

مدل رتبه‌بندی اعتباری هیبریدی با استفاده از الگوریتم‌های ژنتیک و سیستم‌های خبره فازی

(مطالعه موردی: مؤسسه مالی و اعتباری قوامین)

محمدتقی تقوی فرد^۱، فریبا سادات حسینی^۲، محمد خان بابایی^۳

چکیده: سیستم‌های خبره می‌توانند به ساخت مدل‌های رتبه‌بندی اعتباری مشتریان بانک‌ها کمک کنند. در اینجا، انتخاب ویژگی‌های مهم در رتبه‌بندی اعتباری اهمیت دارد. همچنین ممکن است مقادیر ویژگی‌ها، به صورت فازی بیان شوند. مسئله این است، چگونه می‌توان به کمک الگوریتم ژنتیک، انتخاب ویژگی‌ها را بهبود بخشید؛ به گونه‌ای که این ویژگی‌ها به منزله ورودی در سیستم خبره فازی مورد استفاده قرار گیرند. این نوشتار به ارائه مدل رتبه‌بندی اعتباری هیبریدی با ترکیب انتخاب ویژگی‌ها، مبتنی بر الگوریتم ژنتیک و سیستم خبره فازی می‌پردازد. پژوهشی که در این مورد انجام گرفت، از نظر نتایج، کاربردی و از نظر هدف، توصیفی از نوع مطالعه موردی است. برای آموزش و آزمون مدل، از مجموعه داده‌های رتبه‌بندی اعتباری مؤسسه مالی و اعتباری قوامین استفاده شده است. پس از پیش‌پردازش داده‌ها، به کمک الگوریتم ژنتیک ویژگی‌ها انتخاب شدند و از طریق مصاحبه با فردی خبره و به کارگیری منطق فازی، دامنه تغییرات ویژگی‌های منتخب تعیین شده و سپس قوانین فازی رتبه‌بندی اعتباری ایجاد شدند. برای تحلیل داده‌ها از ابزار وکا و ماژول سیستم استنتاج فازی در نرم‌افزار متلب استفاده شده است. نتایج نشان می‌دهد، دقت طبقه‌بندی مدل پیشنهادی نسبت به سایر روش‌های مقایسه‌شده در این مقاله بیشتر است. قوانین فازی ایجادشده این مدل را می‌توان برای رتبه‌بندی اعتباری مشتریان بانکی به کار برد.

واژه‌های کلیدی: الگوریتم ژنتیک، رتبه‌بندی اعتباری، سیستم خبره، منطق فازی.

۱. استادیار، گروه مهندسی صنایع، دانشکده مدیریت و حسابداری، دانشگاه علامه طباطبائی، تهران، ایران

۲. کارشناس ارشد مدیریت فناوری اطلاعات، دانشگاه علامه طباطبائی، تهران، ایران

۳. دانشگاه آزاد اسلامی، واحد علوم و تحقیقات، گروه مدیریت فناوری اطلاعات، تهران، ایران

تاریخ دریافت مقاله: ۱۳۹۲/۰۲/۰۴

تاریخ پذیرش نهایی مقاله: ۱۳۹۲/۰۷/۱۰

نویسنده مسئول مقاله: فریبا سادات حسینی

E-mail: hosseini.fariba1@gmail.com

مقدمه

بانک برای شناسایی مشتریان خود در اعطای تسهیلات اعتباری، نیاز اساسی به شناسایی و رتبه‌بندی اعتباری آنها دارد. در حال حاضر مدل‌ها و روش‌های مختلفی برای رتبه‌بندی اعتباری مشتریان بانک وجود دارد. هدف اصلی آنها، طبقه‌بندی مشتریان به طبقه‌های مشتریان خوش حساب و مشتریان بدحساب است؛ بنابراین رتبه‌بندی اعتباری، در دامنه وسیع مسائل طبقه‌بندی قرار می‌گیرد (لی و چن، ۲۰۰۵؛ یانگ، لی و ژو، ۲۰۱۱ و لین، ۲۰۰۹).

به‌طور کلی روش‌های رتبه‌بندی اعتباری را می‌توان به دو گروه عمده روش‌های پارامتریک (مانند پروبیت، لوجیت، تحلیل تمایزی و رگرسیون لجستیک) و روش ناپارامتریک و داده‌کاوی (مانند درختان تصمیم‌گیری، شبکه‌های عصبی و سیستم‌های خبره) تقسیم‌بندی کرد (البرزی، محمدپور زرنندی و خان‌بابایی، ۱۳۸۹). برخی از مدل‌های ترکیبی نیز در اعتبارسنجی مورد استفاده قرار گرفته‌اند، مثل مدل عصبی فازی، مدل فازی ماشین بردار پشتیبان و مدل ترکیبی شبکه‌های عصبی (ژائو و همکاران ۲۰۱۱).

روش‌های آماری قدیمی، مانند تحلیل تمایزی، رگرسیون لجستیک، تحلیل پروبیت و لوجیت، از ابتدای ظهور رتبه‌بندی اعتباری مورد استفاده قرار گرفتند. سپس رویکردها و روش‌های متنوع دیگری چون الگوریتم‌های فازی، الگوریتم‌های ژنتیک، شبکه‌های عصبی و سیستم‌های خبره ظهور کردند (لی و چن، ۲۰۰۵). یکی از مشکلات استفاده از روش‌های قدیمی آماری این است که برای ورودی‌هایی با ابعاد بزرگ مناسب نبوده و در اغلب آنها فرض بر این است که رابطه خطی میان متغیرها وجود دارد؛ در حالی که معمولاً این رابطه غیرخطی است (خالصی و شکوهی، ۱۳۸۹).

در این پژوهش از مدل ناپارامتریک و هیبریدی سیستم خبره فازی و الگوریتم انتخاب ویژگی‌ها مبتنی بر الگوریتم ژنتیک، برای ساخت مدل پیشنهادی رتبه‌بندی اعتباری مشتریان بانکی استفاده می‌شود. سیستم خبره می‌تواند به بانک‌ها در مورد تصمیم‌گیری اعطای تسهیلات اعتباری به مشتریان کمک کند و یک سیستم پشتیبان تصمیم باشد. در اینجا نیز، سیستم خبره رتبه‌بندی اعتباری به کارشناسان و مدیران بانکی کمک می‌کند تا در خصوص اعطای تسهیلات اعتباری به متقاضیان، تصمیم‌گیری کنند. همچنین با استفاده از این سیستم، می‌توان به نگهداری دانش فرد خبره در بانک‌ها کمک کرد.

در مقاله‌ای (البرزی، محمدپور زرنندی و خان‌بابایی، ۱۳۸۹) از روش‌های انتخاب ویژگی‌ها مبتنی بر الگوریتم ژنتیک برای انتخاب ویژگی‌های مهم و ساخت درختان تصمیم‌گیری در رتبه‌بندی اعتباری مشتریان بانک‌ها استفاده شد. مدل پیشنهادی این مقاله، از یک رویکرد هیبریدی الگوریتم‌های انتخاب ویژگی‌ها، خوشه‌بندی، طبقه‌بندی و الگوریتم ژنتیک بهره می‌برد.

مسئله اصلی این مقاله، ساخت یک سیستم خبره است که بتواند به نحو مناسبی، رتبه‌بندی اعتباری مشتریان خوب و بد بانک‌ها را انجام دهد. اما چگونه می‌توان با استفاده از الگوریتم ژنتیک و انتخاب ویژگی‌ها، به ساخت سیستم خبره فازی برای رتبه‌بندی اعتباری مشتریان بانک‌ها و مؤسسه‌های مالی پرداخت؟ نظرات کارشناسان و خبرگان تسهیلات اعتباری بانک‌ها، ممکن است ماهیت دقیقی نداشته باشد. در این خصوص می‌توان برای استنتاج قوانین در یک سیستم خبره رتبه‌بندی اعتباری، از منطق فازی بهره جست. پس مدل پیشنهادی در این مقاله به ساخت سیستم خبره رتبه‌بندی اعتباری فازی می‌پردازد.

معمولاً انبوهی از اطلاعات تکراری و ویژگی‌های اعتباری مشتریان در پایگاه‌های داده انباشته می‌شود که ممکن است، موجب کاهش دقت طبقه‌بندی و افزایش پیچیدگی مدل رتبه‌بندی اعتباری شود. در این خصوص می‌توان با استفاده از روش‌های انتخاب ویژگی، به تعیین ویژگی‌های مهم در رتبه‌بندی اعتباری پرداخت. به کارگیری الگوریتم‌های انتخاب ویژگی در مدل‌های طبقه‌بندی دارای فواید زیر است:

۱. کاهش هزینه استخراج ویژگی‌ها؛
 ۲. بهبود دقت طبقه‌بندی (آبدو، ۲۰۰۹)؛
 ۳. کاهش زمان یادگیری الگوریتم طبقه‌بندی؛
 ۴. کاهش تعداد مثال‌های مورد نیاز برای یادگیری در مدل طبقه‌بندی؛
 ۵. کاهش تعداد ویژگی‌ها برای افزایش سرعت محاسبات (ژائو و همکاران، ۲۰۱۱).
- در طراحی سیستم خبره فازی رتبه‌بندی اعتباری، می‌توان از الگوریتم انتخاب ویژگی‌ها برای کاهش پیچیدگی سیستم و افزایش دقت طبقه‌بندی مشتریان بانکی بهره برد. از سوی دیگر، الگوریتم‌های انتخاب ویژگی فرض می‌کنند روابط بین ویژگی‌ها خطی بوده و از هم مستقل هستند. بنابراین تعامل و وابستگی بین ویژگی‌ها را در محاسبات خود در نظر نمی‌گیرند. همچنین این الگوریتم‌ها در هنگام انجام فرایند جست‌وجو، ممکن است در بهینه محلی قرار گیرند. در این الگوریتم‌ها فقط از برخی معیارها برای انتخاب ویژگی استفاده می‌شود (خان‌بابایی، ۱۳۸۸). الگوریتم‌های ژنتیک، به دلیل قابلیتشان برای رسیدن به راه‌حل‌های دقیق یا تخمینی در نواحی جست‌وجوی بسیار بزرگ دنیای واقعی، در مدت زمان مناسب، از پُرکاربردترین روش‌ها برای انتخاب ویژگی هستند (سای و چن، ۲۰۱۰). همچنین به دلیل آنکه مسئله انتخاب ویژگی از نوع مسائل سخت است (شهرابی و ذوالقدرشجاعی، ۱۳۸۸)، می‌توان از الگوریتم ژنتیک در انتخاب ویژگی‌ها استفاده کرد. الگوریتم ژنتیک هنگام انتخاب ویژگی‌ها، چند متغیر را همزمان مورد

بررسی قرار می‌دهد. چنانچه ویژگی‌ها تک‌به‌تک ارزیابی شوند، تعامل بین آنها در نظر گرفته نمی‌شود (البرزی، محمدپورزندی و خان‌بابایی، ۱۳۸۹).

در این مقاله به ارائه یک مدل پیشنهادی مناسب رتبه‌بندی اعتباری مشتریان بانک‌ها و مؤسسه‌های مالی با استفاده از رویکرد هیبریدی انتخاب ویژگی‌ها مبتنی بر الگوریتم ژنتیک و سیستم خبره فازی، پرداخته می‌شود.

پیشینه پژوهش

پژوهش‌های متنوعی در خصوص به‌کارگیری روش‌های آماری، غیرپارامتریک و هوش مصنوعی برای پشتیبانی از فرایندهای رتبه‌بندی اعتباری بانک‌ها انجام گرفته است. برای نمونه می‌توان به کاربرد شبکه‌های رگرسیون لجستیک، شبکه عصبی، درخت تصمیم‌گیری، تحلیل تمایزی، ماشین بردار پشتیبان و طبقه‌کننده‌های ترکیبی اشاره کرد (البرزی، محمدپورزندی و خان‌بابایی، ۱۳۸۹). همچنین، در برخی پژوهش‌ها از سیستم‌های خبره برای رتبه‌بندی اعتباری مشتریان بانکی استفاده شده است. در مقاله کومارا در سال ۲۰۰۶ (آیدو، ۲۰۰۹)، از یک سیستم خبره برای سنجش اعتبار در مؤسسه‌های مالی و صادر کردن پذیره‌نویسی استفاده شده است. پی. سینها و ژاو در سال ۲۰۰۸ (اسلیوا و همکاران، ۲۰۱۱)، به‌منظور بهبود طبقه‌بندی برای وام‌دادن با روش‌های داده‌کاوی، به ترکیب سیستم خبره با این روش‌ها پرداختند و دانش خبره را با این سیستم دریافت کرده و به‌منزله ورودی روش‌های داده‌کاوی مورد استفاده قرار دادند. در مقاله بن دیوید و فرانک (۲۰۰۹) به مقایسه ماشین یادگیری و سیستم خبره در امر طبقه‌بندی و مسائل رگرسیونی پرداخته شده است.

پژوهش‌های متعددی به کاربرد روش‌های الگوریتم ژنتیک در انتخاب ویژگی‌ها پرداخته‌اند (سایر حوزه‌های علوم و کسب‌وکار) که از نتایج آنها می‌توان در این مقاله برای ساخت مدل رتبه‌بندی اعتباری مشتریان بانکی استفاده کرد. از جمله آنها می‌توان به ترکیب طبقه‌کننده‌ها مبتنی بر الگوریتم ژنتیک، به‌کارگیری الگوریتم‌های ژنتیک برای انتخاب متغیرهای ورودی و در ترکیب روش‌های انتخاب ویژگی (البرزی، محمدپورزندی و خان‌بابایی، ۱۳۸۹) اشاره کرد.

یکی از رویکردها برای بهبود عملکرد طبقه‌بندی، به‌کارگیری مدل‌های هیبریدی به جای استفاده از طبقه‌کننده‌های مجزا است که به‌تازگی در رتبه‌بندی اعتباری مشتریان بانکی مورد توجه پژوهشگران واقع شده است (سای و وو، ۲۰۰۸). در مدل‌های هیبریدی به جای استفاده از یک طبقه‌کننده، از روش‌های مختلف پارامتریک و ناپارامتریک طبقه‌بندی و خوشه‌بندی برای بهبود عملکرد طبقه‌بندی استفاده می‌شود. برخی از این مدل‌ها در رتبه‌بندی اعتباری عبارتند از:

روش ماشین بردار پشتیبان هیبریدی (چن، ما و ما، ۲۰۰۹)، مدل هیبریدی رگرسیون لجستیک و شبکه‌های عصبی (سینها و ژائو، ۲۰۰۸)، مدل هیبریدی روش‌های خوشه‌بندی و طبقه‌بندی (سای و وو، ۲۰۰۸).

برخی از موارد برجسته مدل پیشنهادی این مقاله، در مقایسه با مدل‌های رتبه‌بندی اعتباری دیگر که به آنها اشاره شد، عبارتند از: ۱. استفاده از روش‌های آماده‌سازی و پیش‌پردازش داده‌ها برای ساخت مدل رتبه‌بندی اعتباری پیشنهادی؛ ۲. به‌کارگیری روش انتخاب ویژگی مبتنی بر الگوریتم ژنتیک، در انتخاب ویژگی‌های مهم رتبه‌بندی اعتباری برای طراحی و ساخت سیستم خبره رتبه‌بندی اعتباری؛ ۳. استفاده از منطق فازی و الگوریتم ژنتیک برای انتخاب ویژگی‌های مناسب رتبه‌بندی اعتباری مشتریان بانکی؛ ۴. به‌کارگیری رویکرد هیبریدی (ترکیب الگوریتم‌های انتخاب ویژگی مبتنی بر الگوریتم ژنتیک، منطق فازی و سیستم خبره، در ساخت مدل رتبه‌بندی اعتباری مشتریان بانکی) به جای استفاده از یک طبقه‌کننده؛ ۵. استفاده از منطق فازی در بیان مقادیر ویژگی‌های رتبه‌بندی اعتباری برای ساخت قوانین سنجش اعتبار مشتریان در سیستم خبره مدل پیشنهادی.

در برخی از مطالعات و مقاله‌ها به استفاده از الگوریتم‌های انتخاب ویژگی در ساخت مدل‌های طبقه‌بندی اشاره شده است، اما پس از جست‌وجو و بررسی مطالعات، درخصوص استفاده از روش‌های انتخاب ویژگی مبتنی بر الگوریتم ژنتیک در طراحی و ساخت سیستم خبره فازی رتبه‌بندی اعتباری، مطالعه‌ای دیده نشد.

روش‌شناسی پژوهش

پژوهش پیش رو، از نظر نتایج، کاربردی و از نظر هدف، توصیفی و از نوع مطالعه موردی است. برای جمع‌آوری اطلاعات، از داده‌های موجود در بانک اطلاعاتی مؤسسه مالی و اعتباری قوامین و پرونده‌های اعتباری مشتریان استفاده شد. پایگاه داده موجود، شامل ۱۱۳۳ رکورد و ۲۱ ویژگی است (با احتساب ویژگی هدف، یعنی «معوقات»). مجموعه داده‌ها که برای ساخت مدل رتبه‌بندی اعتباری در این پژوهش به کار می‌رود، مشتریان حقیقی سال‌های ۱۳۸۸ و ۱۳۸۹ شهر تهران هستند که از بانک تقاضای وام کرده‌اند. متغیرهای مستقل این مجموعه، شامل کد حوزه (اسمی)، کد شعبه (اسمی)، شماره درخواست (اسمی)، نوع وام (اسمی)، کد نوع درخواست (اسمی)، جنسیت (اسمی)، شماره شناسنامه (اسمی)، تاریخ تولد (اسمی)، نرخ وام (عددی)، مقدار وام (عددی)، تعداد اقساط (عددی)، تاریخ درخواست وام (اسمی)، تاریخ واریز وام (اسمی)، تاریخ اولین قسط (اسمی)، تاریخ آخرین قسط (اسمی) و مانده بدهی (عددی)، سابقه کار (عددی)، درآمد

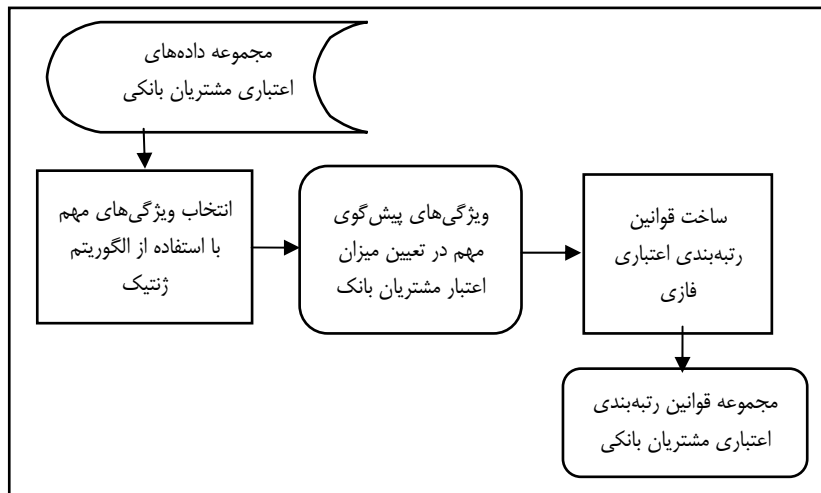
(عددی)، وضعیت مسکن (اسمی)، میانگین موجودی حساب در سه ماهه اخیر (اسمی) و متغیر وابسته آن معوقات (اسمی) است.

اگر تعداد وام‌های معوق یک مشتری سه وام یا بیشتر از آن باشد، فرد جزء مشتریان بدحساب و اگر کمتر از سه باشد، از دسته مشتریان خوش حساب است. برای آماده‌سازی داده‌ها از روش‌های زیر استفاده شد:

- حذف ویژگی‌هایی که دارای مقادیر یکتا هستند؛
- حذف برخی ویژگی‌ها، به دلیل عدم ارائه دانش؛
- حذف تراکنش‌هایی با مقادیر مفقود؛
- بررسی وجود اختلال در مجموعه داده؛
- تغییر قالب مقادیر ویژگی‌هایی که تاریخ دارند.

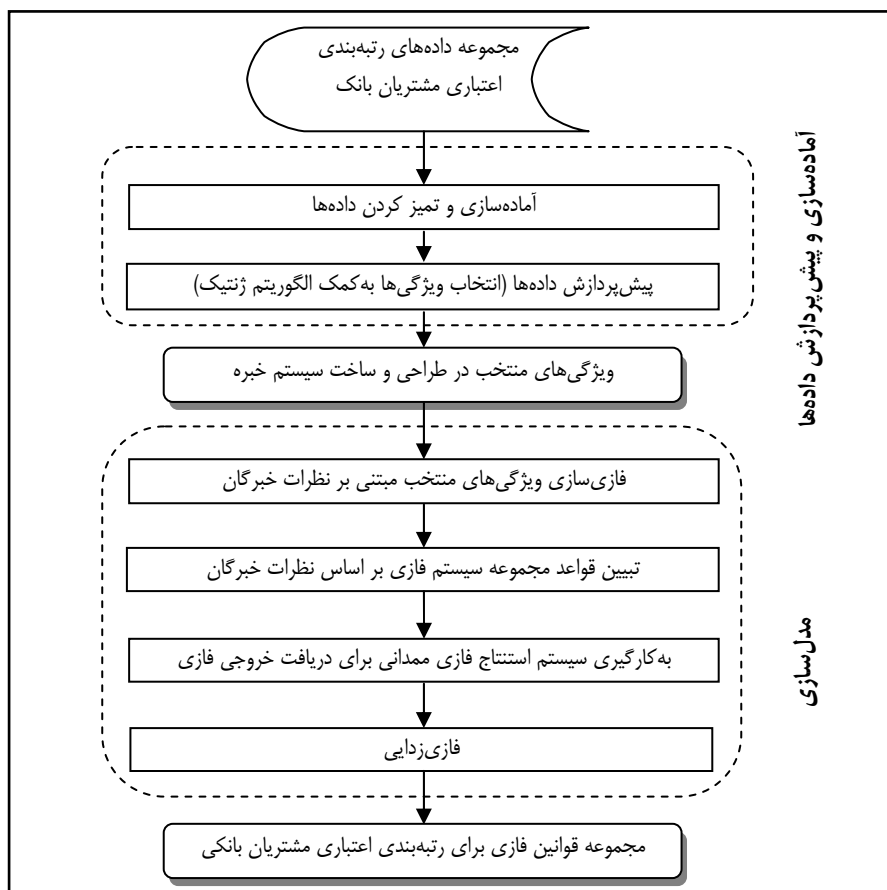
مدل پیشنهادی

ابتدا با استفاده از الگوریتم ژنتیک، ویژگی‌های مهم برای تعیین میزان اعتبار مشتریان بانکی انتخاب شده‌اند و سپس با در نظر گرفتن این مجموعه ویژگی، به ساخت قوانین رتبه‌بندی اعتباری فازی پرداخته شد و در نهایت، مجموعه قوانین رتبه‌بندی اعتباری مشتریان بانکی برای ارزیابی مشتریان به دست آمد. شکل ۱ مراحل کلی ساخت سیستم خبره فازی رتبه‌بندی اعتباری مشتریان بانکی را نشان می‌دهد.



شکل ۱. نمودار مراحل کلی در ساخت سیستم خبره فازی رتبه‌بندی اعتباری مشتریان بانکی

همچنین شکل ۲، به فرایند ساخت و آزمون مدل پیشنهادی سیستم خبره فازی رتبه‌بندی اعتباری مشتریان بانکی می‌پردازد. ابتدا به آماده‌سازی و تمیز دادن مجموعه داده‌های رتبه‌بندی اعتباری مشتریان بانک پرداخته می‌شود. سپس در پیش‌پردازش داده‌ها، انتخاب ویژگی‌ها به کمک الگوریتم ژنتیک صورت می‌گیرد. ماحصل این کار، ویژگی‌های رتبه‌بندی اعتباری منتخب در طراحی و ساخت سیستم خبره است. سپس، مبتنی بر نظرات خبرگان به فازی‌سازی ویژگی‌های منتخب پرداخته می‌شود. در ادامه، قوانین مجموعه سیستم فازی بر اساس نظرات خبرگان تبیین می‌شوند و به کمک سیستم استنتاج فازی ممدانی، خروجی فازی دریافت می‌شود و در پایان، پس از انجام عملیات فازی‌زدایی، مجموعه قوانین فازی برای رتبه‌بندی اعتباری مشتریان بانکی تولید می‌شود.



شکل ۲. فرایند ساخت و آزمون مدل پیشنهادی سیستم خبره فازی رتبه‌بندی اعتباری مشتریان بانکی

در انتخاب ویژگی‌ها به کمک الگوریتم ژنتیک، هر کروموزوم نشان‌دهنده مجموعه ویژگی‌های رتبه‌بندی اعتباری است. از آنجا که مسئله اصلی در انتخاب ویژگی، انتخاب یا عدم انتخاب یک ویژگی است، شیوه کدگذاری در انتخاب ویژگی، کدگذاری صفر و یک است که به ترتیب نشان‌دهنده عدم حضور و حضور هر ویژگی در مجموعه ویژگی‌ها است. روش انتخاب کروموزوم‌ها، چرخ گردان است. عملگر تقاطع از نوع تک‌نقطه‌ای صفر و یک است. عملگر جهش برای عدم قرار گرفتن در جواب‌های بهینه محلی، روش تک‌نقطه‌ای صفر و یک است. همچنین، عملگر جایگزینی کروموزوم‌های قبلی با کروموزوم جدید بر پایه شایستگی است. شرط خاتمه یافتگی الگوریتم، رسیدن به تعداد معینی تکرار است. رویکرد انتخاب ویژگی، رویکرد فیلتر و تابع ارزیاب، دقت طبقه‌بندی با تابع طبقه‌بنده‌کننده C4.5 است. در انتخاب ویژگی‌ها برای تقسیم مجموعه داده به آموزش و آزمون، از روش اعتبارسنجی متقاطع استفاده می‌شود. پس از انتخاب ویژگی‌ها، از طریق مصاحبه باز با کارشناسان و خبرگان، مقادیر فازی و حدود این مقادیر، توابع عضویت ویژگی‌ها و قواعد سیستم خبره فازی تعیین شد. تابع عضویت ویژگی‌ها، فازی‌ساز گوسین است. این فازی‌ساز نسبت به فازی‌سازهای دیگری چون مثلثی و دوزنقه‌ای، از انعطاف‌پذیری بیشتری برخوردار بوده و به مسائل جهان واقعی نزدیکتر است. چون نوع مسئله در این پژوهش به گونه‌ای است که تابع عضویت خروجی نمی‌تواند تابع ریاضی باشد و انعطاف‌پذیری سیستم‌های فازی در سوگنو وجود ندارد، سیستم فازی ممدانی برای تعیین خروجی به کار می‌رود. برای فازی‌زدایی خروجی‌ها، غیرفازی‌ساز مرکز ثقل به کار گرفته شد. در تعیین رتبه اعتباری مشتریان، خروجی سیستم خبره فازی شامل سه مقدار فازی کم، متوسط و زیاد است.

یافته‌های پژوهش

در این بخش به ارائه نتایج و یافته‌های پژوهش پرداخته می‌شود.

مطالعه موردی

مدل پیشنهادی سیستم خبره فازی در مجموعه داده‌های مؤسسه مالی و اعتباری قوامین، مورد ارزیابی و آزمون قرار گرفت. برای تجزیه و تحلیل داده‌های رتبه‌بندی اعتباری و رسیدن به نتایج لازم، از نرم‌افزارهای اکسل، وکا و متلب (ماژول سیستم استنتاج فازی) استفاده شد. در اینجا کارشناسان رتبه‌بندی اعتباری، مبتنی بر قضاوت‌های ذهنی خود، به اعطای تسهیلات اعتباری اقدام می‌کنند. پس از آماده‌سازی و تمیز کردن داده‌ها، تعداد ویژگی‌ها و تراکنش‌ها، به ترتیب برابر با ۱۲ و ۱۰۷۹ به دست آمد. سپس به کمک روش انتخاب ویژگی‌ها مبتنی بر الگوریتم ژنتیک، به

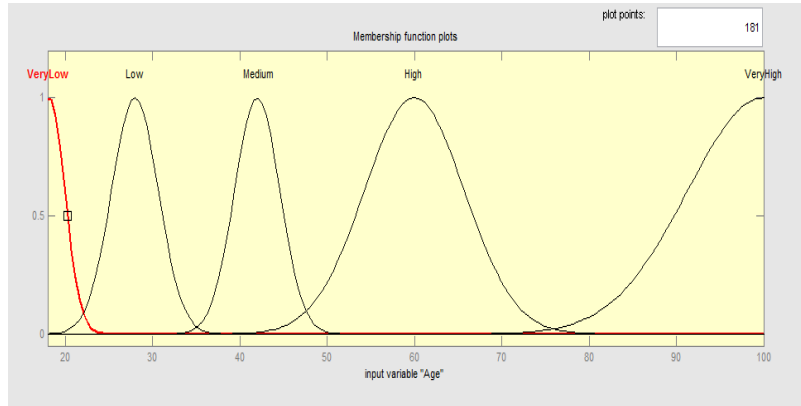
انتخاب نه ویژگی مهم در رتبه‌بندی اعتباری پرداخته شد. مقادیر پارامترهای الگوریتم ژنتیک در انتخاب ویژگی‌ها به قرار زیر است.

تعداد نسل و جمعیت اولیه: ۲۰؛ تابع ارزیاب: دقت طبقه‌کننده درخت تصمیم‌گیری C4.5؛ نرخ تقاطع: ۰/۹؛ نرخ جهش: ۰/۰۱؛ عدد تصادفی: ۱؛ عدد اعتبارسنجی در انتخاب ویژگی‌ها: ۱۰. مقادیر پارامترهای تابع ارزیاب درخت تصمیم‌گیری C4.5 (تابع ارزیابی الگوریتم ژنتیک در مدل پیشنهادی) به صورت پیش فرض در ابزار وکا موجود است. بعد از اجرای الگوریتم انتخاب ویژگی‌ها مبتنی بر الگوریتم ژنتیک، ویژگی‌های سن، سابقه کار، درآمد، میانگین موجودی حساب سه‌ماهه اخیر، میزان وام، نرخ وام، تعداد اقساط، جنسیت و وضعیت مسکن انتخاب شدند. این ویژگی‌ها، نتایج حاصل از اجرای الگوریتم انتخاب ویژگی‌ها مبتنی بر روش جست‌وجوی الگوریتم ژنتیک هستند. به گفته دیگر، در فرایند جست‌وجوی الگوریتم ژنتیک، مقدار نهایی کد این ویژگی‌ها برابر عدد یک شده است که به معنای انتخاب این ویژگی‌ها است. با توجه به اینکه مدل پیشنهادی این پژوهش، به طبقه‌بندی و رتبه‌بندی اعتباری مشتریان بانکی می‌پردازد، تابع ارزیابی در روش جست‌وجوی الگوریتم ژنتیک، تابع ارزیابی مدل طبقه‌بندی درخت تصمیم‌گیری C4.5 است تا بدین ترتیب دقت طبقه‌بندی نهایی مدل پیشنهادی افزایش یابد.

بعد از انتخاب ویژگی‌ها، از طریق مصاحبه با فرد خبره، به تعیین محدوده یا دامنه تغییرات ویژگی‌های منتخب برای طراحی سیستم خبره رتبه‌بندی اعتباری، پرداخته شد. سپس برای هر یک از متغیرهای ورودی به سیستم خبره با توجه به نظرات خبره، مقادیر زبانی تعریف شد. همچنین مقادیر پارامترهای مربوط به توابع عضویت آنها نیز مشخص شد. تابع فازی‌سازی ویژگی‌ها، تابع فازی گوسین است. در اینجا برای نمونه، مقادیر زبانی و مقادیر پارامترهای تابع فازی گوسین ویژگی سن در جدول ۱ آورده شده است. در شکل ۳ نیز، نمودار تابع عضویت پنج‌تایی گوسین برای ویژگی سن مشخص است.

جدول ۱. مقادیر زبانی و مقادیر پارامترهای تابع فازی گوسین ویژگی سن

تعریف	مقادیر زبانی	پارامترهای مسئله (σ, c)
Very low	VL	[۱/۷۶۵ , ۱۸/۲]
Low	L	[۲/۶۹ , ۲۸]
Medium	M	[۲/۶۰۴ , ۴۲]
High	H	[۵/۷۳۲ , ۶۰]
Very high	VH	[۸/۷۰۸ , ۱۰۰]

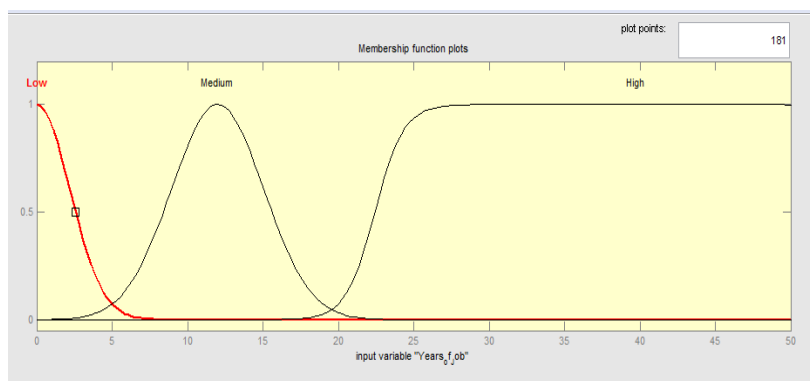


شکل ۳. نمودار تابع عضویت پنج تایی گوسین برای ویژگی سن

همان طور که در جدول ۱ و شکل ۳ مشاهده می شود، تعداد پنج مقدار زبانی برای ویژگی «سن» در نظر گرفته شده است. برای ویژگی «سابقه کار»، همان گونه که در جدول ۲ و شکل ۴ مشخص است، تابع عضویت برای مجموعه فازی High از نوع سیگموئیدی است. در این ویژگی، تعداد سال های ۲۵ به بالا همگی با تابع عضویت یک، جزء سابقه کاری بالا هستند.

جدول ۲. مقادیر زبانی و مقادیر پارامترهای تابع فازی گوسین ویژگی سابقه کار

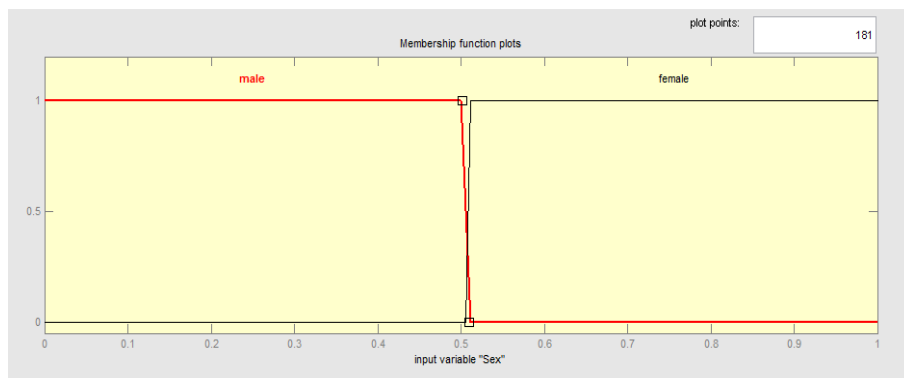
تعریف	مقادیر زبانی	پارامترهای مسئله (σ, c)
Low	L	[۲/۱۹۱, ۰]
Medium	M	[۳/۰۵۸, ۱۲]
High	H	[۱/۰۴, ۲۲/۴۲, ۱/۱۷, ۵۵/۲۳]



شکل ۴. نمودار تابع عضویت سه تایی سیگموئیدی برای ویژگی سابقه کار

مدل رتبه‌بندی اعتباری هیبریدی با استفاده از الگوریتم‌های ژنتیک و... ۴۱

برای ویژگی «جنسیت» هم تابع عضویت به‌صورت نمودار شکل ۵ است که حالت قطعی دارد. ویژگی «وضعیت مسکن» نیز همانند ویژگی جنسیت حالت قطعی دارد.



شکل ۵. نمودار تابع عضویت دوتایی crisp برای ویژگی جنسیت

همان‌طور که پیش از این بیان شد، نه ویژگی رتبه‌بندی اعتباری منتخب به‌منزله متغیرهای ورودی سیستم خبره در نظر گرفته شدند. هر کدام از این نه ویژگی می‌توانند چندین مقدار زبانی بگیرند که به این ترتیب و با این روش، تعداد ۱۰۱۲۵۰ قانون مختلف به‌دست می‌آید. تعداد حالت‌های ممکن قوانین به‌صورت زیر محاسبه شده است:

$$۱۰۱۲۵۰ = ۳ \times ۳ \times ۵ \times ۵ \times ۳ \times ۵ \times ۳ \times ۵ \times ۲$$

اما در عمل ممکن نیست که این همه قانون تولید و استفاده شود. بنابراین با نظر خبره حالت‌های بیشینه و حد وسط قوانین انتخاب شدند. قوانین منتخب به نحو مطلوبی حالت‌های مختلف را پوشش داده و برای رتبه‌بندی اعتباری مشتریان مناسب هستند. به این ترتیب تعداد ۲۵۲ قانون به‌دست آمد. برای نمونه یکی از قوانین در شکل ۶ نمایش داده شده است.

اگر (جنسیت مرد باشد) و (سن متوسط باشد) و (سابقه کار زیاد باشد) و (درآمد خیلی زیاد باشد) و (وضعیت مسکن فرد از نوع مالک) باشد و (میانگین موجودی حساب خیلی کم باشد) و (میزان وام خیلی کم باشد) و (نرخ وام کم باشد) و (تعداد اقساط زیاد) باشد، آنگاه (درجه ریسک اعتباری متقاضی پایین است).

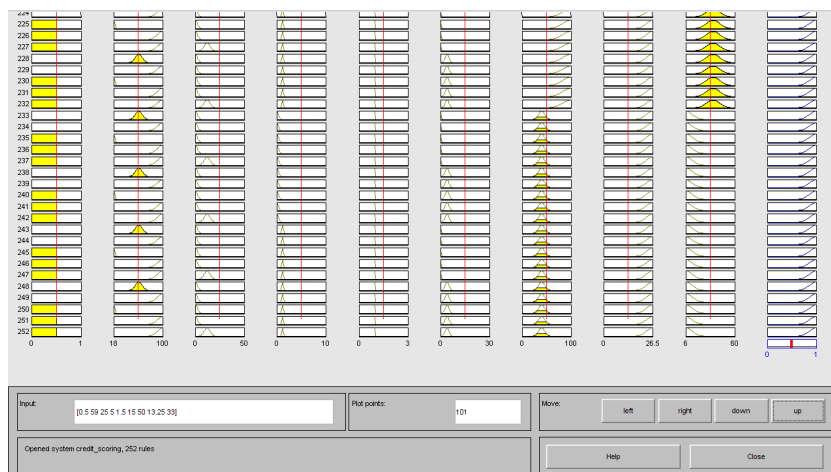
شکل ۶. نمونه قانون به‌دست‌آمده در رتبه‌بندی اعتباری مشتریان به‌کمک سیستم خبره فازی پیشنهادی

هر قانون فازی از چند مقدم تشکیل شده که در آنها مقدار عضویت به مجموعه‌های فازی مشخص است. می‌بایست یک عملگر فازی روی تمام آنها اعمال شود تا یک عدد که نمایانگر

نتیجه کلی تمام مقدم‌ها (ورودی‌ها) است، به‌دست آید. در شکل ۶ مشاهده می‌شود، عملگر فازی بین متغیرهای ورودی عملگر «و» است. در این عملگر مینیمم مقادیر تابع فازی بین همه مقدم‌ها انتخاب می‌شود. پس از آنکه متغیرهای ورودی به اعداد فازی تبدیل شدند و قوانین فازی تبیین شد، نوبت به استنتاج فازی می‌رسد. مقادیر زبانی که در مرحله قبل به متغیرهای ورودی اختصاص داده شده‌اند، بعضی از قوانین را فعال کرده و متغیر خروجی محاسبه می‌شود. سیستم استنتاج فازی در دو مرحله این کار را انجام می‌دهد:

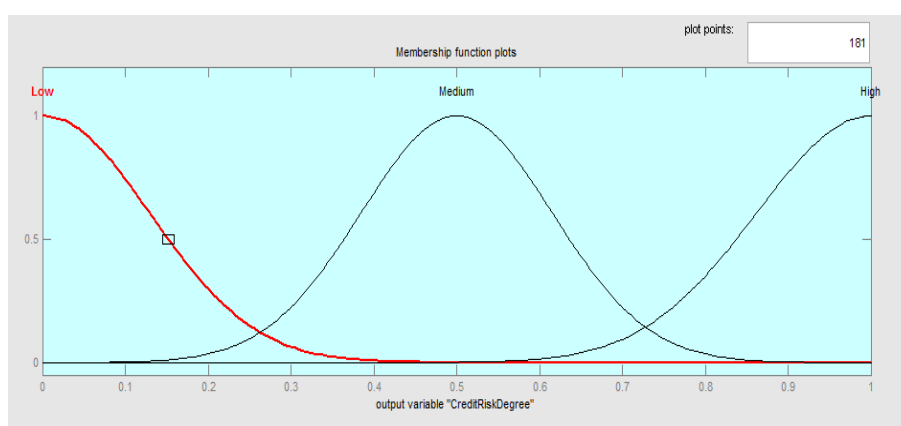
۱. اعمال استلزام: خروجی هر قانون، یک مجموعه فازی است که با یک تابع عضویت ارائه می‌شود. با توجه به عدد به‌دست‌آمده از عملگر «و» روی مقدم‌ها (ورودی‌ها)، این مجموعه فازی تغییر پیدا کرده و به‌منزله خروجی، ارائه می‌شود. استلزام به‌کار گرفته شده در این سیستم، استلزام ممدانی است.

۲. جمع کردن تمامی خروجی‌ها: چون تصمیم‌گیری باید براساس آزمایش تمام قوانین FIS به‌صورت موازی انجام گیرد، بایستی نتیجه تمام قوانین به‌گونه‌ای با هم جمع شوند و مجموعه فازی خروجی واحدی ایجاد کنند. در سیستم ممدانی، در این قسمت اجتماع نتایج قوانین به‌منزله خروجی است. عملگر بیشینه‌کردن در این سیستم انتخاب می‌شود. نمودار شکل ۷، موتور استنتاج سیستم خبره فازی رتبه‌بندی اعتباری را نشان می‌دهد. پس از این مرحله، به فازی‌زدایی خروجی پرداخته می‌شود. در اینجا از غیرفازی‌ساز مرکز ثقل استفاده می‌شود. عدد به‌دست‌آمده از غیرفازی‌ساز، تعیین‌کننده درجه ریسک اعتباری مشتری است.



شکل ۷. موتور استنتاج سیستم خبره فازی رتبه‌بندی اعتباری

با توجه به شکل ۸، توابع عضویت برای متغیر خروجی، تابع عضویت گوسین است. اگر عدد به دست آمده که خروجی غیرفازی‌ساز است، در مجموعه فازی Low، High و Medium قرار گیرد، درجه ریسک مشتری به ترتیب بالا (مشتریان بدحساب)، پایین (خوش حساب) و متوسط (نیاز به تصمیم‌گیری با احتیاط بیشتر، اخذ ضمانت بیشتر از مشتری، در صورت ارائه تسهیلات به وی) است.



شکل ۸. تابع عضویت سه‌تایی گوسین برای متغیر خروجی (درجه ریسک اعتباری مشتری)

در سیستم طراحی شده، می‌توان با وارد کردن مقدار متغیرهای (ویژگی‌های) ورودی برای یک مشتری جدید، سطح ریسک اعتباری وی را پیش‌بینی کرد. برای آزمودن سیستم خبره فازی رتبه‌بندی اعتباری، اطلاعات مربوط به ویژگی‌های اعتباری صد نفر از مشتریان که در پایگاه داده موجود بود، ورودی‌های سیستم در نظر گرفته شدند و در سیستم لحاظ شدند. نتیجه نشان داد، سیستم با دقت ۸۸/۴ درصد، مشتریان را به درستی طبقه‌بندی می‌کند. در ادامه به مقایسه دقت طبقه‌بندی سیستم خبره فازی رتبه‌بندی اعتباری پیشنهادی با مدل‌های طبقه‌بندی درختان تصمیم‌گیری در داده‌های پرداخته می‌شود (جدول ۳). همان‌طور که در جدول ۳ مشاهده می‌شود، دقت طبقه‌بندی مشتریان در سیستم خبره فازی پیشنهادی، نسبت به مدل‌های طبقه‌بندی درختان تصمیم‌گیری بیشتر است. پس می‌توان از سیستم خبره فازی پیشنهادی برای رتبه‌بندی اعتباری مشتریان بانک‌ها و مؤسسه‌های مالی و اعتباری، همچون قوامین استفاده کرد.

جدول ۳. مقایسه دقت طبقه‌بندی سیستم خبره فازی پیشنهادی با الگوریتم‌های درختان تصمیم‌گیری

الگوریتم رتبه‌بندی اعتباری	سیستم خبره فازی پیشنهادی	درخت تصمیم‌گیری C5	درخت تصمیم‌گیری CART	درخت تصمیم‌گیری CHAID	درخت تصمیم‌گیری QUEST
دقت طبقه‌بندی مشتریان	٪۸۸/۴۰	٪۸۶/۳۶	٪۸۱/۵۴	٪۸۵/۳۴	٪۸۰/۴۳

نتیجه‌گیری و پیشنهادها

بانک‌ها در اعطای تسهیلات اعتباری به مشتریان خود، نیازمند رتبه‌بندی اعتباری آنها هستند. سیستم‌های خبره فازی می‌توانند به طبقه‌بندی و رتبه‌بندی مشتریان بپردازند. مسئله اصلی طراحی و ساخت یک سیستم خبره فازی است که بتواند به شکل مطلوبی مشتریان را طبقه‌بندی کند. هدف ارائه یک مدل مناسب رتبه‌بندی اعتباری مشتریان بانک‌ها (در اینجا مؤسسه مالی و اعتباری قوامین) برای اعطای تسهیلات اعتباری متناسب با هر درجه از ریسک اعتباری است. این مدل بر پایه رویکرد هیبریدی روش انتخاب ویژگی، مبتنی بر الگوریتم ژنتیک و سیستم خبره فازی است. از الگوریتم ژنتیک برای انتخاب ویژگی‌های رتبه‌بندی اعتباری استفاده شد. ویژگی‌های منتخب، ورودی‌های سیستم خبره فازی در نظر گرفته شدند و در ساخت قوانین رتبه‌بندی اعتباری مورد استفاده قرار گرفتند. دقت طبقه‌بندی مشتریان در سیستم خبره فازی پیشنهادی بالاتر از دقت طبقه‌بندی مدل‌های درختان تصمیم‌گیری مقایسه‌شده در این مقاله است. می‌توان از مدل هیبریدی سیستم خبره فازی پیشنهادی در این پژوهش، برای ساخت و استخراج قوانین رتبه‌بندی اعتباری مشتریان بانک‌ها استفاده کرد.

با توجه به پیشینه پژوهش و مدل ارائه‌شده، برخی پیشنهادها عبارتند از:

- انتخاب ویژگی با الگوریتم‌های ابتکاری دیگر، مثل پخت شبیه‌سازی شده، بهینه‌سازی کلونی مورچگان، بهینه‌سازی ذرات پراکنده؛
- توسعه مدل پیشنهادی با به‌کارگیری روش‌های طبقه‌بندی داده‌کاوی، همچون شبکه‌های عصبی و ماشین بردار پشتیبان، در ساخت مدل طبقه‌بندی پیشنهادی و رتبه‌بندی اعتباری مشتریان بانکی در کنار استفاده از سیستم خبره فازی.

همچنین پیشنهادهای کاربردی برای بانک‌ها به شرح زیر هستند:

- به‌کارگیری روش‌های هوش مصنوعی، همچون الگوریتم ژنتیک و منطق فازی در طراحی و ساخت مدل‌های رتبه‌بندی اعتباری در کنار استفاده از روش‌های قضاوتی در تعیین اعتبار مشتریان؛

۲. توسعه مدل پیشنهادی از طریق ساخت نرم‌افزار کاربردی سیستم خبره رتبه‌بندی اعتباری؛

۳. طراحی و ساخت یک سیستم پشتیبانی تصمیم‌گیری برای رتبه‌بندی اعتباری مشتریان بانک. در پایگاه مدل این سیستم، می‌توان از مدل پیشنهادی در این مقاله برای ساخت و تولید قوانین رتبه‌بندی اعتباری بهره جست.

منابع

البرزی، م.؛ محمد پورزندی م. ا.؛ خان‌بابایی، م. (۱۳۸۹). به‌کارگیری الگوریتم ژنتیک در بهینه‌سازی درختان تصمیم‌گیری برای اعتبارسنجی مشتریان بانک‌ها. *مجله مدیریت فناوری اطلاعات*، ۲(۴): ۳۸-۲۳.

خالصی، ن.؛ شکوهی، ا. ح. (۱۳۸۹). ارائه روشی جدید برای اعتبارسنجی مشتریان بانکی با استفاده از تکنیک‌های داده‌کاوی. *مجموعه مقالات چهارمین کنفرانس داده‌کاوی ایران*. تهران: دانشگاه شریف، (نهم و دهم آذر ماه).

خان‌بابایی، م. (۱۳۸۸). به‌کارگیری الگوریتم ژنتیک در بهینه‌سازی درختان تصمیم‌گیری برای اعتبارسنجی مشتریان بانک‌ها. *پایان‌نامه کارشناسی ارشد*. دانشگاه آزاد اسلامی واحد علوم و تحقیقات تهران.

شهرابی، ج.؛ ذوالقدرشجاعی، ع. (۱۳۸۸). *داده‌کاوی پیشرفته - مفاهیم و الگوریتم‌ها*. تهران: جهاد دانشگاهی، واحد صنعتی امیرکبیر.

Abdou, H.A. (2009). Genetic programming for credit scoring: The case of Egyptian public sector banks. *Expert systems with applications*, 36 (9): 11402-11417.

Ben-David, A. & Frank, E. (2009). Accuracy of machine learning models versus "hand crafted" expert systems – A credit scoring case study. *Expert Systems with Applications*, 30 (3): 507-518.

Chen, W., Ma, C. and Ma, L. (2009). Mining the customer credit using hybrid support vector machine technique, *Expert Systems with Applications*, 36 (4): 7611-7616.

Kumra, R., Stein, R.M. & Assersohn, I. (2006). Assessing a knowledge-based approach to commercial loan underwriting. *Expert Systems with Applications*, 30 (3): 507-518.

Lee, T.S. & Chen, I.F. (2005). A two-stage hybrid credit scoring model using artificial neural networks and multivariate adaptive regression splines. *Expert systems with applications*, 28 (4): 743-752.

- Lin, S. L. (2009). A new two-stage hybrid approach of credit risk in banking industry, *Expert Systems with Applications*, 36 (4): 8333.
- P.Sinha, A. & Zhao, H. (2008). Incorporating domain knowledge into data mining classifiers: An application in indirect lending. *Expert Systems with Applications*, 46(1): 287-299.
- Silva, S.F. da., Ribeiro, M.X., Batista Neto, J.D.E.S., Traina-Jr, C. & Traina, A.J.M. (2011). Improving the ranking quality of medical image retrieval using a genetic feature selection method. *Decision support systems*, 51 (4):810-820.
- Tsai, C.F. and Chen, M. L. (2010). Credit Rating by Hybrid Machine Learning Techniques, *Applied Soft Computing*, 10 (2): 374-380.
- Tsai, C.F., & Wu, J.W. (2008). Using neural network ensembles for bankruptcy prediction and credit scoring. *Expert Systems with Applications*, 34 (4): 2639-2649.
- Yang, W., Li, D., & Zhu, L. (2011). An improved genetic algorithm for optimal feature subset selection from multi-character feature set. *Expert systems with applications*, 38 (3): 2733-2740.
- Yoshida, H., Leardi, R., Funatsu, K. & Varmuza, K. (2001). Feature selection by genetic algorithms for mass spectral classifiers. *Analytica Chimica Acta*, 446 (1-2): 483-493.
- Zhao, M., Fu, Ch., Ji, L., Tang, K. & Zhou, M. (2011). Feature selection and parameter optimization for support vector machines: A new approach based on genetic algorithm with feature chromosomes. *Expert systems with applications*, 38 (5): 5197-5204.