

طبقه‌بندی اعتباری مشتریان حقوقی یک بانک ایرانی با استفاده از ^۱KFP-MCOC

محمدعلی رستگار* و سجاد جبینی**

تاریخ وصول: ۱۳۹۶/۱/۲۷ تاریخ پذیرش: ۱۳۹۶/۱۲/۸

چکیده

ریسک اعتباری، یکی از مهم‌ترین ریسک‌های بانک‌داری است که ناشی از عدم پرداخت به موقع اصل و فرع تسهیلات اعطایی است. اهمیت اندازه‌گیری این ریسک، به این دلیل است که عدم اندازه‌گیری و مدیریت آن، منجر به افزایش حجم مطالبات مشکوک‌الوصول می‌شود و زیان‌های غیر منتظره‌آتی را برای بانک‌ها به همراه دارد. در پژوهش حاضر به ارائه‌ی یک مدل مبتنی بر بهینه‌سازی خطی و غیرخطی پرداخته شده است. این مدل ریاضی، به دنبال یافتن یک ابرصفحه‌ی جداکننده است تا بتواند به کمک آن، ۸۵ مشتری خوش‌حساب و بدحساب موجود در بورس اوراق بهادار تهران را که از یک بانک ایرانی وام گرفته‌اند، از یکدیگر تفکیک نماید؛ همچنین در این راه، از توابع کرنل، فازی‌سازی داده‌ها و ضرایب جریمه به منظور بهبود مدل ریاضی استفاده شده است. نتایج این پژوهش، بیان‌گر این است که مدل بهینه‌سازی خطی با تابع کرنل سیگموئیدی، بهترین نتایج را در مقایسه با مدل‌های بهینه‌سازی خطی و غیرخطی با کرنل‌های خطی، چندجمله‌ای، سیگموئیدی و پایه‌شعاعی می‌دهد. مدل مذکور مشتریان بدحساب را کاملاً شناسایی می‌کند؛ لیکن در خصوص مشتریان خوش‌حساب و بدحساب با هم، کارایی ۸۰ درصدی دارد.

طبقه بندی JEL: C61, C65, E51, G32

واژه‌های کلیدی: ریسک اعتباری، بهینه‌سازی، تابع کرنل، فازی‌سازی، ضرایب جریمه

^۱ Kernel, Fuzzification, Penalty Factors - Multi Criteria Optimization Classifier

* استادیار رشته‌ی مهندسی مالی، گروه مهندسی مالی، دانشکده‌ی مهندسی صنایع و سیستم‌ها، دانشگاه تربیت مدرس، تهران، ایران. (نویسنده‌ی مسئول)، (Ma_rastegar@modares.ac.ir).

** کارشناس ارشد مهندسی مالی-گروه مهندسی مالی، دانشکده مهندسی صنایع و سیستم‌ها، دانشگاه تربیت مدرس، تهران، ایران.

۱- مقدمه

ریسک اعتباری، یکی از ریسک‌های عمده‌ی بانکی است که تعریف آن عبارت است از: احتمال عدم توانایی یا عدم تمایل وام‌گیرنده در پرداخت تعهدات مربوط به تسهیلات خود. تعریفی که کمیته بال^۲ از ریسک اعتباری ارائه می‌دهد، به این شرح است: «ریسک اعتباری عبارت است از امکان بالقوه این که قرض‌گیرنده از بانک در اجرای تعهدات خود در مقابل بانک در مدت مشخصی ناتوان شود».

اعطای تسهیلات، بانک‌ها را با ریسک اعتباری مواجه می‌کند که مدیران ریسک باید از طریق اتخاذ راهکارهایی این ریسک را مدیریت نمایند. اولین گام برای مدیریت ریسک اعتباری، اندازه‌گیری آن است. یکی از اولین گام‌های اندازه‌گیری این ریسک، تفکیک مشتریان بانکی است. این تفکیک گاه می‌تواند به صورت خوشه‌بندی در قالب گروه‌های مختلف با سطح‌های متفاوت ریسک باشد؛ یا به مانند این پژوهش می‌تواند به صورت یک مسأله‌ی طبقه‌بندی باشد.

در طول ۵۰ سال اخیر، محققان ریسک اعتباری، روش‌های موثری را برای پیش‌بینی نکول یا ورشکستگی قرض‌گیرندگان توسعه داده‌اند. گام‌های مؤثر در این حوزه توسط بیور^۳ (۱۹۶۷) و آلتمن^۴ (۱۹۶۸) برداشته شد که مدل‌هایی مبتنی بر حسابداری و با اتخاذ نسبت‌های مالی به منظور پیش‌بینی ورشکستگی توسط آن‌ها ایجاد شد. پس از ایشان، اولسن^۵ (۱۹۸۰) به منظور نشان دادن مشکل مدل آلتمن، یک مدل لاجیت شرطی ارائه داد. مزایای عملی مدل لاجیت شرطی این بود که هیچ فرضی لازم نداشت؛ بنابراین اولسن نمونه‌هایی شامل بنگاه‌های ورشکسته و غیرورشکسته را با استفاده از ۷ نسبت مالی و دو متغیر مجازی بررسی کرد.

از میان روش‌های اندازه‌گیری ریسک اعتباری و طبقه‌بندی اعتباری تکنیک‌های آماری و بهینه‌سازی در ابتدا شکل گرفت. اصلی‌ترین اشکال روش‌های آماری دقت ناکافی آن‌هاست؛ در حالی که مزیت اصلی آن‌ها، سادگی این مدل‌ها است. مدل‌های مبتنی بر بهینه‌سازی نیز دارای این مزیت هستند که همواره به دنبال بهترین جواب از میان فضای مورد جستجو هستند. با گذشت زمان، روش‌هایی از نوع جعبه‌ی سیاه^۶ برای پشتیبانی تصمیمات مالی ظهور کرد. این روش‌ها، شامل شبکه‌های عصبی مصنوعی با توپولوژی‌های متنوع، ماشین‌های بردار پشتیبان^۷، الگوریتم‌های ژنتیک، برنامه‌ریزی ژنتیک، سیستم‌های ایمنی مصنوعی^۸ و ... است.

^۲ Basel

^۳ Beaver

^۴ Altman

^۵ Ohlson

^۶ Black Box

^۷ Support Vector Machines (SVMs)

^۸ Artificial Immune Systems (AISs)

در این پژوهش ما به دنبال آن هستیم تا به کمک یک ابرصفحه‌ی جداکننده، مشتریان حقوقی وام‌گیرنده از یکی از بانک‌های ایرانی را به دو طبقه‌ی خوش حساب و بدحساب تقسیم‌بندی کنیم. مشتریان بانک مذکور، ۸۵ شرکت متعلق به بورس اوراق بهادار تهران هستند. این کار بر مبنای بهینه‌سازی خطی و غیرخطی صورت پذیرفته‌است که در بخش‌های بعد به تفصیل مورد بررسی قرار خواهد گرفت.

۲- پیشینه‌ی تحقیق

همان‌گونه که در مقدمه‌ی پژوهش حاضر هم اشاره شد، پژوهش‌های زیادی در حوزه‌ی رتبه‌بندی و دسته‌بندی مشتریان صورت گرفته است که در دسته‌هایی نظیر مدل‌های آماری، بهینه‌سازی، ساختاری و مبتنی بر هوش مصنوعی انجام شده است. از طرفی با توجه به شباهت این پژوهش به مدل‌های مبتنی بر هوش مصنوعی به مرور ادبیات در این حوزه پرداخته‌ایم.

تکنیک ماشین‌های بردار پشتیبان بر مبنای نظریه‌ی یادگیری آماری^۹ و بهینه‌سازی در پژوهش‌هایی نظیر کورتس و وپنیک (۱۹۹۵)^{۱۰}، وپنیک (۱۹۹۵)^{۱۱}، محبوب شد. ایده‌ی اصلی تکنیک ماشین‌های بردار پشتیبان، جدا کردن نمونه‌ها از یکدیگر در قالب کلاس‌های متفاوت است که این کار با عبور دادن و انطباق یک ابرصفحه انجام می‌شود. این ابرصفحه به طور هم‌زمان حاشیه‌ی بین کلاس‌ها را پیشینه و طبقه‌بندی اشتباه را کمینه می‌سازد؛ همچنین برای داده‌هایی که به صورت خطی قابل جدا شدن نیستند، در پژوهش‌هایی نظیر کریستیانینی و شاولیور (۲۰۰۰)^{۱۲}، و هامل (۲۰۰۹)^{۱۳} از تابع کرنل استفاده شد تا این مشکل مرتفع شود. (ژانگ، گائو، و شی، ۲۰۱۴)^{۱۴}

یاپ، اونگ، و حسین (۲۰۱۱)^{۱۵} در پژوهش خود از تکنیک‌های داده‌کاوی برای امتیازدهی اعتباری استفاده کردند؛ همچنین یو و همکاران (۲۰۱۱)^{۱۶} از روش حداقل مربعات ماشین‌های بردار پشتیبان برای این کار استفاده کرد. آن‌ها از طراحی آزمایش‌ها برای انتخاب پارامترهای روش ماشین‌های بردار پشتیبان استفاده کردند که در نهایت این مدل نتایج رضایت‌بخشی ارائه داد.

^۹ Statistical Learning Theory

^{۱۰} Cortes & Vapnik

^{۱۱} Vapnik

^{۱۲} Cristianini & Shawe-Taylor

^{۱۳} Hamel

^{۱۴} Zhang, Gao, & Shi

^{۱۵} Yap, Ong & Hussain

^{۱۶} Yu et al

در پژوهش‌هایی نظیر اولافسون، لی، و وو (۲۰۰۸)^{۱۷} و یا کورن، دائننز، و جردن (۲۰۱۲)^{۱۸} علاقه‌ی فزاینده‌ای در خصوص ترکیب بهینه‌سازی و داده‌کاوی وجود داشته است.

نیکلیس، دامپوس و زوپونیدیس (۲۰۱۴)^{۱۹} در پژوهش خود برای رتبه‌بندی اعتباری از رویکردهای یادگیری ماشینی و از میان این رویکردها از مدل ماشین‌های بردار پشتیبان استفاده کرده‌اند که داده‌های حسابداری را با رویکرد مبتنی بر اختیارات بلک^{۲۰}، شولز^{۲۱} و مرتون^{۲۲} ترکیب می‌کند. در این پژوهش، مدل ماشین‌های بردار پشتیبان خطی و غیرخطی بر روی یک مجموعه داده از شرکت‌های یونانی اجرا شده است که به عقیده‌ی آنان دقت بالایی در امتیازدهی اعتباری دارد و در مقایسه با مدل‌های رگرسیونی، عملکرد بهتری از خود نشان داده است.

ژانگ و همکاران (۲۰۱۴) در پژوهشی به کمک استفاده از مدل‌های بهینه‌سازی چند هدفه و با ترکیب آن با تابع کرنل^{۲۳}، ضرایب جریمه تابع هدف و فازی نمودن داده‌ها، مدلی را برای طبقه‌بندی اعتباری ارائه کردند. این مدل دو هدفه میان اهداف همپوشانی کلاس‌های مختلف و فاصله‌ی نقاط ورودی از حدود تصمیم، موازنه‌ای ایجاد می‌کند. در نهایت آن‌ها این مدل را بر روی داده‌های آلمانی، استرالیایی و آمریکایی اجرا کردند و آن را با مدل‌هایی نظیر ماشین‌های بردار پشتیبان فازی و غیرفازی، شبکه‌های عصبی مصنوعی با توابع کرنل مختلف و... مقایسه کردند که نتایج حاکی از برتری مدل مذکور در اکثر مقایسه‌ها بود.

هریس (۲۰۱۵) در پژوهش خود تکنیک ماشین‌های بردار پشتیبان خوشه‌بندی شده^{۲۴} را برای توسعه کارت‌های امتیازی اعتباری معرفی و استفاده کرد. او در مقاله‌ی خود به این موضوع اشاره داشته است که ایراد مدل‌های سابق مبتنی بر ماشین‌های بردار پشتیبان غیرخطی این بوده است که با افزایش حجم داده‌ها، هزینه‌ی محاسبات را بالا می‌برد. بر طبق این گفته، مطالعه‌ی وی روش مذکور را با سایر ماشین‌های بردار پشتیبان غیرخطی و همچنین مدل‌های رگرسیونی مقایسه نموده است و نشان داده که روش وی در طبقه‌بندی، در حالی که هزینه‌ی محاسباتی کمتری دارد، عملکرد خوبی از خود نشان داده است.

¹⁷ Olafsson, Li, & Wu

¹⁸ Corne, Dhaenens, & Jourdan

¹⁹ Niklis, Doumpos, & Zopunidis

²⁰ Black

²¹ Scholes

²² Merton

²³ Kernel Function

²⁴ Clustered Support Vector Machines (CSVM)

همان‌گونه که در بررسی پژوهش‌ها دیده شد، با گذشت زمان، تمایل پژوهشگران از سمت استفاده از مدل‌های ساختاری و رگرسیونی، به سمت مدل‌های مبتنی بر هوش مصنوعی متمایل شد. همان‌طور که گرزالچانی و رودزینسکی^{۲۵} (۲۰۱۶)، بیان نمودند، با گذشت زمان روش‌ها از نوع جعبه سیاه برای پشتیبانی تصمیمات مالی ظهور کرد. این روش‌ها شامل طیف وسیعی از تکنیک‌ها، نظیر شبکه‌های عصبی مصنوعی با توپولوژی‌های متنوع (آتیا، ۲۰۰۱^{۲۶}؛ خشمان، ۲۰۰۹^{۲۷}؛ مالهترا و مالهترا، ۲۰۰۳^{۲۸})، ماشین‌های بردار پشتیبان (لی، ۲۰۰۷^{۲۹}؛ اسکبش و استکینگ، ۲۰۰۵^{۳۰})، و... است.

از میان جدیدترین مدل‌ها، ماشین‌های بردار پشتیبان، محبوبیت زیادی بین محققان پیدا کرده است و هر پژوهش یا به دنبال بسط و توسعه‌ی مدل خود و یا بهبود مدل‌های طبقه‌بندی پیشین بوده است؛ همچنین مطالعات زیادی وجود دارد که مدل ماشین‌های بردار پشتیبان را با سایر مدل‌ها مقایسه کرده و نشان داده‌اند عملکرد این روش بهتر از روش‌هایی نظیر رگرسیون، شبکه‌های عصبی مصنوعی و ... داشته است. از طرفی ژانگ و همکاران (۲۰۱۴) نشان دادند که مدل‌های مبتنی بر ماشین‌های بردار پشتیبان نارسایی‌هایی دارند که باید در مدل‌های ریاضی رفع شوند. برخی از این نارسایی‌ها، عبارتند از: حساسیت ماشین‌های بردار پشتیبان به انواع اخلال، داده‌های پرت، غیر نرمال بودن داده‌ها، عدم جدایی‌پذیری خطی داده‌ها، اهمیت غیر یکسان طبقه‌ها نسبت به یکدیگر و ... از این رو در پژوهش خود از توابع کرنل، فازی‌سازی داده‌ها، و ضرایب جریمه بدین منظور استفاده کردند.

در نهایت باید عنوان کرد که ایده کلی پژوهش حاضر به این طریق شکل گرفته است که به دنبال یک ابرصفحه جداکننده باشیم تا انحرافات از این ابرصفحه حداقل شود و نقاط تا حد امکان از این ابرصفحه دور باشند و پس از آن افزونه‌های ذکر شده را به آن خواهیم افزود.

در ادامه به بیان روش‌شناسی تحقیق و ارائه‌ی مدل ریاضی مورد استفاده پرداخته خواهد شد و پس از آن به ارائه‌ی نتایج تحقیق و جمع‌بندی نهایی آن خواهیم پرداخت.

²⁵ Gorzalczy. & Rudziński

²⁶ Atiya

²⁷ Khashman

²⁸ Malhotra & Malhotra

²⁹ Lee

³⁰ Schebesch & Stecking

۳- روش‌شناسی تحقیق

در این بخش ابتدا به ارائه داده‌های تحقیق و سپس به ارائه‌ی مدل ریاضی مورد استفاده خواهیم پرداخت. در این تحقیق، عمل طبقه‌بندی اعتباری را بر روی مشتریان یکی از بانک‌های ایرانی انجام خواهیم داد. مشتریانی که قصد طبقه‌بندی آن‌ها را داریم، ۸۵ شرکت موجود در بورس اوراق بهادار تهران هستند. این ۸۵ شرکت، توسط بانک ایرانی مذکور و با نظر و تصمیم کارشناسان از رتبه ۱ تا رتبه ۸۵، رتبه‌بندی شده بود؛ به طوری که رتبه‌ی ۱، خوش‌حساب‌ترین مشتری و رتبه‌ی ۸۵ بدحساب‌ترین مشتری را نشان می‌داد. مدل ریاضی این پژوهش از نوع طبقه‌بندی دو وضعیتی است و از طرف دیگر داده‌های ارائه شده از سوی بانک به صورت رتبه‌بندی است؛ لذا باید این رتبه‌های تجربی در دو کلاس خوش‌حساب و بدحساب قرار گیرند. در خصوص تعداد اعضای هر کلاس، با استفاده از پرسشی که از ۱۰ کارشناس امور اعتباری بانک به عمل آمد، به این جمع بندی رسیدیم که باید ۷۰٪ از داده‌ها را به کلاس اعتباری خوب و ۳۰٪ آن‌ها را به کلاس اعتباری بد اختصاص دهیم؛ به عبارت دیگر ۶۰ شرکت را خوش‌حساب و ۲۵ شرکت را بدحساب تلقی می‌کنیم.

بخش دوم اطلاعات، صورت‌های مالی این شرکت‌هاست که قرار است به کمک آن‌ها عمل طبقه‌بندی اعتباری انجام گیرد. به دلیل آن که مفهوم ریسک اعتباری یک سازمان مفهومی آنی نیست و بازه‌ای از عملکرد آن سازمان را شامل می‌شود؛ لذا با توجه به اینکه رتبه‌های اعتباری بانک مذکور، مربوط به سال ۱۳۹۳ است، صورت‌های مالی ۱۲ ماهه مربوط به دوره‌ی ۵ ساله ۱۳۸۹ تا ۱۳۹۳ در این تحقیق مورد بررسی قرار گرفته است. داده‌های خامی که در این بخش جمع‌آوری شده است، ۱۶ آیتم موجود در ترازنامه و صورت حساب سود و زیان این شرکت‌هاست. این ۱۶ آیتم عبارتند از: حساب‌های دریافتی، موجودی مواد و کالا، دارایی‌های جاری، دارایی‌های ثابت، کل دارایی‌ها، بدهی‌های جاری، کل بدهی‌ها، حقوق صاحبان سهام و سود انباشته که از ترازنامه‌ی این شرکت‌ها جمع‌آوری شده است و سرمایه، فروش، سود ناخالص، سود عملیاتی، هزینه‌های مالی، سود قبل از مالیات و سود خالص پس از مالیات که از صورت حساب سود و زیان آن‌ها اخذ شده است. این داده‌های خام جمع‌آوری شده به منظور محاسبه‌ی نسبت‌های مالی مورد استفاده قرار می‌گیرد.

پس از استخراج اقلام ذکر شده از ترازنامه و صورت حساب سود و زیان، نسبت‌های مالی به دست آمده از این نسبت‌ها جهت استفاده برای طبقه‌بندی اعتباری، عبارتند از: نسبت جاری، نسبت آنی، دوره‌ی متوسط وصول مطالبات، نسبت سود عملیاتی به فروش، نسبت سود ویژه به فروش، بازده حقوق صاحبان سهام، بازده مجموع دارایی‌ها، نسبت بدهی، نسبت حقوق صاحبان سهام، گردش

دارایی ثابت، گردش مجموع دارایی‌ها، نسبت بدهی به حقوق صاحبان سهام، نسبت گردش سرمایه جاری، سرمایه در گردش به کل دارایی‌ها، نسبت پوشش بهره، نسبت سود انباشته به کل دارایی‌ها. دلیل استفاده از این نسبت‌ها به اختصار به شرح ذیل است:

نسبت جاری توان پاسخ‌گویی به تعهدات شرکت در کوتاه مدت را نشان می‌دهد که معیار مناسبی جهت بررسی رتبه‌ی اعتباری یک شرکت است. نسبت آنی در صورتی که معادل یک و بیشتر باشد، بیان‌گر آن است که شرکت، هیچ‌گونه مشکلی در پرداخت بدهی‌های جاری خود ندارد. دوره‌ی وصول مطالبات نشانگر این است که شرکت در وصول مطالبات خود پیگیری مؤثری را به انجام می‌رساند یا خیر که این موضوع در وضعیت اعتباری شرکت تأثیرگذار است. نسبت سود عملیاتی به فروش و سود ویژه به فروش، میزان موفقیت شرکت در ایجاد سود حاصل از عملیات را نشان می‌دهد. نسبت‌های بازده حقوق صاحبان سهام و بازده مجموع دارایی‌ها، موفقیت شرکت را در تحصیل سود بیشتر برای سهامداران و همچنین موفقیت در استفاده از دارایی نشان می‌دهد که این خود در ایجاد یک رتبه‌ی اعتباری خوب برای شرکت بی‌تأثیر نیست. نسبت بدهی و نسبت حقوق صاحبان سهام نسبتی است که بالا بودن نسبت بدهی و پایین بودن نسبت حقوق صاحبان سهام بیانگر مخاطره‌آمیز بودن شرکت از دیدگاه بستانکاران است که نسبت بسیار مهمی جهت تعیین وضعیت اعتباری شرکت است.

بالا بودن نسبت بدهی به حقوق صاحبان سهام، بیان‌گر این است که شرکت بستانکاران بیشتری دارد و این موضوع در طبقه‌بندی اعتباری یک شرکت جهت اخذ تسهیلات، فاکتور مهمی است. نسبتی مانند نسبت پوشش بهره، بیانگر این است که شرکت تا چه اندازه می‌تواند هزینه‌های بهره را پرداخت کند که باز هم در تعیین وضعیت اعتباری از اهمیت زیادی برخوردار است. به همین ترتیب سایر نسبت‌ها مانند نسبت گردش سرمایه جاری، نسبت سود انباشته به کل دارایی‌ها و سرمایه در گردش به کل دارایی‌ها، از جمله نسبت‌هایی است که بیانگر استفاده درست از سرمایه شرکت، سود و زیان‌های متحمل شده و همچنین وجود زیان عملیاتی در شرکت است که همگی تأثیر عمده‌ای در تعیین وضعیت اعتباری یک شرکت می‌گذارند.

برای ساختن یک مدل طبقه‌بندی اعتباری دو وضعیتی یعنی وضعیت اعتباری خوب و بد، در ابتدا مجموعه T به عنوان داده‌های آموزش در نظر گرفته می‌شود. این مجموعه، به صورت زوج‌های مرتبی است که به صورت $T = \{(x_1, y_1), \dots, (x_n, y_n)\}$ است که در آن n تعداد مشتریان بانکی است. از طرف دیگر هر x_i برداری از شاخص‌های مالی در نظر گرفته شده برای طبقه‌بندی اعتباری است که در ادامه، این شاخص‌ها معرفی خواهند شد و همچنین $x_i \in \mathbb{R}^d$.

به عبارت دیگر تعداد شاخص‌های مالی که برای هر مشتری در نظر گرفته شده است، برابر با d است. عضو دوم زوج‌های مرتب، y_i ها هستند که بیانگر خوب یا بد بودن وضعیت اعتباری تجربی هر مشتری است که توسط بانک گزارش شده است؛ به عبارت دیگر $y_i \in \{-1, 1\}$ است؛ به طوری که وضعیت اعتباری خوب با عدد -1 و وضعیت اعتباری بد با عدد 1 نشان داده می‌شود؛ همچنین لازم به ذکر است که به ازای $i = 1, \dots, m$ مقدار y_i برابر با -1 و به ازای $i = m + 1, \dots, n$ مقدار y_i برابر با 1 است؛ به عبارت دیگر تعداد مشتریان دارای وضعیت اعتباری خوب، m تا و تعداد مشتریان دارای وضعیت اعتباری بد، $n - m$ تا است.

ایده‌ی کلی برای حل مسأله‌ی طبقه‌بندی اعتباری، پیدا کردن یک ابرصفحه‌ی جداکننده با معادله‌ی کلی $w^T x - b = 0$ است که در آن بردار w بیانگر وزن هر شاخص مالی، و b بیانگر میزان آریبی^{۳۱} ابرصفحه است. به کمک این ابرصفحه نخست می‌توان وضعیت‌های اعتباری خوب و بد را از یکدیگر جدا کرد؛ و سپس در صورت ورود یک مشتری جدید به سیستم تقاضای اعتبار، خوب یا بد بودن وضعیت اعتباری آن را پیش‌بینی کرد. در همین راستا، باید به مدل‌سازی این مسأله از طریق تعریف توابع هدف و محدودیت‌های آن پرداخت.

با معلوم بودن بردار w و مقدار b در ساده‌ترین وضعیت داریم:

اگر $w^T x_i < b$ ، آنگاه $y_i = -1$

اگر $w^T x_i > b$ ، آنگاه $y_i = 1$

قواعد شرطی فوق، به صورت سخت‌گیرانه به مسأله نگاه می‌کند؛ زیرا در عمل امکان دارد که داده‌ای دارای اعتبار خوبی باشد؛ اما به دلیل اینکه اندکی از مرز خود تجاوز کرده است، جزو اعتبارهای بد شناسایی شود و یا برعکس. بدین منظور قواعد شرطی فوق را با تغییر و سهل‌گیری مرزها، اصلاح و معرفی می‌کنیم:

اگر $w^T x_i < b + \alpha_i$ ، آنگاه $y_i = -1$

اگر $w^T x_i > b - \alpha_i$ ، آنگاه $y_i = 1$

شرط‌های جدید، نسبت به نوع ابتدایی آن سخت‌گیری کمتری دارند و به بردارهای x_i این اجازه را می‌دهند تا با میزانی از انحراف مشخص، باز هم وضعیت اعتباری‌شان به درستی شناسایی شود. در روابط بالا، این میزان انحراف از ابرصفحه‌ی مذکور، با α نمایش داده شده است. در حقیقت به هر بردار (هر مشتری) این اجازه داده می‌شود تا با مقداری از انحراف α ، در طبقه‌ی درست خود قرار گیرد. به دلیل اینکه زیاد شدن انحراف به مسأله لطمه وارد می‌کند و عملاً با بازتر شدن کران‌های تصمیم‌گیری، امکان شناسایی اشتباه وضعیت‌های اعتباری وجود دارد؛ بنابراین تابع هدف را به صورت کمینه‌سازی تابع نرم مرتبه p این

³¹ Bias

انحرافات معرفی می‌کنیم. نُرم مرتبه p یک بردار مانند X با n عضو به صورت زیر تعریف می‌شود:

$$\|x\|_p = \sqrt[p]{\sum_{i=1}^n |x_i|^p} \quad (1)$$

در این تحقیق، مدل‌های پایه و ابتدایی به ازای نرم مرتبه ۱ و ۲ طراحی شده‌اند. فرم خطی مدل بهینه‌سازی تحقیق در ادامه ارائه شده است.

$$\text{Min } C \sum_{i=1}^n \alpha_i \quad (2)$$

s.t:

$$w^T x_i < b + \alpha_i \quad \forall i = 1, \dots, m \quad (\forall y_i = -1) \quad (3)$$

$$w^T x_i > b - \alpha_i \quad \forall i = m + 1, \dots, n \quad (\forall y_i = 1) \quad (4)$$

$$\alpha_i \geq 0 \quad \forall i = 1, \dots, n \quad (5)$$

همچنین تابع هدف فرم غیرخطی آن که فاصله‌ها به صورت نُرم اقلیدسی هستند، در ذیل ارائه شده است.

$$\text{Min } C \|\alpha\|_2 \quad (6)$$

پس از ساخت اولیه‌ی مدل، محدودیت‌های نامساوی فوق را به تساوی تبدیل می‌کنیم و پس از آن دو محدودیت اصلی را به یک محدودیت تبدیل می‌نماییم و مسأله را برای هر دو فرم خطی و غیرخطی بازنویسی می‌کنیم. ضمناً به دلیل آن که هدف ثانویه، بیشینه‌سازی β_i ها است، و از طرفی تابع هدف ابتدایی به صورت کمینه‌سازی است، بیشینه‌سازی $\sum_{i=1}^n \beta_i$ را به صورت کمینه‌سازی $-\sum_{i=1}^n \beta_i$ مدل‌سازی می‌کنیم؛ لذا برای مدل خطی داریم:

$$\text{Min } C \sum_{i=1}^n \alpha_i - \sum_{i=1}^n \beta_i \quad (7)$$

s.t:

$$w^T x_i - b = y_i(\beta_i - \alpha_i) \quad \forall i = 1, \dots, n \quad (8)$$

$$\alpha_i, \beta_i \geq 0 \quad \forall i = 1, \dots, n \quad (9)$$

همان‌گونه که در پژوهش ژانگ و همکاران (۲۰۱۴) اشاره شد؛ به دلیل وجود این سه نارسایی که عبارتند از: یکسان نبودن اشتباه در طبقه‌بندی هر یک از دو کلاس، تفکیک‌ناپذیری خطی داده‌ها، و داده‌های پرت. در ادامه‌ی بخش به مدل مورد بررسی، به ترتیب افزونه‌های ضرایب جریمه، تابع کرنل، و فازی‌سازی را می‌افزاییم تا این نارسایی‌ها مرتفع شوند.

اولین افزونه‌ی این مدل، وجود ضرایب جریمه در تابع هدف انحرافات است. از آنجایی که داشتن انحراف، امری نامطلوب است، با افزودن ضریب جریمه به تابع هدف کمینه‌سازی انحرافات، مدل از مثبت شدن انحرافات جلوگیری می‌کند؛

همچنین به دلیل اینکه طبقه‌بندی اشتباه برای مشتریان بدحساب با خوش‌حساب از حساسیت یکسانی برخوردار نیست، انحرافات هر دسته را با ضرایب جداگانه مورد جریمه قرار می‌دهیم و برای اشتباه در طبقه‌بندی مشتریان بدحساب جریمه‌ی بیشتری در نظر می‌گیریم. دومین افزونه مدل توابع کرنل است. برای جداسازی خطی، توابع کرنل متنوعی وجود دارد که از میان آن‌ها در این پژوهش از توابع کرنل خطی $K(x,y)=x^T y$ ، چندجمله‌ای $K(x,y)=(x^T y+c)^d$ ($c \geq 0, d \geq 2$)، کرنل سیگموئیدی تابع پایه شعاعی $K(x,y)=\exp\left(-\frac{\|x-y\|_2^2}{2\sigma^2}\right)$ ($\sigma > 0$) و کرنل سیگموئیدی $K(x,y)=\tanh(ax^T y+r)$ استفاده و عملکرد آن‌ها با یکدیگر مقایسه خواهد شد.

در نهایت سومین افزونه این مدل، فازی‌سازی داده‌هاست. یک تابع تصمیم ایده‌آل در یک مسأله‌ی طبقه‌بندی، هر داده‌ی ورودی را به درستی در یک کلاس از پیش تعریف شده طبقه‌بندی می‌کند. در دنیای واقعی به دلیل وجود جزء اخلاص، داده‌های پرت و غیرنرمال، عملکرد مدل تصمیم‌گیری به شدت تضعیف می‌شود؛ علاوه بر این سهم هر داده‌ی ورودی در طبقه‌بندی، باید با سایر داده‌ها متفاوت باشد و واضح است که ارزش داده‌ها با یکدیگر یکسان نیستند (ژانگ و همکاران، ۲۰۱۴). از این رو در این مدل، داده‌های ورودی با درجه‌ی عضویت فازی S_i که عددی بین صفر و یک است، در یک کلاس قرار می‌گیرند. روابطی که برای فازی‌سازی داده‌ها به کار رفته، در ادامه ارائه شده است. اگر $\bar{\Phi}_{y_i}$ را میانگین کلاس و Γ_{y_i} را شعاع کلاس y_i بنامیم، داریم:

$$\bar{\Phi}_{y_i} = \sum_i y_i \phi(x_i) / \sum_i y_i \quad (10)$$

$$r_{y_i} = \max \|\phi(x_i) - \bar{\Phi}_{y_i}\|_2 \quad (11)$$

آن‌گاه درجه‌ی عضویت فازی S_i برای هر داده‌ی ورودی، یک تابع خطی از میانگین و شعاع کلاس است که عبارت است از:

$$S_i = 1 - \frac{d(\phi(x_i), \bar{\Phi}_{y_i})}{(r_{y_i} + \delta)} \quad (12)$$

که در آن $\delta > 0$ یک ثابت ناچیز است که از صفر شدن S_i جلوگیری می‌کند؛ همچنین $d(\phi(x_i), \bar{\Phi}_{y_i})$ عبارت است از:

$$\begin{aligned} & d(\phi(x_i), \bar{\Phi}_{y_i}) \\ &= \|\phi(x_i) - \bar{\Phi}_{y_i}\|_2 = \sqrt{\phi^2(x_i) - 2\phi(x_i)\bar{\Phi}_{y_i} + \bar{\Phi}_{y_i}^2} \\ &= \sqrt{\phi(x_i)^T \phi(x_i) - 2 \frac{\sum_j y_j \phi(x_j)^T \phi(x_i)}{\sum_j y_j} + \frac{\sum_j \sum_k y_j y_k \phi(x_j)^T \phi(x_k)}{\sum_j y_j \sum_k y_k}} \\ &= \sqrt{K(x_i, x_i) - 2 \frac{\sum_j y_j K(x_i, x_j)}{\sum_j y_j} + \frac{\sum_j \sum_k y_j y_k K(x_j, x_k)}{\sum_j y_j \sum_k y_k}} \quad (13) \end{aligned}$$

در نهایت، فرم مدل اصلی به صورت زیر خواهد بود:

$$\text{Min } C_1 \sum_{i=1}^m s_i \alpha_i + C_2 \sum_{i=m+1}^n s_i \alpha_i - \sum_{i=1}^n \beta_i \quad (14)$$

s.t:

$$\sum_{j=1}^n \lambda_j y_j K(x_j, x_i) - b = y_i (\beta_i - \alpha_i) \quad \forall i=1, \dots, n \quad (15)$$

$$\alpha_i, \beta_i \geq 0 \quad \forall i=1, \dots, n \quad (16)$$

$$0 \leq \lambda_j \leq C_1, \text{ for } y_j = -1, \quad 0 \leq \lambda_j \leq C_2, \text{ for } y_j = 1 \quad (17)$$

ضمن آن که تابع هدف فرم غیرخطی نهایی این مدل نیز به صورت زیر در

می‌آید:

$$\text{Min } C_1 \sqrt{\sum_{i=1}^m (s_i \alpha_i)^2} + C_2 \sqrt{\sum_{i=m+1}^n (s_i \alpha_i)^2} - \sum_{i=1}^n \beta_i \quad (18)$$

پس از حل مدل، مقادیر λ و b به دست می‌آیند که به کمک آن‌ها تابع تصمیم‌گیری برای هر داده‌ی ورودی جدید x که به تازگی وارد سیستم تقاضای اعتبار شود، به دست می‌آید که عبارت است از:

$$f(x) = \text{sign}(w^T \phi(x) - b) = \text{sign}(\sum_{j=1}^n \lambda_j y_j K(x_j, x) - b) \quad (19)$$

در این پژوهش قصد داریم تا پس از تغییر دادن برخی پارامترها، شرایط و ... آن‌ها را با یکدیگر مقایسه کنیم تا بهترین شرایط جهت پیش‌بینی طبقه‌ی اعتباری مشتریان انتخاب شود. برای مقایسه‌ی این حالات، باید ابتدا معیارهای ارزیابی مناسبی را تعریف کرد تا نتیجه‌ی مقایسه، برای تصمیم‌گیری قابل اتکا باشد. در پژوهش ژانگ و همکاران (۲۰۱۴)، چندین معیار ارزیابی برای ارزیابی طبقه‌بندی اعتباری استفاده شده که در این پژوهش قصد داریم از آن‌ها استفاده کنیم. نتیجه‌ای که مدل ریاضی، اعتبار یک مشتری را پیش‌بینی می‌کند، ممکن است اعتبار بد (Positive)، یا اعتبار خوب (Negative) باشد. حال در این راه، چهار حالت زیر ممکن است اتفاق بیفتد که نمودار ۱ شمای کلی از انواع این پیشامدها را ارائه می‌دهد. حال در ادامه به بیان رابطه‌ی معیارهای مورد استفاده خواهیم پرداخت.

نمودار ۱: شمای کلی از انواع پیشامدهای ممکن در طبقه بندی دودویی

		پیش‌بینی	
		بد	خوب
واقعی	بد	TP	FN
	خوب	FP	TN

مأخذ: یافته‌های تحقیق

معیارهای ارزیابی مورد استفاده به ترتیب عبارتند از: صحت، حساسیت، تشخیص، دقت، امتیاز F1، ضریب همبستگی متیو، و سطح زیر منحنی مشخصه‌ی عملکرد که روابط آن در ادامه و به همین ترتیبی که بیان شد، ارائه شده است.

$$\text{صحت} = \frac{TP+TN}{TP+FN+TN+FP} \quad (۲۰)$$

$$\text{حساسیت} = \frac{TP}{TP+FN} \quad (۲۱)$$

$$\text{قابلیت تشخیص} = \frac{TN}{TN+FP} \quad (۲۲)$$

$$\text{دقت} = \frac{TP}{TP+FP} \quad (۲۳)$$

$$\text{امتیاز اف ۱} = \frac{2TP}{2TP+FN+FP} \quad (۲۴)$$

$$\text{ضریب همبستگی متیو} = \frac{TP \times TN - FP \times FN}{\sqrt{(TP+FN)(TP+FP)(TN+FP)(TN+FN)}} \quad (۲۵)$$

همچنین منحنی مشخصه‌ی عملکرد، یک نمودار دو بُعدی است که محور طول‌های آن را نرخ FP، و محور عرض‌های آن را نرخ TP تشکیل می‌دهد. سطح زیر این منحنی، معیار مدنظر ما است.

$$\text{FP rate} = \frac{FP}{FP+TN} \quad (۲۶)$$

$$\text{TP rate} = \frac{TP}{TP+FN} \quad (۲۷)$$

در این پژوهش از روش اعتبارسنجی متقاطع ۵ لایه‌ای استفاده شده است که تعداد ۷۵ داده را برای امر آموزش و اعتبارسنجی در نظر گرفتیم که در هر بار اجرای مدل، ۶۰ داده را برای آموزش و ۱۵ داده را برای اعتبارسنجی به کار گرفتیم؛ همچنین ۱۰ داده را برای تست نهایی مدل کنار گذاشتیم تا عملکرد مدل را بر روی داده‌هایی که در هیچ‌جای مدل دیده نشده‌اند، بررسی کنیم.

همچنین به منظور تحلیل حساسیت مدل‌ها، پارامترهای زیر با مقادیر متغیر انتخاب شدند تا بهترین مدل با دقت بیشتری به دست آید:

- انتخاب مقادیر ۱، ۱۰، و ۱۰۰ برای پارامتر C در تابع هدف خطی و غیرخطی اولیه

- انتخاب زوج‌های مرتب (۱و۲)، (۱و۵)، (۱و۱۰)، (۵و۱۰)، (۵و۲۰)، (۱۰و۲۰)، (۱۰و۱۰۰) و (۵۰و۱۰۰) برای (C_1 و C_2)
 - انتخاب زوج‌های مرتب (۰و۲)، (۰و۳)، (۱و۲)، و (۱و۳) برای پارامترهای (c و d) در تابع کرنل چندجمله‌ای
 - انتخاب مقادیر ۰/۱ و ۱ برای پارامتر σ در تابع کرنل پایه شعاعی
 - انتخاب زوج‌های مرتب مقادیر (۰/۱- و ۰/۱)، و (۰/۱- و ۰/۱) برای پارامترهای (a و I) در تابع کرنل سیگموئیدی
- در نهایت بین میانگین نتایج معیارهای ارزیابی عملکرد از آزمون تحلیل واریانس کروسکال-والیس^{۳۲} به منظور فهمیدن وجود اختلاف معنادار بین میانگین مدل‌ها و از آزمون مقایسه زوجی LSD فیشر به منظور دانستن اینکه اختلاف در کدام معیار ارزیابی وجود دارد، استفاده شد.

۴- ارائه‌ی نتایج تحقیق

پس از اجرای مدل‌های ریاضی، حالات مختلف اجرای مدل‌ها از نظر معیار صحت که معیاری جامع و بیانگر درصد پیش‌بینی‌های درست هر مدل است، مرتب شده است و بهترین مدل در هر دسته جهت اجرا بر روی داده‌ها تست انتخاب شدند. آن جایی که هر جایگشت، خود دارای ۵ مرحله اعتبارسنجی بود، عمل پیش‌بینی داده‌های تست، در ۵ سری انجام شد و نتایج این نمونه‌ها برای معیار صحت و حساسیت در قالب جدول ۱ و جدول ۲ ارائه شده است.

نتایج مدل بیانگر آن است که مدل بهینه‌سازی خطی با کرنل پایه شعاعی، در ۵ معیار بهتر از سایر مدل‌ها عمل کرده است. با این وجود پیش‌بینی درست مشتریان بدحساب به مراتب مهم‌تر از مشتریان خوش حساب است؛ لذا ما به دنبال مدلی هستیم که مشتریان بدحساب را نیز به خوبی پیش‌بینی کند. با توجه به اینکه دو معیار صحت و حساسیت این خواسته را برای ما محقق می‌سازند؛ لذا در بین مدل‌های بهینه‌سازی خطی، مدل با کرنل سیگموئیدی دارای بالاترین مقادیر صحت و حساسیت است که مقادیر آن به ترتیب ۸۰٪ و ۱۰۰٪ می‌باشد.

³² Kruskal-Wallis

جدول ۱: مقادیر معیار صحت در بهترین مدل‌های بهینه‌سازی خطی

نام مدل / شماره نمونه	۱	۲	۳	۴	۵
خطی ساده	۰/۵	۰/۵	۰/۶	۰/۴	۰/۷
خطی با کرنل خطی	۰/۵	۰/۷	۰/۶	۰/۵	۰/۶
خطی با کرنل چندجمله‌ای	۰/۸	۰/۶	۰/۵	۰/۴	۰/۷
خطی با کرنل پایه شعاعی	۰/۷	۰/۷	۰/۶	۰/۶	۰/۵
خطی با کرنل سیگموئیدی	۰/۸	۰/۷	۰/۷	۰/۵	۰/۳

مأخذ: یافته‌های تحقیق

جدول ۲: مقادیر معیار حساسیت در بهترین مدل‌های بهینه‌سازی خطی

نام مدل / شماره نمونه	۱	۲	۳	۴	۵
خطی ساده	۰/۲	۰/۲	۰/۲	۰/۲	۰/۶
خطی با کرنل خطی	۰/۸	۰/۴	۱	۰/۸	۰/۸
خطی با کرنل چندجمله‌ای	۰/۸	۰/۸	۰/۴	۰/۴	۱
خطی با کرنل پایه شعاعی	۰/۸	۱	۰/۸	۱	۱
خطی با کرنل سیگموئیدی	۱	۰/۸	۱	۰/۴	۰/۶

مأخذ: یافته‌های تحقیق

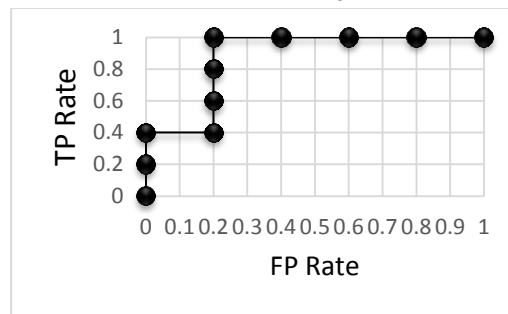
به عبارت دیگر این مدل، همه‌ی مشتریان بدحساب را به‌طور صحیح پیش‌بینی می‌کند و کلیه‌ی مشتریان را با صحت ۸۰٪ طبقه‌بندی می‌کند. ضرایب جریمه این مدل برتر، عبارت است از $C_1 = 1$ و $C_2 = 2$ و همچنین پارامترهای تابع کرنل آن عبارت است از $a = 0.1$ و $r = -0.1$.

همچنین منحنی مشخصه‌ی عملکرد برای این مدل برتر، در ذیل و در قالب نمودار ۲ ارائه شده است. همان‌گونه که ملاحظه می‌شود، مساحت زیر آن 0.88 است. در خصوص معیار AUC هرچه این عدد به ۱ نزدیک‌تر باشد، نتایج پیش‌بینی درست‌تر و قابل اتکاتر است.

همچنین آزمون تحلیل واریانس کروسکال والیس انجام شد که مقدار بحرانی این آزمون برابر با $\chi^2_{\alpha, a-1}$ است که در آن α سطح معنادار بودن، و a تعداد تیمارها (تعداد مدل‌ها) است؛ لذا مقدار بحرانی برابر با $\chi^2_{0.05, 4} = 9.487$ که در صورتی که میزان آماره H از این مقدار بیشتر باشد، فرض صفر رد می‌شود و در غیر این صورت دلیلی بر رد فرض صفر وجود ندارد.

نمودار ۲: بهترین منحنی مشخصه عملکرد میان بهترین مدل‌های خطی (دارای کرنل

سیگموئیدی - مساحت ۰/۸۸)



مأخذ: یافته‌های تحقیق

نتایج تحقیق نشان داد که در بین مدل‌های خطی، فرض برابری میانگین معیارهای حساسیت، تشخیص و سطح زیر منحنی مشخصه عملکرد در سطح معنادار بودن ۰/۰۵ رد می‌شود. این امر نشان‌گر این است که بین میانگین حداقل دو مدل اختلاف معناداری وجود دارد که این خود نشان دهنده‌ی این است که در طبقه‌بندی مشتریان بانکی بهبودی حاصل شده است. در میان مدل‌های غیرخطی نیز تفاوت معنادار بین میانگین معیارهای دقت و امتیاز F1 وجود دارد.

اکنون از آن‌جا که در بین معیارها، معیار حساسیت برای ما از اهمیت بالایی برخوردار است، برای اینکه بدانیم که کدام یک از مدل‌ها در این معیار با یکدیگر اختلاف معنادار دارند، باید آزمون LSD فیشر را برای دو به دو مدل‌ها انجام داد. این آزمون یک آزمون برای مقایسه‌ی مشاهدات زوجی است که شبیه به آزمون t است؛ با این تفاوت که پارامترهای موجود در آماره‌ی آن، از آزمون کروسکال-ولیس گرفته شده است. با توجه به اینکه ۵ مدل داریم، تعداد مقایسه‌های دو به دو، برابر با انتخاب ۲ شی از ۵ شی است که برابر با ۱۰ حالت می‌شود؛ لذا ابتدا آزمون دوطرفه را بین مدل‌های خطی انجام می‌دهیم که نتایج آن در ذیل ارائه شده است. ناحیه‌ی پذیرش این آزمون دو طرفه، برابر با $[-۲/۰۸۶۰, ۲/۰۸۶۰]$ است.

پس از انجام آزمون، اختلاف معناداری بین مدل‌های خطی ساده با کرنل خطی، چندجمله‌ای، پایه شعاعی، و سیگموئیدی ملاحظه می‌شود که پس از آزمون یک طرفه مدل‌های کرنل‌دار، نسبت به مدل خطی ساده، ایجاد بهبود کرده‌اند و به کمک آن می‌توان مشتریان را به درستی تفکیک کرد؛ البته نتایج حاکی از آن است که بین کرنل‌های مختلف، از نظر اختلاف معناداری در معیار حساسیت مشاهده نمی‌شود؛ اگرچه به‌منظور ارائه‌ی یک تابع تصمیم واحد برای طبقه‌بندی مشتریان، گفته شد که کرنل سیگموئیدی بهتر از سایرین عمل کرده است.

پس از اجرای آزمون LSD یک طرفه در سطح معنادار بودن ۰/۰۵، با تعداد درجات آزادی ۲۰، و به تبع آن ناحیه پذیرش $(+\infty, -۱/۷۲۵]$ جدول ۳ حاصل شد که بیانگر برتر بودن مدل‌های دوم نسبت به مدل اول در هر یک از ۴ مورد بود.

جدول ۳: نتایج آزمون LSD یکطرفه برای معیار تشخیص

مدل اول	مدل دوم	آماره	نتیجه آزمون
خطی	کرنل چندجمله‌ای	-۳/۰۳۷۷	دلیلی بر پذیرش فرض صفر وجود ندارد.
خطی	کرنل سیگموئیدی	-۲/۵۷۴۸	دلیلی بر پذیرش فرض صفر وجود ندارد.
چندجمله‌ای	کرنل چندجمله‌ای	-۴/۳۶۸۵	دلیلی بر پذیرش فرض صفر وجود ندارد.
چندجمله‌ای	کرنل سیگموئیدی	-۳/۱۸۲۳	دلیلی بر پذیرش فرض صفر وجود ندارد.

مأخذ: یافته‌های تحقیق

در معرفی مدل پایه‌ای این پژوهش، دو مدل ساده با تابع هدف خطی و غیرخطی ارائه شد. نتایج به دست آمده از اجرای آزمون‌های تحلیل واریانس و آزمون فرض مقایسه‌ی زوجی میانگین معیارهای ارزیابی عملکرد حاکی از آن است که مدل توسعه یافته‌ی ما در این تحقیق، نسبت به مدل پایه‌ای که دارای تابع هدف خطی و بدون هیچ‌گونه افزونه‌ای بود، دارای بهبود در صحت پیش‌بینی بوده؛ اما نسبت به مدل پایه‌ای که دارای تابع هدف غیرخطی بود، بهبودی ایجاد نکرده است. مقادیر دقیق برترین پارامترهای مدل توسعه یافته‌ی ما در بخش بعد ارائه شده است.

۵- جمع‌بندی و نتیجه‌گیری

در این پژوهش به ارائه‌ی مدلی به‌منظور طبقه‌بندی اعتباری مشتریان یک بانک ایرانی پرداخته‌ایم. این مشتریان شامل ۸۵ شرکت موجود در بورس اوراق بهادار تهران‌اند که از یک بانک ایرانی وام گرفته‌اند. نتایج این تحقیق نشان داد که در بین همه‌ی مدل‌های خطی و غیرخطی، و دارای کرنل‌های خطی، چندجمله‌ای، سیگموئیدی و پایه‌شعاعی، مدل با کرنل سیگموئیدی دارای ضرایب جریمه $C_1 = 1$ و $C_2 = 2$ و دارای پارامترهای تابع کرنل $a = 0.01$ و $r = -0.01$ بهترین مدل موجود بود که مشتریان بدحساب را با صحت ۱۰۰٪ و کل مشتریان را با صحت ۸۰٪ برای ما پیش‌بینی می‌کند؛ همچنین جدول ۴ به ما نشان می‌دهد که میانگین دو معیاری که برای ما بیشترین اهمیت را دارند، برای مدل‌های خطی، مربوط به تابع کرنل پایه‌شعاعی است.

جدول ۴: مقادیر میانگین معیارهای صحت و حساسیت برای همه‌ی مدل‌های اجرا شده

مدل/معیار ارزیابی	خطی		غیرخطی	
	صحت	حساسیت	صحت	حساسیت
ساده	٪۵۴	٪۲۸	٪۶۲	٪۵۶
کرنل خطی	٪۵۸	٪۷۶	٪۵۰	٪۸۰
کرنل چندجمله‌ای	٪۶۰	٪۶۸	٪۵۰	٪۴۰
کرنل پایه شعاعی	٪۶۲	٪۹۲	٪۵۰	٪۰
کرنل سیگموئیدی	٪۶۰	٪۷۶	٪۵۰	٪۶۰

مأخذ: یافته‌های تحقیق

نتایج این پژوهش، نتایج حاصل از مقاله‌ی ژانگ و همکاران (۲۰۱۴) را تأیید می‌کند؛ به طوری که مدل بهینه‌سازی با کرنل، فازی‌سازی و ضرایب جریمه باعث بهبود نتایج و برتری آن نسبت به سایر مدل‌ها می‌شود. نتایج پژوهش حاضر نیز حاکی از آن است که توسعه‌ی مدل بر روی مدل خطی ساده تأثیر مثبتی داشته و باعث بهبود آن شده است؛ ولی بر روی مدل غیرخطی ساده نه تنها تأثیر مثبتی نداشته؛ بلکه آن را بدتر هم کرده است.

در پایان، با توجه به مشکلات عمیق سیستم بانکی کشور ناشی از منجمد بودن بخش بزرگی از دارایی‌های بانک‌ها که در قالب تسهیلات معوق، سررسید گذشته و مشکوک‌الوصول است، اهمیت مطالعات کمی و مدل‌سازی ریاضی با دقت بالا برای شناسایی مشتریان بدحساب برای اتخاذ تصمیمات اعطای تسهیلات نمایان‌تر شده است؛ لذا در صورتی که جهت طبقه‌بندی اعتباری مشتریان از مدل اشاره شده در این پژوهش استفاده شود، نتیجه‌ی خوبی حاصل خواهد شد؛ چرا که نارسایی مدل‌های پیشین شامل: داده‌های پرت، یکسان نبودن طبقه‌بندی اشتباه مشتریان و جدایی‌ناپذیری خطی مشتریان را در مدل لحاظ شده است.

فهرست منابع

1. Altman, E. I. (1968). Financial Ratios, Discriminant Analysis and The Prediction of Corporate Bankruptcy. *The Journal of Finance*, 23: 589-609.
2. Atiya, A. F. (2001). Bankruptcy Prediction For Credit Risk Using Neural Networks: A Survey And New Results. *IEEE Transactions on Neural Networks*, 12: 929-935.
3. Cortes, C. & V. Vapnik. (1995). Support-Vector Networks. *Machine Learning*, 20: 273-297.
4. Cristianini, N. & J. Shawe-Taylor. (2000). *An Introduction to Support Vector Machines*. Cambridge University Press Cambridge.
5. Gorzałczany, M.B. & F. Rudziński. (2016). A Multi-Objective Genetic Optimization for Fast, Fuzzy Rule-Based Credit Classification with Balanced Accuracy and Interpretability. *Applied Soft Computing*, 40: 206-220.
6. Hamel, L. (2009). *Knowledge Discovery With Support Vector Machines*, A John Wiley & Sons. Inc., Publication, 75.
7. Harris, T. (2015). Credit Scoring Using The Clustered Support Vector Machine. *Expert Systems with Applications*, 42: 741-750.
8. Khashman, A. (2009). A Neural Network Model for Credit Risk Evaluation. *International Journal of Neural Systems*, 19: 285-294.
9. Khashman, A. (2011). Credit Risk Evaluation Using Neural Networks: Emotional Versus Conventional Models. *Applied Soft Computing*, 11: 5477-5484.
10. Lee, Y. C. (2007). Application of Support Vector Machines to Corporate Credit Rating Prediction. *Expert Systems with Applications*, 33: 67-74.
11. Malhotra, R. & D. K. Malhotra. (2003). Evaluating Consumer Loans Using Neural Networks. *Omega*, 31: 83-96.
12. Min, J. & Y. Lee. (2007). A Practical Approach to Credit Rating. *Journal of Expert Systems With Applications*.
13. Niklis, D., M. Doumpos & C. Zopounidis. (2014). Combining Market and Accounting-Based Models for Credit Scoring Using a Classification Scheme Based on Support Vector Machines. *Applied Mathematics and Computation*, 234: 69-81.
14. Schebesch, K. B. & R. Stecking. (2005). Support Vector Machines For Classifying and Describing Credit Applicants: Detecting Typical and Critical Regions. *Journal of the Operational Research Society*, 56: 1082-1088.
15. Yap, B.W., S.H. Ong & N.H.M. Husain. (2011). Using Data Mining To Improve Assessment of Credit Worthiness Via Credit Scoring Models. *Expert Systems with Applications*, 38:13274-13283.

-
16. Zhang, Z., G. Gao & Y. Shi. (2014). Credit Risk Evaluation Using Multi-Criteria Optimization Classifier with Kernel, Fuzzification And Penalty Factors. *European Journal of Operational Research*, 237: 335-348.

