دوره ۷ ، شماره ۲

مجله فناوری اطلاعات در طراحی مهندسی
Information Technology on Engineering Design
http://ited.sinaweb.net

طراحی یک سیستم چندعاملی امتیازدهی اعتباری با بهره گیری از یادگیری جمعی

احمد قدس الهی* ، دکتر حمیدرضا ناجی، اشکان امیرمدحی

دانشجوی دکتری مدیریت بازرگانی، دانشگاه تهران – دکتری کامپیوتر، دانشگاه شهید بهشتی – کارشناس ارشد مدیریت صنعتی، دانشگاه علوم و تحقیقات تهران

چکیده:

امتیازدهی اعتباری، اصلی ترین تکنیک تحلیلی در ارزیابی ریسک اعتباری متقاضیان برای موسسات مالی می باشد. در این پژوهش که یک پژوهش کاربردی- توصیفی است، سعی شده است برای امتیاز دهی اعتباری یک مدل ترکیبی، مبتنی بر یادگیری جمعی طراحی شود به گونه ایی که یک مدل خوشه بندی و یک مدل دسته بندی را با یکدیگر ترکیب می کند. در مدل ترکیبی پژوهش برای بهبود عملکرد دسته بندی از رویکرد یادگیری جمعی بهره گرفته شده است که بر این اساس، از ده ماشین بردار پشتیبان در قالب سیستم های چند عاملی نرم افزاری استفاده شده است. با ترکیب نتایج عامل ها، خروجی نهایی مدل، حاصل خواهد شد. آموزش و تست مدل از طریق مجموعه داده جهانی از داده های تقاضای اعتبار در بانکی از آلمان انجام شده است. با توجه به معیارهای اعتبارسنجی مدل که برای مقایسه نتایج پیش بینی آن در مورد وضعیت اعتباری مشتریان متقاضی وام و مدل های مشابه بکار گرفته شده است، نتایج نشان داد که سیستم چندعاملی امتیازدهی اعتباری با ۸۱/۴۲ درصد صحت و محدوده زیر نمودار ۷۷/۳۷، می تواند بالاترین صحت و برترین عملکرد را نسبت به مدل های مشابه، به ارمغان آورد.

واژه‌های کلیدی:

ریسک اعتباری، سیستم چند عاملی، یادگیری جمعی، ماشین بردار پشتیبان

*عهده دار مکاتبات

دانشگاه تهران، گروه علوم اجتماعی و رفتاری، تهران، ایران

A.Ghodselahi@ut.ac.ir

۱- مقدمه

بانک ها با ریسک های مختلفی مواجهه هستند، اما مهمترین ریسک در حوزه مالی و بانکی، ریسک اعتباری است که به دلیل متعهد نبودن وام گیرنده به قرارداد بازپرداخت خود می باشد. وجود این ریسک ناشی از این احتمال است که وام گیرنده قادر به پرداخت منظم اصل و فرع بدهی خود نباشد و یا در پرداخت آن کوتاهی کند[۱]. وام دهی بدون ارزیابی و پیش بینی موجب افزایش مشتریان بدحساب و وام های معوق شده است. بحران مالی ۲۰۰۷-۲۰۰۸ این امر را مشهود ساخته است که درک ریسک از طریق بازار های مالیامری مهم برای بانک ها در توسعه قابلیت ها و توانایی هایشان به منظور افزایش سرمایه های می باشد[۲] . ریسک اعتباری موجب مشکلات مالی در بانک ها خواهد شد و ارزیابی آن نیازمند تکنیک های مدلسازی پیشرفته می باشد که مرتبط به منبع عدم قطعیت ناشی شده می باشد. بر این اساس مدلسازی ریسک اعتباری موسسات مالی و بخصوص بانک ها، امری مهم و ضروری است[۳]. یکی از تکنیک های مدیریت ریسک اعتباری که از دیر باز تا کنون برای مدیریت ریسک اعتباری بکار گرفته شده است، تکنیک امتیازدهی اعتباری می باشد. در عمل اصلی ترین روش ارزیابی ریسک اعتباری فرد وام گیرنده، ارزیابی امتیاز اعتباری او می باشد[۴]. مبتنی بر یک پایگاه داده قدیمی، وام گیرندگان به افرادی که وام خود را در موعد مقرر پرداخته اند و یا به پرداخت آن در تاریخ مقرر متعهد نبوده اند، تقسیم بندی می شوند[۵]. تکنیک امتیازدهی اعتباری به برخی از موسسات مالی همچون بانک ها و شرکت های صادر کننده کارت های اعتباری این امکان و توانایی را می دهد که بتوانند در مورد دادن اعتبار به فرد متقاضی، بر اساس یک سری از معیار های از پیش تعیین شده، تصمیم بگیرند، که آیا اعتبار درخواست شده داده شود یا خیر. این تکنیک با استفاده از داده های قدیمی و تفکیک آماری، سعی دارد که اثرات خصوصیات مختلف متقاضی را بروی عدم بازپرداخت وام تفکیک کند. اساسا می توان بیان کرد که تصمیم در مورد اعطای اعتبار درخواست شده به متقاضی می تواند برای بانک بسیار مهم باشد و کسب و کار و سودآوری بانک را بطور چشمگیری تحت تاثیر قرار دهد. اگر وام گیرنده در پرداخت خود دچار مشکل شود، بانک ضرر می کند، از سوی دیگر اگر بانک به افرادی که دارای اعتبار خوبی هستند، اعتبار ندهد، مجددا دچار ضرر و زیان خواهد شد[۶]. مزایای امتیاز دهی اعتباری شامل کاهش هزینه تحلیلات اعتباری، تسریع در امر تصمیم گیری مربوط به اعتبارات و الویت بندی اعتباری می باشد[۷]. روش های آماری قدیمی، که برای ایجاد مدل

های امتیازدهی اعتباری بکار گرفته شده اند، عبارتند از: پروبیت، لجیت، رگرسیون خطی، تحلیل متمایز کننده خطی، تحلیل متمایز کننده نمایی، توبیت، درخت های صفر و یکی،و رگرسیون لجستیک. با وجود استفاده از این روش ها برای امتیازدهی اعتباری، می توان بیان کرد که قابلیت آن ها در تمایز مشتری خوب از بد هم چنان دارای مشکل است. دو روش رایج که در امتیازدهی اعتباری استفاده شده است، عبارتند از :تحلیل متمایز کننده و رگرسیون لجستیک . به دلیل رابطه خطی، هر دو روش دارای عدم صحت کافی می باشند[۸]. در دوره های اخیر روش های جدیدی نیز برای توسعه سیستم های امتیازدهی اعتباری بکار گرفته شده است[۱۶]. به دلیل اینکه بسیاری از محققان در حوزه امتیازدهی اعتباری بروی داده های جهانی کار می کنند، نتایج حاصل از مدل های پیش بینی کننده قابل قیاس با سایر مدل ها می باشد و نتایج حاصل از این روش ها در افزایش صحت و دقت مدل دسته بندی کننده نسبت به مدل های خطی بکار رفته منعکس شده است. توسعه مدل امتیازدهی اعتباری و طراحی سیستم با هدف افزایش صحت و بهبود عملکرد سیستم در امتیازدهی اعتباری متقاضیان انجام می گیرداین امر در حالیست که افزایش صحت پیش بینی حتی به مقدار کم منجر به نتایج مالی و صرفه جویی های مالی چشمگیر برای بانک می شود. یک مدل صحیح و کارا، موسسات مالی را قادر می سازد تا منابع خود را به مشتریانی اختصاص دهند که اقساط خود را بصورت منظم بازپرداخت می کنند و از سوی دیگر از اعطای وام به متقاضیان بد جلوگیری می نماید. در بخش بعدی ادبیات پژوهش مورد بررسی قرار خواهد گرفت و تحقیقاتی که در این زمینه انجام گرفته است مرور خواهد شد. سپس در بخش روش شناسی پژوهش نحوه انجام پژوهش تبیین خواهد شد. در بخش چهارم اجزای مدل پیشنهادی پژوهش برای امتیازدهی اعتباری تشریح خواهد شد و با استفاده از این مدل در بخش پنجم نتایج بکارگیری آن ارائه می شود. در نهایت نتیجه گیری کلی از پژوهش حاضر در بخش ششم ارائه می شود.

۲. مرور ادبیات پژوهش

همانطور که ذکر شد، در پاسخ به رشد صنعت اعتباری، مدل های امتیازدهی اعتباری به صورت فعال در حال توسعه و ایجاد است و تحقیقات وسیعی در حوزه یادگیری ماشینی به منظور ارزیابی ریسک اعتباری انجام شده است و نتایج خوب و موثری حاصل شده است که همگی نمایانگر قابلیت و توانایی این ابزارها در ارزیابی ریسک اعتباری و بطور خاص امتیازدهی اعتباری می باشد. بیزن و همکاران[۲۲] مقایسه ایی را بین چندین تکنیک

دسته بندی بروی هشت مجموعه داده واقعی انجام دادند. آن ها از ماشین بردار پشتیبان و کمترین مربعات ماشین بردار پشتیبان به عنوان مدل های دسته بندی استفاده کردند. نتایج تحقیق آن ها نشان داد که ماشین بردار پشتیبان از نظر عملکردی برترین مدل دسته بندی می باشد. شین و همکاران[۱۷] از ماشین بردار پشتیبان برای پیش بینی ورشکستگی استفاده کردند و نتایج نشان داد که این تکنیک صحت بالاتری را در مقایسه با روش های مشابه ایجاد می کند. سیه[۱۸] از یک رویکرد ترکیبی کاوش در طراحی مدل امتیازدهی اعتباری استفاده کرده است. به دلیل وجود نمونه های غیر نمایانگر از یک رویکرد دو مرحله ایی استفاده شده است که از نقشه های خود سازمانده برای تعیین خوشه ها استفاده شده و از الگوریتم K-Means برای یافتن پاسخ نهایی. هوانگ و همکارانش[۷] با بیان اینکه توسعه های اخیر در تکنیک های داده کاوی مثل شبکه های عصبی، برنامه نویسی ژنتیک و ماشین بردار پشتیبان می توانند عملیات دسته بندی را بدون محدودیت های روش های قدیمی انجام دهند، یک مدل امتیازدهی اعتباری مبتنی بر ماشین بردار پشتیبان ارائه می کنند. به منظور ساخت مدل مناسبی جهت امتیازدهی اعتباری، در این تحقیق از سه استراتژی مبتنی بر ماشین بردار پشتیبان استفاده شده است. این سه استراتژی بکار گرفته شده، با عملکرد شبکه های عصبی، برنامه نویسی ژنتیک و درخت تصمیم مقایسه شده است. ژانگ و همکاران[۲۴] در مقاله خود یک مدل ترکیبی امتیازدهی اعتباری ایجاد کرده اند که مشکلات امتیازدهی اعتباری را از طریق بکارگیری مزایای برنامه ریزی ژنتیک و ماشین بردار پشتیبان برطرف می کند. برای تست مدل از دو مجموعه داده واقعی استفاده شده است. نتایج نشان داد که عملکرد مدل ترکیبی از ماشین بردار پشتیبان، برنامه ریزی ژنتیک، درخت تصمیم، رگرسیون لجستیک و شبکه های عصبی پس انتشار برتر می باشد و صحت بیشتری در دسته بندی دارد. نانی و لومینی[۱۵] اثرات استفاده از یادگیری جمعی را نسبت به مدل های دسته بندی منفرد می سنجند. نتایج بررسی های آن ها نشان داد که استفاده از رویکرد جمعی موجب افزایش صحت مدل دسته بندی می شود. ونگ و همکاران[۲۰] مقایسه ایی بین روش های جمعی امتیازدهی اعتباری انجام داده اند. در این تحقیق، عملکرد سه روش رایج یادگیری مورد بررسی قرار گرفت. این سه روش عبارتند از بگینگ[۱]، بوستینگ[۲] و استکینگ[۳]. نتایج حاکی از افزایش صحت و بهبود عملکرد هر یک از مدل های دسته بندی از طریق بکارگیری یادگیری جمعی بود. در این تحقیق بگینگ و استکینگ

---

[۱] Bagging
[۲] Boosting
[۳] Stacking

درخت تصمیم بهترین نتایج را ایجاد کرد. با توجه به تحقیقات انجام گرفته در حوزه امتیازدهی اعتباری و همچنین روند توسعه مدل های امتیازدهی اعتباری از مدل های خطی به سمت مدل های غیر خطی، نحوه بکارگیری و ترکیب مدل های دسته بندی کننده می تواند در بهبود عملکرد پیش بینی و دسته بندی تاثیرگذار باشد. بنابراین با وجود توسعه مدل هایی برای امتیازدهی اعتباری، ایجاد مدل هایی کارا که از قابلیت پیش بینی با صحت بیشتر و عملکرد بالاتر نسبت به مدل های قبلی برخوردارند، می تواند منجر به نتایج چشمگیر مالی برای کاربران آن ها شود که این امر اهمیت توسعه مدل های صحیح تر و کاراتر را به وضوح تایید می کند.

۳. روش شناسی پژوهش

این پژوهش از نقطه نظر هدف، کاربردی و نوع آن، توصیفی می باشد. با توجه به متغیر مستقل مدل امتیازدهی و متغیر وابسته صحت، در این پژوهش فرض می شود که سیستم چندعاملی امتیازدهی اعتباری که مشخصات آن در این پژوهش تشریح شد، متقاضیان اعتبار را با صحت بالاتری دسته بندی می کند. بنابراین هدف پژوهش بروی افزایش و بهبود صحت و عملکرد مدل امتیازدهی متمرکز می باشد. از مجموعه داده های بانکی در آلمان[۴] برای آموزش و تست مدل پژوهش، استفاده شده است. این پایگاه داده مشتمل بر هزار نمونه می باشد که هر یک از نمونه ها توسط بیست خصیصه توصیف شده اند. خصیصه های داده ها در جدول یک ارائه شده است. از داده های این بانک در تحقیقات بسیاری استفاده شده است[۸، ۱۱، ۱۲].

۴. سیستم چند عاملی امتیازدهی اعتباری

توسعه یک مدل امتیازدهی مناسب و کارا، فعالیت مهمی برای کاهش هزینه ها و انجام تصمیم گیری کارا و موثر می باشد. با افزایش مستمر مقیاس و پیچیدگی موسسات مالی و بانک ها و همراه با سرعت تقاضای نقل و انتقالات، نیاز است که تکنیک های کارآمدتری برای مدیریت ریسک بکار گرفته شود و میزان در معرض ریسک اعتباری قرار گرفتن، به صورت سریع و در زمان کم پایش شود. همانطور که ذکر شد، با حجم قابل ملاحظه ایی از وام ها، حتی بهبود کوچکی هم در صحت امتیازدهی اعتباری می تواند ریسک اعتباری موسسه را کاهش دهد و منجر به پس انداز های

[۴] این داد ها به صورت فایل الکترونیک در وب سایت *www.ics.**uci**.edu* به صورت عمومی قابل دریافت می باشد

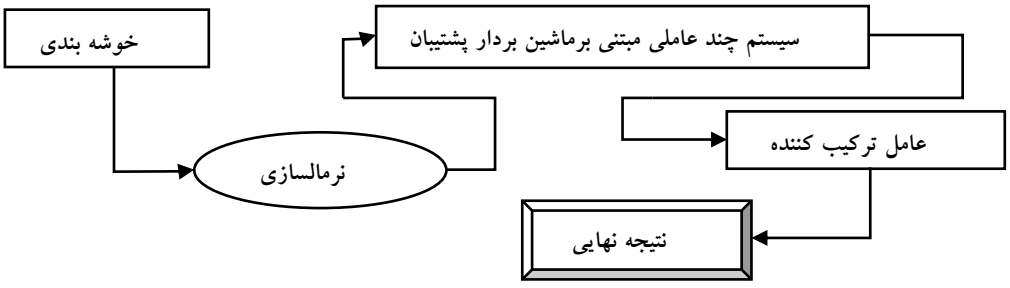
چشم گیر آتی برای بانک شود. در تحقیقات گذشته، در حوزه های دسته بندی و بازشناسی الگوها، افزایش صحت مدل دسته بندی، از طریق بکارگیری مدل های ترکیبی به اثبات رسیده است[۱۹، ۲۴، ۲۳]. در مدل ترکیبی پژوهش ، از دو تکنیک خوشه بندی و دسته بندی استفاده شده است. در اولین گام، خوشه بندی فازی داده های پژوهش انجام می گیرد. علت بکارگیری خوشه بندی فازی فراهم نمودن داده های آموزشی مختلف بر اساس میزان خصوصیات مشترک آن ها و برطرف نمودن عدم قطعیت موجود در داده های پژوهش می باشد. خوشه بندی یک تکنیک دسته بندی غیر نظارتی است که از ویژگی های نمونه ها برای دسته بندی آن ها به گروه هایی با خصوصیات مشترک استفاده می کند. خوشه بندی داده های پژوهش کمک خواهد کرد که خصوصیات مشابه نمونه ها را پیدا شود و نمونه ها به گروه های همگن و مشابه گروه بندی شود و در نتیجه پیش بینی ها و دسته بندی های ناصحیح مدل دسته بندی کننده را که به دلیل نمونه های آموزشی نامناسب رخ می دهد، کم کند. گاهی اوقات حتی با یک مدل دسته بندی بسیار کارا، قابلیت مدل برای پیش بینی یک نمونه جدید محدود می شود. این محدودیت به دلیل الگوهای نامناسب دسته بندی است که در داده های آموزشی بروز می کند و عدم قطعیت موجود در این داده ها، پیچیدگی یادگیری را افزایش می دهد. اگر بتوان کیفیت نمونه ها را بهبود بخشید، توانایی مدل نیز افزایش می یابد. از خوشه بندی به عنوان فرایند پیش پردازش داده ها استفاده شده است تا کیفیت داده های ورودی ماشین بردار پشتیبان بهبود داده شود و در نتیجه از بکارگیری مدل جمعی ماشین بردار پشتیبان، عملکرد بالاتری حاصل شود. در گام دوم خوشه های ایجاد شده از طریق روشی ساده اما کارا نرمال سازی شده اند. می توان با نرمال سازی داده های ورودی مدل دسته بندی، قابلیت پیش بینی و شناسایی الگو را در مدل دسته بندی افزایش داد. یکی از مشکلاتی که در رابطه با استفاده از مدل های دسته بندی در حوزه امتیازدهی اعتباری در تحقیقات قبلی مشاهده می شود بحث نرمال سازی داده ها می باشد.

جدول ۱. خصیصه های مورد بررسی در مورد متقاضیان اعتبار

وضعیت حساب جاری	وضعیت شغلی از نظر اقامت	سابقه اعتباری	مدت زمان عملکرد حساب
وضعیت شغلی	وضعیت حساب پس انداز	میزان اعتبار درخواستی	هدف از درخواست اعتبار
مدت زمان اقامت در محل کنونی	درصد اقساط	مشخصات فردی و جنسیت	وضعیت شغلی ضامن

دارایی ها	سن	پرداخت های دیگر اقساط	نوع مالکیت واحد مسکونی
تعداد اعتبارات اخذ شده در این بانک	وضعیت شغلی	تعداد ضامن	مالکیت تلفن

در امتیازدهی اعتباری، ارزش های عددی (داده های ورودی) نمایانگر خصوصیات متقاضی اعتبار می باشند که هر متقاضی ارزش های متفاوتی را دارد و در صورت اجرای یک نرمال سازی ساده (به عنوان مثال تقسیم هر داده بر بزرگترین داده مجموعه کلی داده ها) اطلاعات موثر و مفیدی را از دست خواهیم داد. در مجموعه داده ها بیشینه عددی برابر با ۱۸۴۲۴ می باشد. در صورتیکه کلیه داده ها را بر این عدد بمنظور انجام نرمالسازی داده ها، تقسیم کنیم، ممکن است اطلاعات مفیدی را که در داده ها وجود دارد از دست بدهیم، به این دلیل که با اینکار اکثر داده به سمت صفر میل می کند و به صفر نزدیک می شود. این کار منتهی به آموزش ناکافی شبکه می شود و توانایی مدل دسته بندی را در پیش بینی امتیاز اعتباری متقاضیان جدید بطور صحیح کاهش می دهد. بنابراین نرمال سازی داده های ورودی متقاضی اعتبار بایستی با دقت انجام گیرد، در حالیکه مفاهیم هر کدام از خصایص حفظ می شود. نحوه انجام نرمال سازی داده ها به آن صورت است که در این مرحله داده های ورودی بصورت جداگانه به اعدادی بین صفر و یک نرمال می شوند. این امر از طریق یافتن بیشه عددی در هر ستون (خصیصه) از مجموعه ها و تقسیم باقی ورودی های هر ستون به آن مقدار بیشینه انجام می گیرد. در حوزه دسته بندی، استفاده از چندین مدل دسته بندی کننده به جای یک مدل دسته بندی منجر به بهبود عملکرد و صحت دسته بندی می شود. مدل هایی که از چندین مدل دسته بندی کننده برای انجام دسته بندی استفاده می کنند، به عنوان مدل های جمعی و یا مدل های مبتنی بر یادگیری جمعی شناخته می شوند. در این مدل ها تصمیم گیری بر اساس ترکیب نتایج مجموعه ایی از مدل های دسته بندی صورت می گیرد که در این حالت ترکیب نتایج آن ها منجر به حصول عملکرد بهتری نسبت به استفاده منفرد از آن ها می شود.



شکل ۱. مدل پیشنهادی امتیازدهی اعتباری

مدل های مبتنی بر یادگیری جمعی، قابلیت دستیابی به پیش بینی های صحیح تر و با قابلیت های تعمیم بالاتر را به کاربران ارائه می کند. این مدل ها قادرند تا به صورت موثر و کارا از تنوع و گوناگونی دانش اعضاء مدل استفاده کنند تا خطای واریانس کاهش یابد. به عبارت دیگر، به منظور کاهش خطای انحراف توسط یک مدل دسته بندی، از چند مدل دسته بندی استفاده می شود. تصمیمات این مدل ها در نهایت توسط یک روش جمعی ترکیب خواهد شد. هدف اصلی یادگیری جمعی، ترکیب مجموعه ایی از مدل هاست که هر کدام وظیفه مشابهی را انجام می دهند تا یک مدل جمعی کلی تر و بهتر به دست آید که هم صحت بیشتری دارد و هم تصمیمات پایا و مطمئن تری نسبت به مدل های فردی ارائه می دهد. ایده اصلی در سیستم هایی با چند مدل دسته بندی، ترکیبی از مدل های دسته بندی است که در این صورت سیستم با ترکیب نتایج این مدل های دسته بندی، صحت و کارایی برتری نسبت به مدل دسته بندی منفرد به دست خواهد آورد. مدل هایی که مبتنی بر یادگیری جمعی هستند، از هر مدل دسته بندی منفرد، از طریق گردآوری و جمع آوری تصمیمات همه مدل های دسته بندی بهتر و برتر عمل می کنند. عملکرد مدل جمعی وابسته به این است که هر یک از مدل های دسته بندی بکار گرفته شده در مدل جمعی به خوبی آموزش ببینند. بنابراین تنوع و گوناگونی بین مدل های یادگیرنده بایستی حفظ شود. ترکیب مدل های دسته بندی در صورتی موثر و کارا عمل خواهد کرد که این مدل ها با یکدیگر متفاوت باشند. این تنوع را می توان از راه های مختلفی ایجاد کرد (معماری های گوناگون، مجموعه داده های آموزشی مختلف یا اختصاص تصادفی آن ها و یا بر اساس تحلیلات خوشه بندی). ایجاد واحد ترکیب کننده و حفظ تنوع میان دسته بندی کننده ها، دو عامل اصلی برای ساخت مدل جمعی است. نکته کلیدی رویکرد یادگیری جمعی مبتنی بر گرفتن نظر خبرگان قبل از تصمیم گیری نهایی است. این کار نشعت گرفته از فرایند تصمیم گیری در جامعه است. اثبات شده است که گرفتن نظر و ایده گروهی از خبرگان منجر به تصمیمات بهتری نسبت به تصمیمات یک فرد خبره می شود. بر اساس یادگیری جمعی و استفاده از چندین مدل دسته بندی، از ده ماشین بردار پشتیبان به عنوان داوران سیستم امتیازدهی اعتباری، در قالب ده عامل نرم افزاری در یک سیستم چند عاملی

استفاده شده است. مدل های تحت عامل یک پارادایم مناسب برای ایجاد محاسبات سریع و موازی ایجاد می کنند. این امر بدین دلیل است که آن ها دارای قابلیت هایی همچون انعطاف پذیری، قابلیت پیمانه ایی بودن، همزمانی، قابلیت پاسخگویی، استدلال، یادگیری، استقلال، ارتباط و همکاری می باشند. در مدل پژوهش برای هر خوشه یک عامل جداگانه در نظر گرفته شده است که این عامل وظیفه پردازش مورد نظر و تعریف شده برای آن عامل را روی خوشه اختصاص یافته به آن انجام می دهد. بدین طریق سیستم قادرخواهد بود که به طور همزمان وظایف مختلف را پردازش نماید. لازم به ذکر است که علاوه بر پردازش موازی در این سیستم، عوامل مختلف می توانند هر کدام عملیات مختلفی را روی خوشه مربوط به خود انجام دهند. این دو خصوصیت، قابلیت بسیار بالایی به مدل پیشنهادی، هم در زمینه سرعت عملیاتی و هم انعطاف پذیری جهت انجام عملیات مختلف روی یک بانک اطلاعاتی را می دهد. در گام نهایی، نتایج ایجاد شده توسط داوران توسط روشی مناسب ترکیب خواهد شد که این امر، یکی از موارد مهم و تاثیر گذار بر روی خروجی و عملکرد مدل جمعی می باشد. با توجه به بکارگیری روش خوشه بندی فازی، برای ترکیب نتایج از درجه عضویت نمونه ها استفاده شده است، به گونه ایی که برای هر عامل وزن اهمیتی متناسب با مجموع درجه عضویت اعضای خوشه متناظر آن عامل در نظر گرفته شده است. به عنوان نمونه وزن عامل یک برابر است با مجموع درجه عضویت اعضای خوشه یک، به همین ترتیب نیز برای خوشه های دیگر عمل شده است. در این پژوهش میزان صحت کلی هر روش به عنوان معیار برتری آن مطرح خواهد شد. علاوه بر صحت کلی، صحت نوع اول و صحت نوع دوم نیز برای هر روش محاسبه شده است. هم چنین، به عنوان معیار ارزیابی دیگری که در رابطه با عملکرد مدل های دسته بندی در تحقیقات مختلف بکار گرفته می شود، از معیار محدوده زیر نمودار برای مقایسه عملکرد روش های ذکر شده در این پژوهش استفاده شده است[۲۱].

۵. نتایج پژوهش

در گام ابتدایی از رویکرد خوشه بندی فازی برای تقسیم داده های این پژوهش به تعداد خوشه های معین استفاده شده است. برای این منظور از روش خوشه بندی FCM[۵] استفاده شده است که تعداد ده خوشه از مجموعه داده های پژوهش، استخراج شد. با توجه به فازی بودن خوشه بندی، ممکن است که یک داده حداقل به یک خوشه و حداکثر به ده خوشه

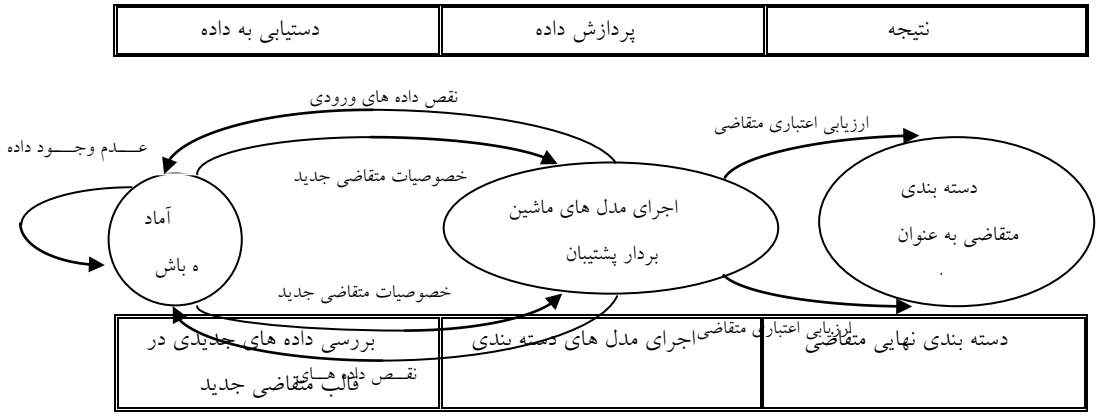
[۵] Fuzzy C-Mean

تعلق داشته باشد. به منظور ایجاد یک مدل دسته بندی ماشین بردار پشتیبان، اقدام به تست مدل با موتورهای استنتاج مختلف شد تا صحیح ترین مدل در جهت ایجاد دسته بندی های صحیح تر مشخص شود. نتایج مربوط به تست مدل با موتورهای استنتاج مختلف در جدول دو ارائه شده است و این نتایج با دو روش ترکیب نتایج تصمیمات اعضای مدل جمعی محاسبه شده است. روش اکثریت آراء پر کاربرد ترین روش ترکیب نتایج در حوزه دسته بندی و مدل های جمعی است و به طور وسیعی در تحقیقات اخیر مورد استفاده قرار گرفته شده است. در این پژوهش روش ترکیب جدیدی بکار گرفته شد که از درجه عضویت خوشه برای تعیین اهمیت هر یک از عامل های نرم افزاری سیستم، استفاده می کند. با توجه به نتایج ارائه شده در جدول دو، مشهود است که برترین صحت کلی مربوط به ماشین بردار پشتیبان با موتور استنتاج چند جمله ایی می باشد. با توجه به جدول دو، بکارگیری روش ترکیب نتایج درجه عضویت، صحت بالاتری را نسبت به روش اکثریت آراء ایجاد کرده است. سیستم چندعاملی بر اساس ده عامل با ساختار ذکر شده و موتور استنتاج چند جمله ایی طراحی شد. نمودار حالت سیستم چندعاملی پژوهش در شکل دو ارائه شده است. در این نمودار وضعیت عملکرد عامل تبیین شده است. برای عامل نرم افزاری سه مرحله اساسی مفروض است: دستیابی به داده، پردازش داده و نتیجه. در هر یک از این مراحل عامل مربوطه وظایفی را انجام خواهد داد: بررسی داده های جدید در قالب متقاضی جدید، اجرای مدل دسته بندی و دسته بندی نهایی متقاضی. چرخه عملکرد عامل در هر یک از مراحل در شکل دو تشریح شده است. عامل نرم افزاری پس از انجام دسته بندی نهایی متقاضی به حالت آماده به کار درمی آید. یکی از معروف ترین معماری های عامل ها در سیستم های چند عاملی معماری عقاید، خواسته ها، طرح ها می باشد.

جدول۲. نتایج پیاده سازی موتورهای استنتاج متفاوت ماشین بردار پشتیبان

نوع موتور استنتاج	**روش ترکیب نتایج**	
	اکثریت آراء	**درجه عضویت**
	صحت کلی (٪)	***صحت کلی (٪)***

RBF(C:10, RBF Gamma:0.1)	77.5	78.93
Polynomial(C:10, Gamma:1)	79.64	81.42
Sigmoid(C:10, Gamma:1)	49.64	56.42
Linear(C:10, Gamma:1)	71.78	71.07



شکل۲. نمودار حالت سیستم چندعاملی

عوامل نرم افزاری این پژوهش بر مبنای معماری عقاید، خواسته ها و طرح ها که شامل سه بخش مختلف زیر است طراحی شده است:

۱. عقاید: تعیین کننده پارامترها و مقادیر و پیش فرض های مورد نیاز و واقعیات های دریافتی از محیط است.

۲. خواسته ها: اهداف یک عامل است که باید از طریق اجرای فرآیند ها و طرح هایی از پیش تعریف شده به آن ها دست یابد.

۳. طر ح ها: شامل فرآیند ها و الگورتیم های اجرائی یک عامل است که وظایف یک عامل را تعیین می کند[۱۴].

برای هر یک از عوامل عقاید، خواسته ها و طرح هایی به شرح جدول سه متصور است. در نهایت، سیستم چندعاملی امتیازدهی اعتباری با ده عامل نرم افزاری ماشین بردار پشتیبان و روش ترکیب نتایج درجه عضویت تست شد. نتایج مدل های دیگر که در ادبیات پژوهش مرور شد، برای مقایسه با نتیجه سیستم پیشنهادی، مورد استفاده قرار گرفت. نتایج مدل های مختلف در جدول چهار ارائه شده است. با توجه به نتایج ارائه شده در جدول چهار، آشکار است که سیستم چندعاملی امتیازدهی اعتباری که مبتنی بر یادگیری جمعی است، بالاترین میزان صحت کلی را در بین روش های دیگر دارا می باشد. بگینگ، بوستینگ و

استکینگ نیز مدل هایی هستند که در حوزه یادگیری جمعی مطرح می شوند. در بسیاری از تحقیقات از این مدل های جمعی برای مسائل مربوط به دسته بندی استفاده شده است و نتایج خوبی ارائه شده است.

جدول۳. معماری عقاید، خواسته ها و طرح های عامل های پژوهش

طرح ها	خواسته ها	عقاید		
پارامتر تنظیم (C)	ارزیابی	وضعیت حساب جاری	مشخصات	دارایی ها
دقت رگرسیون (ε)	اعتباری و دسته	مدت زمان اقامت در محل کنونی	فردی و	سن
نوع موتور استنتاج	بندی متقاضی در	تعداد اعتبارات اخذ شده در این	جنسیت	درصد
گاما	دو دسته خوب یا	بانک	وضعیت	اقساط
بایاس	بد	وضعیت شغلی از نظر اقامت	شغلی ضامن	وضعیت
درجه		پرداخت های دیگر اقساط	نوع	شغلی
		مدت زمان عملکرد حساب	مالکیت واحد	تعداد
		وضعیت حساب پس انداز	مسکونی	ضامن
		میزان اعتبار درخواستی	سابقه	مالکیت
		هدف از درخواست اعتبار	اعتباری	تلفن
				وضعیت
				شغلی

نتایج مربوط به مدل های جمعی ذکر شده، که در تحقیق وانگ و همکاران[۲۰] و با داده های مشابه این پژوهش انجام گرفته، به منظور مقایسه با مدل پیشنهادی، در جدول چهار ارائه شده است. با وجود نتایج خوب و موفقی که این مدل ها حاصل کرده اند، نتایج جدول چهار، حاکی از این امر است که مدل پیشنهادی این پژوهش می تواند صحتی بالاتر(۸۱.۴۲) از این مدل ها ارائه دهد. علاوه بر این، تفاوت چشمگیری میان معیار محدوده زیر نمودار[۶] سیستم چندعاملی با روش های دیگر وجود دارد که نشان از عملکرد برتر این سیستم در دسته بندی متقاضیان اعتبار دارد. این معیار به خوبی می تواند رفتار کلی یک مدل دسته بندی را توصیف کند.

۶. نتیجه گیری

[۶] Area Under the Curve (AUC)

در این پژوهش مدلی ترکیبی مبتنی بر یادگیری جمعی به منظور امتیازدهی اعتباری ارائه شد. هدف این مدل، دسته بندی متقاضیان اعتبار به دو دسته متقاضیان "خوب" و "بد" بر اساس یک سری از معیار ها و خصوصیات فرد متقاضی می باشد. این مدل از ترکیب دو مدل دسته بندی و خوشه بندی حاصل شد، که برای خوشه بندی از خوشه بندی فازی و از ماشین بردار پشتیبان به عنوان مدل دسته بندی استفاده شد. هم چنین برای بهبود صحت و عملکرد مدل از یادگیری جمعی بهره گرفته شده است، که یک پارادایم قوی یادگیری ماشینی است. بر اساس رویکرد جمعی ده ماشین بردار پشتیبان به عنوان داور و در قالب عامل نرم

جدول۴. نتایج مدل های مختلف امتیازدهی اعتباری بروی داده های بانک آلمان

توضیحات	محدوده زیر نمودار	صحت(%)			مدل
		کلی	نوع دوم	نوع اول	
	۶۶/۲	/۹۱ ۶۵	/۹۱ ۶۴	/۴۹ ۶۷	تحلیل متمایز کننده
	/۶۷۵ ۶۶	۷۱	/۶۷ ۸۴	/۶۸ ۴۸	رگرسیون لجستیک
	۶۴/۲۴	/۳۵ ۷۰	/۵۵ ۷۸	/۹۳ ۴۹	درخت تصمیم
	۶۲/۸۸	/۵ ۶۸	/۲۹ ۸۶	/۴۷ ۳۹	شبکه RBF
	/۶۰۵ ۶۲	۷۱	/۵۸ ۹۷	/۶۳ ۲۷	ماشین بردار پشتیبان منفرد
	۷۰/۳۷	۷۴	/۴۸ ۸۵	/۲۶ ۵۵	پرسپترون چندلایه ایی منفرد
]۲۰[	۶۷/۳۱	/۹۲ ۷۴	/۳۴ ۸۶	/۲۸ ۴۸	درخت تصمیم (بگینگ)
	۶۷/۸	/۵۶ ۷۵	/۲۰ ۸۷	/۴۰ ۴۸	شبکه عصبی (بگینگ)
	۶۶/۶۴	/۹۳ ۷۵	/۸۶ ۸۹	/۴۲ ۴۳	ماشین بردار پشتیبان (بگینگ)
	/۹۲۵ ۶۱	/۷۷ ۷۲	/۹۶ ۸۲	/۸۹ ۴۰	درخت تصمیم (بوستیگ)

شبکه عصبی (بوستیگ)	/۲۰ ۴۹	/۶۳ ۸۳	/۳ ۷۳	/۴۱۵ ۶۶	
ماشین بردار پشتیبان (بوستیگ)	/۶۲ ۴۵	/۴۴ ۸۹	/۳ ۷۶	۶۷/۵۳	
استکینگ	۴۵	/۲۴ ۸۹	/۹۷ ۷۵	۶۷/۱۲	
سیستم چندعاملی امتیازدهی	/۶۶ ۶۶	/۰۸ ۸۸	**/۴۲** **۸۱**	**۷۷/۳۷**	

افزاری طراحی شد. متفاوت با روش های قبلی ترکیب نتایج که در تحقیقات قبلی در حوزه یادگیری جمعی بکارگرفته شده است، در این پژوهش روش ترکیب جدیدی بکارگرفته شد، که منجر به بهبود صحت سیستم امتیازدهی اعتباری در پژوهش گردید. مدل های دسته بندی ماشین بردار پشتیبان از طریق مجموعه داده های واقعی بانکی در آلمان مورد آموزش و تست قرار گرفتند. به منظور اعتبار سنجی مدل پیشنهادی، از چند مدل قدیمی و جدید استفاده شد. معیار ارزیابی مدل های مختلف، محدوده زیر نمودار و میزان دسته بندی درست متقاضیان، و یا به عبارتی میزان صحت کلی هر یک از روش ها می باشد. نتایج نشان داد که مدل پیشنهادی بسیار موثر و کارا است و منجر به سیستمی شد که با بهره گیری از عامل های نرم افزاری به صورت مناسب قادر به دسته بندی ورودی ها با صحت بالا می باشد. در این پژوهش نشان داده شد که استفاده از یادگیری جمعی و یا به عبارتی استفاده از چندین مدل دسته بندی به جای یک مدل دسته بندی، می تواند منجر به بهبود عملکرد دسته بندی شود. هم چنین، روش ترکیب نتایج درجه عضویت عملکرد برتری نسبت به روش اکثریت آراء، در ترکیب نتایج اعضای مدل جمعی از خود ارائه داد. در میان مدل های ذکر شده، برترین صحت متعلق به سیستم چندعاملی امتیازدهی اعتباری می باشد که مبتنی بر رویکرد یادگیری جمعی است و با روش درجه عضویت، نظر هر یک از داوران را با یکدیگر ترکیب می کند. با انجام ارزیابی های مختلف از نظر صحت و عملکرد مشخص شد که سیستم چندعاملی امتیازدهی اعتباری، بیشترین صحت و برترین عملکرد را در امتیازدهی اعتباری حاصل خواهد کرد و می تواند با دقت و صحت بیشتری نسبت به مدل های دیگر، بانک ها و موسسات مالی را برای اعتبارسنجی متقاضیان اعتبار یاری کند .

فهرست منابع

[۱]. اداره مطالعات و کنترل ریسک بانک تجارت . مدیریت ریسک در بانکداری . ویرایش سوم. تهران : بانک تجارت با همکاری نشر جهانگیر؛ ۱۳۸۶.

[2] Y, Altunbas, L, Gambacorta, & D, Ibanez, "Bank risk and monetary policy", vol 6, 121- Journal of financial stability 129, 2010.

[3] S, Lin, "A new two-stage hybrid approach of credit risk in banking industry", Expert system with application, vol 36, 8333-8341, 2009.

[4] P, Ya-qiong, "study on evaluation of consumer credit's risks of commercial banks", International Conference on Wireless Communications (1-4), 2007.

[5] X, Li, W, Ying, J, Tuo, B, Li, et al., "Application of classification trees to consumer credit scoring methods in commercial banks", IEEE international conference on systems, Man and Cybernetics (4112-4117), 2004.

[6] L, Yu, S, Wang, & K, Lai, "An intelligent-agent-based fuzzy group decision making model for financial multicriteria decision support: the case of credit scoring" European journal of operational research, vol 195, 942-959, 2009.

[7] C, Huang, M, Chen, & C, Wang, "Credit scoring with a data mining approach based on support vector machines", Expert systems with applications, vol 33, 847-856, 2007.

[8] Sustersic, M, Mramor, D, & Zupan, J Consumer credit scoring models with limited data Expert system with application 2009 vol 36: 4736-4744

[9] Y, Wang, "Building credit scoring systens based on support-based support vector machine", Fourth international conference on natural computation (323-327), 2008.

[10] L, Huang, & Y, Dai, "support vector machine approach for prediction of T cell epitopes", Third asia-pacific bioinformatics conference, (312-328) Singapore, 2010.

[11] A, Khashman, "A Neural Networks for Credit Risk Evaluation: Investigation of Different Neural Models and Learning Schemes", Expert System with Applications, vol 37, 6233-6239, 2010.

[12] E, Angelini, G, Tollo, & A, Roil, "neural network approach for credit risk evaluation", The quarterly review of economics and finance, vol 48, 733-755, 2088.

[13] J, Min, Y, Lee, "Bankruptcy prediction using support vector machine with optimal choice of kernel function parameters", Expert sustems with applications, vol 28, 603-614, 2005.

[14] H, Naji, B, Wells, & L, Etzkorn, "Creating an Adaptive Embedded System by Applying Multi Agent Techniques to Reconfigurable Hardware", Future Generation Computer Systems, vol 20: 1055-1081, 2004.

[15] L, Nanni, & A, Lumini, "An experimental comparison of ensemble of classifiers for bankruptcy prediction and credit scoring", Expert systems with applications, vol 36, 3028-3033, 2009.

[16] S, Piramuthu, "Financial credit-risk evaluation with neural and neurofuzzy systems", European journal of operational research, vol 112, 310-321, 1999.

[17] K, Shin,T, Lee, & H, Kim, "An application of support vector machines in bankruptcy prediction model", Exper systems with pplications, vol 28, 127-135, 2005.

[18] N, Hsieh "Hybrid mining approach in design of credit scoring model Expert systems with applications", vol 28, 655-665, 2005.

[19] C, Tsai, & M, Chen, “Credit rating by hybrid machine learning techniques”, Applied soft computing, vol 10, 374-380, 2010.

[20] G, Wang, J, Hao, J, Ma, & H, Jiang, “A comparative assessment of ensemble learning for credit scoring” Expert systems with applications, vol 38, 223-230, 2011.

[21] T, Fawcett, “ROC Graphs: Notes and practical considerations for researchers HP laboratories Palo Alto: Intelligent enterprise technologies laboratory 2004.

[22] B, Baesens, T, Van Gestel, S, Viaene, M, Stepanova, “Benchmarking state-of-art classification algorithm for credit scoring”, Journal of operational research society, vol 54: 627-635, 2003.

[23] W, Chen, C, Ma, & L, Ma “Mining the customer credit using hybrid support vector machine technique”, Expert systems with applications, vol 36, 7611-7616, 2009.

[24] D, Zhang, M, hifi, Q, Chen, & W, Ye, “A hybrid credit scoring model based on genetic programming and support vector machines” Fourth international conference on natural computation, (8-12) Jinan, 2008