالاترین

^۱. Pre-test

^۲. Probability Proportional of Size –p.p.s



نرخ‌های عدم بازپرداخت مربوط به شغل‌های آزاد و افراد بازاری (خواروبارفروش، خیاط و...) و پایین‌ترین نرخ عدم بازپرداخت در رابطه با کارفرماهای اقتصادی و شغل‌های دولتی (مرغداری‌ها، دامداری‌ها، شرکت‌های توزیعی و تولیدی تا پزشکان، کارمندان و ...) می‌باشد برای تعیین میزان ارتباط دو متغیر مورد بررسی پژوهش از معیار همبستگی استفاده می‌شود. تحلیل همبستگی ابزار آماری است که به وسیله آن می‌توان درجه ارتباط خطی دو متغیر را اندازه‌گیری نمود.

فرضیه اول: بین جنسیت متقاضی تسهیلات و درجه ریسک اعتباری متقاضیان ارتباط وجود دارد.

H₀: بین جنسیت و درجه ریسک اعتباری متقاضیان رابطه معنی‌داری وجود ندارد.

H₁: بین جنسیت و درجه ریسک اعتباری متقاضیان رابطه معنی‌داری وجود دارد.

جدول ۲- ضریب همبستگی بین جنسیت و ریسک اعتباری

متغیر	سطح خطا	سطح معنی‌داری	ضریب همبستگی	تصمیم‌گیری	تأیید فرضیه
جنسیت و درجه ریسک	۰/۰۵	۰/۰۱۴	۰/۰۱	p < ۰/۰۵	H ₁

فرضیه دوم: بین شغل متقاضی و درجه ریسک اعتباری متقاضیان رابطه معناداری وجود دارد

H₀: بین شغل متقاضی و درجه ریسک اعتباری متقاضیان رابطه معنادار وجود ندارد

H₁: بین شغل متقاضی و درجه ریسک اعتباری متقاضیان رابطه معنادار وجود دارد.

جدول ۳- ضریب همبستگی بین شغل و ریسک اعتباری

متغیر	سطح خطا	سطح معنی‌داری	ضریب همبستگی	تصمیم‌گیری	تأیید فرضیه
شغل و درجه ریسک	۰/۰۵	۰/۰۴۶	-۰/۰۸	p < ۰/۰۵	H ₁

۱۵- برازش مدل‌ها

برای برازش مدل رتبه‌بندی اعتباری از سه روش مدل رگرسیون لجستیک، شبکه عصبی و ساختار درختی استفاده شده است. برای انتخاب بهترین مدل دقت آن‌ها را در طبقه‌بندی مشتریان خوش‌حساب و بدحساب با هم مقایسه می‌کنیم بطوریکه روشی که کمترین خطا را در طبقه‌بندی داشته باشد به عنوان مدل کارا انتخاب می‌شود.

۱۶- مدل رگرسیون لجستیک

با استفاده از خروجی نرم‌افزار PASW SPSS چنین نتیجه می‌شود که از ۵۲۴ پرونده مورد بررسی تنها ۵۲۲ مورد در تحلیل رگرسیون لجستیک مورد استفاده قرار گرفته‌اند. همچنین خروجی حاصل از آماره آزمون omnibus برای مدل نشان می‌دهد که چون p-value به دست آمده از این آزمون کوچک‌تر از ۰,۰۵ می‌باشد بنابراین برازش مدل رضایت‌بخش خواهد بود. همچنین R² به دست آمده از برازش مدل در حدود ۱۱٪ می‌باشد به عبارت دیگر ۱۱ درصد از تغییرات متغیر وابسته توسط متغیرهای مستقل بیان می‌شود. دقت طبقه‌بندی مدل به طور صحیح در حدود ۶۶,۹٪ یا تقریباً ۷۰٪ می‌باشد. همچنین از ۷ متغیر دخیل شده در روش مدل رگرسیون لجستیک ۶ متغیر روی متغیر وابسته تأثیرگذار بودند. بنابراین مدل بدست‌آمده به صورت زیر می‌باشد:

$$Y = ۰,۰۰۱ * (\text{مبلغ}) + ۰,۲۱۷ * (\text{مدت}) + ۰,۱۶۸ * (\text{سود}) - ۰,۱۲۲ * (\text{جریمه}) - ۰,۲۰۱ * (\text{تعداد}) - ۰,۲۵۵ * (\text{شغل}) + ۰,۷۶۴$$

با توجه به مدل در می‌یابیم که متغیرهای شغل، تعداد و جریمه با متغیر وابسته رابطه عکس دارند. همچنین متغیر جنسیت در مدل رگرسیون لجستیک در رتبه‌بندی مشتریان تأثیرگذار نمی‌باشد.

بنابراین متغیرهای تأثیرگذار بر مطالبات معوق به صورت زیر می‌باشد.

۱. مدت تسهیلات
۲. تعداد تسهیلات
۳. شغل متقاضی تسهیلات
۴. سود تسهیلات
۵. نرخ جریمه تسهیلات
۶. مبلغ تسهیلات



مهم‌ترین متغیر تاثیر گذار روی متغیر وابسته در این روش متغیر مدت تسهیلات و کمترین متغیر تاثیرگذار روی متغیر وابسته متغیر مبلغ تسهیلات می‌باشد.

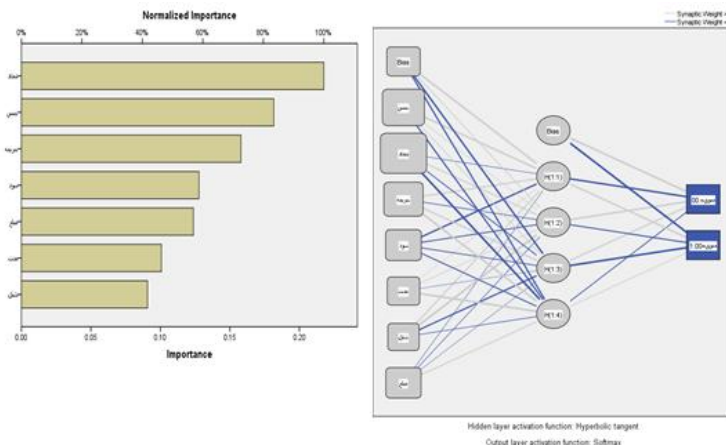
۱۷- مدل شبکه عصبی

برای رتبه‌بندی اعتباری اشخاص حقیقی از روش شبکه‌های عصبی که به عنوان شبکه‌های هوشمند شناخته شده‌اند استفاده می‌کنیم. در این مرحله نیز هدف مشخص کردن تعداد لایه‌های میانی با حداقل خطا و بیشترین درصد طبقه‌بندی درست مشتریان می‌باشد. به بیان دیگر مهم‌ترین تحلیل شبکه عصبی، ساختار شبکه با حداقل لایه میانی و کمترین خطا می‌باشد.

در این روش که با نرم‌افزار PASW SPSS مورد تحلیل قرار گرفت از ۵۲۴ مورد پرونده، ۳۵۵ مورد به عنوان نمونه Training و ۱۶۷ مورد پرونده به عنوان نمونه Testing دسته‌بندی شد. ۲ مورد بقیه برای پیش‌بینی در آینده استفاده می‌شود. برای پیش‌بینی مدل از نمونه Training و برای ارزیابی مدل از نمونه Testing استفاده می‌شود.

۱۷-۱-۱- ساختار شبکه عصبی پرسپترون چند لایه

پس از برازش مدل، معماری شبکه عصبی حاصل شامل ۷ لایه ورودی، ۴ لایه میانی و یک لایه خروجی می‌باشد. تابع عملکرد مورد استفاده در تعیین تعداد لایه میانی تابع Softmax و تابع خطای مورد استفاده خطای آنتروپی می‌باشد. خطای آنتروپی به دست آمده از این مدل برابر ۱۸۴،۵۹۹ در نمونه Training و مقدار این خطا در نمونه Testing برابر ۲۷،۴۲ می‌باشد. همچنین در نمونه Training برابر ۲۹،۹٪ و در نمونه Testing در صد پیشگویی نادرست برابر ۳۷،۱٪ می‌باشد.



شکل ۴: مدل شبکه عصبی و ترتیب اهمیت متغیرها در مدل

بنابراین دقت طبقه‌بندی مدل در نمونه Training در شبکه عصبی برابر ۷۰،۱٪ می‌باشد.

ترتیب اهمیت متغیرهای مستقل مطابق شکل ۲ در مدل ساختار شبکه عصبی به صورت زیر می‌باشد.

۱. تعداد تسهیلات دریافتی، ۲. جنسیت متقاضیان، ۳. نرخ جریمه، ۴. نرخ سود تسهیلات، ۵. مبلغ تسهیلات دریافتی، ۶. مدت تسهیلات دریافتی، ۷. شغل متقاضی تسهیلات

همچنین با توجه به مطالب بالا چنین نتیجه می‌شود که مدل ساختار شبکه عصبی از همه متغیرهای مستقل مسئله برای پیش‌بینی متغیر وابسته استفاده می‌کند.

مهم‌ترین متغیر تاثیرگذار روی متغیر درجه ریسک اعتباری در این روش متغیر تعداد تسهیلات و کمترین متغیر تاثیرگذار روی آن متغیر شغل متقاضی تسهیلات می‌باشد.

۱۸- برازش مدل ساختار درختی

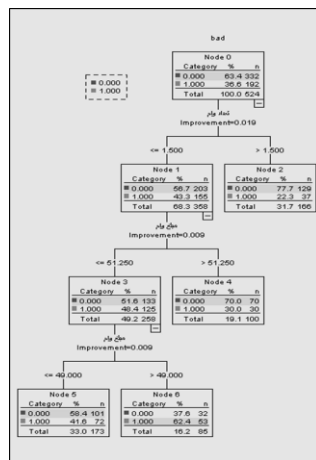
یکی دیگر از مدل‌های پیش‌بینی متغیر وابسته، ساختار درختی می‌باشد که در این مدل نیز دقت طبقه‌بندی موارد درست بررسی می‌شود.



با توجه به خروجی نرم‌افزار PASW SPSS چنین استنباط می‌شود که تعداد متغیرهای اصلی مسأله که ۷ متغیر می‌باشند در این ساختار از ۶ متغیر تعداد تسهیلات، مدت تسهیلات، جنسیت متقاضی تسهیلات، شغل متقاضی تسهیلات و نرخ جریمه برای طبقه‌بندی و پیش‌بینی استفاده می‌کند.

در این روش برای پیش‌بینی متغیر وابسته از ساختار درختی با مدل رشد Crit استفاده شده است. همچنین تعداد گره‌های مورد استفاده در این ساختار برابر ۷ و تعداد خروجی‌ها برابر ۴ می‌باشد. خطای معیار به دست آمده در این روش برابر ۰,۰۱۶ و در ۳۲,۶٪ موارد طبقه‌بندی نادرست صورت گرفته است. یعنی ساختار درختی در حالت کلی ۶۷,۴٪ موارد را درست طبقه‌بندی کرده است. همچنین ترتیب اهمیت متغیرهای مستقل به صورت زیر می‌باشد:

۱. مبلغ تسهیلات، ۲. مدت تسهیلات، ۳. تعداد تسهیلات، ۴. شغل متقاضی تسهیلات، ۵. جنسیت متقاضی تسهیلات، ۶. نرخ جریمه
- مهم‌ترین متغیر تأثیرگذار روی متغیر وابسته در این روش متغیر مبلغ تسهیلات و کمترین متغیر تأثیرگذار روی آن متغیر وابسته متغیر نرخ جریمه تسهیلات می‌باشد.



شکل ۵: خروجی مدل ساختار درختی

بنابراین به این نتیجه می‌رسیم که در بین سه روش رتبه‌بندی روش شبکه عصبی مصنوعی با توجه به دقت بیشتر در طبقه‌بندی روش کارآمدتری می‌باشد.

جدول ۴- مقایسه سه روش مدل‌بندی درجه ریسک اعتباری متقاضیان

مدل	شبکه عصبی	رگرسیون لجستیک	ساختار درختی
دقت طبقه بندی	۷۰,۱٪	۶۶,۹٪	۶۷,۴٪
تعداد متغیرهای دخیل در مدل	۷	۶	۶ متغیر
مهم‌ترین متغیر مستقل در مدل	مبلغ وام	مدت وام	مبلغ وام

۱۹- نتیجه‌گیری و پیشنهادات

در حالت کلی با توجه به تحلیل‌های انجام گرفته مطالبات معوقه با متغیرهای مستقل جنسیت، تعداد تسهیلات دریافتی، مبلغ تسهیلات دریافتی، شغل، نرخ سود، نرخ جریمه رابطه معنی‌داری دارد که در مدل رگرسیون لجستیک، شبکه‌های عصبی و ساختار درختی این مطلب نیز تأیید شد بطوریکه روش‌های داده‌کاوی ساختار درختی و شبکه‌های عصبی از همه متغیرهای مستقل موجود در مدل استفاده کردند و به این نتیجه نیز رسیدیم که روش شبکه عصبی در مدل‌بندی مطالبات معوقه کارا تر می‌باشد. این نتایج در راستای کار محمدرپور با عنوان "اندازه‌گیری ریسک اعتباری در یک بانک تجاری" (محمدرپور، ۱۳۸۷) می‌باشد بطوریکه در تحقیق ایشان در میان متغیرهای توضیحی نهایی نوع تسهیلات، نرخ سود بانکی، مدت تسهیلات و جنسیت رابطه مستقیم و مثبت با افزایش مطالبات معوق دارند و متغیرهای سن، تحصیلات و نوع وثیقه رابطه غیرمستقیم و منفی با افزایش مطالبات معوق دارند.»

با توجه به آزمون فرض‌های انجام شده، نتایج حاصل از این فرضیات به صورت زیر بیان می‌شوند.

