

طبقه بندی متقاضیان تسهیلات اعتباری بانکی با استفاده از داده کاوی و منطق فازی

محمد تقی تقوی فرد*
احمد نادعلی**

چکیده

در این پژوهش، هدف، بهره گیری از ابزارهای داده کاوی و منطق فازی برای طبقه بندی مشتریان تسهیلات اعتباری می باشد به طوریکه ابهامات و عدم قطعیت را در خصوص طبقات مشتریان و نیز متغیرهای تاثیر گذار در رفتار آنها را پوشش دهد. روش کار بدین شکل می باشد که طبق یک فرایند استاندارد داده کاوی، داده های مشتریان سابق بانک سامان جمع آوری و پالایش شده و سپس طبقات و متغیرهایی که قابلیت فازی کردن داشتند، طبق نظر کارشناسان بانک و با توجه به اصطلاحات کلامی آنها برای این متغیرها، فازی شده و با استفاده از تکنیک درخت تصمیم فازی، داده های نهایی مدلسازی گردیدند. همچنین داده های غیر فازی نیز با چند الگوریتم دیگر مدلسازی شدند. نتیجه حاصل شده نشان داد که درخت تصمیم فازی نتایج بهتری را به لحاظ دقت تفکیک مشتریان نسبت به درختهای سنتی، شبکه های

* استادیار، دانشکده مدیریت و حسابداری، دانشگاه علامه طباطبایی تهران (نویسنده مسئول)

Email: dr.taghavifard@gmail.com

** کارشناسی ارشد مدیریت فناوری اطلاعات، دانشگاه آزاد اسلامی واحد علوم و تحقیقات تهران

عصبی و روش‌های آماری از قبیل رگرسیون لجستیک و شبکه‌های بیزین دارد؛ ولی نسبت به مدل‌های درخت ژنتیکی و ماشین بردار پشتیبان دقت کمتری دارد. از طرفی تکنیک درخت تصمیم فازی نسبت به عملکرد پیش‌بینی کارشناسان اعتبارسنج بانک نیز قدرت پیش‌بینی بهتری حاصل نموده است. واژگان کلیدی: امتیازدهی اعتباری، طبقه‌بندی، داده کاوی، منطق فازی، درخت تصمیم فازی.

مقدمه

بانک‌ها به عنوان 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 ریسک 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 رتبه‌بندی اعتباری مشتریان" است. [۶] با بهره‌گیری از تحلیل اطلاعات مربوط به مشتریان بانک با استفاده از فرایند داده کاوی می‌توان به رتبه‌بندی اعتباری متقاضیان وام و طبقه‌بندی آنها به مشتریان خوش حساب و بد حساب، بدون قضاوت شخصی و براساس سیستم‌های هوشمند پرداخت. از آنجائیکه مساله تخصیص وام مساله‌ای غیر قطعی و نیمه ساختارمند است و در ارزیابی درخواست وام از عوامل کیفی و کمی توأم با هم استفاده می‌شود؛ لذا یکی از روش‌های برخورد با چنین مسایلی می‌تواند بهره‌گیری از "منطق فازی" باشد. در گذشته نیز در زمینه رتبه‌بندی اعتباری مشتریان تحقیقاتی صورت گرفته است. در این میان می‌توان به مطالعه "فیشر" (در سال ۱۹۳۶)، که پایه‌های روش امتیازدهی اعتباری نیز بنا شده است و اولین سیستم ارزیابی تقاضای اعتبار می‌باشد، اشاره نمود [۱۹]. این مقاله به بررسی قابلیت تشخیص گروه‌ها در یک جامعه کارخانه دارد براساس پنج معیار تجربی موقعیت شغلی فرد، صورت درآمد، صورت

مالی (ترازنامه)، ضامن ها یا وثایق و اطلاعات مربوط به بازپرداخت وام دریافتی از بانک ها می پردازد. در سال ۱۹۴۱ "دوراند" به منظور دانستن پارامترهای مهم از نظر وام دهندگان و ویژگی هایی که از نظر آماری مهم بودند، از تحلیل تمایزی^۱ و با تکیه بر نتایج فیشر انجام گرفت، استفاده کرد [۱۷]. این کار وی، حرکتی برای بهبود چارچوب نظری تعیین اهمیت یک معیار مطمئن بود. وی همچنین پیشنهاداتی برای تحلیل ریسک اعتباری ارائه داد. او در مطالعه خود، سیستم مبتنی بر امتیازدهی را، برای طبقه بندی افراد متقاضی دریافت وام جهت خرید ماشین دست دوم، مورد استفاده قرار داد. وی از پارامترهای، شغل یا سمت شغلی متقاضی وام، تعداد سال هایی که در این سمت شغلی بوده است، تعداد سال های سپری شده در آدرس جاری، حساب های بانکی، بیمه نامه های عمر و پس انداز، جنسیت و میزان اقساط ماهانه ای که در حال حاضر می پردازد، در مدل خود استفاده نمود "بوگس" در مقاله سال ۱۹۶۷، اولین فردی است که استفاده از ابزارهای پیچیده چند متغیره آمار نمود که به بهبود بیش از پیش مدل های دقیق امتیازدهی اعتباری منجر گردید [۱۱]. از جمله مطالعات دیگر می توان به مقاله، "بی ور" (۱۹۶۷)، در زمینه برآورد موفقیت و شکست شرکتها با استفاده از برخی شاخص های مالی اشاره نمود [۹]. همچنین "آلتمن" (۱۹۶۸) به عنوان یکی از پیشگامان بحث اعتبارسنجی، تلاش بسیاری برای یافتن یک رابطه معنی دار بین متغیرهای حسابداری یک شرکت و احتمال عدم توانایی در پرداخت دیون این شرکت در آینده انجام داد و رابطه ای معروف به Z-Score را ارائه داد. این روش مبتنی بر تحلیل ممیزی خطی بین شرکت های خوب و بد بود [۷]. دیکن (۱۹۷۲) در زمینه استفاده از روش تحلیل ممیزی به منظور ارزیابی عوامل شکست شرکتها با استفاده از ۱۴ نسبت مالی به عنوان متغیر مستقل [۱۵] و همچنین ارزیابی عملکرد شرکتها با استفاده از همین مدل (۱۹۸۹) اقدام نموده است [۱۶]. مورگان (۱۹۹۴) [۲۶] در زمینه طراحی مدل اندازه گیری ریسک اعتباری و تریسی (۱۹۹۸) در زمینه طراحی مدل ارزش در خطر^۲ برای تخمین تابع چگالی احتمال عدم بازپرداخت [۳۳]، اشاره نمود.

1- Discriminate Analysis
2- Value-at-risk

از منطق فازی نیز در تحقیقات جدید بهره گرفته شده است ولی هیچکدام مستقیماً از درخت تصمیم فازی برای طبقه بندی اعتباری مشتریان تسهیلات بانکی استفاده نموده‌اند و کارایی آن در مقایسه با روشهای دیگر آزمون نشده است. ولی تحقیق پیش رو سعی در بررسی طبقه بندی فازی متقاضیان تسهیلات بانکی با استفاده از درخت تصمیم فازی به عنوان یکی از تکنیک‌های داده کاوی دارد. هدف از این تحقیق شناسایی عوامل تاثیرگذار بر رفتار اعتباری مشتریان متقاضی تسهیلات اعتباری بانکی با توجه به پایگاه داده‌های موجود در بانک سامان و جستجوی الگوهای موجود بین داده‌های مشتریانی که قبلاً تسهیلات اعتباری دریافت نموده‌اند با استفاده از روش طبقه بندی متقاضیان تسهیلات اعتباری بانکی با استفاده از داده کاوی و منطق فازی می‌باشد.

در ادامه ابتدا ادبیات موضوع تشریح می‌گردد و سپس در قسمت روش تحقیق به بیان مراحل طی شده در انجام این تحقیق توضیح می‌گردد و سپس در بخش چارچوب نظری تحقیق به بیان داده‌های تحقیق و درخت تصمیم فازی پرداخته خواهد شد و سپس در بخش یافته‌های تحقیق به بیان دقت مدل و ارزیابی اعتبار مدل و همچنین کارایی مدل در ارتباط با سایر مدلها و کارایی مدل در مقایسه با عملکرد کارشناسان اعتبار سنج بانک پرداخته خواهد شد.

ادبیات موضوع

داده کاوی و دسته بندی فازی

داده کاوی، عبارت از اقتباس یا استخراج دانش از مجموعه‌ای از داده‌ها است [۱۸]. به بیان دیگر، داده کاوی فرایندی است که با استفاده از تکنیکهای هوشمند، دانش را از مجموعه‌ای از داده‌ها استخراج می‌کند. دانش استخراج شده در قالب مدلها، الگوها یا قواعد ارائه می‌شود. این الگوها، مدلها و قواعد اشکال مختلفی از برای ارائه دانش استخراج شده هستند. این دانش می‌تواند ملاک تصمیم‌گیری‌های آتی، عملکردهای بعدی و یا تغییرات لازم در سیستم قرار گیرند [۴].

یکی از روشهای اکتشاف دانش که عموماً در داده کاوی به کار می‌رود روش طبقه بندی و استفاده از الگوریتمهای درخت تصمیم می‌باشد درخت‌های تصمیم می‌توانند قواعد قابل فهمی را تولید کنند و حتی در یک درخت بزرگ یا پیچیده هم، یک مسیر را به راحتی می‌توان طی کرد و این باعث می‌شود که تفسیر دسته بندی‌ها یا پیش بینی‌ها نسبتاً آسان باشد. الگوریتم‌های مختلفی جهت ساخت درخت‌های تصمیم معرفی شده است، که از روش‌های معروف آن می‌توان به روش‌های ^۱AID، ^۲CHAID، ^۳CART، ^۴OC، ^۵ID3، ^۶C4.5، ^۷C5، QUEST^۸ و SAS algorithms اشاره کرد [۵].

وجود الفاظ زبانی^۹ در منطق فازی و قابل تفسیر بودن آنها توسط بشر یکی از مهمترین مزایای سیستم‌های فازی است [۲۴]. از آنجایی که قابل تفسیر بودن نتایج برای یک سیستم یادگیری یک مزیت به‌شمار می‌آید، به‌کارگیری منطق فازی در فرایند استنتاج و پردازش سیستم‌های داده‌کاوی از این نظر نقطه قوتی برای آن محسوب می‌شود [۳۵].

درخت‌های تصمیم، تصمیمات نمادینی را به نمونه‌ها منتسب می‌کنند. با این وجود، اگر چه روش‌های گوناگونی برای ساخت درخت‌های تصمیم ارائه شده‌اند، و اگر چه به‌کارگیری این روش‌ها در قلمروهای نمادین با موفقیت‌های بسیار خوبی همراه بوده است، ولیکن درخت‌های تصمیم نمادین، برخی مواقع از بازدهی مناسبی برخوردار نخواهند بود به‌عنوان مثال هنگامی که یک تصمیم عددی مورد نیاز می‌باشد، یا هنگامی که تصمیم‌گیری عددی می‌تواند پردازش‌های بعدی را بهبود بخشد [۲۲]؛ و یا وجود آشفتگی‌هایی در مجموعه آموزشی از قبیل اغتشاش یا فقدان ارزش خصیصه‌ها در توصیف نمونه‌ها [۲۷]؛ و یا هنگامی که نیاز به کلاس‌بندی فازی باشد [۱۴].

دسته‌بندی فازی نوعی دسته‌بندی است که در آن به جای استفاده از مقادیر

1- Automatic Interaction Detection
2- Chi-Squared Automatic Interaction Detection
3- Classification and Regression Trees
4- Oblique Classifier
5- Iterative Dichotomizer tree
6- Quick, Unbiased, Efficient, Statistical Tree
7- Linguistic Terms

عددی (واقعی) از مقادیر فازی استفاده می‌شود [۱۳]. منطق فازی با به کار بردن مجموعه‌هایی که قابلیت همپوشانی دارند، کارایی سیستم دسته‌بندی و سیستم‌های تصمیم‌یار را بهبود بخشیده است [۲۰]. همچنین با فراهم کردن دقت بیشتر دسته‌بندی [۳۲] و فرآیند تصمیم‌گیری توانسته است تفسیر پذیری نتایج را بهبود بخشد [۳۱]. یکی از روشهای دسته‌بندی فازی، به کارگیری روش ایجاد درخت تصمیم فازی است. یک درخت تصمیم فازی دارای ساختاری مشابه با درخت تصمیم عادی است با این تفاوت که داده‌هایی که با آن سروکار دارد، فازی هستند. البته مجموعه قواعد دسته‌بندی فازی می‌تواند توسط فرد خبره تعریف شود و یا با به کارگیری روشهای مختلف یادگیری ماشین از یک مجموعه از اشیاء استخراج گردد.

امتیاز دهی اعتباری

امتیازدهی اعتباری، نظامی است که به وسیله آن بانک‌ها و موسسات اعتباری با استفاده از اطلاعات حال و گذشته متقاضی، احتمال عدم بازپرداخت وام توسط وی را ارزیابی نموده و به او امتیاز می‌دهند. به عبارت دیگر امتیاز دهی به معنی کمی نمودن احتمال نکول در آینده است [۳۴]. امتیاز دهی اعتباری یک ابزار عینی برای مدیریت ریسک است که مشتریان اعتباری را بی طرفانه و براساس آمار و اطلاعات رتبه‌بندی می‌نماید. درحالی که روش‌های قدیمی برای ارزیابی مشتریان عمدتاً ذهنی و متکی بر دیدگاه مسئول (یا مسئولین) پرداخت وام می‌باشند [۱]. صاحب‌نظران تعاریف نسبتاً مشابهی درباره رتبه‌بندی اعتباری ارائه نموده‌اند. فلدمن، رتبه‌بندی اعتباری را تخصیص یک معیار کمی بصورت واحد یا نمره، به یک قرض گیرنده بالقوه، جهت ارائه تخمینی از عملکرد آتی وی می‌داند [۸]. از نظر موسسه "استاندارد اند پور" "سیستم رتبه‌بندی، اظهار نظر در مورد ارزش اعتباری یک بدهکار بر اساس عوامل ریسک است". به نظر موسسه مودیز^۲ رتبه‌بندی عبارت از "اظهار نظر در مورد توانایی آتی بدهکار و تعهد حقوقی منتشر کننده

1- Standard and poor
2- Moodys

اوراق برای انجام پرداخت‌های به موقع اصل و بهره، روی یک اوراق بهادار با درآمد ثابت و مشخص است [۲۵].
به طور کلی فنون اندازه‌گیری ریسک اعتباری را می‌توان به دو گروه عمده تقسیم بندی نمود [۲۱]:

۱- الگوهای امتیاز دهی اعتباری غیر پارامتری

- برنامه ریزی ریاضی
- درخت طبقه بندی
- الگوی نزدیک ترین همسایه
- فرآیند تحلیل سلسله مراتبی
- سیستم خبره
- شبکه‌های عصبی مصنوعی و الگوریتم ژنتیک

۲- الگوهای امتیازدهی اعتباری پارامتری

- مدل احتمالی خطی
 - مدل تحلیل ممیزی
 - مدل لوجیت
 - مدل پروبیت
- موسسات اعتباری و بانکها می‌بایست با توجه به پیچیدگی فعالیتها و محیط اقتصادی پیرامونشان، مدل‌های مناسب جهت ارزیابی امتیازدهی اعتباری مشتریان را انتخاب نمایند.

یکی از عوامل موفقیت تصمیمات اعتباری، کیفیت اطلاعات از لحاظ دقیق، کامل و به روز بودن است. عمده‌ترین اطلاعات مورد نیاز، اطلاعات راجع به تاریخچه پرداخت‌ها، اطلاعات در مورد اوراق بهادار، اطلاعات اعتباری مشخصی در مورد مالک یا مالکین بنگاه، اطلاعات مالی بنگاه، اطلاعات در مورد شاخصه‌های اقتصادی و ... است. اما از پرکاربردترین این اطلاعات برای اداره وام و اعتبار بانک، نسبت‌های مالی است [۲].

به منظور امتیاز دهی اعتباری از معیارهای مختلفی بهره گرفته می‌شود از جمله

آنها معیار 5C شامل شخصیت، ظرفیت، سرمایه، وثیقه، شرایط و ضوابط تسهیلات یا اعتبارات^۱ و معیار LAPP شامل نقدینگی، فعالیت، سودآوری و توان بالقوه و یا معیار 5P شامل مردم، محصول، حمایت و پرداختها و شمای کلی آینده^۲ می‌توان اشاره کرد [۱].

از مزایای امتیاز دهی اعتباری برای مشتریان می‌توان به فرایند اعتباری بسیار ساده‌تر و پاسخ در یک چارچوب زمانی کوتاه‌تر و کاهش میزان اطلاعات مورد نیاز و دسترسی سریعتر و آسان‌تر به اعتبار موقعی که مشتریان به آن نیاز دارند، اشاره کرد و از مزایای امتیاز دهی اعتباری برای بانکها می‌توان به کاهش هزینه‌های ارزیابی وام‌ها و ارائه اعطاء وام استاندارد در بانک و افزایش کارایی اعطاء وام اشاره کرد.

روش تحقیق

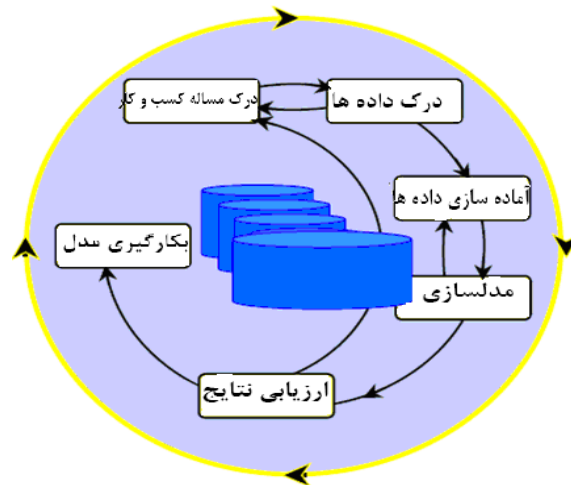
با توجه به ماهیت تحقیق که استفاده از داده کاوی جهت طبقه‌بندی مشتریان بانک می‌باشد؛ این تحقیق از نوع داده محور^۳ می‌باشد. پایه اصلی تحقیق حاضر بر کشف دانش از پایگاه داده‌های بانک مورد مطالعه نهاده شده است از اینرو استاندارد جهانی^۴ CRISP-DM جهت انجام فرایند تحقیق مورد استفاده قرار گرفته است. که در شکل ۱ این مراحل که شامل درک مسئله کسب و کار، درک داده‌ها، آماده سازی داده‌ها، مدل‌سازی، ارزیابی نتایج و بکارگیری مدل می‌باشد به همراه ارتباط بین مراحل مشخص می‌باشد.

1- Condition or terms of loans

2- Perspective

3- Data Oriented

4- CRoss Industry Standard Process for Data Mining



شکل ۱. فرایند داده کاوی مدل [۱۲] CRISP-DM

نگرش این تحقیق از پایین به بالا می باشد. به این شکل که از کار با داده ها شروع شده و سعی بر آن است تا مواردی را که قبلاً آگاهی نسبت به آنها وجود نداشته است، کشف کرده و برای آنها قوانینی ساخته شود. بنابراین از روش کشف دانش استفاده شده است در این تحقیق پس از جمع آوری داده های مشتریان سابق بانک مورد نظر از پایگاه های داده مربوطه و پس از آن، پالایش داده ها، به شناسایی متغیرهای تاثیرگذار در رتبه بندی مشتریان پرداخته شد که این کار از طریق مصاحبه با کارشناسان امر و مستندات علمی، انجام گرفته است. سپس متغیرهایی که قابل فازی شدن بودند، با تابع عضویتی که در نظر گرفته شد، به فرمت متغیر فازی تبدیل گردیدند. بعد از این مرحله برای تمامی مشتریان نمونه نهایی، با توجه به تعریفی که از خوش حساب بودن یا بد حساب بودن مشتریان وجود دارد، یک بر حسب طبقه با همان تعریف در نظر گرفته شده است. در مرحله بعد با استفاده از تکنیک درخت تصمیم فازی و با کمک نرم افزارهای مربوطه، مشتریان بر اساس ویژگی هایشان طبقه بندی گردیدند. سپس طبقه بندی با الگوریتم های دیگر نیز صورت گرفت و با میزان دقت طبقه بندی الگوریتم درخت تصمیم فازی مقایسه صورت گرفته است. در انتها قوانین و الگوهای موجود در داده های کل مشتریان را بر اساس طبقات تعریف

- شده پیدا نموده و به عنوان چارچوبی جهت پیش بینی اعتبار متقاضیان جدید ارائه می‌گردد تا قرار گرفتن متقاضیان را در طبقات تعریف شده، پیش بینی شوند.
- مراحل اجرایی این تحقیق به صورت زیر قابل خلاصه شدن می‌باشد:
- ۱- جمع آوری داده از پایگاه داده‌های موجود (پرونده‌های اعتباری مشتریان سابق بانک مورد نظر و سیستم‌های عملیاتی کامپیوتری بانک)؛
 - ۲- شناسایی عوامل تاثیر گذار در رفتار اعتباری مشتریان که در پایگاه داده‌های مورد بررسی، موجود می‌باشند؛
 - ۳- تعیین شاخص‌هایی برای تعریف طبقات مشتریان خوب (خوش حساب)، متوسط و مشتریان بد (بد حساب)؛
 - ۴- فازی نمودن متغیرها (متغیرهای مستقل که قابل فازی شدن هستند و نیز طبقات مشتریان)
 - ۵- تقسیم داده‌های نمونه به دو مجموعه داده‌های آموزشی و داده‌های تست؛
 - ۶- ساخت قوانین با استفاده از داده‌های آموزشی با تکنیک درخت تصمیم فازی (ایجاد درخت تصمیم فازی)؛
 - ۷- آزمون مدل با مجموعه داده‌های تست؛
 - ۸- ارائه الگوی کشف شده از طبقه بندی مشتریان؛
 - ۹- سنجش اعتبار مدل و مقایسه با مدل‌های طبقه بندی.

چارچوب نظری تحقیق

داده‌های تحقیق

روش گردآوری داده‌ها هم به صورت "مشاهده" و هم از طریق "مصاحبه" بوده است. بدین صورت که برای دستیابی به اطلاعات مشتریان بانک از انواع پایگاه داده‌های موجود مانند پرونده‌ها و سیستم‌های کامپیوتری از روش مشاهده و ثبت اطلاعات مورد نیاز در چک لیست استفاده شده است. همچنین برای شناسایی متغیرهای تاثیر گذار در رفتار اعتباری مشتریان و متغیرهای کلامی از مصاحبه آزاد با خبرگان بانک مورد مطالعه استفاده شده است.

داده‌های موجود، در سه پایگاه ذخیره شده بود : ۱- سیستم عملیاتی بانکی (سیستم پردازش تراکنش) ۲- پرونده‌های تسهیلات اعتباری موجود در بایگانی ۳- فایل‌های شخصی کارشناسان.

از آنجا که پایگاه داده مربوط به پرونده‌های اعتباری، مبنای جستجو در دو پایگاه داده دیگر بود بنابراین در گام اول این پایگاه مورد پاکسازی اولیه قرار گرفته سپس داده‌های باقی مانده در آن مبنای جستجو و جمع‌آوری اطلاعات از دو پایگاه دیگر قرار گرفتند و در نهایت همه داده‌ها در یک پایگاه ادغام گردیدند تا در مراحل بعدی پیش پردازش شوند

جامعه آماری تحقیق متشکل از، ۴۱۸ مورد از مشتریان حقوقی بانک سامان می‌باشند که در دوره زمانی سال مالی ۱۳۸۵ و ۱۳۸۶ از بانک، تسهیلات اعتباری دریافت نموده‌اند.

با توجه به اینکه فقط این تعداد داده در دسترس محقق قرار گرفت و دسترسی به کل داده‌های مشتریان حقوقی بانک جهت نمونه‌گیری، امکان پذیر نبود و بنابر اظهارات مدیران بخش اعتباری بانک مورد مطالعه، بقیه داده‌ها بصورت پراکنده و غیر منسجم در اختیار شعبات سراسر کشور می‌باشد؛ بنابراین نمونه‌گیری در خصوص این جامعه آماری صورت نگرفته است و بعد از پالایش داده این تعداد، سرانجام ۱۲۰ مورد داده جهت ساخت مدل نهایی مورد استفاده قرار گرفته است.

این داده‌ها در یک پایگاه داده در نرم افزار Excel ذخیره شده‌اند. خصیصه‌های مشتریان که همان متغیرهای مستقل تحقیق و طبقه مشتری که همان متغیر وابسته تحقیق می‌باشند، فیلدهای این پایگاه داده را تشکیل می‌دهند. عناوین این فیلدها شامل ۱۹ خصیصه مشتریان شامل سن شرکت، نوع شرکت، میزان سرمایه، زمینه فعالیت، تحصیلات مدیر، سن مدیر عامل، نوع قرارداد وام، نرخ بهره وام، مبلغ وام، مدت وام، نوع مصرف وام، نوع باز پرداخت، نسبت سود آوری، تعداد کارکنان، سابقه همکاری، قدرت وثایق، نسبت جاری، نسبت بدهی، دوره وصول مطالبات و ۱ طبقه کلاس می‌باشند که متغیرهای اسمی با شقوق مربوطه به صورت جدول ۱ می‌باشد.

جدول ۱. شقوق متغیرهای اسمی

متغیرهای اسمی	شقوق مربوطه
زمینه فعالیت	بازرگانی، خدماتی، تولیدی
نوع شرکت	تعاونی، سهامی عام، سهامی خاص، با مسولیت محدود
نوع قرارداد وام	مضاربه، مشارکت مدنی، مضاربه عام، جعاله، سلف، خرید دین، فروش اقساطی
سطح تحصیلات مدیر	پایین (دیپلم، پایین تر)، متوسط (فوق دیپلم و لیسانس)، بالا (فوق لیسانس و بالاتر)
نوع مصرف وام	بازرگانی داخلی، تولید - صنعت، واردات
نوع باز پرداخت وام	یکجا، قسطی

در ادامه فرایند آماده سازی، برای پاکسازی و پیش پردازش داده ها، دو عملیات مهم کاهش داده و اعمال تغییرات در شکل داده ها بر روی پایگاه داده رابطه ای صورت گرفت.

پاکسازی داده ها در سه بخش اصلی ۱- اصلاح اشتباهات کاربر، ۲- یک شکل نمودن داده ها و ۳- فازی نمودن متغیرها، صورت گرفته است. تشخیص اشتباهات کاربران در ورود اطلاعات با مشاهده داده های غیر قابل قبول و تطبیق با بقیه داده ها و مقایسه داده های مشترک در پایگاه داده های مختلف صورت گرفته است. برای یکپارچه نمودن داده های داخل برخی فیلدها با یکدیگر و همچنین قابلیت پردازش دادن به بعضی داده ها در مدل لازم بود تا داده ها یک شکل گردند. از جمله فیلدهایی که در این مرحله تغییرات بر روی آنها صورت گرفته عبارت بودند از فیلد "مدت وام" که بصورت ماه و سال بود و همگی به روز تبدیل شدند؛ یا فیلد سنوات فعالیت شرکت بصورت سال تاسیس بود که بصورت تعداد سنوات فعالیت تغییر یافت.

از آنجا که یکی از پایه های اصلی این تحقیق بهره گیری از منطق فازی می باشد، بنابراین در این تحقیق کلیه متغیرهای وابسته طبقات و برخی متغیرهای مستقل که قابلیت فازی شدن داشتند، به صورت متغیرهای فازی تبدیل گردیده اند.

فازی نمودن متغیرها در سه مرحله اصلی ۱- جدا نمودن متغیرها، ۲- تعیین اصطلاحات کلامی و طیف متغیرها در مقادیر کلامی و ۳- تعیین تابع عضویت صورت گرفته است.

در مرحله جدا نمودن متغیرها، متغیرهای غیر اسمی که قابل فازی کردن هستند از کل متغیرها جداسازی شدند سپس در مرحله بعد به تعیین اصطلاحات کلامی و طیف متغیرها در مقادیر کلامی با عنایت به مطالعاتی که در گزارشات کارشناسان بانک موجود در پرونده‌های بایگانی شده و نیز با اطلاع یافتن از کار کارشناسان، به این جمع بندی رسیده شد که مقادیر کلامی بکارگرفته شده برای طبقات، بصورت سه وضعیت شامل خوب، متوسط و بد و برای خصیصه‌ها، سه وضعیت شامل بالا، متوسط و پایین قابل تفسیر می‌باشد. در گام بعدی از کارشناسان خواسته شد تا طیف‌هایی از طبقات مشتریان و خصیصه‌ها را که با این سه متغیر کلامی در تصمیم‌گیری‌هایشان لحاظ می‌نمایند، بیان کنند. که در جدول‌های ۲ و ۳ قابل مشاهده می‌باشند.

جدول ۲. طیف متغیرها در مقادیر کلامی

	پایین	متوسط	بالا
سابقه فعالیت شرکت	۱-۵	۶-۱۴	۱۵-۵۴
میزان سرمایه ثبتی	۱-۱۰/۰۰۰	۱۰۰/۰۰۰-۱/۰۰۰/۰۰۰	۱۰۰/۰۰۰-۱/۰۰۰/۰۰۰
سن مدیر عامل	۲۰-۳۵	۳۶-۴۵	۴۶-۸۲
نرخ بهره وام	۴-۱۴	۱۵-۱۹	۲۰-۲۸
مبلغ مصوب وام	۰/۱-۱/۰۰۰	۱۰۰۰-۱۰/۰۰۰	۱۰/۰۰۰-۲۰۰/۲۰۰
مدت وام	۱۸-۱۷۹	۱۸۰-۳۶۵	۳۶۶-۱۴۶۰
قوت وثایق	۰/۵-۲/۹	۳-۳/۹	۴-۵
سابقه همکاری	۰-۱	۲-۴	۵-۷
نسبت جاری	۰-۱/۲	۱/۲-۱/۴	۱/۴-۵/۷
نسبت بدهی	۰-۰/۶۵	۰/۶۵-۰/۸	۰/۸-۱/۳۳
دوره وصول مطالبات	۰-۹۰	۹۰-۱۵۰	۱۵۰-۱۴۶۳
نسبت سودآوری	۰-۰/۱	۰/۱-۰/۲	۰/۲-۰/۶۴
تعداد کارکنان	۴-۴۹	۵۰-۱۹۹	۲۰۰-۱۲۶۹

جدول ۳. طیف طبقات در مقادیر کلامی

طبقه مشتری بد	طبقه مشتری متوسط	طبقه مشتری خوب
۱۸۰-۷۰۰	۶۰-۱۸۰	۰-۶۰
تأخیر در بازپرداخت اقساط وام (روز)		

تابع عضویتی که در این تحقیق برای تعیین درجات عضویت بکار گرفته شده است، "دوزنقه‌ای" می‌باشد. علت این انتخاب این بود که اولاً نوع برخورد کارشناسان برای طیف‌های مقادیر کلامی به اینصورت بود که فاصله مشخصی از یک متغیر که مربوط به یک مقدار کلامی است هیچگونه تفاوتی وجود ندارد. بنابراین تابع عضویت دوزنقه‌ای در این خصوص بیشتر صدق می‌کند. از طرفی سابقه مطالعات قبلی در مورد مباحث مالی و اعتبارسنجی، نشان دهنده بیشترین استفاده از این تابع عضویت در تحلیل‌های مشابه با تحقیق حاضر می‌باشد.

بر همین اساس با کمک خبرگان بانک، مقادیر متغیر که متناسب با درجات عضویت صفر و یک هستند را برای نقاط به ترتیب پایین‌ترین و بالاترین درجات عضویت تعیین نموده و بقیه مقادیر دیگر، بین صفر و یک در نظر گرفته شده است. با توجه به میزان "تأخیر در بازپرداخت اقساط وام" به عنوان مبنا در هر رکورد برای هر یک از مشتریان، یک درجه تعلق به کلاس‌های هر یک (خوب، متوسط یا بد) در نظر گرفته می‌شود. و بدین ترتیب طبقات نیز حالت قطعی نداشته و بنابر این فازی تعیین می‌گردند.

درخت تصمیم فازی

الگوریتم‌های متفاوتی برای ایجاد درخت تصمیم فازی ارائه شده‌اند. ایجاد درخت تصمیم فازی، در اکثر روش‌های ارائه شده برای آن، از چهار مرحله اصلی زیر تشکیل می‌شود:

الف: فازی‌سازی داده‌های آموزشی

ب: ایجاد یک درخت تصمیم فازی

ج: استخراج قواعد از درخت تصمیم

د: به کارگیری قواعد فازی در دسته‌بندی که هریک از مراحل فوق در روشهای مختلف، از نظر کلی مشابهت‌هایی با هم دارند.

در [۲۹]، روش ID3 برای مدل‌سازی فازی به کار برده شده است. در این تحقیق، ID3 به شکل کلاسیک آن به کار برده شده و از همان روش تعیین بهترین متغیر ورودی و محدوده فواصل خروجی استفاده شده است با این تفاوت که فواصل خروجی، نواحی فازی هستند. در [۲۳] روشی ارائه شده است که بسیار به روش ID3 کلاسیک مشابهت دارد، تنها تفاوت موجود در آن علاوه بر به کارگیری اقلام فازی، معیار انتخاب مشخصه به کار گرفته شده در آن است.

در [۱۰] روشی توسط برد ارائه شده است که در آن از الگوریتم ID3 برای دسته‌بندی استفاده می‌شود. مقادیر ورودی گسسته یا عددی هستند و فازی سازی همانند مطالعات [۲۸] و [۲۳] تنها بر روی مشخصه‌های عددی انجام می‌شود. معیار انتخاب مشخصه نیز در آن مبتنی بر آنروپی است. تنها تفاوت عمده آن در روش فازی‌سازی آن است. فازی‌سازی آن به گونه‌ای است که مجموعه نمونه‌های حاصل، دارای حجم مشخصی هستند. در [۳۰] از روشی به نام Fuzzy-ID3 استفاده شده است. راه حل بدین ترتیب است که پیش از آنکه استنتاج بروی داده‌ها انجام شود، فازی‌سازی بر روی آنها صورت می‌پذیرد. فازی‌سازی تنها در مواردی که ورودی یا خروجی شامل مقادیر عددی یا پیوسته است انجام شده و برای داده‌هایی که حالت گسسته دارند، فازی‌سازی انجام نمی‌شود.

پس از اتمام مرحله مطالعاتی، روش Fuzzy ID3 برای مدل‌سازی انتخاب گردید. دو دلیل اصلی برای انتخاب این روش می‌توان ذکر کرد: نخست اینکه این روش بسط طبیعی روش ID3 به قلمروی تئوری فازی بوده و به لحاظ شهودی به خوبی قابل درک می‌باشد. دیگر اینکه این روش بیش از روش‌های موجود دیگر مورد اقبال واقع و به کار گرفته شده است. الگوریتم درخت تصمیم فازی مورد استفاده، همان الگوریتم Fuzzy ID3 می‌باشد که در مقاله جانیکوف [۲۲] به آن اشاره شده

است که از داده‌های فازی، درخت ایجاد می‌کند. برای اعمال تکنیک درخت تصمیم فازی در تحقیق حاضر از نرم افزار FDT، استفاده شده است [۳]. نحوه عملکرد بدین شکل است که برای انجام عملیات مدلسازی در این نرم افزار نیاز به نوشتن دو نوع فایل اطلاعاتی دارد، فایل داده و فایل دستوری که این کار با توجه به پایگاه داده نهایی و نوع عملکرد نرم افزار صورت پذیرفت.

یافته‌های تحقیق

ارزیابی اعتبار مدل

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 به این صورت می‌باشد که داده‌ها به دو مجموعه داده‌های آموزشی و داده‌های آزمایشی تقسیم می‌گردند. هدف از اینکار اینست که با تعداد داده‌های آموزشی، الگوریتم انتخابی، دانشی را حاصل می‌کند. ولی اینکه نتایج حاصله تا چه میزان دارای اعتبار هستند باید توسط نتایج داده‌های جدید و قدرت پیش بینی الگوریتم در مورد داده‌هایی که تا کنون با آن مواجه نبوده، آزمون شوند. از این جهت داده‌های تست به عنوان داده‌های ناظر به الگوریتم، داده می‌شوند و نتایج حاصله می‌توانند میزان صحت مدل را ارزیابی نمایند.

صحت طبقه بندی یا تفکیک داده‌های تست در طبقه‌ها، معیار ارزیابی اعتبار و صحت مدل می‌باشد. که در این تحقیق از "اعتبارسنجی متقابل با ۱۰ بار تکرار" استفاده شده است. این روش اعتبارسنجی مجموعه داده‌ها را به ۱۰ قسمت تقسیم نموده و هر بار ۹۰ درصد از داده‌ها را به عنوان مجموعه داده آموزشی و ۱۰ درصد را به عنوان مجموعه داده تست انتخاب نموده و میزان دقت طبقه بندی را می‌سنجد. این فرایند ۱۰ بار صورت می‌گیرد و در نتیجه از کلیه درجات دقت میانگین گرفته شده و به عنوان دقت نهایی مدل ارائه می‌گردد. با بهره‌گیری از این روش دیگر نگرانی از

تصادفی بودن انتخاب دو مجموعه داده آموزشی و تست وجود نخواهد داشت. درخت حاصل شده از اعمال الگوریتم، درخت بسیار بزرگی است که نیاز به هرس کرده دارد. بنابراین از روش "پیش هرس درخت" استفاده شده، بدین صورت که قبل از اعمال الگوریتم، حداقل تعداد نمونه هایی که کاربر می خواهد تا درون هر برگ جای گیرد را تعیین می نماید. میزان دقت درخت به لحاظ طبقه بندی صحیح مشتریان با توجه به میزان هرس های تست شده بصورت جدول ۴ ذیل می باشد.

جدول ۴. میزان دقت درخت تصمیم فازی

حداقل تعداد نمونه در برگهای نهایی	۲	۳	۴	۵	۶	۷
دقت	٪۵۸/۲	٪۵۵/۵	٪۵۵/۸	٪۵۸	٪۵۵/۳	٪۵۹/۵

در درخت ایجاد شده به تعداد مسیرهایی که از ریشه (بالاترین گره) تا یک برگ دسترسی وجود داشته باشد، می توان قاعده استخراج نمود.

کارایی مدل درخت تصمیم فازی در مقایسه با سایر مدلها

برای سنجش کارایی مدل ارائه شده توسط درخت تصمیم فازی، صحت نتایج این مدل با الگوریتم های طبقه بندی دیگری مورد مقایسه قرار گرفته اند. این الگوریتم ها عبارتند از: تعدادی از درخت های تصمیم سنتی، شبکه های عصبی، ماشین بردار پشتیبان، شبکه های بیزین، رگرسیون لجستیک و درخت ژنتیکی. ظاهراً به نظر می رسد که دقت پیش بینی طبقات در الگوریتم ارائه شده و الگوریتم های دیگر طبقه بندی که مورد بررسی قرار گرفتند، پیش بینی متوسطی است و از وضعیت بسیار مطلوبی برخوردار نیستند؛ بنابراین در این بخش نشان داده می شود که یکی از مهمترین عوامل تاثیرگذار به جز داده های نویزی، تعداد بسیار کم نمونه مورد بررسی در مقابل تعداد زیاد خصیصه ها می باشد. از این رو در این قسمت تمامی الگوریتم ها را با نصف تعداد نمونه یعنی ۶۰ نمونه مورد بررسی قرار

گرفته و با صحت الگوریتم‌های ۱۲۰ تایی به صورت ماتریس‌های متقابل^۱ در جدول ۵ مورد مقایسه قرار گرفته است. بنابراین می‌توان نتیجه گرفت که تعداد داده‌های بیشتر، پتانسیل بیشتری برای ارائه یک مدل با دقت بالاتر دارد. بنابراین اگر نمونه از ۱۲۰ عدد افزایش چشمگیری داشته باشد می‌تواند دقت مدل بهبود یابد. از طرفی نتایج الگوریتم‌ها برای نمونه ۶۰ تایی، نشان دهنده پایایی نتایج تحقیق می‌باشد.

جدول ۵. مقایسه نتایج صحت الگوریتم‌ها در نمونه‌های ۱۲۰ تایی و ۶۰ تایی

الگوریتم	Fuzzy ID3	C4.5 Tree	REP Tree	BF Tree	Random Tree	FT
صحت نمونه ۶۰	٪۴۸	۳۳/۳۳	۴۱/۶۷	٪۴۰	۳۱/۶۷	۴۳/۳۳
صحت نمونه ۱۲۰	٪۵۸/۲	٪۴۷/۵	٪۵۳/۳۳	٪۵۰/۸۳	٪۴۵	٪۴۵/۸۳
الگوریتم	SVM	RBF (NN)	Logistic	Bayes Net	MLP(NN)	GA Tree
صحت نمونه ۶۰	٪۵۳/۳۳	٪۴۱/۶۷	٪۴۰	٪۴۶/۶۷	٪۴۳/۳۳	٪۵۱
صحت نمونه ۱۲۰	٪۶۵	٪۴۸/۳۳	٪۴۶/۶۷	٪۵۱/۶۷	۴۹/۱۷	٪۶۴/۱۷

درخت حاصل شده از مدل‌سازی، به عنوان دانش استخراج شده از تحقیق تلقی می‌گردد. که در مرحله قبل نتایج حاصل از صحت پیش بینی طبقات مشتریان با چند مورد از الگوریتم‌های دسته بندی دیگر مورد مقایسه قرار گرفتند. حال با توجه به مقایسات صورت گرفته، تکنیکی که از بقیه پیش بینی دقیق تری را انجام می‌دهد برای استفاده بهره وران در بانک مورد مطالعه پیشنهاد می‌گردد.

کارایی مدل در مقایسه با عملکرد کارشناسان اعتبارسنج بانک

کارشناسان بانک با توجه به کلیه اطلاعات اخذ شده از مشتریان، در انتهای هر

1- Confusion Matrix

پرونده نظر کارشناسی خود را در خصوص نوع مشتری یادداشت نموده اند تا مسولان بخش اعتبارات در خصوص اعطا یا عدم اعطای تسهیلات تصمیم نهایی را اخذ کنند. در جدول ۶ می توان میزان صحت پیش بینی های کارشناسان را مشاهده نموده و با عملکرد مدل مورد استفاده در تحقق به لحاظ کارایی تفکیک مشتریان مورد مقایسه قرار داد.

جدول ۶. عملکرد کارشناسان اعتبارسنج بانک در خصوص پیش بینی طبقه مشتریان

کارشناسان اعتبار سنجی بانک		
نمونه های طبقه بندی شده بطور صحیح	۴۶	۳۸/۳٪
نمونه های طبقه بندی شده به طور غلط	۷۴	۶۱/۶٪
کل نمونه ها	۱۲۰	۱۰۰٪

نتیجه گیری

تحقیق پیش رو یک تحقیق کاربردی است که با یک مطالعه موردی در بانک سامان مورد بررسی قرار گرفته است و الگوگیری از فرایند انجام تحقیق و نتایج آن می تواند برای کلیه بانک ها و موسسات مالی و اعتباری دیگر هم مفید واقع شود. در این تحقیق از داده کاوی و منطق فازی برای طبقه بندی مشتریان بانک بهره گرفته شده است.

نتایج مستقیم حاصل از اجرای تحقیق را می توان به صورت زیر خلاصه نمود:

- ۱- درخت تصمیم فازی میزان سهم نمونه ها در هر طبقه را در هر برگ مشخص نموده است و با این نوع دانش هر مشتری هم می تواند به هر سه طبقه خوب، بد و متوسط (با درجات متفاوت) تعلق داشته باشد.
- ۲- درخت تصمیم دقت بالاتری نسبت به الگوریتم های درخت تصمیم سنتی مورد استفاده در تحقیق دارد.
- ۳- الگوریتم درخت تصمیم فازی در مقایسه با درخت ژنتیکی و ماشین بردار پشتیبان کارایی پایین تری به لحاظ دقت تفکیک مشتریان در سه طبقه مورد نظر دارد.
- ۴- استفاده از داده کاوی و منطق فازی در قالب مدل درخت تصمیم فازی، نتایجی

به مرتب بهتر از عملکرد کارشناسان اعتبارسنجی بانک مورد مطالعه را حاصل نمود. که این نشان دهنده کارایی بالای مدل مورد استفاده تحقیق در مقایسه با کارشناسان اعتبارسنجی بانک می باشد که بر اساس تجربه و به روش قضاوتی به پیش بینی در خصوص نوع مشتریان می پردازند. از این رو نتایج تحقیق می تواند بصورت یک سیستم سایه ای به موازات کار کارشناسان اعتبارسنجی، کاربرد داشته باشد. بدین شکل که نتایج طبقه بندی و ارزیابی سیستم داده کاوی، بر ارزیابی کارشناسان اعتبارسنجی تاثیر گذار باشد.

از مزایای این تحقیق اعمال مفاهیم کلامی مورد استفاده کارشناسان برای متغیرهای موثر در رفتار اعتباری و طبقات مشتریان با استفاده از منطق فازی و همچنین کاهش خطاهای انسانی و جلوگیری از فرایند قضاوتی کارشناسان خبره بوسیله بکارگیری الگوهای کشف شده این تحقیق در یک سیستم هوشمند و پاسخگویی سریع تر به متقاضیان تسهیلات اعتباری می باشد. از جمله محدودیتهای این تحقیق، عدم وجود اطلاعات منسجم در یک سیستم اطلاعاتی کامپیوتری و عدم وجود وحدت رویه در ثبت اطلاعات مشتریان و ثبت داده ها در یک سری زمانی مشخص، وجود اطلاعات ناقص و غلط در داده های مورد بررسی که داده ها را با نویز بسیاری مواجه ساخته می باشد.

تشکر و سپاسگذاری

از مدیریت محترم بانک سامان و همکاریهای ارزشمند مدیران و کارکنان معاونت اعتبارات این بانک که این پژوهش را مورد حمایت مالی و معنوی قرار دادند، کمال تشکر را داریم.

منابع و مأخذ

۱. روئین تن، پونه، (۱۳۸۴) پایان نامه "بررسی عوامل موثر بر ریسک اعتباری مشتریان حقوقی بانک (بررسی موردی بانک کشاورزی)"، کارشناسی ارشد، دانشگاه شهید بهشتی.
۲. سبزواری، حسن، (۱۳۸۴) "برآورد و مقایسه مدل امتیازدهی اعتباری لاجیت و روش AHP (مطالعه موردی: مشتریان حقوقی بانک پارسیان)"، پایان نامه کارشناسی ارشد، دانشگاه صنعتی شریف.
۳. گودرزی، محمد رضا، (۱۳۸۳) "استفاده از منطق فازی در یادگیری درخت تصمیم فازی و هرس کردن آن در محیط‌های مغشوش"، پایان نامه کارشناسی ارشد، دانشگاه علم و صنعت.
۴. مروج، مصطفی، (۱۳۸۳) "افزودن قابلیت داده‌کاوی فازی به بانک‌های اطلاعاتی رابطه‌ای"، پایان نامه کارشناسی ارشد، دانشگاه صنعتی امیر کبیر.
۵. مشکانی، علی؛ ناظمی، عبدالرضا، (۱۳۸۸) "داده‌کاوی کاربردی"، دانشگاه آزاد اسلامی واحد نیشابور.
۶. موسوی، ابتهاج، (۱۳۸۶) "بررسی عوامل موثر بر رتبه بندی اعتباری مشتریان"، پایان نامه کارشناسی ارشد، موسسه عالی بانکداری ایران.
7. Altman E.I., (1968) "Financial ratios discriminate analysis and the prediction of corporate Bankruptcy", The Journal of finance 23.
8. Banasik, J., Crook, J. and Thomas, L., (2003) "Sample selection bias in credit scoring models", Journal of the Operational Research Society, vol.54, pp. 822-832.
9. Beaver W.H., (1967) "Financial ratios as Predictors of Failure", Journal Of Accounting Reserch.
10. Bird R., (2003) "Fuzzy data analysis method for large volume of data", Project report in support of degree of master of engineering, University of Bristol, Department of Engineering Mathematics.
11. Boggess W.B., (1967) "Screen-Test your credit Risks", journal Of Accounting Research 4.
12. Chapman. Pete & et al. (1999) "CRISP-DM 1.0: Step-by-step data mining guide", <http://www.crisp-dm.org/CRISPwP-0800.pdf>.
13. Chen J.H., Ho S.Y., (2002) "Intelligent Multi-Objective Evolutionary Algorithm for Editing Minimum Reference Set", Proceedings of the 6th

- Pacific-Asia Conference on Knowledge Discovery and Data Mining, Communications of the Institute of Information and Computing Machinery, V. 5, No. 2, pp. 4-13.
14. Chiang I.J., and Hsu J.Y., (2002) "Fuzzy Classification Trees for Data Analysis," Fuzzy Sets and Systems, vol.130, no.1, pp.87-99.
 15. Deakin E.B, (1972) "A Discriminate analysis of predictors of business failure", journal Of Accounting Research 10(1).
 16. Deakin E.B, (1989) "Rational Economics Behavior and lobbying on Accounting Issues: Evidence from the Oil and Gas Industry", the Accounting Review 66(1).
 17. Durand D, (1941) "Risk element in consumer installment lending", National bureau of economic research, New york, vol.study#8.
 18. Edward F.R , Mishkin F.S.,(1995) "the decline of traditional banking: implication for financial stability and regulatory policy", Federal reserve bank of New York policy Review, pp.27-45.
 19. Fisher R.A, (1936) "The Use-of multiple measurement in Taxonomic problem", Annals of Eugenics.
 20. Gehrke J., Ramakrishnan R., Ganti V., RainForest (1998) "A Framework for Fast Decision Tree Construction of Large Datasets", Proc. 24th International Conference on Very Large Data Bases (VLDB).
 21. Kiss, France, (2003) "credit scoring process from a knowledge management prospective", Budapest University of Technology And Economics, pp:96-108.
 22. Janikow C.Z., (1998) "Fuzzy Decision Trees: Issues and Methods," IEEE Trans. on Systems, Man, and Cybernetics, vol.28, no.1, pp.1-14.
 23. Meier A.Savary C., Schindler G., Veryha Y, (2001) "Database Schema with Fuzzy Classification and Classification Query Languag", Proc. Of the International Congress on Computational Intelligence-Mehods and Applications. Bangor U.K.
 24. Mikut R., Jens J., (2004) "Interpretability in Data-based Learning of Fuzzy Systems", Fuzzy Set and Systems, Elsevier.
 25. Moon, C. G. and Stotsky, J. G., (1993) "Testing the Differences Between the Determinants of Moody's and Standard & Poor's Ratings: An Application of Smooth Simulated Maximum Likelihood Estimation", Journal of Applied Econometrics, Vol. 8, No. 1, pp. 51-69.
 26. Morgan, Guaranty, (1994) "Risk Metrics Technical Document" ,2nd Edition,New York: Morgan Guaranty.
 27. Olaru C. and Wehenkel L.,(2003) "A Complete Fuzzy Decision Tree Technique," Fuzzy Sets and Systems, vol.138, no.2, pp.221-254.
 28. Shafer J. ,Agrawal R., Mehta M.,(1996) "SPRINT: A Scalable Parallel Classifier for Data Mining", Proc. 22nd International Conference on

- Very Large Data Bases (VLDB).
29. Sahraoui H., Boukadoum M., Lounis H., (2000) "Using Fuzzy Threshold Values for Predicting Class Libraby Interface Eolution", In Proceedings of the 4th International ECOOP workshop on Quantitative Approaches in Object-Oriented Software Engineering, Nice (France).
 30. Sattler K., Dunemann O.,(2001) "SQL Database Primitives for Decision Tree Classifiers", In Proc. of the 10th ACM CIKM Int. Conf. on Information and Knowledge Management, November 5-10, Atlanta, Georgia, USA.
 31. Weber R., (1992) "Automatic Knowledge Acquisition for Fuzzy Control Applications", Proc. Internat. Symp. on Fuzzy Systems (Iizuka, Japan, 12 15 July 1992), pp. 9-12.
 32. Weber R., (1992) "Fuzzy-ID3: A Class of Methods for Automatic Knowledge Acquisition", Proc. 2nd Internat. Conf. on Fuzzy Logic & Neural Networks, pp.265-268.
 33. William F.Treacy &Mark S.Carey, (1998) "Credit risk rating at large U.S.bank," Federal Reserve Bulletin, Board of Governors of Federal reserve System (U.S).
 34. Yang, Lui, (2001) "New Issues In Credit Scoring Applications", George-August ,University Gottingen, Institute For Wirtschafts informatics, pp.3 .
 35. Zimmerman H.J.,(1992) "Fuzzy set theory and its application",Kluwer Academic Publishers, London.