

## داده کاوی صورت های مالی جهت اعطای تسهیلات مالی

امیررضا کیقبادی\* و حید خدایمی\*\*

تاریخ دریافت: ۱۳۹۱/۱۰/۱۵ تاریخ پذیرش: ۱۳۹۱/۱۲/۱۷

### چکیده

بررسی، سنجش و اندازه گیری اعتبار مشتریان در مؤسسات اعتباری، امروزه یکی از مهمترین تصمیم های مالی بشمار می آید. نحوه تصمیم گیری در خصوص اعطای تسهیلات به مشتریان از این جهت دارای اهمیت می باشد که عدم ارزیابی دقیق مشتریان می تواند منجر به مطالبات سررسید گذشته و معوق با کاهش توان تسهیلات دهی بانک ها و در نهایت سوخت شدن مطالبات بانک ها گردد. این پژوهش با هدف مدل سازی اعتبارسنجی مشتریان در بانک به روش های شبکه عصبی، درخت تصمیم و ماشین بردار پشتیبان انجام می شود. بدین منظور اطلاعات و داده های مالی و کیفی یک نمونه تصادفی ۳۰۰ تایی (۲۱۸ مشتری خوش حساب و ۸۲ مشتری بدحساب) از شرکت های حقوقی را که در سال های ۸۹ و ۹۰ از بانک ملی ایران شعب شهر تهران تسهیلات اعتباری دریافت نموده اند، مورد بررسی قرار می گیرد. در این تحقیق پس از بررسی پرونده های اعتباری هریک از مشتریان، ۳۱ متغیر توضیح دهنده مورد ارزیابی قرار گرفت و نتایج ضمن دلالت بر تأیید نظریه های اقتصادی و مالی نشان می دهد که تکنیک های داده کاوی جهت اعتبارسنجی مشتریان از کارایی بالایی برخوردار می باشد و همچنین عملکرد پیش بینی الگوی شبکه عصبی به مراتب بهتر از سایر الگوها است.

**واژه های کلیدی:** اعتبارسنجی - داده کاوی - ماشین بردار پشتیبان - درخت تصمیم - شبکه های عصبی.

---

\* بورسیه دکتری و عضو هیئت علمی دانشگاه آزاد اسلامی واحد تهران مرکزی، نویسنده اصلی و مسئول مکاتبات. a.keyghobadi@iauctb.ac.ir

\*\* کارشناس ارشد حسابداری، دانشگاه آزاد اسلامی واحد تهران مرکزی.

## ۱- مقدمه

بانک‌ها به‌عنوان بخش اصلی نظام مالی همواره با ریسک‌های متفاوتی روبرو هستند که یکی از عمده‌ترین آن‌ها ریسک اعتباری است. حجم قابل ملاحظه‌ای از تسهیلات اعطایی سوخت شده یا معوقه بانک‌ها، گویای فقدان مدل‌های مناسب اندازه‌گیری اعتباری و سیستم‌های مدیریت ریسک در شبکه بانکی است. یکی از مهمترین ابزارهایی که بانک‌ها برای مدیریت و کنترل ریسک اعتباری بدان نیازمند هستند، "سیستم اعتبارسنجی مبتنی بر صورت‌های مالی مشتریان" است. با بهره‌گیری از تحلیل اطلاعات مربوط به مشتریان بانک با استفاده از فرآیند داده کاوی<sup>۱</sup> می‌توان به اعتبارسنجی متقاضیان وام و طبقه بندی آن‌ها به مشتریان خوش حساب و بدحساب، بدون قضاوت شخصی و براساس سیستم‌های هوشمند پرداخت. از طرفی با رشد سریع صنعت بانکداری مدل‌های اعتبارسنجی نیز به طور وسیعی برای ارزیابی تصمیم‌های اعطا یا عدم اعطای تسهیلات به مشتریان مورد استفاده قرار می‌گیرد. برای این امر مدیران بانک‌ها نیاز به تحلیل صحیح از داده‌های مشتریان دارند تا بر اساس آن تصمیم‌های مناسبی را برای تخصیص مناسب اعتبارات به متقاضیان اتخاذ نمایند. و بانک‌ها و مؤسسات اعتباری جهت کاهش ریسک اعتباری خود ملزم به شناسایی متقاضیان وام می‌باشند.

در این مقاله ابتدا پیشینه ای از بررسی‌های صورت گرفته در زمینه اعتبارسنجی و رتبه‌بندی مشتریان مطرح شده و سپس چارچوب نظری و مدل مفهومی پژوهش ارائه شده است. در بخش بعدی پژوهش به اختصار به متدولوژی به کار رفته جهت اعتبارسنجی پرداخته شده است. موضوع‌های این بخش در بر گیرنده، فرضیه‌ها، روش پژوهش، جامعه آماری، شیوه‌های نمونه گیری، روش‌های تحلیل داده‌ها و آزمون فرضیات می‌باشد و در نهایت در بخش پایانی مقاله، یافته‌های پژوهش و نتیجه‌گیری و

پیشنهادها برای استقرار مدل‌های اعتبارسنجی با استفاده از تکنیک‌های داده کاوی ارائه شده است.

## ۲- مبانی نظری و پیشینه پژوهش

### ۲-۱- مبانی نظری

ریسک اعتباری عبارت است از احتمال تعویق، مشکوک بودن یا عدم وصول اصل و سود تسهیلات مالی که در قالب تسهیلات به متقاضیان ارائه می‌گردد. از مهمترین رویکرد جهت کاهش خسارت های ریسک اعتباری، شناسایی، سنجش درجه و طبقه بندی اعتباری مشتریان میباشد که اصطلاحاً به آن اعتبارسنجی گویند.

اساس کار اعتبارسنجی بر مبنای داده کاوی بنیان نهاده شده است و در فرآیند اعتبارسنجی، مشتریان بر اساس شاخص‌ها و ویژگی‌های مختلفی مورد ارزیابی و رتبه‌بندی قرار می‌گیرند. این رتبه یا امتیاز در واقع نمایانگر اعتبار مالی مشتری است که بانک می‌تواند بر اساس آن نسبت به ارائه خدمات به مشتری خیلی سریع و دقیق‌تر تصمیم‌گیری کند.

تعریف‌های متفاوتی از داده کاوی وجود دارد ولی تعریفی که در اکثر مراجع به اشتراک ذکر شده عبارت است از "استخراج اطلاعات و دانش و کشف الگوهای پنهان از یک پایگاه داده‌های بسیار بزرگ و پیچیده". هدف داده کاوی این است که ارتباطات و الگوهای معتبر، جدید، قابل استفاده، و قابل درک را از میان داده‌ها استخراج کند.

یکی از مباحث مهم در صنعت بانکداری تشخیص توانایی یا ارزیابی قدرت شرکت‌ها در بازپرداخت بدهی، جهت کاهش خسارت‌های ناشی از ناتوانی آنان در بازگرداندن تسهیلات دریافتی است. که برخی از مزایای آن عبارت است از: ۱- کاهش هزینه تحلیل ۲- تصمیم‌گیری سریع ۳- تضمین تسهیلات و حذف ریسک‌های احتمالی.

### ۴- تعیین اولویت در مجموعه اعطاء تسهیلات

در نتیجه ما می‌توانیم از مدل‌های مختلفی برای ارزیابی وضعیت مالی مشتریان استفاده کنیم که این مدل‌ها به دو گروه اصلی تقسیم می‌شوند که عبارتند از:

گروه اول: مدل‌های پارامتریک: شامل ۱- مدل احتمال خطی<sup>۱</sup> ۲- مدل لاجیت و پروبیت<sup>۲</sup> ۳- مدل‌های بر مبنای تحلیل ممیزی<sup>۳</sup>

گروه دوم: مدل‌های ناپارامتریک: شامل ۱- برنامه ریزی ریاضی<sup>۴</sup> ۲- شبکه‌های عصبی<sup>۳</sup> ۳- درخت‌های تصمیم<sup>۴</sup> ۴- مدل نزدیکترین همسایگی<sup>۵</sup> ۵- فرآیند تحلیل سلسله مراتبی<sup>۶</sup> ۶- سیستم‌های خبره<sup>۷</sup> ۷- الگوریتم ژنتیک<sup>۸</sup>

متدهای آماری که در تحقیقات گذشته مورد استفاده قرار گرفته‌اند مدل‌های تحلیل ممیزی و رگرسیون لجستیک بوده‌اند. از جمله مشکلات عمده این متدها این است که برای ورودی‌هایی با ابعاد بزرگ و سائز نمونه کوچک مناسب نیستند و در اغلب این متدها فرض بر این است که رابطه خطی میان متغیرها وجود دارد در حالی که معمولاً این رابطه غیر خطی است. از این رو فرآیند مدل‌سازی خودکار دشوار است. در مدل‌های ایستا معمولاً در زمانی که محیط در حال تغییر است، محکوم به شکست بوده و از این رو مدل‌ها ممکن است دوباره ایجاد شوند.

علاوه بر این در سال‌های اخیر مدل‌های کلاسیکی هوش مصنوعی نیز در رتبه‌بندی اعتباری مورد استفاده قرار گرفته‌اند که شامل شبکه‌های عصبی، برنامه نویسی ژنتیک، نزدیک‌ترین همسایگی، ماشین بردار پشتیبان و مدل‌های درخت تصمیم می‌باشند.

تکنیک‌های داده کاوی مدرن در زمینه علم اطلاعات سهم بسزایی داشتند و می‌توان آن‌ها را با مدل‌های سنجش اعتبار تطبیق داد. متخصصان و محققان همواره به دنبال مدلی هستند که بتواند اندکی دقت پیش‌بینی را بالا ببرد این اندک تغییرات کوچک در دقت پیش‌بینی نقش بسزایی را در کاهش ریسک اعطای تسهیلات به متقاضیان وام دارد.

1. Liner Probability Model
2. Logit & Probit Model
3. Discriminate Analysis Model
4. Mathematical Planning
5. Nearest Neighbors Model
6. Analytical Hierarchy Process
7. Expert System
8. Genetic Algorithm

در سال‌های گذشته محققان همواره به دنبال مدلی بوده اند که دقت و کارایی خوبی داشته باشد. در ارزیابی متدها این مسأله که کدام زیر مجموعه از متغیرها موجود جهت پیش بینی انتخاب شود و انتخاب تصادفی متغیرها که بتواند دقت و کارایی مدل را افزایش دهد، همواره مورد توجه و بررسی محققان بوده است.

روش‌های گسترده ای در حوزه های ریاضی، آمار، اقتصادسنجی و پژوهش در عملیاتی همچون: برنامه ریزی ریاضی، شبیه سازی احتمالی و قطعی، شبکه های عصبی مصنوعی، تحلیل بقاء. نظریه ی بازی ها، تحلیل ممیزی، تحلیل لوجیت و تحلیل پروبیت در توسعه الگویی برای اندازه گیری دقیق ریسک اعتباری سهمیم بوده اند. همچنین پیشرفت نظریه های بازارهای مالی مانند نظریه ی آربیتراژ، نظریه ی قیمت گذاری اختیار معامله و مدل قیمت گذاری دارایی های سرمایه‌ای همگی در توسعه مدل‌های دقیق اندازه گیری ریسک اعتباری نقش مؤثری را ایفا کرده اند. (آلتمن و همکاران، ۲۰۰۲)

در این پژوهش سه روش برای ارزیابی مشتریان بانک از نقطه نظر اعتبار آن‌ها، مورد استفاده قرار خواهد گرفت. همچنین سعی شد تا با استفاده از یک مجموعه داده، مدلی مناسب برای پیش بینی وضعیت اعتباری مشتریان جدید طراحی شود. مدلی که بتواند با کم‌ترین خطا مشتریان را اعتبارسنجی کند. از آنجایی که پیشرفت صحت حتی به میزان کم می‌تواند منجر به کاهش هزینه های کلان برای بانک در زمینه ریسک اعتباری شود، در این پژوهش از روش‌های ماشین بردار پشتیبان<sup>۱</sup>، درخت تصمیم<sup>۲</sup> و شبکه عصبی<sup>۳</sup> برای اعتبارسنجی مشتریان استفاده می‌شود. همچنین در این مقاله سعی شده تا با بررسی کارایی تکنیک‌های داده کاوی، مناسب‌ترین مدل برای اندازه گیری ریسک اعتباری مشتریان در جهت اعطای تسهیلات مالی طراحی و تبیین شود.

1. Support Vector Machine
2. Decision Tree
3. Neural Network

## ۲-۲- پیشینه خارجی

طراحی مدلی برای اندازه‌گیری و درجه‌بندی ریسک اعتباری برای نخستین بار در سال ۱۹۰۹ به وسیله جان موری<sup>۱</sup> روی اوراق قرضه انجام شد. برخی از محققان متوجه شباهت زیاد اوراق قرضه و تسهیلات اعطایی شدند و اندازه‌گیری ریسک عدم پرداخت اصل و سود وام‌ها را بررسی نمودند.

در همین راستا اعتبارسنجی روشی برای شناخت گروه‌های مختلف جامعه است و زمانی مفید می‌باشد که شخص نمی‌تواند ویژگی‌هایی که گروه‌ها را از هم تفکیک می‌کند، مشاهده نماید. این تکنیک یک تکنیک آماری است که توسط "فیشر"<sup>۲</sup>، در سال ۱۹۳۶ بیان گردید. در سال ۱۹۴۱، "دوران"<sup>۳</sup> متوجه شد که می‌توان از این تکنیک و تکنیک‌های مشابه برای تفکیک قرض‌گیرندگان خوب و بد استفاده نمود. با ظهور کارت‌های اعتباری در سال ۱۹۶۰، ارزش اعتبارسنجی بیشتر شد. زمانی که بانک‌ها از این تکنیک استفاده کردند، متوجه شدند که این تکنیک بهتر از طرح‌های قضاوتی است. در سال ۱۹۶۶ برای تعیین ورشکستگی شرکت‌ها، مدل رگرسیون لجستیک به وسیله بی و ر<sup>۴</sup> بکار گرفته شد. بعدها از این مدل برای اندازه‌گیری ریسک اعتباری اوراق قرضه منتشر شده شرکت‌ها استفاده شد. در سال ۱۹۸۰ موفقیت اعتبارسنجی در کارت‌های اعتباری منجر به این شد که بانک‌ها به استفاده از روش‌های اعتبارسنجی به دیگر خدمات خود مثل وام به اشخاص ثالث تشویق گردند. زیرا تا قبل از آن به خانه و کسب و کارهای کوچک وام می‌دادند. موفقیت در امر محاسبات منجر به این شد که تکنیک‌های دیگر نیز به کمک اعتبارسنجی بیانند مانند رگرسیون لجستیک و برنامه ریزی خطی و درخت تصمیم‌گیری. اخیراً نیز تکنیک‌های هوش مصنوعی مانند سیستم‌های خبره و شبکه‌های عصبی به جمع این روش‌ها اضافه شده‌اند.

1. Jon Mory
2. Fisher
3. Durand
4. Beaver

اثرات متفاوت اندازه شرکت و نوع صنعت بر سودآوری ..... ۱۸۵

گوکاسیان و سیمان<sup>۱</sup> (۲۰۰۷) در " استراتژی هایی برای پیش بینی نکول در قرارداد اجاره تجهیزات " با استفاده از ۲۵۰ هزار قرارداد اجاره در طول دوره ی زمانی ۲۰۰۲ تا ۲۰۰۵ و بکارگیری سه روش رگرسیون لجستیک، تحلیل تمایزی و شبکه عصبی به نتایج زیر رسیدند: درجه ی رتبه- بندی ترکیبی<sup>۲</sup> پی نت (یک سیستم رتبه بندی اعتباری)، متغیرهای جمعیت شناسی سنتی<sup>۳</sup>، عقود اجاره قبلی شرکت<sup>۴</sup> و سابقه ی استقراض<sup>۵</sup> پیش بینی کننده های برجسته ی ریسک اعتباری در هر سه مدل طبقه بندی یاد شده بوده اند. نتیجه بیانگر این بود که بر خلاف انتظارات، تحلیل تمایزی پیش بینی دقیق تری نسبت به دو مدل دیگر ارائه داده است.

آلمر و بروفسکی<sup>۶</sup> (۱۹۸۸) برای " پیش بینی توانایی پرداخت وام ها از مدل شبکه عصبی چند لایه پرسپترون<sup>۷</sup> " استفاده نموده اند. متغیرهای استفاده شده در این مدل همان متغیرهای بکار گرفته شده توسط آلتمن (نسبتهای کل دارایی / سرمایه در گردش، کل دارایی / سود انباشته، کل دارایی / درآمد قبل از بهره و مالیات، ارزش دفتری بدهی ها / ارزش دفتری حقوق صاحبان سهام، کل دارایی / کل فروش) بوده و نتایج حاکی از این بود که قدرت پیش بینی مدل پرسپترون بیشتر از مدل های نمره دهی اعتباری بوده است.

### ۲-۳ - پیشینه داخلی

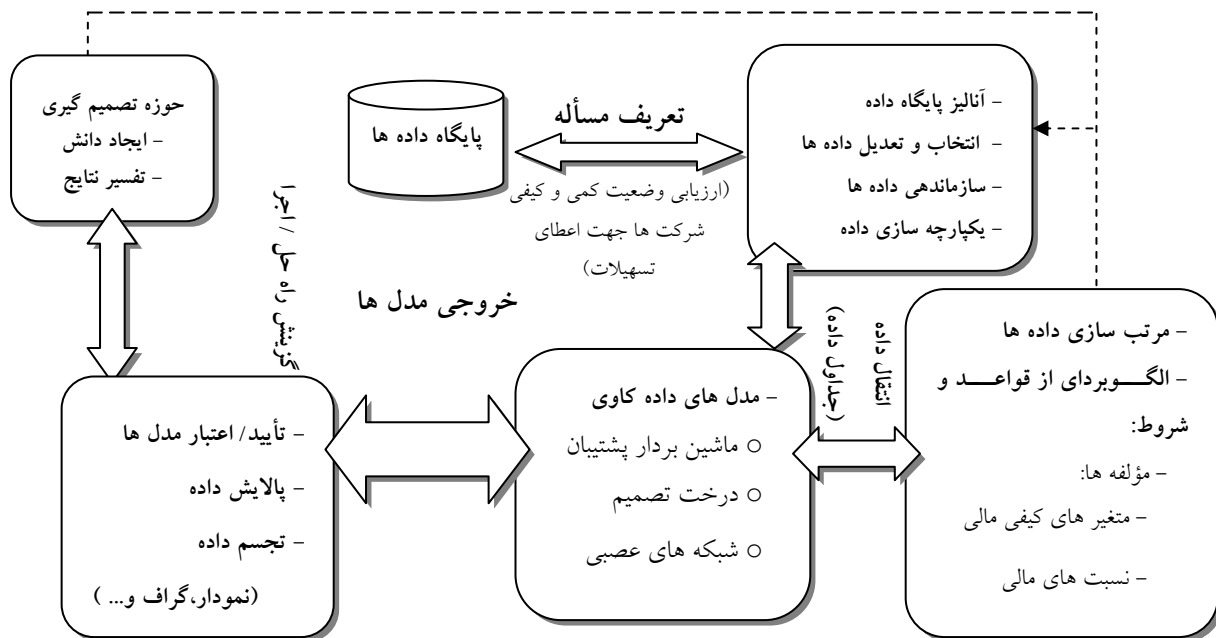
رضایی و آقاییگی (۱۳۸۶) در پژوهشی با نام " اعتبارسنجی مشتریان اعتباری بانک ملی بر اساس رگرسیون لجستیک " به سنجش اعتبار کلیه مشتریان حقوقی تسهیلات اعتباری بانک ملی طی سال های ۸۲ لغایت ۸۵ که فعالیت آن ها تولیدی بود، پرداختند.

1. Goukasian And Seaman
2. Composite Paynet Rating Score
3. Conventional Demographic Variables
4. Companys Prior Leasing
5. Borrowing History
6. Elmer And Borowski
7. Multi Layer Perceptron

در بدو امر تعداد ۶۱ متغیر تعیین و متعاقباً پس از استخراج کامل کلیه اطلاعات موجود از بانک‌های اطلاعاتی و پالایش اطلاعات بدست آمده نهایتاً تعداد ۱۶ متغیر جهت مدل‌سازی انتخاب گردید که عبارتند از: نوع طرح، داشتن هم‌گروه، داشتن تعهدات قبلی، نوع شرکت، مدت زمان تنفس، میزان تسهیلات، میزان سرمایه شرکت، سابقه فعالیت شرکت، میزان سهم متقاضی در سرمایه‌گذاری، نسبت دارایی جاری، دوره گردش موجودی بر حسب روز، دوره وصول مطالبات بر حسب روز، بازده فروش شرکت، نسبت مالکانه و نسبت بدهی. نتیجه تحقیق نشان می‌دهد، فرآیند اعتباردهی در بانک ملی، فرآیند قضاوتی است همچنین متغیرهای داشتن تعهدات قبلی، نسبت دارایی جاری، نسبت مالکانه و نسبت بدهی، بر میزان بدحسابی و خوش-حسابی متقاضیان تسهیلات متناسب با ضرایب برآورد شده تأثیرگذار می‌باشند.

تهرانی و فلاح شمس (۱۳۸۵) در "طراحی و تبیین مدل ریسک اعتباری در نظام بانکی کشور" با استفاده از داده‌های اعتباری ۳۱۶ مشتری حقوقی بانک‌های کشور و با استفاده از مدل‌های احتمال خطی، لجستیک و شبکه‌های عصبی مصنوعی اقدام به طراحی و آزمون کارایی مدل ریسک اعتباری پرداختند. نتایج حاکی از این بود که ارتباط بین متغیرها در مدل پیش‌بینی ریسک اعتباری به صورت خطی نبوده و تابع‌های نمایی و سیگموئید، مناسب‌ترین مدل‌های پیش‌بینی ریسک اعتباری است و بیشترین کارایی برای پیش‌بینی ریسک اعتباری به ترتیب مربوط به شبکه‌های عصبی مصنوعی و مدل لجستیک می‌باشد.





نمودار ۱-مدل مفهومی پژوهش

\* منبع: یافته‌های پژوهشگر

در مدل مطرح شده در نمودار ۱ به این صورت عمل می‌شود که در ابتدا داده‌های تحقیق در یک پایگاه جمع آوری می‌گردند، سپس بر روی این داده‌ها مسأله تحقیق (ارزیابی وضعیت کمی و کیفی شرکت‌ها جهت اعطای تسهیلات) بیان می‌گردد. بعد از بیان مسأله داده‌ها مورد تجزیه و تحلیل قرار می‌گیرند تا داده‌هایی که بر روی هدف تأثیر چندانی ندارند و به‌عنوان کاندید انتخاب گردیده بودند تعدیل و یکپارچه سازی گردند بعد از سازمان دهی اطلاعات، این داده‌ها به صورت جداولی به‌عنوان ورودی برای مدل‌های تحقیق (ماشین بردار پشتیبان، درخت تصمیم و شبکه عصبی) انتخاب می‌گردند. در داخل هریک از این مدل‌ها مجدداً داده‌ها بر اساس دستورالعمل هریک از مدل‌ها، مرتب و خوشه‌بندی می‌شوند. خروجی که از این مدل‌ها حاصل می‌گردد به دو

صورت می‌باشد: یکی اینکه منجر به تأیید اعتبار مدل (بر اساس مجموعه داده‌های آزمون) می‌گردد و دیگری اینکه تجزیه و تحلیل داده‌ها که همان هدف می‌باشد به صورت نمودار یا گراف (بسته به قواعد هر مدل) نمایش داده می‌شود. این خروجی‌ها برای اجرای تحقیق و گزینش بهترین مدل در اختیار حوزه تصمیم‌گیری قرار می‌گیرد که در این حوزه به تفسیر نتایج پرداخته می‌شود و باعث ایجاد یک دانش جدیدی می‌گردد که با یک سری داده‌های جدیدتر و متفاوت از حالتهای قبل دوباره تعدیل می‌گردند و وارد مدل می‌شوند تا بهترین نتیجه و کارا ترین مدل انتخاب گردد.

### ۳- فرضیه‌های پژوهش

بررسی فرضیات تحقیق در پژوهش‌های کاربردی از اهمیت ویژه‌ای برخوردار است در این پژوهش هدف اصلی تحقیق، اعتبارسنجی مشتریان بانکی می‌باشد که درصد دریافت تسهیلات اعتباری می‌باشند. فرضیات تحقیق متناسب با این هدف مورد توجه قرار گرفته و بررسی گردید.

#### ۳-۱- فرضیه اصلی

مدل‌های منتج از تکنیک‌های داده کاوی جهت اعتبارسنجی مبتنی بر صورت‌های مالی از کارآیی مناسبی برخوردار می‌باشند.

فرضیه‌های فرعی به شرح زیر می‌باشند:

**فرضیه فرعی ۱:** مدل منتج از تکنیک ماشین بردار پشتیبان جهت اعتبارسنجی مبتنی بر صورت‌های مالی از کارآیی مناسبی برخوردار است.

**فرضیه فرعی ۲:** مدل منتج از تکنیک درخت تصمیم جهت اعتبارسنجی مبتنی بر صورت‌های مالی از کارآیی مناسبی برخوردار است.

**فرضیه فرعی ۳:** مدل منتج از تکنیک شبکه‌های عصبی جهت اعتبارسنجی مبتنی بر صورت‌های مالی از کارآیی مناسبی برخوردار است.

#### ۴- ملاک کارآمدی مدل‌ها

در این تحقیق با توجه به نظر کارشناسان امر در حوزه داده کاوی، پرس و جو از خبرگان و مسئولان تصمیم‌گیری در بانک مبنای کارآیی مناسب برای مدل‌ها ۷۰٪ تعیین شد به این معنا که مدل‌هایی که درجه اعتبارشان از ۷۰٪ بیشتر باشد جهت اعتبارسنجی از کارآیی مناسبی برخوردار می‌باشند و مدل‌هایی که درجه اعتبارشان از ۷۰٪ کمتر باشد جهت اعتبارسنجی ناکارآمد می‌باشد.

#### ۵- روش پژوهش

این تحقیق با توجه به نتایجی که می‌تواند به همراه داشته باشد یک تحقیق بنیادی<sup>۱</sup> است. زیرا درصدد شناسایی عوامل مختلف تأثیرگذار بر رفتار دریافت کنندگان تسهیلات اعتباری و مدل‌سازی این رفتارها است و از طرف دیگر با توجه به کاربرد این تحقیق برای پیش بینی رفتار مشتریان بانک و برای مسائل اجرایی (در سیستم بانکی) به کار گرفته می‌شود، یک تحقیق کاربردی<sup>۲</sup> می‌باشد. از نظر روش تحقیق با توجه به ماهیت پژوهش در حوزه علوم مالی، تحقیق حاضر از روش پیمایشی<sup>۳</sup> استفاده می‌کند. در روش تحقیق پیمایشی نمونه ای از کل جامعه مورد پژوهش با استفاده از تکنیک‌های مناسب انتخاب شده و با بررسی و تحلیل نمونه یک نتیجه کلی حاصل می‌شود. در این پژوهش، اطلاعات گردآوری شده با استفاده از نرم افزارهای موجود جهت داده کاوی بر مبنای مدل‌های مختلف در این حوزه مورد تجزیه و تحلیل قرار خواهد گرفت. پایه اصلی تحقیق حاضر، بر کشف دانش از پایگاه داده‌های بانک مورد مطالعه نهاده شده است. از این رو جهت انجام فرآیند تحقیق از مراحل شامل درک مسئله کسب و کار، درک داده‌ها، آماده سازی داده‌ها، مدل‌سازی، ارزیابی نتایج، بکارگیری

1. Basic Research
2. Applied Research
3. Field Research

مدل و به همراه ارتباط بین مراحل مشخص می‌باشد، مورد استفاده قرار گرفته است. مراحل اجرایی و گام‌های اساسی در اجرای این پژوهش به صورت زیر قابل خلاصه شدن می‌باشد:

- ۱- جمع‌آوری داده از پایگاه داده‌های موجود (پرونده‌های تسهیلات اعطایی سابق بانک مورد نظر و سیستم‌های عملیاتی کامپیوتری بانک)
- ۲- شناسایی عوامل (متغیرهای) تأثیرگذار در رفتار شرکت‌ها جهت بازپرداخت بدهی که در پایگاه داده‌های مورد بررسی، موجود می‌باشد
- ۳- تعیین شاخص‌هایی برای تعریف طبقات: شرکت‌های خوب (دارای توان بازپرداخت بالا) و شرکت‌های بد (عدم توانایی در بازپرداخت)
- ۴- تقسیم داده‌های نمونه به دو مجموعه داده‌های آموزشی و داده‌های آزمون
- ۵- ساخت مدل‌ها با استفاده از داده‌های آموزشی
- ۶- آزمون مدل‌ها با مجموعه داده‌های آزمون
- ۷- بررسی دقت و سنجش اعتبار مدل‌ها در تعیین وضعیت اعتباری مشتریان
- ۸- ارائه بهترین الگو جهت ارزیابی وضعیت مشتریان

از آنجا که روش ارائه شده در هر تحقیقی باید به لحاظ اعتبار، مورد سنجش قرار گیرد، بنابراین در این تحقیق نیز با عنایت به اینکه روش تحقیق از نوع "داده محور" می‌باشد، روش اعتبارسنجی مدل‌ها به این صورت می‌باشد که داده‌ها به دو مجموعه آموزشی و داده‌های آزمون (تست) تقسیم می‌گردند. صحت طبقه‌بندی و تفکیک داده‌های آزمون در طبقه‌ها، معیار ارزیابی اعتبار و صحت مدل می‌باشد. که در این تحقیق از "اعتبارسنجی متقابل مدل با تکرار" استفاده شده است. این روش اعتبارسنجی مدل مجموعه داده‌ها را به ده قسمت تقسیم نموده و هر بار ۷۵ درصد از داده‌ها را به عنوان مجموعه داده آموزشی و ۲۵ درصد را به عنوان مجموعه داده آزمون انتخاب نموده و میزان دقت طبقه‌بندی را می‌سنجد. این فرآیند ده بار صورت می‌گیرد و در نتیجه از کلیه درجات دقت میانگین گرفته شده و به عنوان دقت نهایی مدل ارائه می‌گردد. که در نهایت از سه روش

مذکور برای ارزیابی مشتریان بانک از نقطه نظر اعتبار آن‌ها سعی شده است تا با استفاده از یک مجموعه داده، مدلی مناسب برای پیش بینی وضعیت اعتباری مشتریان جدید طراحی شود. مدلی که بتواند با کم‌ترین خطا مشتریان را اعتبارسنجی کند چون که پیشرفت صحت حتی به میزان کم می‌تواند منجر به کاهش هزینه های کلان برای بانک در زمینه ریسک اعتباری شود.

## ۶- جامعه و نمونه آماری پژوهش

با توجه به اینکه هدف تحقیق اعتبارسنجی مشتریان می‌باشد، در این پژوهش جامعه آماری شامل شرکت‌های وام گیرنده که در شعبات بانک ملی استان تهران در طی سال های ۸۹ و ۹۰ از این بانک، تسهیلات دریافت کرده و اصل و سود آن را با بانک‌ها عودت داده یا نداده اند، می باشند. دلیل انتخاب واحدهای اقتصادی به‌عنوان جامعه آماری، در دسترس بودن داده‌های مالی موثق و حسابرسی شده آن‌ها می‌باشد. جامعه آماری متشکل از مشتریان خوش حساب (ریسک اعتباری کمتر) و مشتریان بدحساب (ریسک اعتباری بالاتر) می‌باشد.

با توجه به اینکه دسترسی به کل داده‌های بانک امکان پذیر نبوده و بنا بر اظهارات مدیران بانک مورد مطالعه، داده‌ها به صورت پراکنده و غیر منسجم در اختیار شعبات می‌باشد، بنابراین بر اساس یک نمونه گیری تصادفی و روش کوکران<sup>۱</sup> با حجم جامعه محدود در خصوص این جامعه آماری، تعداد ۳۴۵ مشتری انتخاب گردید. که فرمول آن در رابطه (۱) آمده است.

رابطه ۱- فرمول کوکران با حجم جامعه محدود

$$n = \frac{\frac{z^2 pq}{d^2}}{1 + \frac{1}{N} \left( \frac{z^2 pq}{d^2} - 1 \right)}$$

که در این فرمول داریم:

حجم جامعه آماری (تعداد شرکت های دریافت کنندگان

۱۹۲ ..... پژوهش‌های حسابداری مالی و حسابرسی، سال پنجم، شماره هفدهم، بهار ۱۳۹۲

تسهیلات مالی در سالهای ۸۹ و ۹۰ (N= ۳۳۰۰) و احتمال برخورداری از صفت مورد نظر: مشتریان خوش حساب (p= ۰.۵) و احتمال عدم برخورداری از صفت مورد نظر: مشتریان بد حساب (q=۰.۵) و انحراف یا خطای مطلوب (d= ۰.۰۵) و درجه یا ضریب اطمینان ۹۵ درصد (Z=۱.۹۶)

$$n = \frac{(1.96)^2 \times \frac{(0.5 \times 0.5)}{(0.05)^2}}{1 + \frac{1}{3300} (1.96)^2 \times \frac{(0.5)(0.5)}{(0.05)^2} - 1} = 344.19$$

در این صورت داریم:

از این تعداد فوق فقط ۳۰۰ پروند واجد شرایط بودند. مسئله ای که درباره ی این مشتریان وجود داشت عدم وجود صورت‌های مالی در پرونده آن‌ها بود.

## ۷- متغیرهای پژوهش و توصیف برخی داده‌های گردآوری شده

انتخاب متغیرهایی که با احتمال نکول وام گیرنده رابطه مشخصی داشته باشند، یکی از مراحل مهم تحقیق است. در این پژوهش با استفاده از نتایج تحقیقات داخلی و خارجی مرتبط با موضوع و ادبیات موضوع، متغیرهای متعددی در دو حوزه متغیرهای کیفی و مالی و حوزه نسبت های مالی مورد بررسی قرار می‌گیرد که عبارت است از: الف) پارامترهایی که هریک از مشتریان برای دریافت تسهیلات به بانک ارائه می دهند و در پرونده آن‌ها موجود است مثل نوع شرکت (تعاونی، سهامی عام، سهامی خاص و با مسئولیت محدود)، موضوع فعالیت شرکت (صنعتی، خدماتی و بازرگانی)، سابقه فعالیت شرکت، میزان سرمایه شرکت، مبلغ وام، سطح تحصیلات مدیر عامل، وضعیت مالکیت محل فعالیت، وضعیت مالیاتی، اعتبار شرکت نزد بانک (خوش حساب یا بدحساب بودن) و نسبتهای مالی از جمله نسبت های سودآوری، نسبت های اهرمی، نسبت های نقدینگی و نسبت های فعالیت و کارایی.

با توجه به تعداد متغیرها، به منظور تعیین مدل بهینه و بالا بردن دقت مدل و از

سویی محدودیت های روش های کاربردی در رابطه با تعداد متغیرهای توضیحی، براساس مطالعات پیشین متغیرهای فوق انتخاب گردیدند و در مدل قرار خواهند گرفت. از سویی دیگر، از آن جایی که بسیاری از متغیرها، از صورت های مالی و اطلاعات پایه ای آن استخراج می شوند، ممکن است به صورت دو به دو با یکدیگر همبستگی داشته باشند برای جلوگیری از عدم همپوشانی آنها، بر اساس نظر کارشناسان امر، تعدادی از این متغیرهای به هم وابسته و متغیرهایی که تأثیر قابل توجهی در خروجی سیستم ندارند، حذف می گردند. از این رو متغیرهای شناخته شده در بدو امر متغیرهای کاندید تلقی شده و به عنوان ورودی بکار گرفته می شوند. که در ادامه در جدول ۱ به توصیف برخی از آنها اشاره شده است.

**اعتبار:** یکی دیگر از متغیرها می باشد که به دو دسته تقسیم می گردد:

- ۱- مشتریان خوش حساب: مشتریانی هستند که یا هیچگونه تأخیری در پرداخت اقساط خود نداشته و یا حداکثر ۲ ماه تأخیر دارند.
  - ۲- مشتریان بدحساب: مشتریانی هستند که بدهی آنها به سرفصل مطالبات سررسید گذشته (بین ۳ تا ۶ ماه بعد از سررسید) و یا معوق (از ۷ تا ۱۸ ماه) و یا مشکوک الوصول (بیش از ۱۸ ماه) منتقل شده باشد.
- ضمناً شرکت هایی که پس از سررسید گذشته یا معوق شدن تسهیلات مورد نظر و پس از گذشت مدتی بیش از ۲ ماه نسبت به تأدیه تعهدات خود اقدام می نمایند، نیز جزء گروه مشتریان بدحساب می باشند.

جدول ۱ - توصیف برخی متغیرهای کیفی به تفکیک خوش حساب و بدحساب

متغیرهای کیفی	شرح	فراوانی	درصد	خوش حساب	بدحساب
زمینه فعالیت	صنعتی	۱۱۲	٪ ۳۷	۸۳	۲۹
	خدماتی	۹۷	٪ ۳۳	۶۲	۳۵
	بازرگانی	۹۱	٪ ۳۰	۷۴	۱۷
وضعیت مالیات	پرداخت شده	۱۴۸	٪ ۵۰	۸۹	۵۹
	پرداخت نشده	۱۳	٪ ۴	۶	۷
	در دست رسیدگی	۴۸	٪ ۱۶	۳۰	۱۸
	عدم افشاء	۹۱	٪ ۳۰	۶۹	۲۲
نوع شرکت	سهامی عام	۲۰	٪ ۷	۱۴	۶
	سهامی خاص	۲۳۴	٪ ۷۸	۱۵۲	۸۲
	بامسئولیت محدود	۴۶	٪ ۱۵	۴۱	۵
مدرک تحصیلی	دیپلم	۱۹	٪ ۶	۱۱	۸
	فوق دیپلم	۲۹	٪ ۱۰	۲۰	۹
	لیسانس	۲۰۰	٪ ۶۷	۱۵۲	۴۸
	فوق لیسانس	۴۰	٪ ۱۳	۳۵	۵
	دکترآ	۱۲	٪ ۴	۱۰	۲

\* منبع: یافته‌های پژوهشگر

با توجه به موضوع مورد مطالعه لازم است متغیرهایی در نظر گرفته شود که با بررسی آن‌ها بتوان مالی شرکت‌های مورد بررسی و در نهایت رتبه‌بندی آن‌ها را از این لحاظ تعیین کرد. در واقع باید متغیرهایی انتخاب شود که دارای بیشترین رابطه با توان مالی است. به منظور شناسایی متغیرهایی که بیشترین را با توان مالی دارند، پژوهش



اثرات متفاوت اندازه شرکت و نوع صنعت بر سودآوری ..... ۱۹۵

که تاکنون در این زمینه انجام شده است مورد بررسی قرار گرفت که در جدول ۲ به متغیرهای مورد استفاده در برخی از آن تحقیقات داخلی اشاره شده است.

جدول ۲ - متغیرهای مورد استفاده در برخی از تحقیقات داخلی

متغیرهای مورد استفاده	موضوع تحقیق	محقق - سال
نوع شرکت، سابقه همکاری با بانک، سابقه فعالیت، مبلغ تسهیلات، سوابق اعتباری مشتریان، نسبتهای مالی استخراج شده از صورت‌های مالی شرکت‌ها	مدل ریسک اعتباری و رتبه‌بندی مشتریان حقوقی	تقوی، لطفی و سهرابی (۱۳۸۶)
مبلغ وام، سرمایه شرکت، داشتن گزارش حسابرسی، نسبت جاری، نسبت آنی، سابقه همکاری با بانک، سابقه داشتن بدهی‌های معوق به بانک، نوع شرکت	عوامل مؤثر بر ریسک اعتباری مشتریان بانکی	عرب مازار و رویین تن (۱۳۸۵)
تعداد کل حسابهای بانکی، نوع فعالیت، نوع تسهیلات، سابقه همکاری، نسبتهای مالی	بررسی عوامل مؤثر بر ریسک اعتباری اشخاص حقوقی بانک‌ها	میرزایی، نظریان و باقری (۱۳۸۹)
نوع مصرف وام، زمینه فعالیت، نوع قرارداد وام، سطح تحصیلات مدیر، نوع بازپرداخت، نوع شرکت، تعداد کارکنان، نسبتهای مالی شرکت‌ها	مقایسه الگوریتم‌های درخت ژنتیکی و کلونی مورچگان جهت دسته‌بندی مشتریان اعتباری بانک	نادعلی، نصرت آبادی و عباسیان (۱۳۸۹)

<p>میزان تسهیلات، داشتن گزارش حسابرسی، داشتن مجوز کسب، مالکیت محل فعالیت، سابقه بازپرداخت، نسبتهای مالی</p>	<p>مدیریت ریسک اعتباری مشتریان حقوقی در بانک‌های تجاری</p>	<p>صفری، ابراهیمی و شیخ (۱۳۸۸)</p>
---	--	--

\* منبع: یافته‌های پژوهشگر (ادامه جدول ۲)

## ۸- روش‌های تحلیل داده‌ها و آزمون فرضیه‌ها

در این تحقیق پس از جمع آوری داده‌های شرکت‌های دریافت‌کننده تسهیلات بانک مورد نظر از پایگاه داده مربوطه و پس از آن، پالایش داده‌ها، به شناسایی متغیرهای تأثیرگذار در ارزیابی وضعیت مالی شرکت‌ها پرداخته که این کار از طریق مصاحبه با کارشناسان و مستندات علمی، انجام می‌گردد. بعد از این مرحله برای تمامی مشتریان نمونه انتخاب شده، با توجه به تعریفی که از خوش حساب یا بدحساب بودن شرکت‌ها بر اساس وضعیت مالی ارزیابی شده توسط صورت‌های مالی وجود دارد، یک برچسب مربوط به آن بخش (طبقه) با همان تعریف در نظر گرفته می‌شود.

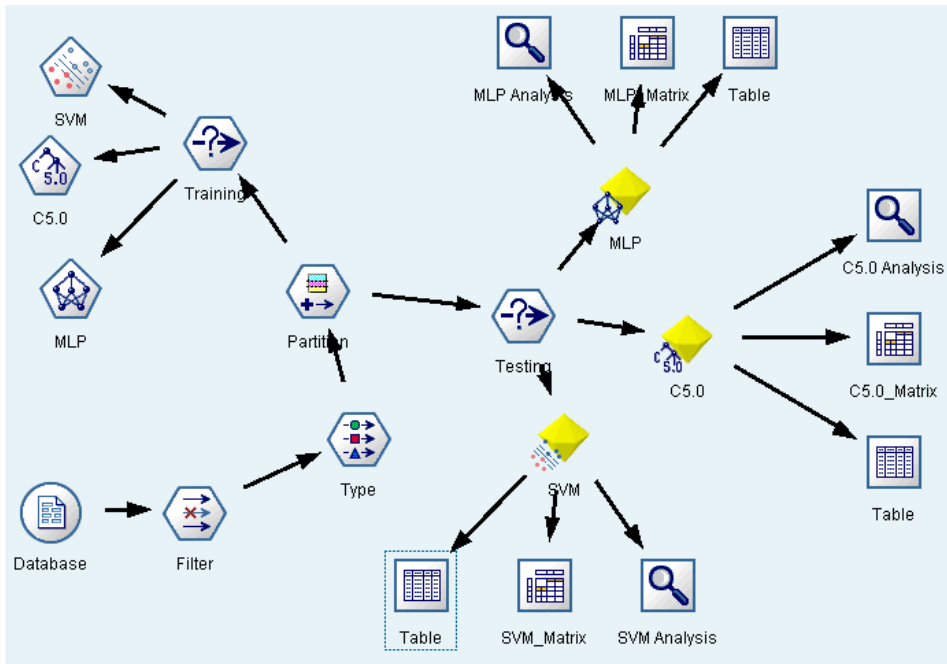
در مرحله بعد با استفاده از تکنیک‌های داده کاوی مورد نظر و با کمک نرم افزارهای مربوطه، شرکت‌ها را بر اساس ویژگی‌های شان و اطلاعات صورت‌های مالی آن‌ها ارزیابی نموده، بعد با تسهیلات اعطاء شده تطبیق می‌دهیم تا میزان تفاوت در تصمیم‌گیری در اعطای تسهیلات توسط بانک‌ها مشخص شود که در نهایت این مدل جهت تصمیم‌گیری راجع به اعطاء یا عدم اعطای تسهیلات استفاده خواهد شد.

برای ساخت مدل لازم است ابتدا تکنیک مدل‌سازی و نرم افزارهای لازم برای اعمال تکنیک‌ها انتخاب شود. برای اعمال تکنیک‌ها در تحقیق حاضر از نرم افزارهای

SPSS Clementine و Microsoft Excel V 2007 و Microsoft SQL Server V 2008

V 12 استفاده شده است.

در نمودار ۲ نحوه ساخت الگوریتم‌های مورد مطالعه نمایش داده شده است.



نمودار ۲- نحوه ساخت الگوریتم در نرم افزار SPSS Clementine

\* منبع: یافته‌های پژوهشگر

در رابطه با نحوه ساخت مدل‌ها در هریک از فرضیه‌های مربوطه توضیحاتی ارائه شده است. اما در مورد مدل‌های تحقیق به اختصار در ذیل توضیحاتی ارائه شده است:

### الف) ماشین بردار پشتیبان

متد دیگری که در سال‌های گذشته در زمینه ارزیابی وضعیت اعتبار مشتریان بسیار مورد توجه محققان بوده است، ماشین بردار پشتیبان است. این متد یک متد ریاضی قوی در دسته بندی است. متدی شبیه به شبکه های عصبی که به جای خط جدا کننده بدنبال بهترین خط جدا کننده است که دارای حداکثر حاشیه باشد. یعنی بهترین خط جدا کننده ای که با نزدیکترین نقطه کم‌ترین فاصله را داشته

۱۹۸ ..... پژوهش‌های حسابداری مالی و حسابرسی، سال پنجم، شماره هفدهم، بهار ۱۳۹۲

باشد (کیونگ، تایک، هیونگ<sup>۱</sup>، ۲۰۰۵، ۱۳۰). در ماشین بردار پشتیبان انتخاب بهینه ویژگی‌ها در دقت مدل تأثیر بسزایی دارد از این رو در سال‌های گذشته بسیاری از تحقیقات بر روی ماشین بردار پشتیبان، بر روی انتخاب بهینه پارامترها متمرکز می‌باشند.

تکنیک‌های یادگیری ماشینی بخشی از علم هوش مصنوعی به شمار می‌آیند. هوش مصنوعی به طور مؤثر و موفقی در گستره وسیعی از شاخه‌های تجاری همچون بازاریابی، حسابداری، سیستم‌های اطلاعاتی مدیریت و مدیریت تولید بکار گرفته می‌شود. در اکثر تحقیقات از روش‌های یادگیری ماشینی برای پیش‌بینی تغییرات آتی سهام، بحران‌های مالی، ورشکستگی، شناسایی کلاهبرداری کارت‌های اعتباری بکار گرفته می‌شود.

### ب) درخت تصمیم

درخت تصمیم یکی از ابزارهای قوی و متداول برای دسته‌بندی و پیش‌بینی می‌باشد. در طبقه‌بندی نمونه به روش درخت تصمیم‌گیری، به این صورت عمل نمی‌شود که امتیازی به هر شاخص داده شده و سپس مقادیر شاخص‌ها با هم جمع شود. در مقابل در این روش، مشتریان به گروه‌هایی تقسیم می‌شوند که هر گروه در داخل خود از نظر ریسک قصور در اعتبار کاملاً همگن می‌باشد و بین گروه‌ها نیز از حیث ریسک قصور در اعتبار، ناهمگنی زیادی مشاهده می‌شود. درخت تصمیم‌گیری بر خلاف شبکه‌های عصبی به تولید قانون می‌پردازد. یعنی درخت تصمیم‌گیری پیش‌بینی خود را در قالب یک سری قوانین توضیح می‌دهد در حالیکه در شبکه‌های عصبی تنها پیش‌بینی نهایی بیان می‌شود و چگونگی آن در خود شبکه پنهان باقی می‌ماند. (کارپتین<sup>۲</sup>، ۱۹۹۹، ۱۱۰)

درختان تصمیم‌گیری تنها مقادیر ویژگی‌های گسسته را توسط متغیر هدف (ویژگی کلاس یا ویژگی وابسته) پیش‌گویی می‌کند. این پیش‌گویی توسط متغیر کلاس که

---

1. Shin S. Kyung, Lee S. Taik, & Kim J. Hyun  
2. Corporation

ویژگی هدف یا ویژگی وابسته نیز نامیده می‌شود، صورت می‌گیرد. مقادیر ویژگی هدف، وابسته به مقادیر متغیرهای (ویژگی‌های) مستقل (توصیف کننده) و وجود آن‌ها در ساختار درخت تصمیم‌گیری است (هیگر و پاو<sup>۱</sup>، ۲۰۰۳، ۶).

### ج) شبکه عصبی

این شبکه‌ها ابزارهایی هستند که در موارد پیش بینی، خوشه بندی و تخمین کاربرد دارند. جذابیت شبکه‌های عصبی در این است که آن‌ها بوسیله مدل‌سازی ارتباطات عصبی مغز انسان در کامپیوترهای دیجیتال پوی برای فاصله موجود ایجاد می‌کنند.

برای شناخت الگوهایی که در داده‌ها وجود دارند بسیار مفید هستند، خصوصاً در مواقعی که نوع رابطه بین هدف (به‌عنوان مثال ارزیابی وضعیت اعتباری وام‌گیرنده) و متغیرهای ورودی (مثلاً ویژگی‌های جمعیتی، نسبت‌های مالی، روند اقتصادی، کیفیت مدیریت و...) نامشخص و یا پیچیده باشد. یک شبکه عصبی بر اساس سه ویژگی داده‌های ورودی، وزن‌ها و لایه‌های پنهان مشخص می‌شود (ساندرز و آلن، ۲۰۰۲، ۱۲). برای ایجاد شبکه‌های عصبی چندین روش وجود دارد که شبکه عصبی پرسپترون، بویژه پرسپترون چند لایه<sup>۲</sup>، در زمره‌ی کاربردی‌ترین شبکه‌های عصبی مصنوعی می‌باشند. این شبکه‌ها می‌توانند با گزینش شماره لایه‌ها و سلول‌های عصبی (نرون‌ها)، که اغلب زیاد نیستند، یک نگاهت غیر خطی را با دقت دلخواه انجام دهد. همچنین این روش به‌عنوان "طبقه بند جامع"<sup>۳</sup> شناخته می‌شود به این دلیل که از لحاظ تئوریک قادر به طبقه بندی هر فرآیند تصمیم‌گیری است.

شبکه‌های عصبی روش محاسبه‌ای متفاوت با روش‌های متداول می‌باشند. محاسباتی که با روش‌های معمولی انجام می‌شود از نوع برنامه ریزی شده است و در آن‌ها الگوریتم‌ها و مجموعه‌هایی از قواعد به کار می‌روند تا مسأله را حل کنند. در این روش‌ها اگر الگوریتمی در دست نباشد یا الگوریتم پیچیده باشد، راه حلی برای

1. D heygere, T,Goethals, P.L & Pauw , N.D

2. MLP or Back Propagation

3. Universal Classifier

مسأله وجود نخواهد داشت. اما در محاسباتی که شبکه عصبی اجرا می‌کند به الگوریتم و مجموعه قواعد نیاز نیست.

## ۹- نتایج آزمون و یافته‌های فرضیه اصلی

فرضیه اصلی مطرح می‌کند که مدل‌های منتج از تکنیک‌های داده کاوی جهت اعتبارسنجی مبتنی بر صورت‌های مالی از کارایی مناسبی برخوردار می‌باشند. این فرضیه که هدف آن بررسی کارایی تکنیک‌های داده کاوی در اعتبارسنجی مشتریان اعتباری بانک به منظور ارزیابی آن‌ها می‌باشد از اهمیت ویژه‌ای در این تحقیق برخوردار است. زیرا در صورتیکه با استفاده از این مدل‌ها می‌توان قبل از اعطای تسهیلات، مشتریان را ارزیابی به تبع آن می‌توان تسهیلات اعتباری را به مشتریان واجد شرایط اعطا نمود. به منظور آزمون این فرضیه ابتدا با استفاده از پیشینه پژوهشی و با استفاده از ایده‌های محقق متغیرهای کیفی و کمی و نسبت‌های مالی انتخاب گردیدند. سپس با استفاده از نرم افزار SPSS Clementine کارایی مدل‌ها در پیش بینی وضعیت اعتباری مشتریان بانک بررسی گردید.

نتایج حاصل از بررسی در هر سه مدل، کارایی این تکنیک‌ها را در اعتبارسنجی نشان داد. از آنجایی که تکنیک‌های داده کاوی از رابطه خطی پیروی نمی‌کنند مانند روش‌های آماری و این محدودیت را برطرف نموده است و با توجه به میزان صحت طبقه بندی آن‌ها، این فرضیه تأیید می‌گردد. بدین معنی که با استفاده از تکنیک‌های داده کاوی می‌توان مشتریان را بر اساس صورت‌های مالی اعتبارسنجی نمود.

### ۹-۱- روش آزمون فرضیه فرعی اول

فرضیه فرعی ۱: مدل منتج از تکنیک ماشین بردار پشتیبان جهت اعتبارسنجی مبتنی بر صورتهای مالی از کارایی مناسبی برخوردار است.

برای آزمون این فرضیه از نرم افزار SPSS Clementine استفاده گردید. بدین

اثرات متفاوت اندازه شرکت و نوع صنعت بر سودآوری ..... ۲۰۱

صورت که ابتدا داده‌های خام را جمع آوری و طبقه بندی نموده، سپس در قالب SQL وارد نرم افزار گردید. بعد فیلتر را در راه آن گذاشته تا پاکسازی و پیش پردازش روی داده‌های انجام شود. بعد بر اساس یک تایپ متغیرهای ورودی و خروجی را مشخص کرده و بر اساس ایجاد پارتیشن، مقدار داده‌های آزمون و آموزش را مشخص می‌شود که در این تحقیق ۷۵ درصد را به‌عنوان داده‌های آموزش و ۲۵ درصد را برای داده‌های آزمون انتخاب می‌گردد. بعد الگوریتم ماشین بردار پشتیبان (SVM) را ساخته و در آن هدف (تعیین وضعیت اعتباری) و متغیرهای ورودی را آورده سپس الگوریتم را اجرا نموده تا مدل ماشین بردار ساخته شود. سایر مشخصات پارامترهای این الگوریتم عبارت است از: نوع کرنل: آر بی اف (RBF) - پارامتر تنظیم: ۱۰ - صحت رگرسیونی: ۰.۱ - گاما: ۱ - گاما آر بی اف: ۰.۱ - بایاس: ۰ - درجه: ۳

#### ۹-۱-۱- یافته ها و نتایج آزمون فرضیه فرعی اول

میزان صحت مدل ماشین بردار پشتیبان (SVM) در مجموعه داده‌های آموزش ۶۹.۱۴٪ می‌باشد. درصد‌های مربوط به آزمون مربوط به صحت طبقه بندی و معیاری برای ارزیابی اعتبار و صحت مدل می‌باشد که در این تکنیک ۷۴.۵۰٪ می‌باشد. همچنین تعداد داده‌های آموزش مدل ۴۴۴ تراکنش و تعداد داده‌های آزمون در این مدل ۱۴۹ تراکنش برآورد شد.

پیش بینی های مورد انتظار برای ۱۴۹ مشتری در مجموعه داده‌های آزمون برای مدل ماشین بردار پشتیبان در جدول ۳ مشاهده می‌شود.

جدول ۳ - عملکرد پیش بینی مدل ماشین بردار پشتیبان در دوره آزمون

شرح	خوش حساب	بد حساب	مجموع
موارد تشخیص به‌عنوان خوش حساب	۹۱	۲۱	۱۱۲
موارد تشخیص به‌عنوان بد حساب	۱۷	۲۰	۳۷
مجموع	۱۰۸	۴۱	۱۴۹
تعداد پیش بینی درست	۹۱	۲۰	۱۱۱

۷۴.۵۰	۴۸.۷۸	۸۴.۲۵	درصد دقت مدل
۲۵.۵۰	۵۱.۲۲	۱۵.۷۵	درصد خطای مدل

\* منبع: یافته‌های پژوهشگر (ادامه جدول ۳)

یافته‌های تحقیق حاکی از این است که از مجموع ۱۰۸ مشتری خوش حساب، تعداد ۹۱ مشتری به‌عنوان خوش حساب و ۱۷ مشتری (به اشتباه) به‌عنوان بدحساب تشخیص داده شده‌اند به طوری که دقت الگو در این خصوص ۸۴.۲۵ درصد محاسبه شده است. به همین ترتیب از ۴۱ مشتری بدحساب، ۲۰ مشتری به‌عنوان بدحساب و ۲۱ مشتری (به اشتباه) خوش حساب تشخیص داده شده‌اند و دقت الگو در این حالت حدود ۴۸.۷۸ درصد می‌باشد. بنابراین میانگین دقت پیش‌بینی در این مدل ۷۴.۵۰ درصد است.

با توجه به یافته‌ها، این مدل از میزان صحت بالایی برخوردار می‌باشد چون میزان کارایی در این تحقیق ۷۰٪ تعیین شده است، در نتیجه این فرضیه تأیید می‌شود و از کارایی مناسبی برای اعتبارسنجی برخوردار می‌باشد.

#### ۹-۱-۲- روش آزمون فرضیه فرعی دوم

فرضیه فرعی ۲: مدل منتج از تکنیک درخت تصمیم (C5.0) جهت اعتبارسنجی مبتنی بر صورت‌های مالی از کارایی مناسبی برخوردار است.

درختان تصمیم‌گیری دارای الگوریتم‌های مختلفی هستند از جمله: C5.0 و CHAID و C&R و QUEST و C4.0 و C4.5. در این مقاله از بهترین الگوریتم و آخرین نوع آن یعنی الگوریتم (C5.0) که نتایج بهتری همراه با بالاترین دقت تصمیم‌گیری و صحت را دارد استفاده خواهیم کرد.

الگوریتم درخت تصمیم‌گیری با انتخاب آزمونی شروع می‌شود که بهترین جداسازی را برای دسته‌ها انجام دهد. مهمترین هدف از انجام دسته‌بندی، به دست آوردن مدلی برای پیش‌بینی می‌باشد. بدین منظور از مجموعه‌ای از داده‌ها به نام " داده‌های آموزشی " که مجموعه‌ای از متغیرها و رکوردها است استفاده می‌کنیم. در



اثرات متفاوت اندازه شرکت و نوع صنعت بر سودآوری ..... ۲۰۳

مراحل بعدی، همین کار برای گره های پایین تر با داده های کمتر صورت می گیرد تا بهترین قانون ها حاصل شود. سرانجام درخت آن قدر بزرگ می شود تا دیگر نتوان جداسازی بهتری برای داده های گره انجام داد (آذر، احمدی، و حید بسط، ۲۱، ۱۳۸۹). در این مرحله باید اثربخشی درخت ایجاد شده اندازه گیری شود. برای این کار از یک مجموعه رکوردها یا داده های آزمایشی استفاده می شود که متفاوت با داده های اولیه که درخت را ایجاد کرده اند میباشد. معیاری که اندازه گیری می شود هم عبارت است از درصد داده هایی که به طور صحیح دسته بندی می شوند و دسته پیش بینی شده با دسته واقعی آنها یکسان است.

#### ۹-۱-۲-۱- یافته ها و نتایج آزمون فرضیه فرعی دوم

میزان صحت مدل درخت تصمیم (C5.0) در مجموعه داده های آموزش ۹۴.۳۷٪ می باشد. درصد های مربوط به آزمون مربوط به صحت طبقه بندی و معیاری برای ارزیابی اعتبار و صحت مدل می باشد که در این تکنیک ۹۱.۹۵٪ می باشد. همچنین تعداد داده های آموزش مدل ۴۴۴ تراکنش و تعداد داده های آزمون در این مدل ۱۴۹ تراکنش برآورد شد.

پیش بینی های مورد انتظار برای ۱۴۹ مشتری در مجموعه داده های آزمون برای مدل درخت تصمیم در جدول زیر مشاهده می شود.

جدول ۴ - عملکرد پیش بینی مدل درخت تصمیم در دوره آزمون

شرح	خوش حساب	بد حساب	مجموع
موارد تشخیص به عنوان خوش حساب	۱۰۵	۹	۱۱۴
موارد تشخیص به عنوان بد حساب	۳	۳۲	۳۵
مجموع	۱۰۸	۴۱	۱۴۹
تعداد پیش بینی درست	۱۰۵	۳۲	۱۳۷
درصد دقت مدل	۹۷.۲۲	۷۸.۰۴	۹۱.۹۵
درصد خطای مدل	۲.۷۸	۲۱.۹۶	۸.۰۵

(ادامه جدول ۴)

\* منبع: یافته های پژوهشگر

۲۰۴ ..... پژوهش‌های حسابداری مالی و حسابرسی، سال پنجم، شماره هفدهم، بهار ۱۳۹۲

یافته‌های تحقیق حاکی از این است که از مجموع ۱۰۸ مشتری خوش حساب، تعداد ۱۰۵ مشتری به‌عنوان خوش حساب و ۳ مشتری (به اشتباه) به‌عنوان بدحساب تشخیص داده شده‌اند به طوری که دقت الگو در این خصوص ۹۷.۲۲ درصد محاسبه شده است. به همین ترتیب از ۴۱ مشتری بدحساب، ۳۲ مشتری به‌عنوان بدحساب و ۹ مشتری (به اشتباه) خوش حساب تشخیص داده شده‌اند و دقت الگو در این حالت حدود ۷۸.۰۴ درصد می‌باشد. بنابراین میانگین دقت پیش‌بینی در این مدل ۹۱.۹۵ درصد است.

با توجه به یافته‌ها، این مدل از میزان صحت بالایی برخوردار می‌باشد و قدرت پیش‌بینی آن هم بین سایر مدل‌ها از میزان قابل قبولی برخوردار است و اینکه توانست متغیرهای با اهمیت را شناسایی نماید در نتیجه این فرضیه تأیید شد و از کارآیی مناسبی برای اعتبارسنجی برخوردار می‌باشد.

### ۹-۱-۳- روش آزمون فرضیه فرعی سوم

فرضیه فرعی ۳: مدل منتج از تکنیک شبکه‌های عصبی جهت اعتبارسنجی مبتنی بر صورتهای مالی از کارآیی مناسبی برخوردار است.

مدل انتخابی از نوع پرسپترون چند لایه است این مدل‌ها کاربرد موفقیت آمیزی در حل برخی از مسائل مخصوصاً مسائل رتبه‌بندی داشته‌اند. این مدل جزء شبکه‌های پیشخور چند لایه است که مجموعه‌ای از واحدهای گیرنده، لایه ورودی را تشکیل می‌دهند و یک یا چند لایه پنهانی از گره‌های محاسباتی وجود دارد و همچنین یک لایه خروجی گره محاسباتی در انتها قرار دارد.

تعداد نرون‌ها در لایه ورودی به سادگی برابر با تعداد متغیرها در مجموعه داده‌هایی است که ورودی شبکه را تشکیل می‌دهند. با توجه به هدف تحقیق که طبقه‌بندی متقاضیان به دو گروه است، یک نرون در لایه خروجی به کار گرفته می‌شود که دو کد مختلف را به متقاضیان خوب و بد اختصاص می‌دهد. همچنین شبکه دارای یک لایه میانی است که باید در مورد تعداد نرون‌ها در این لایه تصمیم‌گیری شود.

اثرات متفاوت اندازه شرکت و نوع صنعت بر سودآوری ..... ۲۰۵

### ۹-۱-۳-۱- یافته ها و نتایج آزمون فرضیه فرعی سوم

میزان صحت مدل شبکه عصبی (MLP) در مجموعه داده‌های آموزش ۹۴.۵۵٪ می‌باشد. درصدهای مربوط به آزمون مربوط به صحت طبقه بندی و معیاری برای ارزیابی اعتبار و صحت مدل می‌باشد که در این تکنیک ۹۳.۴۶٪ می‌باشد. همچنین تعداد داده‌های آموزش مدل ۴۴۰ تراکش و تعداد داده‌های آزمون در این مدل ۱۵۳ تراکش برآورد شد. پیش بینی های مورد انتظار برای ۱۴۹ مشتری در مجموعه داده‌های آزمون برای مدل شبکه عصبی در جدول ۵ مشاهده می‌شود.

جدول ۵- عملکرد پیش بینی مدل شبکه عصبی در دوره آزمون

شرح	خوش حساب	بدحساب	مجموع
موارد تشخیص به‌عنوان خوش حساب	۱۰۴	۶	۱۱۰
موارد تشخیص به‌عنوان بدحساب	۴	۳۹	۴۳
مجموع	۱۰۸	۴۵	۱۵۳
تعداد پیش بینی درست	۱۰۴	۳۹	۱۴۳
درصد دقت مدل	۹۶.۲۹	۸۶.۶۶	۹۳.۴۶
درصد خطای مدل	۳.۷۱	۱۳.۳۴	۶.۵۴

\* منبع: یافته‌های پژوهشگر

یافته‌های تحقیق حاکی از این است که از مجموع ۱۰۸ مشتری خوش حساب، تعداد ۱۰۴ مشتری به‌عنوان خوش حساب و ۴ مشتری (به اشتباه) به‌عنوان بدحساب تشخیص داده شده اند به طوری که دقت الگو در این خصوص ۹۶.۲۹ درصد محاسبه شده است. به همین ترتیب از ۴۵ مشتری بدحساب، ۳۹ مشتری به‌عنوان بدحساب و ۶ مشتری (به اشتباه) خوش حساب تشخیص داده شده اند و دقت الگو در این حالت حدود ۸۶.۶۶ درصد می‌باشد. بنابراین میانگین دقت پیش بینی در این مدل ۹۳.۴۶ درصد است.

با توجه به یافته ها، این مدل از میزان صحت بالایی برخوردار می‌باشد و قدرت پیش بینی آن هم بین سایر مدل‌ها از بقیه بالاتر می‌باشد و درجه صحت طبقه بندی آن

۲۰۶ ..... پژوهش‌های حسابداری مالی و حسابرسی، سال پنجم، شماره هفدهم، بهار ۱۳۹۲

اختلاف معناداری با سایر مدل‌ها دارد و اینکه توانست متغیرهای با اهمیت را شناسایی نماید در نتیجه این فرضیه تأیید شد و از کارآیی مناسبی برای اعتبارسنجی برخوردار می‌باشد.

## ۱۰- نتایج مقایسه کارآیی مدل‌ها

برای سنجش کارآیی مدل‌های ارائه شده، صحت نتایج این مدل‌ها در جدول ۶ مقایسه گردیده است:

جدول ۶- مقایسه تطبیقی صحت الگوریتم‌ها

الگوریتم	SVM	C5.0	MLP
صحت	٪ ۷۴.۵۰	٪ ۹۱.۹۵	٪ ۹۳.۴۶

\* منبع: یافته‌های پژوهشگر

همانطور که در (۶) مشخص است، هر سه مدل مورد مطالعه (SVM، C5.0 و MLP) در این تحقیق از لحاظ صحت پیش بینی از قدرت بالایی برخوردار می‌باشد که در این بین الگوریتم شبکه عصبی (MLP) نسبت به دو الگوریتم دیگر از قدرت بالاتری برخوردار است.

## ۱۱- نتیجه گیری

با توجه به اهمیت شناسایی ریسک نکول و یافتن راهکارهایی برای مدیریت و کنترل آن، دسترسی به مدل‌هایی که با دقت بالا بتواند این رخداد را پیش بینی نماید، نقش مهمی به لحاظ اقتصادی و عملی در بقا و سلامت بانک‌ها دارد. در این پژوهش سه مدل مختلف برای پیش بینی ریسک نکول برآورد گردید و یافته‌های حاصل مقایسه شدند.

در حالی که نتایج پژوهش حاضر نشان‌دهنده ی این واقعیت است که امکان پیش بینی ریسک اعتباری مشتریان در هنگام اعطای تسهیلات اعتباری از راه مختصات

اثرات متفاوت اندازه شرکت و نوع صنعت بر سودآوری ..... ۲۰۷

مشتریان به عنوان متغیرهای پیش بین و استفاده آن‌ها در مدل‌های آماری و شبکه‌های عصبی وجود دارد. نتایج حاکی از این است که در بین سه مدل طراحی شده، به ترتیب مدل شبکه عصبی پرسپترون و درخت تصمیم از کارایی بالایی برای پیش بینی ریسک اعتباری مشتریان (یعنی پیش بینی احتمال عدم بازپرداخت مشتریان) برخوردار هستند. به طور کلی نتایج مستقیم حاصل از اجرای تحقیق را می‌توان به صورت زیر خلاصه نمود:

- شبکه عصبی، مشتریان را با دقت ۹۳.۴۶ درصد در دو دسته از پیش تعیین شده طبقه بندی نمود.

- درخت تصمیم، مشتریان را با دقت ۹۱.۹۵ درصد در دو دسته از پیش تعیین شده طبقه بندی نمود.

- ماشین بردار پشتیبان، مشتریان را با دقت ۷۴.۵۰ درصد در دو دسته از پیش تعیین شده طبقه بندی نمود.

- شبکه عصبی دقت بالاتری نسبت به درخت تصمیم و ماشین بردار پشتیبان و سایر الگوریتم‌های طبقه بندی داده کاوی دارد.

- ماشین بردار پشتیبان در مقایسه با شبکه عصبی و درخت تصمیم، کارایی پایین تری به لحاظ دقت تفکیک مشتریان (در دو طبقه مورد نظر) دارد.

- براساس متغیرهای کمی و کیفی می‌توان مشتریان حقوقی بانک را از نظر ریسک اعتباری دسته بندی و امتیازدهی نمود که از بین این متغیرها، مبلغ وام، نسبت جاری، نسبت مالکانه، نسبت بدهی، بازده سرمایه در گردش، نسبت دارایی جاری، نسبت تسهیلات دریافتی کوتاه- مدت به فروش، نسبت تسهیلات دریافتی به دارایی، بازده فروش بیشترین سهم را در تفکیک مشتریان به دو گروه شرکت‌های با ریسک اعتباری بالا و شرکت‌های با ریسک اعتباری پایین دارند.

نتایج حاصل از این تحقیق مبنی بر عملکرد برتر شبکه‌های عصبی نسبت به سایر مدل‌های طبقه بندی، با نتایج تحقیقات قبلی شامل آلمر و بروفسکی (۱۹۹۸)، فلاح

۲۰۸ ..... پژوهش‌های حسابداری مالی و حسابرسی، سال پنجم، شماره هفدهم، بهار ۱۳۹۲

شمس (۱۳۸۵)، عیسی زاده (۱۳۸۷)، انصاری (۱۳۸۸) منطبق است، اما با نتایج گوکاسیان و همکاران (۲۰۰۷)، قدس الهی (۱۳۸۹)، نادعلی (۱۳۸۹)، هداوندی (۱۳۸۸) هماهنگ نیست. ولی در یک موضوع همگی باهم در اشتراک هستند و آن هم برتری تکنیک های داده کاوی نسبت به روش های آماری و قضاوتی می باشد.

در نهایت با توجه به جنبه های کاربردی پژوهش حاضر، پیشنهاد های حاصل از تحقیق به ترتیب زیر مطرح می شود:

۱- به سرمایه گذاران و اعتبار دهندگان به ویژه بانک ها و مؤسسات مالی پیشنهاد می شود که جهت تصمیم گیری در مورد اعطای وام و انتخاب بهترین متقاضی اعتبار از بین چندین گزینه مختلف، از تکنیک شبکه عصبی و نسبت های مالی منتخب در این تحقیق، استفاده نمایند.

۲- با استقرار این سیستم و تعیین میزان ریسک مشتریان، این امکان برای سیستم بانکی فراهم خواهد شد که نرخ سود تسهیلات اعطایی هر مشتری را نه بر اساس بخش اعتباری بلکه بر اساس ریسک هر مشتری مشخص نمایند.

۳- طراحی و برقراری سیستم نرم افزاری مدل شبکه های عصبی و درخت تصمیم در پیش بینی ریسک اعتباری و اتصال آن به پایگاه اطلاعاتی برای تعیین درجه ریسک اعتباری هریک از مشتریان در هر لحظه از زمان جهت استفاده تمامی شعب بانک ها و مؤسسات مالی.

۴- الزامی نمودن بانک ها به اخذ صورت های مالی از مشتریان متقاضی اعتبار (که توسط حسابرس مستقل حسابرسی شده اند) بر اساس دستورالعمل اجرایی، در زمینه اعطای تسهیلات و تهیه گزارش های ویژه مربوط، توسط واحدهای تجاری وام گیرنده.

۵- بورس اوراق بهادار با بهره گیری از رتبه بندی (با استفاده از تکنیک های داده کاوی) می تواند در راستای جلوگیری از بروز بحران های مالی گریبان گیر شرکت ها، سرمایه گذاران و همچنین اعتبار دهندگان و یا برای کل بازار، اقدام کند.

۶- طراحی نرم افزار جهت بکارگیری مدل شبکه های عصبی در بانک های کشور و

## *Archive of SID*

اثرات متفاوت اندازه شرکت و نوع صنعت بر سودآوری ..... ۲۰۹

بازنگری و به روز کردن این مدل‌ها با اعلام اطلاعات مالی جدید مشتریان در هر سال  
مالی.

## منابع

اخباری، مهدیه، (۱۳۸۷)، "رتبه‌بندی اعتباری مشتریان حقوقی بانک‌ها با رویکرد هوش مصنوعی"، پایان‌نامه کارشناسی ارشد، دانشگاه صنعتی اصفهان.  
ایزد پناهی، نگهدار، (۱۳۷۲)، "میزان اتکای بانک‌های تجاری به اطلاعات صورت‌های مالی در اعطای تسهیلات به واحد‌های اقتصادی در ایران"، پایان‌نامه کارشناسی ارشد، دانشگاه علامه طباطبائی.

تهرانی، رضا و میرفیض فلاح شمس، (۱۳۸۴)، "طراحی و تبیین مدل ریسک اعتباری در نظام بانکی کشور"، مجله علوم انسانی و اجتماعی دانشگاه شیراز، ۲ (۴۳): ۶۰-۴۵.

حائری مهریزی، علی اصغر، (۱۳۸۲)، "داده‌کاوی: مفاهیم و روش‌ها و کاربردها"، پایان‌نامه کارشناسی ارشد، دانشکده اقتصاد دانشگاه علامه طباطبائی.

سوری، داود، (۱۳۸۴)، "یک مدل سنجش اعتبار برای مشتریان اشخاص حقوقی یک بانک، کاربردی از روش بیزین"، مجموعه مقالات شانزدهمین همایش بانکداری اسلامی، ۴۸۴-۴۸۸.

مدرس، احمد و سید مرتضی ذکاوت، (۱۳۸۱)، "مدل‌های ریسک اعتباری مشتریان بانک (مطالعه موردی)"، فصل‌نامه حسابرس، ۱۹(۱۴): ۵۸-۵۴.

میرعمادی، سید علی اکبر، (۱۳۸۶)، "بررسی رابطه‌ی بین نسبت‌های مالی (متغیرهای حسابداری) و ریسک اعتباری مشتریان بانک‌ها"، پایان‌نامه کارشناسی ارشد، دانشگاه تربیت مدرس.

Abdi. H.(2007)," Discriminant Correspondence Analysis" in: N.J. Salkind(Ed.):Encyclopedia of Measurement and Statistics.Thousand Oaks(CA):Sage,PP.270-275.

Abdou H. Pointon J.(2008), " Neural Nets Versus Conventional Techniques in Credit Scoring in Egyptian Banking". Expert Systems with Applications.doi:10.1016/j.eswa,P.1.

Cheng-Lung Huang et al.:(2006), " Credit Scoring with a Data Mining



Based on Support Vector Mashines", Expert System with Application 33,PP.847-856.

E. Khandani, Amir. J. Kim, Adlar.and W. Lo, Andrew.(2010)," Consumer Credit-Risk Models via Machine Learning Algorithms".Journal of Banking & Finance 34,PP.2767-2787.

Elmer , P. J & D. M. Borowski.(1988), " An Expert System and Neural Networks Approach To Financial Analysis ", Financial Management ,12(5):pp.66-76.

Goukasian, L. & S. Seaman.(2009)," Strategies for Predicting Equipment Lease Default", Journal of Equipment Lease Financing , 27(1):pp. 1-7.

Hassan Sabzevari, Mehdi Soleymani, Eaman Noorbakhsh,(2007)," A Comparison Between Statistical and Data Mining Methods for Credit Scoring in Case of Limited Available Data", Credit Research Center Conference, UK.s.n,PP.1-7.

Hussein A.Abdou.(2010), " Genetic Programming for Credit Scoring: The Case of Egyptian Public Sector Banks", Expert System with Application,36:pp.11402-11417,doi:10.1016/j.esw.

Nanni L., Lumini A.(2009)," An Experimental Comparison of Ensemble of Classifiers for Bankruptcy Prediction and Credit Scoring". Expert Systems with Applications.Vol.36,PP.1-4.

Weimin Chen, Guocheng Xiang, Youjin Liu, Kexi Wang.(2012)," Credit risk Evaluation by Hybrid Data Mining Technique", Systems Engineering Procedia, Volume3,issue(2012),PP.194-200.

Xu,X,Zhou.C,&Wang,Z.(2008)," Credit Scoring Algorithm based on Link Analysis Ranking with Support Vector Machine",Expert Systems With Applications. Vol.36 Issue 2, March, 2009 ,PP. 2625-2632.

Ye, Jieping. Janardan, Ravi.and Li, Qi.(2011), " Two-Dimensional Linear Discriminant Analysis", Vol.44, No.10-11, PP.2565-2575.