Archive of SID

فصلنامه پژوهشنامه اقتصادی
سال هجدهم، شماره ۷۱، زمستان ۱۳۹۷، صفحات ۲۲۳- ۱۸۵

ارائه مدلی ترکیبی برای خوشه‌بندی، رتبه‌بندی و پیش‌بینی عملکرد تسهیلات اعطایی موسسات مالی-اعتباری از منظر بازپرداخت بدهی تسهیلات (مورد مطالعه: صندوق کارآفرینی امید)

فرزاد اصغری* و فرید احمدی**

تاریخ دریافت: ۱۳۹۶/۱۰/۱۶ تاریخ پذیرش: ۱۳۹۷/۰۵/۲۷

چکیده

هدف این مقاله ارائه مدلی ترکیبی است تا ضمن ارزیابی عملکرد تسهیلات سیستم بانکی از منظر بازپرداخت بدهی تسهیلات، امکان پیش‌بینی وضعیت اعتباری متقاضیان تسهیلات را فراهم اورد.در این راستا در ابتدا با اتخاذ رویکرد مدیریت اعطای تسهیلات توسط بانک‌ها به خوشه‌بندی و رتبه‌بندی ۱۰۰۲۲۴ فقره از تسهیلات صندوق کارآفرینی امید پرداخته شده است. تمامی اطلاعات مربوط به تسهیلات اعطایی به مشتریان فوق از نرم‌افزار بانکداری متمرکز صندوق استخراج شده است و با اتکا به این مجموعه داده کمی ارزشمند و دارای روایی بالا از روش‌های کیفی برای گردآوری داده‌ها استفاده نشده است. در این مقاله از روش تحلیل عاملی «رب پی سی ای» برای طبقه‌بندی و از الگوریتم دو مرحله‌ای «کی-مینز» برای خوشه بندی استفاده می‌شود. همچنین غیر از روش‌های خوشه‌بندی اشاره شده از روش «سی سی ار» نیز برای ارزیابی عملکرد تسهیلات صندوق استفاده شده است. در ادامه با هدف ایجاد زمینه پیش‌بینی وضعیت اعتباری متقاضیان پیش از اعطای تسهیلات به ارائه مدلی برای پیش‌بینی اعتبار با استفاده از دو الگوریتم ماشین بردار پشتیبان و الگوریتم ترکیبی فرا ابتکاری شبکه عصبی-ژنتیک پرداخته شده است. نتایج به‌دست آمده از پیش‌بینی وضعیت اعتباری متقاضیان تسهیلات نشان می‌دهد که مدل به‌دست آمده از روش ترکیبی شبکه عصبی-ژنتیک با میانگین مربعات خطا ۰/۲۳ و ضریب تعیین ۷۸ درصد از صحت پیش‌بینی بیشتری در مقایسه با مدل ماشین بردار پشتیبان برخوردار است. بنابراین، مدل ارائه شده برای پیش‌بینی وضعیت اعتباری در این مقاله، می‌تواند پیش‌بینی به نسبت مناسبی از عملکرد متقاضیان تسهیلات داشته باشد. روشی جدید که در قالب یک نرم‌افزار داده کاوی امکان پیش‌بینی اعتبار متقاضیان از منظر بازپرداخت بدهی تسهیلات را برای موسسات مالی-اعتباری فراهم می‌آورد.

طبقه‌بندی JEL: *G17,G21,C53,C61,C81,C88*
کلید واژه‌ها: *خوشه‌بندی مشتریان، رتبه‌بندی مشتریان، پیش‌بینی وضعیت اعتباری متقاضیان تسهیلات، صندوق کارآفرینی امید، تحلیل عاملی رب پی سی ای، الگوریتم دو مرحله‌ای «کی-مینز».*

---

* کارشناس ارشد، مهندسی فناوری اطلاعات، دانشگاه صنعتی ارومیه
** دانشیار گروه مهندسی فناوری اطلاعات و کامپیوتر دانشگاه صنعتی ارومیه، نویسنده مسئول، پست الکترونیک : f.ahmadi@uut.ac.ir

www.SID.ir

Archive of SID

۱- مقدمه

بانک‌ها، نهادهایی هستند که با مشتریان تعامل مستقیم داشته و عنصر مشتری در این سازمان‌ها اهمیت ویژه‌ای دارد. بدیهی است جهت پیشرو بودن در عرصه رقابت توجه به مشتری و اختصاص خدمات ویژه با تشخیص نیازمندهای آن‌ها و ارائه خدمات درست به آن‌ها محقق خواهد شد (اسلامی، ۱۳۸۷). تحلیل منابع با ارزش داده‌ای در رابطه با مشتریان کنونی بانک‌ها و مشتریان بالقوه همچنین شناخت مشتریان، ایجاد تمایز بین گروه‌های مختلف مشتریان و رتبه‌بندی آن‌ها یکی از مسائل مهم در بانک‌ها و سازمان‌های مشتری‌محور است؛ برای رسیدن به درکی صحیح از مشتریان، سازمان‌ها نیازمند استفاده از مقیاسی هستند که از طریق آن‌ها بتوان میزان اهمیت مشتریان مختلف را سنجش کرد (هشامیان و همکاران، ۱۳۹۱).

یکی از ابزارهای اصلی نظام پولی و بانکی برای توسعه و ایجاد رونق اقتصادی در هر کشور اعطای تسهیلات هدفمند از طریق سیستم بانکی است. این تسهیلات در بخش‌های مختلف اقتصادی شامل تولیدی، صنعتی، کشاورزی، بازرگانی و خدماتی به متقاضیان عرضه می‌شود. در حقیقت عدم شناسایی دقیق و نظارت ناکافی در حین و بعد از فرایند اعطای تسهیلات می‌تواند ضمن هدر رفت منابع، آثار نامناسبی مانند رشد نامتوازن اقتصادی و همچنین نقدینگی در کشور ایجاد کند. بنابراین، بانک‌ها می‌کوشند از طریق شناسایی دقیق جامعه هدف و طبقه‌بندی آن‌ها به اعطای تسهیلات اثربخش اقدام کنند. علاوه بر این، بانک‌ها همچون کسب‌وکارهای دیگر در طول حیات خود با ریسک‌هایی مواجه می‌شوند. یکی از مهم‌ترین آن‌ها ریسک اعتباری است که باید با آن مقابله کنند. اعتبارسنجی به‌عنوان یک تکنیک موثر به شناخت مشتریان خوب و بد پرداخته و با این کار می‌تواند ریسک اعتباری آن‌ها را تعیین کند. تحقیقات زیادی روی مدل‌های اعتبارسنجی در بانک‌ها صورت گرفته است. در ابتدا مدل‌های اعتبارسنجی به‌صورت قضاوتی بودند. سپس روش‌های پارامتریک در اعتبارسنجی مطرح شدند. به تازگی از روش‌های ناپارامتریک نیز در اعتبارسنجی مشتریان بانک‌ها استفاده می‌شود. در این مقاله سعی شده است تا با استفاده از روش داده‌کاوی دانش نهفته در اطلاعات موجود در پایگاه داده‌های حاصل از تعاملات بانک و مشتری را استخراج کرده و از آن در جهت اخذ راهبردهای مدیریت ارتباط با مشتری بهره‌گیری کنیم.

www.SID.ir

Archive of SID

برای ارزیابی کارایی مدل ارائه شده در این مقاله از داده‌های صندوق کارآفرینی امید بهره گرفته شده است. ماموریت صندوق کارآفرینی امید تجهیز سپرده‌های خرد مردم و تلفیق آن با اعتبارات دولتی، شناسایی افراد واجد شرایط، توانمندسازی و تامین مالی ایشان، شبکه‌سازی کسب‌وکارهای بازارمحور و نظارت بر حسن انجام فعالیت‌های دریافت‌کنندگان تسهیلات است (صندوق کارآفرینی امید، ۱۳۹۵). مشتریان صندوق کارآفرین امید توسط نظر خبرگان این صندوق و فقط با استفاده از چند معیار محدود انتخاب می‌شوند. استفاده از الگوریتم‌های داده‌کاوی مانند قواعد همبستگی و استخراج قوانینی جهت شناسایی دقیق اقشار هدف صندوق می‌تواند معیار خوبی جهت تکمیل نظر خبرگان در جهت انتخاب این دسته از مشتریان باشد. بنابراین، در این پژوهش قصد داریم با استفاده از داده‌های صندوق کارآفرینی امید، الگویی جهت خوشه‌بندی و رتبه‌بندی مشتریان صندوق با استفاده از روش‌های تحلیل عاملی قدرتمند، تحلیل پوششی داده‌ها و الگوریتم خوشه‌بندی دومرحله‌ای «کی-مینز» عرضه کنیم. پیش‌بینی وضعیت اعتباری مشتریان صندوق نیز با استفاده از دو الگوریتم یادگیری زبان ماشین و شبکه عصبی پرسپترون چند لایه مورد تحلیل قرار می‌گیرد.

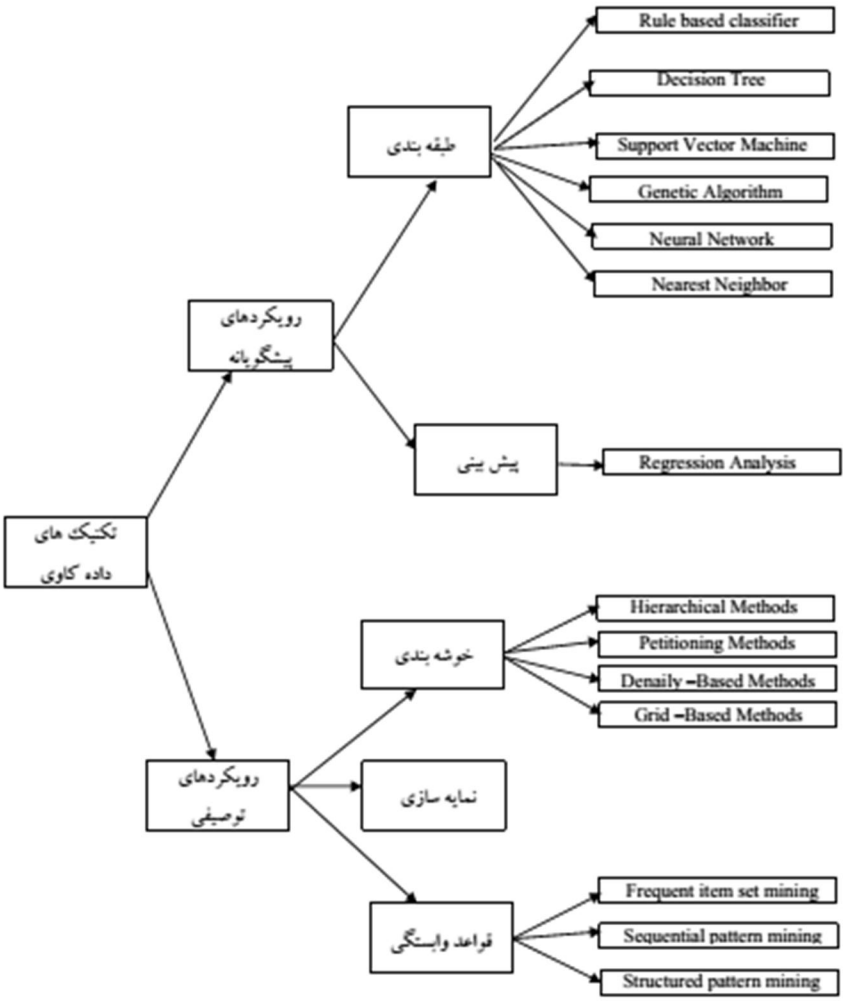
۲- داده‌کاوی در نظام بانکی

در این بخش از مقاله ابتدا به روش‌های داده‌کاوی پرداخته و سپس به پیشینه آن در سیستم بانکداری می‌پردازیم. روش‌های داده‌کاوی در یک تقسیم‌بندی کلی به دودسته روش‌های توصیفی و روش‌های پیش بینانه تقسیم می‌شوند. روش‌های توصیفی به خواص عمومی داده‌ها می‌پردازند. به این ترتیب که الگوهایی قابل تفسیر توسط انسان از داده‌ها استخراج می‌کنند. در رویکرد پیش‌بینانه، هدف پیش‌بینی رفتارهای آینده است. در این روش‌ها از چند متغیر جهت پیش‌بینی مقادیر آینده استفاده می‌شود. تقسیم‌بندی برخی روش‌های داده‌کاوی در نمودار (۱) نشان داده شده است.

www.SID.ir

Archive of SID

نمودار (۱): دسته‌بندی کلی عملکردهای داده‌کاوی(کوزا، ۲۰۱۰)



صنعت بانکداری در جهان تحت تغییرات شدید در طریقه انجام کسب‌وکار هستند. بانک‌های پیشرو از ابزارهای داده‌کاوی برای بخش‌بندی، تعیین سودمندی، دسته‌بندی اعتبار، پیش‌بینی قصور در بازپرداخت‌ها، بازاریابی، تشخیص تراکنش‌های متقلبانه و... استفاده می‌کنند. این ابزار به‌عنوان ابزار رقابتی در بانک شناخته شده است.

www.SID.ir

Archive of SID

۲-۱- مطالعات داخلی

تقوی‌فرد و خواجه‌فرد (۱۳۹۳)، مطالعه‌ای را روی بخش‌بندی خوشه‌ای مشتریان بانک صادرات ایران انجام دادند. آنان با استفاده از اطلاعات مربوط به ۳۰ هزار مشتری بانک صادرات ایران در بازه یک‌ساله ۱۳۸۹ تا ۱۳۹۰ بر اساس تحلیل «ار بی اف» و با استفاده از الگوریتم دو بخشی به خوشه‌بندی مشتریان بانک پرداختند. همچنین آنان به رتبه‌بندی مشتریان بانک با استفاده از الگوریتم C5.0 k نیز پرداختند. نتایج حاکی از کشف ویژگی‌های چهار دسته «مشتریان طلایی»، «مشتریان ارزشمند و وفادار»، «مشتریان با احتمال رویگردانی» و «مشتریان کم‌ارزش» شد.

آجرلو (۱۳۹۲)، پایان‌نامه‌ای با عنوان الگویی جهت تعیین ارزش چرخه عمر مشتریان ارائه داده است. در این تحقیق با استفاده از اطلاعات تراکنشی مربوط به پنج هزار مشتری حقیقی و حقوقی بانک ملت مدلی برای سنجش و اندازه‌گیری ارزش طول عمر مشتری ارائه شده است.

عیسی‌زاده و عریانی (۱۳۹۲) با استفاده از روش نمونه‌گیری خوشه‌ای از بانک کشاورزی تهران، ۲۸۶ شرکت تسهیلات‌گیرنده را مورد بررسی قرار دادند. آنان با استخراج داده‌های نامناسب تنها ۷۵ شرکتی که از فروش اقساطی ۲۴ ماهه تا پایان اردیبهشت ۱۳۹۱ تسهیلات دریافت کرده بودند برای تجزیه و تحلی نهایی انتخاب کردند. در نتیجه کارایی فنی محاسبه و شرکت‌ها رتبه‌بندی شدندکه ۱۵ شرکت آن‌ها دارای بازدهی بالا و مناسب بودند.

امیری (۱۳۹۲)، پایان‌نامه‌ای در راستای بررسی رفتار مشتریان در میزان استقبال از کانال‌های ارتباطی نوین بانکداری الکترونیکی در بانک صادرات ارائه داده است که در آن به تحلیل رفتار مشتری در استفاده از این ابزارها پرداخته‌شده است که با استفاده از تکنیک داده‌کاوی، بخش‌بندی روی مشتریان استفاده‌کننده از این خدمات انجام‌شده است.

امیری (۱۳۹۱) در پایان‌نامه‌ای با عنوان مدیریت دانش مشتریان بانک مهر اقتصاد با استفاده از تکنیک‌های داده‌کاوی به بخش‌بندی مشتریان بانک مهر اقتصاد با هدف کشف ویژگی‌های رفتاری مشابه برای کمک به مدیران این بانک جهت تسهیل اتخاذ سیاست‌های متناسب جهت ارتباط با مشتریان و در نهایت سودآوری برای بانک پرداخت. نتایج حاکی از تقسیم شدن مشتریان با بیشترین شباهت در الگوی رفتاری در پنج گروه مختلف است.

www.SID.ir

Archive of SID

حسینی (۱۳۹۰) از تکنیک‌های هوشمند داده‌کاوی جهت مدیریت ارتباط با مشتریان بانک پارسیان بهره گرفته است. او در پایان‌نامه خود با ترکیب دو دسته از متغیرها، ۳۰ متغیر برای بررسی الگوی رفتاری مشتریان شناسایی کرده و با استفاده از شبکه عصبی کوهنن ماتریس نرون‌های خروجی شبکه ۷ * ۵ را تشخیص داده و مشتریان را در ۱۷ خوشه تقسیم کرد. در ادامه این تحقیق، ارائه سرویس‌های بانکی مناسب به مشتریان بالقوه با استفاده تکنیک‌های داده‌کاوی را مورد بررسی قرارداد.

حسینی و همکاران (۱۳۸۹) در مقاله‌ای به تحلیل اهمیت عملکرد ویژگی‌های خدمت بر پایه بخش‌بندی مشترکان بانک مسکن در استان یزد پرداخته‌اند و برای انجام آن از نقشه‌های خودسازمان‌ده و برای محاسبه اهمیت ویژگی‌ها در هر بخش از شبکه عصبی پیش‌خور چندلایه استفاده شده است و مشترکان بانک مسکن در استان یزد را در سه بخش دسته‌بندی کرده‌اند.

آذر و همکاران (۱۳۸۹) در مطالعه‌ای به بررسی منابع صندوق‌های بانکی پرداختند. آنان برای انجام پژوهش خود جهت پیش‌بینی بازدهی وام‌هایی که به شخص ثالث به‌عنوان وام پرداخت‌شده است از الگوریتم‌های فرااابتکاری شبکه عصبی مصنوعی و ژنتیک استفاده کردند.

البرزی و همکاران (۱۳۸۹) در مقاله‌ای با عنوان به‌کارگیری الگوریتم ژنتیک در بهینه‌سازی درختان تصمیم‌گیری برای اعتبارسنجی مشتریان بانک‌ها به ارائه مدلی برای اعتبارسنجی مشتریان به‌منظور اعطای تسهیلات اعتباری پرداخته‌اند. در ساخت این مدل، فرایند توسعه در شناخت الگو و فرایند «سی ار ای اس پی»[1] به‌کار رفته است.

۲-۲- مطالعات خارجی

تحقیق دیگر در زمینه بانکداری، مطالعه‌ای است که در سال ۲۰۱۵ توسط هنلی، کاپلن و بیسر انجام شده است که در آن مدلی برای محاسبه ارزش چرخه عمر مشتری ارائه شده است. اساس کار در این تحقیق بخش‌بندی مشتریان به گروه‌های متجانس براساس چهار متغیر سن، نوع زندگی، خدمات بانکی مورداستفاده توسط مشتری و سطح‌فعال بودن مشتری است. این تحقیق براساس ترکیبی از مدل زنجیره مارکوف و تحلیل کارت است.

1- CRISP

www.SID.ir

Archive of SID

مطالعه‌ای نیز توسط کومار و همکارانش در سال ۲۰۱۴ انجام شده است که در آن شناسایی مشتریان وفادار براساس مفاهیمی چون طول دوره ارتباط مشتری با سازمان، فراوانی خرید مشتریان و متغیرهای «ار اف ام»[1] مربوط به مشتریان انجام شده است. براساس نتایج حاصل از این مطالعه، ۲۰ درصد مشتریان برتر ازلحاظ «سی ال وی»[2] ۹۵ درصد سود را خلق کرده‌اند.

امین و همکارانش (۲۰۱۳) در پژوهشی به ارزیابی خوشه‌بندی رفتار مشتریان بانک در شرق مالزی پرداختند. آنان با استفاده از تحقیقات تجربی و فاکتورهای سن، جنس، قومیت، درآمد ماهیانه و دین به خوشه‌بندی رفتار مشتریانی که وام‌های اسلامی دریافت کردند، پرداختند. نتایج کار حاکی از تقسیم شدن مشتریان به هفت گروه مختلف است.

هانوکل و همکارانش (۲۰۱۳) در مطالعه‌ای با استفاده از الگوریتم فراابتکاری ژنتیک به پیش‌بینی برنامه‌های تجاری در بانک‌ها پرداختند. روش کار آنان به این صورت بود که از الگوریتم پیش‌بینی ژنتیک در دو مرحله پیش‌بینی و یادگیری استفاده کردند. به این‌صورت که در مرحله، پیش‌بینی پارامترهای دامنه مساله را تخمین زده و در مرحله یادگیری، ارتباط الگوهای داده به دست می‌آمد و نتایج این الگوریتم در دو حوزه مالی: بانک‌های سپرده بازرگانی و پیش‌بینی ورشکستگی بررسی می‌شود.

یانینگ و همکارانش (۲۰۱۲) در پژوهشی به‌کارگیری رویکرد داده‌کاوی در پیاده‌سازی سیستم مدیریت با مشتریان را موردتوجه قرارداده‌اند و مزایای بهره‌گیری از داده‌کاوی را از دیدگاه کسب‌وکار در پذیرش مشتریان بانکداری اینترنتی در قالب یک مدل مفهومی بیان کردند. این مطالعه بیشتر به تحلیل داده‌های کارت‌های اعتباری می‌پردازد.

فنگ و ما (۲۰۰۹) درخصوص کاربرد تکنولوژی داده‌کاوی در مدیریت ارتباط با مشتری، بیان کردند که داده‌کاوی نقش مهمی در بهبود سطح و کارایی مدیریت ارتباط با مشتری در بانک‌های تجاری بازی می‌کند. این مقاله بر کاربرد تکنولوژی داده‌کاوی در مدیریت ارتباط با مشتری بانک‌های اقتصادی تاکید داشته و درنهایت به این نتیجه می‌رسند که تنها با یکپارچه‌سازی سیستم‌های مبتنی بر داده‌کاوی، سیستم‌های بانکداری و

1- RFM
2- CLV

www.SID.ir

Archive of SID

سیستم‌های پشتیبانی تصمیم‌گیری، می‌توان سطح خدمات و سوددهی بانک‌ها را بیشتر از گذشته ارتقا داد.

لی و همکاران (۲۰۱۰)، یک روش ترکیبی جدید برای انتخاب داده‌های ورودی دسته‌بندی مشتریان بانک جهت ارزیابی اعتبار آنان معرفی کرده‌اند. در این مقاله رویکردی ترکیبی که حاصل ترکیب رویکردهای انتخاب ویژگی موجود (آنالیز ترکیبی خطی، تئوری مجموعه‌های ناهموار، درخت تصمیم و روش اف - اسکور[1]) با روش ماشین بردار پشتیبان است برای انتخاب متغیرها معرفی‌شده و در ادامه مقایسه‌ای بین رویکردهای ترکیبی بر پایه اس وی ام[2] به‌منظور انتخاب ویژگی‌ها صورت گرفته است. در این تحقیق از داده‌های معروف مربوط به کارت‌های اعتباری UCI مربوط به کشور آلمان و استرالیا استفاده شد. تمرکز مقاله بر یافتن مرتبط‌ترین ویژگی‌ها با اهداف مدنظر است و بیان می‌شود که مساله مهم نه فقط کاهش حجم متغیرهای ورودی، بلکه حذف نویزهای ورودی نیز است.

بوسنیاک (۲۰۱۰)، ایجاد یک ارتباط محکم و بلندمدت و موثر با مشتریان را هسته اصلی مدیریت ارتباط با مشتری و درک درست از مشتری را اساس افزایش ارزش طول عمر مشتری (CLV) دانسته است. در این مقاله، هدف از بخش‌بندی مشتری، ایجاد گروه‌های سودآور و در حال رشد بر پایه ویژگی‌های مشترک آن‌ها است که سازمان را قادر می‌سازد تا هر گروه را با پیشنهادهایی خاص مورد هدف قرار دهد. مقاله بر بخش‌بندی اعتبار مشتریان در صنعت بانکداری تمرکز دارد و در مطالعه موردی خود از شبکه‌های عصبی چندلایه با بازخورد روبه‌جلو برای بخش‌بندی مشتریان استفاده کرده است.

چن و همکاران (۲۰۱۰) به دسته‌بندی مشتریان جهت کشف تقلب و سواستفاده‌های مالی و شناسایی بعضی از عوامل کلیدی برای تشخیص تقلب خریداران و فروشندگان در معاملات آنلاین توسط کارت اعتباری در بانک پرداخته‌اند. تکنیک مورداستفاده در این تحقیق رگرسیون لاجیت بوده که برخی از ویژگی‌های مربوط به داده‌های اولیه ورودی آن مبلغ اولیه پیشنهادی، مدت‌زمان انجام مزایده، سود مزایده، میزان اعتبار فروشنده و قیمت

1- F-Score
2- SVM

www.SID.ir

Archive of SID

بهایی مزایده بوده است که اعتبار فروشنده به‌عنوان عامل مهم در تشخیص تقلب استفاده‌شده است.

نای و همکاران (۲۰۰۹) از الگوریتم‌های داده‌کاوی جهت ساخت مدلی به‌منظور پیش‌بینی روی‌گردانی مشتریانی که از کارت اعتباری استفاده می‌کنند، استفاده‌شده است. این تحقیق روی پایگاه داده مشتریان یک بانک چینی صورت گرفته و از چهار دسته متغیر اطلاعات مشتری، اطلاعات کارت اعتباری، داده‌های مربوط به ریسک مشتری و اطلاعات مربوط به تراکنش‌ها استفاده شده که درمجموع شامل ۱۳۵ متغیر است. از بین این متغیرها ۹۵ متغیر با توجه به همبستگی بین آن‌ها برای انجام مراحل بعدی انتخاب شدند. در این تحقیق الگوریتم‌های رگرسیون و درخت تصمیم جهت دسته‌بندی مشتریان انتخاب شده و نتایج تحقیق نشان داد که عملکرد مدل‌های رگرسیونی کمی بهتر از عملکرد درخت‌های تصمیم بوده است.

زای و همکارانش (۲۰۰۹) از یک روش یادگیری جدید به نام جنگل‌های تصادفی متوازن بهبودیافته[1] جهت تعدیل مشکل بحث‌برانگیز عدم توازن در توزیع داده‌ها در تحلیل روی‌گردانی مشتری استفاده‌شده است. در این مطالعه اثربخشی رویکرد جنگل تصادفی استاندارد و روش‌های مختلف نمونه‌برداری نیز در پیش‌بینی روی‌گردانی مشتری نیز بررسی‌شده و با عملکرد «ای بی ار اف» مورد مقایسه قرار گرفت.

لین و همکارانش (۲۰۱۱) از تئوری مجموعه‌های ناهموار که یک روش تصمیم‌گیری براساس قوانین است برای استخراج قوانین مرتبط با روی‌گردانی مشتری به‌کار می‌گیرند. سپس از گراف شبکه جریان که یک روش رویکرد وابسته به مسیر است برای کشف قوانین و متغیرهای تصمیم‌گیری استفاده می‌کنند و درنهایت ارتباط بین قوانین و انواع مختلف روگردانی را نشان می‌دهد. این تحقیق روی نمونه‌ای ۲۱ هزار نفری از مشتریان صورت گرفته و آن‌ها را به سه دسته مشتریان پایدار، روی‌گردانی داوطلبانه و روی‌گردان غیرداوطلبانه تقسیم می‌کند. متغیرهای ورودی این تحقیق را متغیرهای جمعیت شناختی، روان‌شناختی و تراکنش‌های مربوط به این گروه از مشتریان تشکیل می‌دهند.

سار و یاگیل(۲۰۱۵) در مطالعه خود از منحنی بازدهی دولتی و گسترش اعتباری برای پیش‌بینی پارامترهای اقتصاد کلان مانند نرخ رشد استفاده کردند. آنان نیز کاربرد دیگری

1- IBRF

www.SID.ir

Archive of SID

برای پیش‌بینی عوامل ورشکستگی بانکداری و شرکت‌های بیمه با استفاده از روش پیش‌بینی فرا ابتکاری نشان می‌دهد.

پوربا و همکارانش در سال (۲۰۱۴) با استفاده از شبکه عصبی مصنوعی و الگوریتم‌های ژنتیک به پیش‌بینی بازدهی وام‌هایی پرداخته اند.

۳- روش تحقیق

با وجود بسط و توسعه کارایی و روش‌های محاسبه آن در دو دهه اخیر تاکنون کمتر مطالعه‌ای در مورد خوشه‌بندی و رتبه‌بندی تسهیلات با الگوریتم‌های خوشه‌بندی به همراه پیش‌بینی وضعیت اعتباری متقاضیان تسهیلات پیش از اعطای تسهیلات با استفاده از الگوریتم‌های ترکیبی فراابتکاری انجام گرفته است. از این رو، در این مقاله به خوشه‌بندی و رتبه‌بندی تسهیلات اعطایی با استفاده از روش الگوریتم‌های خوشه‌بندی و فراابتکاری براساس شاخص‌های بازدهی و روی حجم بسیار بزرگی از داده‌های تسهیلات (بیش از ۱۰۰ هزار مورد تسهیلات) صندوق کارآفرینی امید پرداخته شده است. تمام اطلاعات مربوط به تسهیلات اعطایی به مشتریان فوق از نرم‌افزار بانکداری متمرکز صندوق استخراج شده است و با اتکا به این مجموعه داده کمی ارزشمند و دارای روایی بالا از روش‌های کیفی برای گردآوری داده‌ها استفاده نشده است. این مقاله هم از حیث مدل ارائه شده و جامعه مورد مطالعه و حجم داده‌های بررسی شده نوآور است.

روش انجام تحقیق در این مقاله عبارت بوده است از:

۱- تحقیق و مطالعه میدانی در مورد وضعیت تسهیلات اعطایی صندوق کارآفرینی امید

۲- ارائه مدل خوشه‌بندی و رتبه‌بندی تسهیلات صندوق

۳- انتخاب متغیرها؛ پیش پردازش و نرمال‌سازی داده‌ها

۴- پیاده‌سازی الگوریتم‌های شناسایی‌شده جهت خوشه‌بندی و رتبه‌بندی تسهیلات براساس داده‌های واقعی صندوق کارآفرینی امید

۵- ارزیابی عملکرد تسهیلاتی صندوق کارآفرینی امید براساس نتایج مدل خوشه‌بندی و رتبه‌بندی

www.SID.ir

Archive of SID

۶- ارائه مدل پیش‌بینی وضعیت اعتباری متقاضیان تسهیلات صندوق با استفاده از دو الگوریتم ماشین بردار پشتیبان و الگوریتم ترکیبی فرا ابتکاری شبکه عصبی-ژنتیک و بررسی صحت پیش‌بینی مدل ارائه شده.

کارایی مدل پیش‌بینی: مدل پیش‌بینی کارا به مدلی اطلاق می‌شود که دارای بیشترین صحت در پیش‌بینی باشد. برای ارزیابی صحت مدل پیش‌بینی، اقدام به نمونه‌گیری استاندارد (براساس فرمول کوکران) از مجموعه کل داده‌های تسهیلات کرده و پس از تقسیم‌بندی آن به داده‌های آموزش و تست (آزمون)، مدل پیش‌بینی روی مجموعه داده‌های نمونه، پیاده سازی می‌شود. در ادامه از طریق محاسبه شاخص‌های میانگین مربعات خطا و ضریب تعیین به ارزیابی صحت مدل پیش‌بینی پرداخته می‌شود.

تحلیل پوششی داده‌ها: DEA یا تحلیل پوششی داده‌ها از روش‌های برنامه‌ریزی ناپارامتریک است که در سال‌های اخیر اهمیت زیادی پیدا کرده است. مدل «تحلیل بهره‌وری چندعامله» است که برای اندازه‌گیری عملکرد دسته واحدهای تصمیم‌گیرنده همگن به‌کار می‌رود. از این روش برای ارزیابی عملکرد مدارس، بیمارستان‌ها، بانک‌ها و نیز تخصیص منابع می‌توان استفاده کرد. تحلیل پوششی واحدها، بدون نیاز به ایجاد یک رابطه تابعی خاص میان عوامل تولید و میزان محصول تولید شده به ارزیابی کارایی واحدها نسبت به یکدیگر می‌پردازد. این ویژگی از آن، ابزاری بسیار ساده و مفید ساخته است. اولین مقاله تحلیل پوششی داده‌ها در این سال توسط چارنز وکوپر به چاپ رسید (چارنز و همکاران،۱۹۷۸) و مدل ارائه شده در آن به مدل CCR[1] معروف شد که در این مقاله از این مدل برای ارزیابی عملکرد صندوق کارآفرینی امید بهره‌گیری شده است.

۳-۱- انتخاب متغیرها

شناسایی متغیرها: از گام‌های اساسی در مسیر تحقیق شناسایی متغیرها است. با در نظر گرفتن اهداف تحقیق مطالعه تحقیقات انجام شده در حوزه بخش‌بندی رفتار مشتریان و بررسی رویکردهای سازمانی صندوق کارآفرینی امید، سه دسته متغیر زیر انتخاب شده است:

1- Charnes, Cooper and Rhodes

www.SID.ir

Archive of SID

- داده‌های جمعیت آماری مشتریان: این متغیرها شامل جنسیت، سن، تحصیلات، وضعیت تاهل، شغل و تاریخ تولد، محل منطقه بانکی و... است که در این بخش متغیرهای مورد استفاده در مقاله عبارتند از: سطح سواد مشتریان و نام بخش اقتصادی (شغل) مشتریان.
- پرداخت بدهی تسهیلات توسط مشتریان: پرداخت بدهی تسهیلات مشتری می‌تواند شاخص مناسبی برای بررسی الگوهای رفتاری مشتری باشد و تحقیقات زیادی در این زمینه صورت گرفته است. در این میان متغیرهایی چون تکرار، متغیر پولی، ارزش پولی در مطالعات داده‌کاوی کاربرد گسترده‌ای یافته‌اند. این متغیرها در پایگاه اطلاعاتی بانک وجود دارند و از طریق پردازش داده‌های تراکنشی مشتریان محاسبه می‌شوند. ازاین‌رو، متغیرهای مورد استفاده در این بخش عبارتند از: مقدار مبلغ مصوب برای مشتری، تعداد پرداختی‌های مشتری از شروع قرارداد تا تاریخ مطالعه، سود بانکی و ارزش ترهیمی وثایق پرونده.
- دانش ضمنی خبرگان بانکی: صندوق کارآفرینی امید برخی از مشتریان خود را به‌عنوان مشتریان کلیدی می‌شناسد. مشتریان کلیدی بانکی در هر استان بنا بر نظر خبرگان بانکی با ویژگی‌های خاصی تعریف می‌شوند. منشا این ویژگی‌ها دانش ضمنی خبرگان بانکی در هر استان است که این روش کیفی مبتلا به ضعف‌های خاص خود است.

۳-۲- آماده‌سازی و پیش‌پردازش داده‌ها

در این پژوهش برای جمع‌آوری داده‌ها از داده‌های موجود در پایگاه اطلاعاتی صندوق کارآفرینی امید استفاده‌شده است. به‌طورمعمول ویژگی نمونه‌های مورد مطالعه در یک تحقیق داده‌کاوی دارای بازه تغییرات یکسان نیستند. برای مثال، اگر سن و متغیر پولی را به‌عنوان دو ویژگی یک مشتری در نظر بگیریم، این ویژگی‌ها دارای بازه تغییرات متفاوتی هستند. همچنین مقادیر به‌دست‌آمده برای هر ویژگی نیز به واحد استفاده شده برای اندازه‌گیری آن ویژگی ارتباط مستقیم دارد. بنابراین، برای رفع این مشکل از نرمال‌سازی به روش کمینه- بیشینه[1] استفاده‌شده است.

1- Max-Min

www.SID.ir

Archive of SID

۳-۳- نرمال‌سازی داده‌ها

در این روش با یک تبدیل خطی یا غیرخطی، داده‌ها را در بازه‌ای که توسط کاربر انتخاب می‌شود، قرار می‌گیرند. این بازه در کاربردهای داده‌کاوی معمولا [۱، ۱-] یا [۱، ۰] است. برای قرارگیری داده‌ها در بازه بین صفر و یک می‌توان از رابطه (۱) استفاده کرد که در آن X نشان‌دهنده مقدار متغیر است.

$$X^* = \frac{X - Min(X)}{Range(X)} \qquad (۱)$$

۳-٤- خوشه‌بندی

خوشه‌بندی یا گروه‌بندی، تقسیم اقلام موجود در یک مجموعه داده است که به‌طور طبیعی با هم شباهت دارند. داده‌هایی که با این معیار به‌صورت خوشه‌هایی تفکیک می‌شوند با داده‌های موجود در خوشه‌ای که در آن قرار می‌گیرند، بیشترین شباهت را دارند و یا با داده‌های موجود در سایر خوشه‌ها متفاوتند.

در خوشه‌بندی موضوعات زیر مورد توجه قرار می‌گیرد:

- چه تعداد از خوشه‌ها می‌تواند دانش نهفته درداده‌ها را کشف کند؟ مساله تعداد خوشه‌ها معمولا به‌صورت جداگانه مورد بررسی قرار می‌گیرد.
- معیارهای شباهت و تفاوت داده‌ها چیست؟ این معیارها خود به‌واسطه روش‌های مختلفی محاسبه می‌شود، اما در بیشتر روش‌های خوشه‌بندی موجود از معیار فاصله فضایی دو داده از یکدیگر چون فاصله اقلیدسی، فاصله مینکوفسکی و یا فاصله مانهاتان محاسبه می‌شود.
- بعد از تعیین تعداد خوشه‌ها و معیار شباهت و یا فاصله داده‌ها، داده‌ها با استفاده از چه روشی در تعداد خوشه‌های معین جای گیرند. خوشه‌بندی یک الگوریتم بدون ناظر در داده‌کاوی است، زیرا هیچ صفت منفردی برای هدایت فرآیند Training استفاده نمی‌شود و همه صفات ورودی ارزش یکسان دارند. تفاوت این روش با طبقه‌بندی در این است که در طبقه‌بندی هر داده یک طبقه (کلاس) از پیش مشخص‌شده تخصیص می‌یابد، اما در خوشه‌بندی هیچ اطلاعی از کلاس‌های موجود درون داده‌ها وجود ندارد. به بیان دیگر، خود خوشه‌ها نیز از داده‌ها استخراج می‌شود. با مطالعه و بررسی روش‌های داده‌کاوی

www.SID.ir

Archive of SID

موجود جهت خوشه‌بندی، رتبه‌بندی و کارایی داده‌ها، الگوریتم‌های «رب پی سی ای»، دو مرحله‌ای «کی-مینز» و الگوریتم DEA برای اعمال روی مجموعه داده‌ها انتخاب‌شده‌اند.

٤- تحلیل و نتایج

در این بخش مدل مفهومی مورد استفاده در این مقاله در نمودار (۲)، نشان داده شده است. دیاگرام ارائه شده شامل سه نوع ارزیابی خوشه‌بندی و رتبه‌بندی مشتریان است که هر یک به صورت زیر توصیف می‌شود:

٤-۱- خوشه‌بندی به روش «رب پی سی ای»[1]

مشتریان به روش تحلیل عاملی «رب پی سی ای» از اثرات آن در خروجی‌ها مورد ارزیابی قرار می‌گیرد. روش رب پی سی ای(Robust Principal Component Analysis) ارائه شده وزن هایی بین صفر و یک برای هر مشاهده ارائه می‌دهد که به صورت مستقیم نشان‌دهنده اثرات مشاهدات روی بردار میانگین و ماتریس کواریانس دارد. همچنین «رب پی سی ای» اطلاعات مفیدی در رابطه به کلاس‌بندی داده‌ها ارائه می‌دهد. در خوشه‌بندی «کی-مینز» امتیازها یا مقادیر ویژه ارائه شده توسط «رب پی سی ای» با ترکیب با الگوریتم خوشه‌بندی «کی-مینز» به صورت دومرحله‌ای خوشه‌بندی می‌شود.

برای خوشه‌بندی و شناسایی مشتریان صندوق از دیاگرام ترکیبی نشان داده در نمودار (۲) استفاده شده است. دیاگرام شامل کلاس‌بندی تک مرحله‌ای از طریق تحلیل عامل قدرتمند «رب پی سی ای» و خوشه‌بندی دو مرحله‌ای است. در روش «رب پی سی ای» اطلاعات مشتریان با ابعاد بالا به متغیرهای کوچک‌تری تجزیه می‌شوند که این مولفه‌ها ترکیب خطی از متغیرهای اصلی هستند. هدف اصلی PCA قدرتمند یا «رب پی سی ای» به‌دست آوردن مولفه‌هایی است که از لایه خارجی داده‌ها تاثیر نمی‌پذیرند که ماتریس کواریانس با ماتریس تخمین کواریانس قدرتمند جایگزین شده است. به این ترتیب در تمام روش‌های «رب پی سی ای» هدف پیدا کردن ترکیب خطی از متغیرها حتی برای داده‌های لایه خارجی و نوع و تعداد داده‌های لایه خارجی است. در این مقاله از روش «رب پی سی ای» با در نظر گرفتن هفت متغیر اصلی مساله شامل: ۱- مقدار مبلغ مصوب

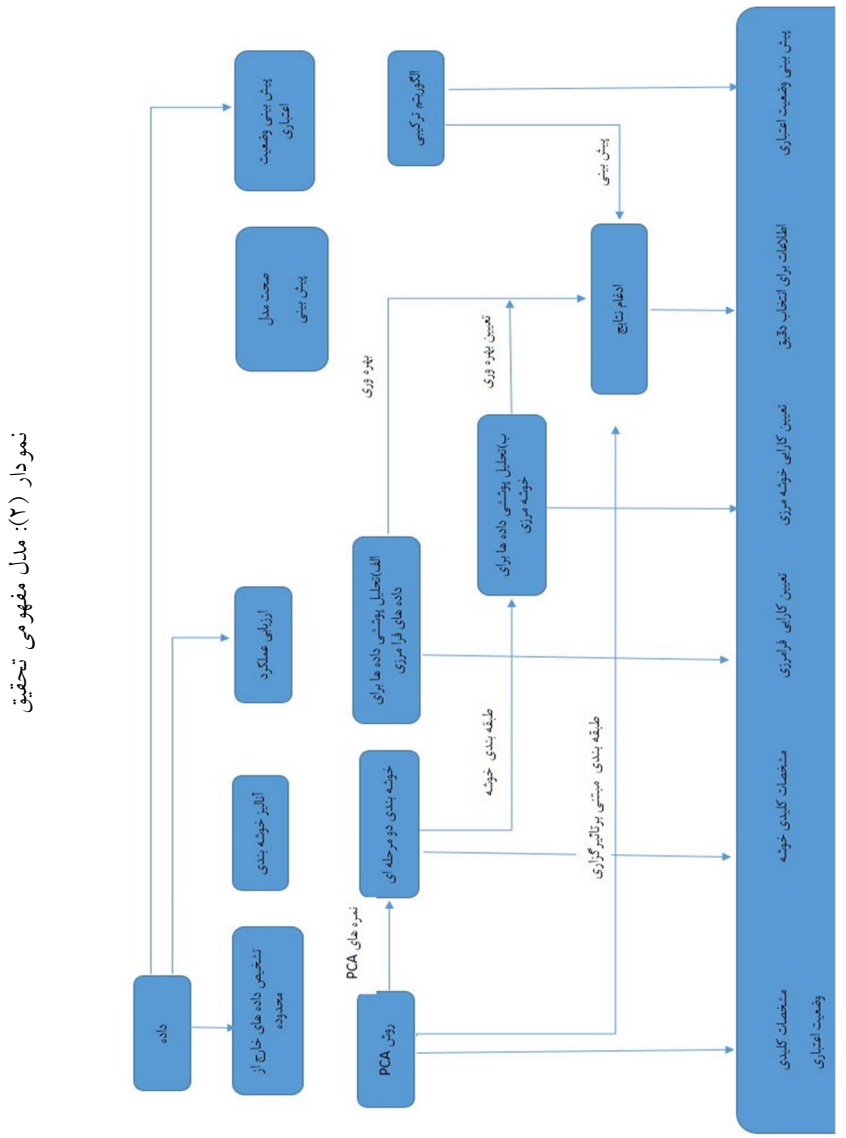
[1] Robust Principal Component Analysis

www.SID.ir

Archive of SID

برای مشتری، ۲- تعداد پرداختی‌های مشتری از شروع قرار داد تا تاریخ مطالعه، ۳- سود بانکی، ۴- ارزش ترهینی وثایق پرونده، ۵- سطح سواد مشتریان، ۶- وضعیت حساب و ۷- نام بخش اقتصادی استفاده شده است.

نمودار (۲): مدل مفهومی تحقیق



از بین متغیرهای اصلی مساله سطح سواد مشتریان، وضعیت حساب و نام بخش اقتصادی متغیرهای کیفی هستند که به صورت کمی به شرح جدول (۱)، (۲) و (۳) کدگذاری شده‌اند.

www.SID.ir

Archive of SID

جدول (۱): کدگذاری سطح سواد مشتریان

سطح سواد	کد
بی‌سواد	۱
زیر دیپلم	۲
دیپلم	۳
لیسانس	٤
فوق لیسانس	٥
دکتری	٦
فوق دکتری	۷
کاردانی	۸
فقه	۹
نامعلوم	۱۰

جدول (۲): کدگذاری وضعیت حساب مشتریان

وضعیت	کد
فعال	۱
معوق	۲
مشکوک‌الوصول	۳
تسویه شده	٤
سررسید گذشته	٥
آماده آزادسازی	٦
نامعلوم	۷

جدول (۳): کدگذاری بخش اقتصادی مشتریان

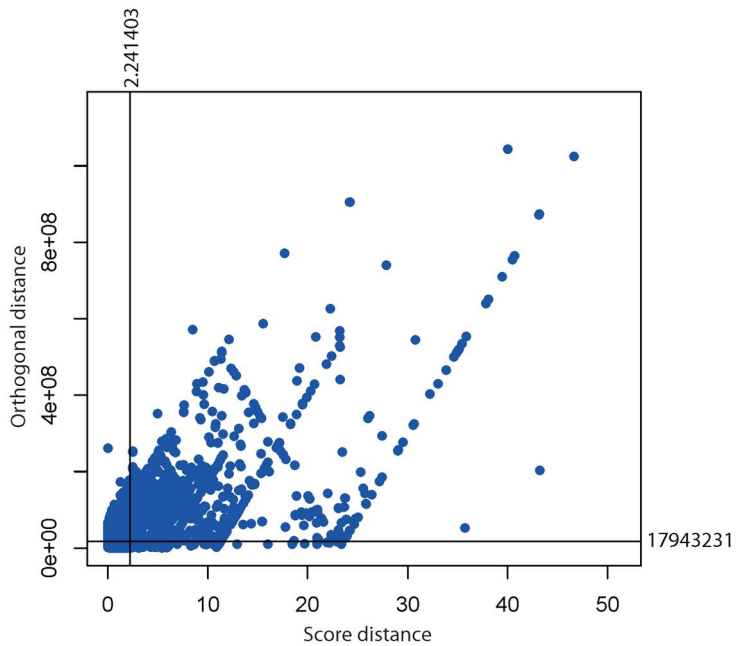
بخش اقتصادی	کد
کشاورزی	۱
تعمیر مسکن	۲
ساخت مسکن	۳
صنعتی	٤
دیگر	٥
مردمی	٦
خدمات	۷
رفاهی	۸

www.SID.ir

Archive of SID

تعداد مولفه‌های k مورد استفاده در «رب پی سی ای» با استفاده از معیار هابرت[1] (۲۰۰۵) با توجه به فاصله مقادیر ویژه اول و دوم تعیین می‌شود. در این مقاله با توجه به فاصله زیاد بین مقدار ویژه اول و دوم $k = ۱$ انتخاب شده است که نتیجه آن تبدیل فضای هفت بعدی از متغیرهای مساله به فضای تک بعدی است که برابر بردار ویژه مشاهدات است. از بردار ویژه مشاهدات به عنوان رتبه برای خوشه‌بندی در مرحله بعد استفاده می‌شود. با توجه به دو معیار فاصله رتبه SD و فاصله عمود OD نمودار تشخیص در نمودار (۳) نشان داده شده است.

نمودار (۳): تشخیص برای مشتریان بانک



با توجه به نمودار (۳)، مشاهدات مساله شامل اطلاعات ۱۰۰۲۲۴ مشتری است که به دو متغیر فاصله رتبه SD و فاصله عمود OD رسم شده است. این نمودار مشتریان را به چهار لایه افراز کرده است که سه لایه شامل اهرم بد (راست بالا)، اهرم خوب (راست پایین) و اهرم عمود (چپ بالا) به عنوان لایه‌های خارجی و افراز مشاهدات منظم (چپ پایین) به عنوان لایه داخلی شناخته شده‌اند. دو خط برش افقی و عمودی به ترتیب با مقادیر

1- Hubert

www.SID.ir

Archive of SID

۲/۲۴۱۴۰۳ و ۱۷۹۴۳۲۳۱ این چهار بخش را از هم مجزا کرده است. دو جدول (۴)و (۵) میانگین متغیرهای مساله برای هر اهرم را نشان می‌دهد. با توجه به جدول (۵)، ۷۳,۲۲۹ نفر از مشتریان شامل ۷۳ درصد در لایه داخلی قرار دارند که به عنوان مشتری منظم محسوب می‌شوند. همچنین ۲۷ درصد از مشتریان در لایه خارجی قرار دارند که ۱۷ درصد از آن‌ها به صورت عمود بر فضای «رب پی سی ای» قرار دارند که به صورت مستقیم با مشاهده داده‌های قابل تشخیص نیستند. همچنین ۲ و ۷ درصد از مشاهدات به ترتیب در دو لایه خارجی اهرم خوب و بد قراردارند.

جدول (٤): میانگین و انحراف معیار متغیرها

	میانگین	انحراف معیار
مبلغ مصوب	۸۹٦٤٥۸٤۲	۹۲۹٤۰٥۳٦
تعداد پرداخت	۲۰/۸٥۱٦۲۲۳۷	۱٥/۲٥٤
سود	۷٤٦۹۸۰٤/۷٦	۱۲۷۲۷۲۳۸
رهن	۱۹۳٤۳۱٥۳٤/٦	۲٤۰۸۱۱۳٥۰
سواد	٦/٤٤۷٤۱۸	۳/٤٥٥۸
حساب	۲/٦۹۷۰۳۹	۱/۷۷۹۷
اقتصادی	٦/۷٤٦۲۸۸	۰/۰۰٦۱۹

جدول (۵): درصد میانگین (میانگین هر خوشه بر میانگین کل) متغیرها برای هر اهرم

انواع هرم	منظم	عمود	خوب	بد
تعداد	۰/۷۳۱	۰/۱۷٤	۰/۰۲٥	۰/۰۷۰
مبلغ مصوب	۰/٦٥٤	۱/۲۷۹	۲/۸۷۱	۳/۲٤۰
تعداد پرداخت	۱/۱٥۷	۰/٦۱٦	۰/٦۲۸	۰/٤٤۹
سود	۰/٦٦۹	۱/۰۸۹	۳/٥٥۸	۳/۳۱۸
رهن	۰/٦۸۹	۰/۸۹۳	۳/٤۱۳	۳/٦٤۸
سواد	۰/۹۹٥	۱/۰۳۱	۰/۹٤۲	۰/۹۹۹
حساب	۱/۱۰٥	۰/٦۹۸	۰/٦۹۰	۰/۷٦۲
اقتصادی	۱/۰۳۸	۰/۹۲٥	۰/۸۸۸	۰/۸۳۱

www.SID.ir

Archive of SID

میانگین مبلغ مصوب برای اهرم منظم یا مشتریان عادی بر اساس دو جدول (۵) و (۶) با ۷۳ درصد فراوانی ۶۵ درصد میانگین مصوب کل وام دریافت شده بانک است. ۷۴ درصد از مشتریان به طور تقریبا مساوی دارای حساب فعال یا تسویه شده هستند. از لحاظ اقتصادی اکثر (۵۹/۸۴ درصد) وام‌ها در مراکز رفاهی دریافت شده و همچنین سطح سواد اکثر مشتریان ثبت نشده است.

مشاهدات اهرم بد همان طور که در جدول (۷) نشان داده شده است، مشاهداتی هستند که میانگین مبلغ مصوبشان ۳/۲ برابر میانگین مبلغ مصوب کل است. همچنین نتایج کلاس‌بندی نشان می‌دهد که ۷ درصد از مشتریان ۵ برابر بیشتر از مبلغ وام مشتریان عادی یا منظم شامل ۷۳ درصد کل مشتریان را شامل می‌شود. همچنین با توجه به جدول، میانگین پرداختی‌های مشتریان اهرم بد کمترین مقدار را نسبت به دیگر اهرم‌ها دارند به طوری که میانگین پرداختی اهرم بد ۳۸ درصد میانگین پرداختی مشتریان عادی را دارند. بیشتر مشتریان اهرم بد از لحاظ اقتصادی در بیشتر(۵۱ درصد) شامل مراکز خدماتی می‌شود. ۵۸/۱۵ درصد مشتریان اهرم بد، حساب فعال داشته و سطح سواد نامعلومی دارند.

جدول (٦): سطح سواد، وضعیت حساب و بخش اقتصادی مشتریان منظم

سطح سواد	درصد افراد	وضعیت حساب	درصد افراد	بخش اقتصادی	درصد افراد
بی‌سواد	۰/۰۹	فعال	۳۹/۰۸	کشاورزی	۵/٦۱
زیر دیپلم	١٤/٦٩	معوق	۵/۱۲	تعمیر مسکن	۰/۳۸
دیپلم	۲۱/٤۸	مشکوک‌الوصول	۲/۰۹	ساخت مسکن	۱/۱۲
لیسانس	۱۱/٤٤	تسویه شده	۳۵/۰٤	صنعتی	۵/۲۲
فوق لیسانس	۰/۲۳	سررسید گذشته	۹/٦٦	دیگر	۱/۳۳
دکتری	۰/۰۱	آماده آزادسازی	۹	مردمی	۱/٤۲
فوق دکتری	۰/۰۰٤	نامعلوم	۰	خدمات	۲۹/۱۰
کاردانی	۵/۱۰	--	--	رفاهی	۵۹/۸٤
فقه	۰/۰۱	--	--	--	--
نامعلوم	٤٤/۹۱	--	--	--	--

www.SID.ir

Archive of SID

جدول (۷): سطح سواد، وضعیت حساب و بخش اقتصادی مشتریان اهرم بد

سطح سواد بد	درصد افراد	وضعیت حساب	درصد افراد	بخش اقتصادی	درصد افراد
بی‌سواد	۰/۰۱	فعال	۵۸/۱۵	کشاورزی	۱۰/۱۹
زیر دیپلم	۱۲/٤۲	معوق	۲/۵۲	تعمیر مسکن	۰
دیپلم	۲۱/۵۲	مشکوک‌الوصول	۰/۵۲	ساخت مسکن	۰
لیسانس	۱۲/۹۸	تسویه شده	۲۰/۲۲	صنعتی	۲۸/٤۸
فوق لیسانس	۳/۲٤	سررسید گذشته	۹/۱۹	دیگر	۰/۹۲
دکتری	۰/۵۵	آماده آزادسازی	۰/۸٤	مردمی	۰/۰۱
فوق دکتری	۰	نامعلوم	۰/۰۸	خدمات	۵۱/۰۸
کاردانی	٤/۷۸	--	--	رفاهی	۹/۳۳
فقه	۰	--	--	--	--
نامعلوم	٤٤/۵۱	--	--	--	--

مشتریان اهرم عمود در فضایی عمود بر فضای تک بعدی ($k=۱$) قرار دارند، یعنی مشتریانی هستند که اگر فضای مساله دو بعدی ($k=۲$) شود، می‌توانند عضو مشتریان منظم باشند. مشتریان اهرم عمود دو برابر میانگین مصوب مشتریان عادی وام دریافت کرده‌اند که میانگین برگشت وام آن‌ها ۵۳ درصد مشتریان منظم است. ۷۱/۷۷ درصد از مشتریان اهرم عمود دارای حساب فعال هستند که از لحاظ اقتصادی به طور تقریبا مساوی بین مراکز خدماتی، رفاهی و صنعتی تقسیم شده است. ویژگی‌های مشتریان اهرم عمود در جدول (۸) نشان داده شده است. ۲ درصد از مشتریان در اهرم خوب –همان‌طور که در جدول (۹) اشاره شده است– مبلغ دریافتی آن‌ها ۴/۳۹ برابر مبلغ میانگین مشتریان عادی است. همچنین از لحاظ برگشت پول، ۵۸ درصد کمتر از مشتریان عادی وام پرداخت کرده‌اند. در این اهرم، ۷۴/۴۶ درصد افراد حساب فعال داشته و بیشتر در مرکز خدماتی سرمایه‌گذاری شده است.

www.SID.ir

Archive of SID

جدول (۸): سطح سواد، وضعیت حساب و بخش اقتصادی مشتریان اهرم عمود

سطح سواد عمود	درصد افراد	وضعیت حساب	درصد افراد	بخش اقتصادی	درصد افراد
بی‌سواد	۰/۰۷	فعال	۷۱/۷۷	کشاورزی	۵/۹۶
زیر دیپلم	۱۳/۶۴	معوق	۳/۳۷	تعمیر مسکن	۰/۰۲
دیپلم	۱۸/۰۲	مشکوک‌الوصول	۱/۹۶	ساخت مسکن	۰/۱۳
لیسانس	۱۲/۰۸	تسویه شده	۱۳/۲۳	صنعتی	۲۳/۲۵
فوق لیسانس	۳/۳۱	سررسید گذشته	۷/۰۷	دیگر	۳/۶۵
دکتری	۰/۳۴	آماده آزادسازی	۲/۲۴	مردمی	۰/۱۴
فوق دکتری	۰	نامعلوم	۰/۰۳	خدمات	۲۸/۸۹
کاردانی	۴/۴۷	--	--	رفاهی	۳۷/۹۵
فقه	۰/۰۰۵	--	--	--	--
نامعلوم	۴۸/۰۶	--	--	--	--

جدول (۹): سطح سواد، وضعیت حساب و بخش اقتصادی مشتریان اهرم خوب

سطح سواد خوب	درصد افراد	وضعیت حساب	درصد افراد	بخش اقتصادی	درصد افراد
بی‌سواد	۰/۰۸	فعال	۷۴/۴۶	کشاورزی	۹/۴۹
زیر دیپلم	۱۴/۹۲	معوق	۳/۷۳	تعمیر مسکن	۰
دیپلم	۲۴/۸۵	مشکوک‌الوصول	۰/۶۹	ساخت مسکن	۰/۰۴
لیسانس	۱۲/۵۷	تسویه شده	۴/۵۰	صنعتی	۱۵/۱۲
فوق لیسانس	۱/۹۵	سررسید گذشته	۱۵/۱۲	دیگر	۰/۲۸
دکتری	۰/۵۷	آماده آزادسازی	۱/۵۰	مردمی	۰
فوق دکتری	۰	نامعلوم	۰	خدمات	۷۲/۷۶
کاردانی	۵/۸۸	--	--	رفاهی	۲/۳۱
فقه	۰	--	--	--	--
نامعلوم	۳۹/۲۴	--	--	--	--

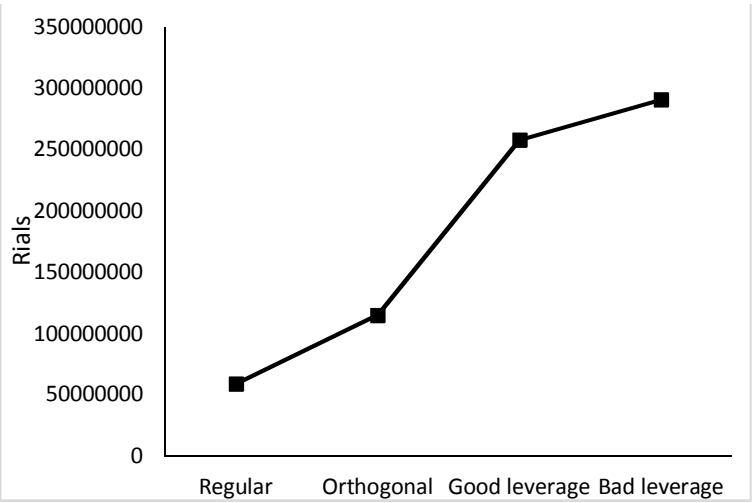
با توجه به نمودارهای مبالغ دریافتی و برگشتی مطابق با نمودارهای (۴) و (۵) به ترتیب اهرم منظم، اهرم عمود، اهرم خوب و اهرم بد بیشترین وام را دریافت کرده اند همچنین با افزایش مبلغ مصوب و حرکت به سمت اهرم بد تعداد قسط‌های برگشتی کاهش می‌یابد.

www.SID.ir

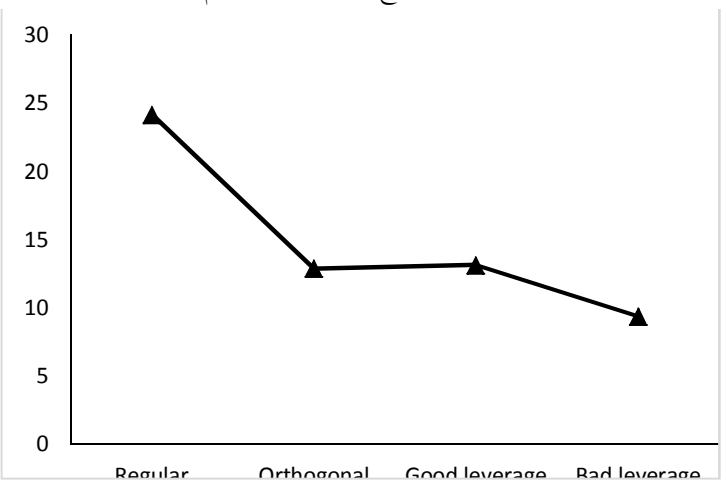
Archive of SID

اگر مشتریان خوش حساب مشتریانی تعریف شود که بیشترین برگشت پول و کمترین وام را دریافت کرده اند در این صورت می‌توان رتبه‌بندی به ترتیب اهرم منظم، اهرم عمود، اهرم خوب و اهرم بد برای خوش حسابی مشتریان انجام داد.

نمودار (٤): تعداد پرداختی‌های هر اهرم



نمودار (٥): مبلغ مصوب در هر اهرم



www.SID.ir

Archive of SID

٤-۲- خوشه‌بندی براساس الگوریتم دو مرحله‌ای کی مینز

در این قسمت از تحقیق از الگوریتم دو مرحله‌ای «کی-مینز» برای خوشه‌بندی مشتریان استفاده شده است. برای این منظور ضریبی از رتبه‌های «رب پی سی ای» به عنوان ورودی خوشه‌بندی به جای داده‌های اصلی استفاده شده است. این پارامتر به صورت رابطه (۲) تعریف می‌شود.

$$y = \text{Eigen vector} \times \sqrt{\text{Eigen value}} \qquad (۲)$$

در این مساله با توجه به اینکه از یک مولفه در «رب پی سی ای» استفاده شده است، مساله شامل یک بردار ویژه و یک مقدار ویژه است. پس از خوشه‌بندی به روش «کی-مینز» به چهار خوشه نتایج به این شرح است: با توجه به جدول (۱۰)، خوشه اول شامل ۳۸ (۰/۰۳ درصد) نفر از مشتریان است که ۱۱/۳۹ برابر میانگین کل وام (یک میلیارد ریال) دریافت کرده‌اند که می‌توانند از مشتریان با وام بالا محسوب شوند. این خوشه از مشتریان از بین خوشه‌های دیگر کمترین پرداخت را داشته‌اند. ۷۹ درصد از این مشتریان حسابی فعال دارند که ۶۸ درصد از این خوشه در بخش صنعتی سرمایه‌گذاری شده‌اند. خوشه چهارم شامل ۹۲/۸ درصد از فراوانی مشتریان است که میانگین وام ۷۴/۲ میلیون ریالی دریافت کرده‌اند. میزان برگشت پول برای این خوشه از مشتریان برابر میانگین تعداد پرداختی‌ها است. با توجه به نمودارهای (۶) و (۷) به تدریج از خوشه اول به چهارم با کاهش وام مصوب، تعداد قسط‌های برگشتی پول افزایش یافته است.

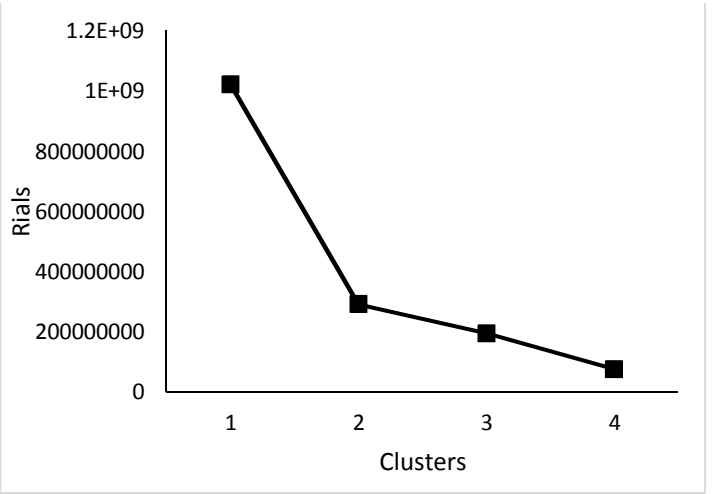
جدول (۱۰): میانگین متغیرها برای هر خوشه

خوشه	۱	۲	۳	٤
تعداد	۳۸	٦۷۷۷	۳۸۲	۹۳۰۲۷
مبلغ مصوب	۱۰۲۱۰۵۲٦۳۲	۲۹۰۵۱٤٦٤۸	۱۹۳٦٦۵۵٦۸/۱	۷٤۲۰٤۹۸۱/۹۱
تعداد پرداخت	۵/۷٦۳۱۵۷۸۹۵	۹/٦۳٦۸۵۹۹٦۸	۱۱/۷۵۳۹۹۲٦۷	۲۱/۷۱۲۱۳۷۳٤
سود	۹۰۰۸۲۷٦۷/۱٦	۲٦۸۳۷۱۳۰/۲۹	۱٦۸۱٤٤۳۸/۱٤	۵۹۸٦۷۸۰/۳۹٤
رهن (میلیون)	٤/۲۸	۰/۸	۰/٤۲	۰/۱۵
سواد	٦/٦۸٤۲۱۱	٦/۰۲۲۵۷٦	٤/۹۸٦۹۱۱	٦/٤۸٤۲٦۸
حساب	۱/٦۸٤۲۱۱	۲/۱۳۱۳۲۷	۱/۹۷۱۲۰٤	۲/۷٤۱٦٤۵
اقتصادی	٤/٦۳۱۵۷۹	۵/۵۵۲۳۰۹	۵/٤۵۰۲٦۲	٦/۸۳۹٤۵۵

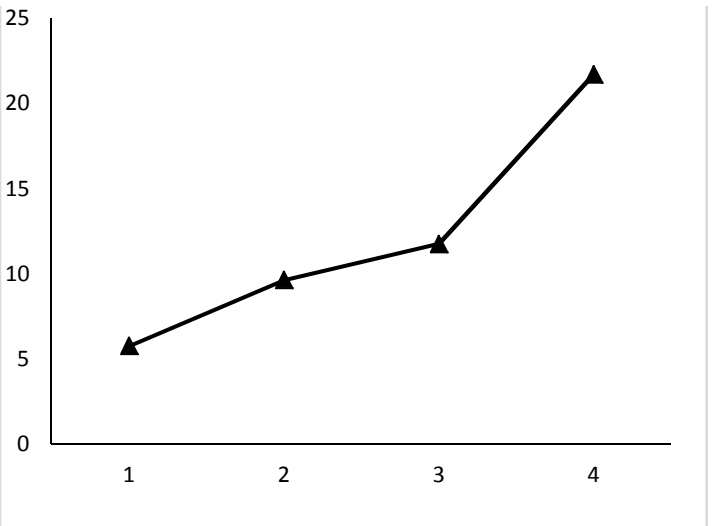
www.SID.ir

Archive of SID

نمودار (٦): مبلغ مصوب در هر خوشه



نمودار (۷): تعداد پرداختی‌ها هر خوشه



اگر مشتریان هم از لحاظ خوشه‌بندی و هم از لحاظ کلاس‌بندی «رب پی سی ای» مجزا شود، مشتریان به ۱۶ بخش افراز می‌شوند. با توجه بـه جـدول (۱۱)، خوشـه اول از مشـتریان که بیشترین وام را دریافت کرده‌اند، همه عضو اهـرم بـد محسـوب مـی‌شـوند. همچنـین دو

www.SID.ir

Archive of SID

خوشه ۲ و ۳ نیز عضو لایه خارجی داده‌ها هستند. همه مشتریان منظم و اهرم عمود عضو خوشه چهارم هستند. نتایج خوشه‌بندی و «رب پی سی ای» تقریبا نتایج مشابه را نشان می‌دهد که همدیگر را تایید می‌کنند.

جدول (۱۱): تعداد مشتریان هر خوشه و هر اهرم به صورت مجزا

خوشه‌ها	اهرم منظم	اهرم عمود	اهرم خوب	اهرم بد	جمع
۱	۰	۰	۰	۳۸	۳۸
۲	۰	۰	۱۸۵٦	٤۹۲۱	٦۷۷۷
۳	۰	۰	۹۰	۲۹۲	۳۸۲
٤	۷۳۲۲۹	۱۷٤٦٤	۵۲۱	۱۸۱۳	۹۳۰۲۷
جمع	۷۳۲۲۹	۱۷٤٦٤	۲٤٦۷	۷۰٦٤	۱۰۰۲۲٤

با توجه به جدول (۱۱) می‌توان مشتریان را به روش ترکیبی به ۱۶ بخش افراز کرد که برای هر بخش سیاست مورد نظر در آن اعمال شود. برای مثال، مشتریان خوشه چهارم و اهرم منظم مشتریان خوش حساب هستند که بیشترین فراوانی از افراد هستند که بیشترین تعداد اقساط را پرداخت کرده‌اند. این افراد که بیشتر در بخش رفاهی سرمایه‌گذاری شده‌اند از اهمیت خاصی برخوردارند، زیرا می‌توانند معرف رفتار کلی صندوق باشند. همچنین مشتریان اهرم عمود در خوشه چهارم مشتریانی با تعداد بیشتر از دیگر اهرم‌های لایه خارجی دارند که رفتاری متفاوت از مشتریان منظم دارند. این مشتریان می‌توانند اگر به تعداد قابل توجهی برسند یکی از محورهای اصلی با رفتاری متفاوت محسوب شوند.

٤-۳- تحلیل پوششی داده‌ها

در این قسمت با استفاده از روش تحلیل پوششی داده‌ها مدل CCR به بررسی عملکرد تسهیلات صندوق کارآفرینی امید پرداخته می‌شود. برای این منظور با استفاده از روش آماری کوکران (رابطه (۳)) به نمونه‌گیری از تعداد کل تسهیلات (۱۰۰۲۲۴ مشتری) پرداخته شد و نمونه مورد بررسی تعداد ۳۸۳ براساس آزمون کوکران به دست آمد که در آن، N حجم جمعیت آماری، z درصد خطای معیار ضریب قابل اطمینان، $q = (۱-p)$ نسبتی از جمعیت فاقد صفت معین و d درجه اطمینان با دقت احتمالی مطلوب است.

www.SID.ir

Archive of SID

$$n = \frac{\frac{z^2 pq}{d^2}}{1 + \frac{1}{N}\left(\frac{z^2 pq}{d^2} - 1\right)} \qquad (۳)$$

سپس از ۶ متغیر ورودی و ۳ متغیر خروجی استفاده شد که متغیرهای ورودی استفاده شده عبارتند از : ۱- مقدار مبلغ مصوب برای مشتری، ۲- بخش اقتصادی (مبتنی بر بخش‌بندی جدول (۳))، ۳- ارزش ترهینی وثایق پرونده، ۴- سطح سواد مشتریان، ۵- نوع تسهیلات (مبتنی بر بخش‌های اشتغال، ضروری، ازدواج، مسکن، وجوه مردمی، وجوه دولتی، تفاهمنام ای و زیربخش‌های هر بخش) و ۶- زمان درخواست تسهیلات توسط مشتری براساس سال پایه ۱۳۸۸ و متغیرهای خروجی استفاده شده در این بخش عبارتند از: ۱- وضعیت حساب، ۲- تعداد پرداختی‌های مشتری از شروع قرارداد تا تاریخ مطالعه و ۳- سود بانکی.

نتایج به‌دست‌آمده از شناسایی عملکرد کل مشتریان با ۳۸۳ مشتری نمونه گرفته با استفاده از روش تحلیل پوششی داده‌ها مدل CCR نشان می‌دهد که ۳۱۱ مشتری ناکارا (۸۱/۱ درصد) و ۷۲ مشتری کارا (۱۸/۸۰ درصد) هستند. جدول (۱۲) نتایج میانگین و انحراف معیار مشتریان کارا در کل داده‌ها را نشان می‌دهد.

جدول ۱۲: نتایج آمارایی مشتریان کارا در کل داده‌ها

ورودی‌ها						
	مبلغ وام	ارزش وثایق	سطح سواد	بخش اقتصادی	زمان درخواست	نوع تسهیلات
Mean	۸۹۲۵۶۹۴۴	۱۸۵۳۰۹۴۴۴	۱/۸۴۷۲	۶/۳۱۹۴	۱۳۹۲/۸۴۷۲	۲۸۱/۶۳۸۹
Std. Dev.	۱۲۴۴۲۰۹۹۶	۲۵۴۶۲۰۸۷۱	۱/۲۹۶۳۶	۲/۲۳۱۸۲	۵۵۳/۴۵۱۴۲	۱۹۷/۹۶۹۲۲
خروجی‌ها						
	تعداد پرداختی		سود		وضعیت حساب	
Mean	۲۹/۲۲۲۲		۹۱۸۲۷۶۷		۵/۰۶۹۴	
Std. Dev.	۱۵/۱۸۶۳		۱۵۱۸۵۹۴۳		۱/۷۴۶۸۴	

ماخذ داده: مجموعه داده تسهیلات صندوق کارآفرینی امید (۱۰۰۲۲۴ مورد تسهیلات پرداختی)

www.SID.ir

Archive of SID

با توجه به نتایج به‌دست آمده از مقایسات آماری در ورودی‌ها و خروجی‌ها به‌کار رفته در این بخش، مشخص است که به دلیل انحراف معیار بالا در همه‌ی ورودی‌ها و خروجی‌ها، یک ناهمگنی بین الگوهای مشتریان وجود داشته است که به این منظور نمی‌توان یک مقایسه بین مشتریان کارا و ناکارا انجام داد. بنابراین، در ادامه مدل ورودی- خروجی CCR برای هر خوشه استفاده‌شده است. مدل‌های معمول تحلیل پوششی داده‌ها یا از مدل ورودی محور (کاهش اندازه ورودی‌ها بدون تغییر در خروجی‌ها) و یا مدل خروجی‌محور (کاهش اندازه خروجی‌ها بدون تغییر در اندازه ورودی‌ها) بهره می‌گیرند. در حالی که در این مقاله از روش کامل‌تر ورودی- خروجی‌محور به مفهوم کاهش اندازه ورودی‌ها به همراه کاهش هم زمان اندازه خروجی‌ها استفاده شده است. در این رویکرد، مقایسه کارایی واحدهای تصمیم‌گیرنده از حجم تبدیل ورودی‌های چندگانه به خروجی‌های چندگانه (تفاضل ترکیب وزنی ورودی‌های هر واحد از ترکیب وزنی خروجی‌های آن واحد) صورت می‌پذیرد. جدول (۱۳) نتایج به‌دست‌آمده برای مشتریان هر خوشه استفاده شده است.

جدول (۱۳): محاسبه کارایی و ناکارایی مشتریان در هر خوشه

	خوشه اول	خوشه دوم	خوشه سوم	خوشه چهارم	مجموع
کارا	۱۲	۳۸	۲٤	۲۸	۱۰۲
ناکارا	۲٦	۳۲٦	۳۵۸	۳۲۵	۱۰۳۵
مجموع	۳۸	۳٦٤	۳۸۲	۳۵۳	۱۱۳۷

ماخذ داده: مجموعه داده تسهیلات صندوق کارآفرینی امید (۱۰۰۲۲٤ مورد تسهیلات پرداختی)

با توجه به جدول (۱۳)، تعداد مشتریان کارا ۱۰۲ نفر از تعداد کل ۱۱۳۷ مشتری در کل خوشه‌ها و تعداد مشتریان ناکارا ۱۰۳۵ نفر براساس ۱۱۳۷ مشتری در کل خوشه‌ها است که نتایج به‌دست‌آمده از جمع مشتریان کارا در خوشه‌ها ۱/۵ برابر تعداد مشتریان کارا در کل داده‌ها است. در این بخش با توجه به تعداد بالای مشتریان در خوشه‌های دوم (۶۷۷۷) و خوشه چهارم (۹۳۰۲۷) با استفاده آزمون آماری کوکران به نمونه‌گیری از تعداد مشتریان پرداخته شده است که نتایج نمونه‌گیری آزمون کوکران تعداد ۳۶۴ نفر برای خوشه دوم و تعداد ۳۵۳ نفر برای خوشه چهارم جهت بررسی کارایی خوشه‌ها انتخاب شدند.

www.SID.ir

Archive of SID

جدول (١٤): توصیف آماری برای کارایی مشتریان بانک در هر خوشه

خوشه	آمار	ورودی						خروجی		
		مبلغ وام	ارزش وثایق	تحصیلات	بخش اقتصادی	زمان درخواست	نوع تسهیلات	سود پرداختی	تعداد پرداختی	وضعیت حساب
١	Mean	١٠٦٦٦٦٦٦٧	٤٠٦٨٠٧٦٧	٦٠٨٣٣٣٤	٤/٠٠٠	١٣٨٩/٢٥٣١	٢٢١/٣٦٢١	٢٣٦٥٤١٠٤	١١/١٦٦٧	٢/٩١٦٧
	Std.Dev	٢٩٩٤٩٤٥٢٣	٥٧٥٠١٩٠١١	١/٣٦٢١	١/٨٠٩٠٧	٤٥٥/٥٣٦٣	١٨٧/٦٣٥١	٤٥٣٦٥٤١٢	١٤/٥٢٧٩٣	٢/٠٢٠٧٣
٢	Mean	٣٣٣٤٢٨٦	٣١٩٨٦٧٥٧	٨٢٥٢١٥٠٠٠	٤/٨٥٧	١٣٩١/٢٦٥١	٦٦٥/٢٣٢	٣٦٧١٤٢٨٥٧	٢٠/١٤٢٩	٣/٣٥٧١
	Std.Dev	٢٤٥٣٦١٦	٦٨٠٦٢٢١٥	٣٣٤٥٩٠٢٠٦	٢/٨٤٠٠٦	٣٦٥١/٥١٤٢	٣٢٨/٣٦٢	٢٣٣٨٦٢٢٠٠٢	٢٤/٧٩٢٦	١/٩٠٥٧٥
٣	Mean	٧٠٢٠٨٣٣	٢٢٨١٥٧٩٦	٤١٩٢٠٨٦٢٥	٣/٨٣٣٣	١٣٨٧/٣٢١	٤٢٥/٥٢٣	٢٠٤٨٧٥٠٠٠	٢٢/٦٢٥٠	٤/٠٠٠
	Std.Dev	٨٥٣٠٣٠٠	١٨٩٥٩٥٤١	٧/٦١١٤٤٣	٢/٨٩٩٢٨	٤٥١/٣٦٢	٤١٢/٣٢٥	١٠٢٠١١٦٤١	٢٣/٧٨١٢٧	٣/٠٢١٦٦
٤	Mean	٨٥٦٣٢٤٦٦	١٧٥٦٣٢١٢	٧٥٣٢١٤٥	٥/٩٦٢	١٣٩٢/٥٣٦٤	٢٧٩/٥٣٤٢	٢٨١٢٣١١٠	٢٠/٨٩٣٢١	٤/٩٣٦
	Std.Dev	١١٣٦٣٢١٦	٢٤٥٤٢٠٦٣٢	٦٥٤٢١٤٣	٢/٢٣٣	٥٥٠/٢٣٦١	١٩٦/٥٣٦١	١٤٥٢٣٦٥١	٢٤/٢٣٥٤٢	١/٦٥٢١

www.SID.ir

Archive of SID

با توجه به جدول (۱۴)، توصیف آماری برای عملکرد تسهیلات بانک در هر خوشه نشان داده شده است. مقایسه تغییرات متغیرهای هر خوشه با توجه به جداول (۱۰)، (۱۲) و (۱۴) نشان می‌دهد که کارایی مشتریان در خوشه‌ها تغییرات کمتری در مقایسه با خوشه‌بندی کل ورودی‌ها و خروجی‌ها دارد. تغییرات رفتار مشتریان در خوشه‌ها با یکدیگر مشابه هستند. همچنین میانگین متغیرهای کارا در هر خوشه متفاوت از میانگین کل در هر خوشه است. همچنین برای خوشه میانگین بازپرداخت برای مشتریان کارا تقریبا دو برابر میانگین بازپرداخت کل مشتریان هر خوشه است. از لحاظ وضعیت حساب مشتریان مقادیر به‌دست آمده برای میانگین مشتریان کارا عددی بزرگ‌تر از کل مشتریان است که نشان از سابقه بیشتر مشتریان کارا نسبت به‌کل مشتریان هر خوشه است.

به صورت کلی و با تایید روش‌های مورد استفاده برای خوشه‌بندی و ارزیابی عملکرد، می‌توان اظهار کرد اغلب تسهیلات اعطایی صندوق کارآفرینی امید در دوره مورد مطالعه -به دلیل پرداخت منظم بدهی تسهیلات (اقساط تسهیلات) و در نتیجه کاهش ریسک اعتباری صندوق- در خوشه تسهیلات با عملکرد مناسب قرار می‌گیرد که حکایت از توفیق نسبی تسهیلات اعطایی این صندوق دارد. از آنجا که اغلب تسهیلات اعطایی صندوق در قالب خرد و متوسط قرار می‌گیرد و اعطای تسهیلات نیز به صورت قرض الحسنه و با سود ۴ درصد انجام می‌پذیرد و به دلیل نظارت مستمر و پویای صندوق بر نحوه هزینه‌کرد تسهیلات دریافتی در حوزه قرارداد تسهیلاتی، امکان انحراف تسهیلات به سمت بخش‌های دیگر کمتر محتمل است، از این رو، بازپرداخت منظم تسهیلات شاید بتواند حکایت از موفقیت نسبی کسب و کار هدف و در نتیجه اثربخشی تسهیلات اعطایی نیز داشته باشد. همچنین مناسب‌ترین عملکرد تسهیلات اعطایی نیز به ترتیب در حوزه‌های های اقتصادی رفاهی، خدمات، مسکن، کشاورزی و صنعتی بوده است.

٤- پیش‌بینی وضعیت اعتباری متقاضیان تسهیلات صندوق

امروزه در دنیای داده‌های بزرگ، اعطای قابلیت پیش‌بینی برای عملکرد متقاضیان تسهیلات از نیازمندی‌های اصلی هر موسسه مالی- اعتباری است. در این بخش برای پیش‌بینی عملکرد متقاضیان تسهیلات براساس بازپرداخت بدهی تسهیلات از دو روش الگوریتم بردار پشتیبان و الگوریتم ترکیبی شبکه عصبی پرسپترون چندلایه- ژنتیک

www.SID.ir

Archive of SID

استفاده می‌شود. در این روش فرضیه اصلی مطرح می‌شود که تکنیک‌های پیش‌بینی داده‌کاوی جهت اعتبارسنجی مبتنی بر صورت‌های مالی از کارایی مناسبی برخوردار هستند. این فرضیه که هدف آن بررسی کارایی تکنیک‌های پیش‌بینی داده‌کاوی در اعتبارسنجی مشتریان اعتباری بانک به‌منظور ارزیابی آن‌ها است از اهمیت ویژه‌ای برخوردار هستند، زیرا با استفاده از این مدل‌ها می‌توان قبل از اعطای تسهیلات مشتریان را ارزیابی و به‌تبع آن می‌توان تسهیلات اعتباری مشتریان واجد شرایط را اعطا کرد. به‌منظور آزمون این فرضیه ابتدا با استفاده از چهار اهرم خوب (خوش‌حساب- خوش‌حساب)، عمود (بدحساب- خوش‌حساب)، منظم (خوش‌حساب- بدحساب) و اهرم بد (بدحساب- بدحساب) و با استفاده از پیشینه پژوهشی به‌دست‌آمده و با استفاده از نرم‌افزار متلب و با استفاده از متغیرهای مورداستفاده در بخش خوشه‌بندی کارایی دو مدل ذکر شده در پیش‌بینی وضعیت اعتباری مشتریان بانک بررسی شد.

نتایج حاصل از بررسی دو مدل، کارایی تکنیک‌ها در اعتبارسنجی را نشان می‌دهد. در این بخش از ۶ متغیر ورودی و ۳ متغیر خروجی استفاده شد که متغیرهای ورودی استفاده شده عبارتند از : ۱- مقدار مبلغ مصوب برای مشتری[1]، ۲- نام بخش اقتصادی[2]، ۳- ارزش ترهینی وثایق پرونده[3]، ۴- سطح سواد مشتریان[4]، ۵- نوع تسهیلات[5]، ۶- زمان درخواست تسهیلات توسط مشتری[6] براساس سال پایه ۱۳۸۸ و متغیرهای خروجی استفاده شده در این بخش عبارتند از: ۱- وضعیت حساب[7]، ۲- تعداد پرداختی‌های مشتری[8] از شروع قرارداد تا تاریخ مطالعه و ۳- سود بانکی[9].

فرض نوع اول: مدل استفاده‌شده از تکنیک ماشین بردار پشتیبان جهت پیش‌بینی عملکرد متقاضیان تسهیلات از منظر بازپرداخت بدهی تسهیلات از کارایی مناسبی برخوردار است.

1- Main Mount
2- Economic Sector
3- Mortgage
4- Education
5- Type of Facility
6- Facility Demand
7- Account Status
8- Payed No
9- Advantage

www.SID.ir

Archive of SID

یافته‌ها و نتایج به‌دست‌آمده از فرضیه اول: یافته‌ها و نتایج به‌دست‌آمده با استفاده از الگوریتم ماشین‌بردار پشتیبان (SVM) نشان می‌دهد که میزان صحت در مجموعه‌های داده‌های آموزش ۶۴ درصد است. درصدهای مربوط به آزمون مربوط به صحت طبقه‌بندی و معیاری برای ارزیابی اعتبار و صحت مدل است که در این تکنیک ۷۴ درصد است. همچنین تعداد داده‌های آموزش مدل ۵۴۷ مشتری و تعداد داده‌های آزمون در این مدل ۳۸۳ مشتری برآورد شد. پیش‌بینی‌های مورد انتظار برای ۳۸۳ مشتری در مجموعه داده‌های آزمون برای مدل ماشین پشتیبان در جدول ۱۵ مشاهده می‌شود.

جدول (۱۵): عملکرد پیش‌بینی مدل ماشین بردار پشتیبان در دوره آزمون

شرح	خوش‌حساب	بدحساب	مجموع
موارد تشخیص به‌عنوان خوش‌حساب	۲۶۵	۸۰	۳۴۵
موارد تشخیص به‌عنوان بدحساب	۱۹	۱۹	۳۸
مجموع	۲۸۴	۹۹	۳۸۳
تعداد پیش‌بینی‌های درست	۲۶۵	۱۹	۲۸۴
درصد دقت مدل	۹۳	۱۹	۷۴
درصد خطای مدل	۰/۰۷	۰/۸۱	۰/۲۶

یافته‌های تحقیق حاکی از این است که از مجموع ۲۸۴ مشتری خوش‌حساب، تعداد ۲۶۵ مشتری به‌عنوان خوش‌حساب و ۱۹ مشتری (به‌اشتباه) به‌عنوان بدحساب تشخیص داده شده‌اند به‌طوری‌که دقت الگو در این خصوص ۹۳ درصد محاسبه شده است. به همین ترتیب از ۹۹ مشتری بدحساب، ۱۹ مشتری به‌عنوان بدحساب و ۸۰ مشتری (به‌اشتباه) خوش‌حساب تشخیص داده شده‌اند و در دقت الگو در این حالت ۱۹ درصد است. بنابراین، میانگین دقت پیش‌بینی در این مدل ۷۴ درصد است. با توجه به یافته‌ها، این مدل از میزان صحت بالایی برخوردار است، چون میزان کارایی در این مقاله ۷۰ درصد تعیین شده است در نتیجه این فرضیه تایید می‌شود و از کارایی مناسبی برای اعتبارسنجی برخوردار است.

فرض دوم: مدل استفاده‌شده از الگوریتم ترکیبی پرسپترون چندلایه و ژنتیک جهت پیش‌بینی عملکرد متقاضیان تسهیلات از منظر بازپرداخت بدهی تسهیلات از کارایی مناسبی برخوردار است.

www.SID.ir

Archive of SID

یافته‌های نتایج به‌دست آمده آزمون فرضیه دوم: یافته‌های نتایج به‌دست آمده آزمون فرضیه سوم نشان می‌دهد که میزان صحت مدل شبکه عصبی- ژنتیک در مجموعه داده‌های آموزش ۹۴/۵۵ درصد است. درصدهای مربوط به آزمون مربوط به صحت طبقه‌بندی و معیاری برای ارزیابی اعتبار صحت مدل است که در این تکنیک ۹۳/۴۶ درصد است. همچنین تعداد داده‌های آموزش مدل ۵۴۷ تراکنش و تعداد داده‌های آزمون در این مدل ۳۸۳ تراکنش برآورد شد. پیش‌بینی‌های مورد انتظار برای ۳۸۳ مشتری در مجموعه داده‌های آزمون برای مدل شبکه عصبی در جدول (۱۶) مشاهده می‌شود.

جدول (۱٦): عملکرد پیش‌بینی مدل شبکه عصبی -ژنتیک در دوره آزمون

شرح	خوش‌حساب	بدحساب	مجموع
موارد تشخیص به‌عنوان خوش‌حساب	۲٦۹	٦۵	۳۳٤
موارد تشخیص به‌عنوان بدحساب	۲۵	۲٤	٤۹
مجموع	۲۹٤	۸۹	۳۸۳
تعداد پیش‌بینی‌های درست	۲٦۹	۲٤	۲۹۳
درصد دقت مدل	۹۱	۲۷	۷۷
درصد خطای مدل	۰/۰۹	۰/۷۳	۰/۲۳

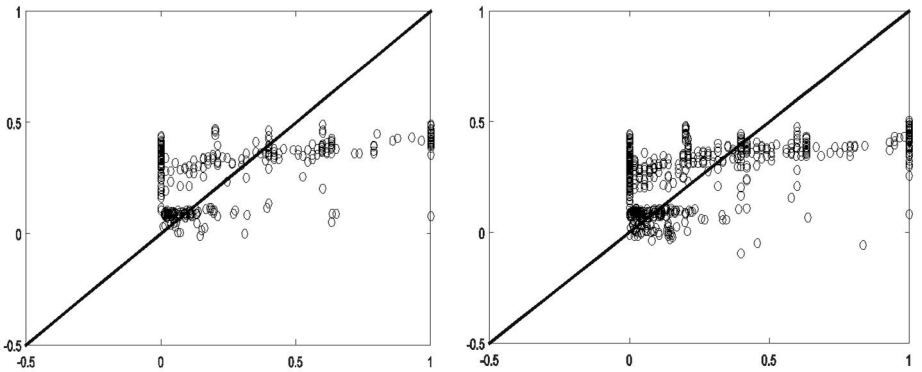
یافته‌های تحقیق حاکی از این است که از مجموع ۲۹۴ مشتری خوش‌حساب، تعداد ۲۶۹ مشتری به‌عنوان مشتری خوش‌حساب و ۲۵ مشتری (به‌اشتباه) به‌عنوان بدحساب تشخیص داده‌شده‌اند به‌طوری‌که دقت الگو در این خصوص ۹۱ درصد محاسبه‌شده است. به همین ترتیب از ۸۹ مشتری بدحساب، ۶۵ مشتری به‌عنوان مشتری به‌عنوان بدحساب و ۲۴ مشتری (به‌اشتباه) خوش‌حساب تشخیص داده‌شده‌اند و دقت الگو در این حالت حدود ۲۷ درصد است. بنابراین، میانگین دقت پیش‌بینی در این مدل ۷۷ درصد است. ضریب تعیین برای این مدل ۰/۷۸ =R^2 به‌دست‌آمده است. ورودی‌های مدل عملکرد بانک یا متغیرهای پیش تعیین‌شده مشتریان هستند و متغیرهای خروجی، متغیرهایی از عملکرد مشتریان را نشان می‌دهد، بنابراین، مدل به‌دست‌آمده می‌تواند پیش‌بینی به نسبت خوبی از عملکرد مشتریان، انجام دهد. با تعیین اولیه متغیرهای خروجی برای مشتریان جدیدالورود می‌توان پیش‌بینی اولیه‌ای از مکان و اولویت مشتریان در هر خوشه یا اهرم به دست آورد. همچنین نمودار (۸

www.SID.ir

Archive of SID

- الف) و (۸- ب) به ترتیب نمودار پراکندگی داده‌ها حول روند را به ترتیب برای بخش آموزش و تست داده‌های مورد استفاده نشان داده شده است.

نمودار (۸): پراکندگی مشتریان در پیش‌بینی آموزش شبکه عصبی- ژنتیک



الف: نمودار پراکندگی آموزش مشتریان

ب: نمودار پراکندگی تست مشتریان

با توجه به یافته‌های این مدل از میزان صحت بالایی برخوردار است و قدرت پیش‌بینی آن در مقایسه با روش ماشین‌بردار پشتیبان بیشتر است و درجه صحت طبقه‌بندی آن اختلاف معناداری با مدل ماشین بردار پشتیبان دارد و اینکه توانست متغیرهایی بااهمیت را شناسایی کند در نتیجه این فرضیه تایید شد و از کارایی مناسبی برای اعتبارسنجی برخوردار است.

٤-١- نتایج مقایسه کارایی مدل‌ها

با توجه به ارائه دو مدل در این مقاله برای پیش‌بینی سنجش اثربخشی تسهیلات اعطایی، نیاز است که مطالعه تطبیقی بین دو مدل از منظر کارایی پیش‌بینی انجام پذیرید. برای سنجش کارایی مدل‌های ارائه‌شده صحت نتایج این مدل‌ها در جدول (۱۷) مقایسه شده است.

www.SID.ir

Archive of SID

جدول (١٦): مقایسه تطبیقی صحت الگوریتم‌ها

الگوریتم	SVM	MLP
صحت	٪ ٧٤ / ٥٠	٪ ٩٣/٤٦

همان‌طور که در جدول (۱۷) مشخص شده است هر دو مدل مورد مطالعه (تکنیک ماشین‌بردار پشتیبان و الگوریتم ترکیبی شبکه عصبی- ژنتیک) در این مقاله از لحاظ صحت پیش‌بینی از قدرت بالایی برخوردار است که در این بین الگوریتم ترکیبی شبکه عصبی پرسپترون چندلایه- ژنتیک نسبت به مدل دیگر از قدرت بالایی برخوردار است. با توجه به اهمیت شناسایی ریسک و یافتن راهکارهایی برای مدیریت و کنترل آن دسترسی به مدل‌هایی که با دقت بالا بتواند این رخداد را پیش‌بینی کند، نقش مهمی به لحاظ اقتصادی و عملی در بقا و سلامت بانک‌ها دارد. در این مقاله دو مدل مختلف (تکنیک ماشین بردار پشتیبان و الگوریتم ترکیبی شبکه عصبی (MLP) –ژنتیک) برای پیش‌بینی ریسک برآورد شد و یافته‌های حاصل مقایسه شدند.

۵- نتیجه‌گیری

هدف این مقاله، ارائه مدلی ترکیبی برای خوشه‌بندی، رتبه‌بندی و پیش‌بینی عملکرد تسهیلات اعطایی موسسات مالی- اعتباری از منظر بازپرداخت بدهی تسهیلات است. جهت این منظور برای خوشه‌بندی تسهیلات اعطایی از داده‌های واقعی تسهیلات صندوق کارآفرینی امید استفاده شد. برای انجام این کار نیاز به یک مدل خوشه‌بندی بود. بنابراین، ابتدا از یک مدل ترکیبی شامل روش کلاس‌بندی «رب پی سی ای»، روش خوشه‌بندی دومرحله‌ای «کی-مینز» و روش تحلیل پوششی داده‌ها استفاده شد. به صورت کلی و با تایید روش‌های مورد استفاده برای خوشه‌بندی می‌توان اظهارنظر کرد اکثریت قریب به اتفاق تسهیلات اعطایی صندوق کارآفرینی امید در دوره مورد مطالعه -به دلیل پرداخت منظم بدهی تسهیلات (اقساط تسهیلات) و در نتیجه کاهش ریسک اعتباری صندوق- در خوشه تسهیلات با عملکرد مناسب قرار می‌گیرد که حکایت از توفیق نسبی تسهیلات اعطایی این صندوق دارد.

www.SID.ir

Archive of SID

از آنجا که اغلب تسهیلات اعطایی صندوق در قالب خرد و متوسط قرار می‌گیرد و اعطای تسهیلات نیز به صورت قرض‌الحسنه و با سود ۴ درصد انجام می‌پذیرد و به دلیل نظارت مستمر و پویای صندوق بر نحوه هزینه‌کرد تسهیلات دریافتی در حوزه قرارداد تسهیلاتی، امکان انحراف تسهیلات به سمت بخش‌های دیگر کمتر محتمل است، بنابراین، بازپرداخت منظم تسهیلات شاید بتواند حکایت از موفقیت نسبی کسب و کار هدف و در نتیجه اثربخشی تسهیلات اعطایی نیز داشته باشد. همچنین مناسب‌ترین عملکرد تسهیلات اعطایی نیز به ترتیب در حوزه‌های های اقتصادی رفاهی، خدمات، مسکن، کشاورزی و صنعتی بوده است. در گام بعدی مدلی برای پیش‌بینی وضعیت اعتباری مشتریان معرفی شد.

نتایج پژوهش حاضر نشان‌دهنده این واقعیت است که امکان پیش‌بینی ریسک اعتباری مشتریان هنگام اعطای تسهیلات اعتباری از راه مختصات مشتریان به‌عنوان متغیرهای پیش‌بینی و استفاده آن‌ها در مدل‌های آماری و شبکه‌های وجود دارد.

در این مقاله با یک رویکرد جدید و ترکیبی از دو روش پیش‌بینی ماشین بردار پشتیبان و الگوریتم ترکیبی الگوریتم شبکه عصبی- ژنتیک برای پیش‌بینی وضعیت اعتباری متقاضیان تسهیلات صندوق استفاده‌شده است. مدل به‌دست‌آمده از روش ترکیبی شبکه عصبی- ژنتیک با میانگین مربعات خطا ۰/۲۳ و ضریب تعیین ۷۸ درصد است. ورودی‌های مدل عملکرد صندوق یا متغیرهای پیش تعیین شده مشتریان هستند و متغیرهای خروجی، متغیرهایی از عملکرد مشتریان را نشان می‌دهد، بنابراین، مدل به‌دست‌آمده می‌تواند پیش‌بینی به نسبت خوبی از عملکرد متقاضیان تسهیلات در مقایسه با تکنیک پیش‌بینی ماشین بردار پشتیبان انجام دهد. با تعیین اولیه متغیرهای خروجی برای مشتریان جدیدالورود می‌توان پیش‌بینی اولیه‌ای از مکان و اولویت متقاضیان تسهیلات در هر خوشه یا اهرم به دست آورد. نتایج حاکی از این است که در بین دو مدل طراحی‌شده به ترتیب مدل ترکیبی شبکه عصبی پرسپترون چندلایه - ژنتیک از کارایی بالایی برای پیش‌بینی ریسک اعتباری متقاضیان تسهیلات (یعنی پیش‌بینی احتمال عدم بازپرداخت بدهی توسط متقاضیان تسهیلات) برخوردار است.

نتایج هر دو بخش خوشه‌بندی و پیش‌بینی حکایت از کارایی مناسب مدل ارائه شده برای خوشه‌بندی و پیش‌بینی اثربخشی تسهیلات بر روی داده‌های بزرگ دارد که می‌تواند بصورت نرم‌افزاری در موسسات مالی- اعتباری دیگر نیز مورد بهره برداری قرار گیرد.

www.SID.ir

Archive of SID

منابع

الف) فارسی

آجرلو، نصیبه (۱۳۹۲)، الگویی جهت تعیین ارزش چرخه عمر مشتریان (CLV) در بانک ملت، پایان‌نامه کارشناسی ارشد، موسسه عالی بانکداری ایران.

اداره آموزش و مدیریت بانک ملی (۱۳۸۷)، *بانکداری الکترونیکی بانک ملی*، تهران، بانک ملی

اسلامی علیشاه، حسن (۱۳۸۷)، «مهارت‌های بازاریابی در شعب بانک‌ها؛ بانک مسکن»، *ماهنامه اقتصادی و اجتماعی*، شماره۹۲.

آذر، عادل، پرویز احمدی و محمدوحید سبط (۱۳۸۹)، «طراحی مدل انتخاب نیروی انسانی با رویکرد داده کاوی (استخدام داوطلبان آزمون های ورودی یک بانک تجاری در ایران)»، *مدیریت فناوری اطلاعات*، ۲ (۴)، ۲۲-۳.

البرزی، محمود، محمدابراهیم محمد پورزرندی و محمد خان بابایی (۱۳۸۹)، «به‌کارگیری الگوریتم ژنتیک در بهینه‌سازی درختان تصمیم گیری برای اعتبارسنجی مشتریان بانک‌ها»، *مدیریت فناوری اطلاعات*، ۲(۴)، ۲۳-۲۸.

امیری، فهیمه (۱۳۹۲)، بررسی رفتار مشتریان در استفاده از ابزارهای بانکداری الکترونیک با رویکرد داده کاوی (مورد کاوی: بانک صادرات ایران)، پایان‌نامه کارشناسی ارشد، دانشکده علوم اجتماعی و اقتصاد دانشگاه الزهرا (س).

امیری، سیدعلی (۱۳۹۱)، مدیریت دانش مشتریان بانک مهر اقتصاد با استفاده از تکنیک‌های داده‌کاوی، مازندران، پایان‌نامه کارشناسی ارشد، دانشگاه علوم و فنون مازندران.

تقوی‌فرد، محمدتقی و سعید خواجوند (۱۳۹۳)، «بخش‌بندی خوشه‌ای مشتریان بانک (مورد مطالعه: بانک صادرات ایران)»، *دو فصلنامه علمی- پژوهشی کاوش‌های مدیریت بازرگانی*، سال پنجم، شماره ۹، ۶۴ — ۳۹.

حسینی بامکان، سیدمحمد (۱۳۹۰)، پایان‌نامه کارشناسی ارشد، به‌کارگیری تکنیک‌های داده کاوی جهت بهبود مدیریت ارتباط با مشتری در صنعت بانکداری (مطالعه موردی: بانک پارسیان)، استاد راهنما، دکتر محمد رضا تقوا، دانشکده حسابداری و مدیریت، دانشگاه علامه طباطبایی.

www.SID.ir

Archive of SID

حسینی، سید صفدر، حبیب جهانگرد و حلیمه شهبازی (۱۳۸۸)، «پیش‌بینی تقاضای پول در افق ۱۴۰۴ در ایران (کاربرد الگوی سری زمانی)»، *پژوهشنامه اقتصادی*، ۱۰، ۳۸، ۸۶-۶۷.

حسینی و همکاران (۱۳۸۹)، «تحلیل اهمیت- عملکرد ویژگی‌های خدمت بر پایه بخش‌بندی مشتریان با رویکرد داده‌کاوی (پژوهشی در بانک مسکن در استان یزد)»، *مدیریت فناوری اطلاعات*، (۱۳)، ۷۰-۴۵.

فلاح شمس، میرفیض و حمید مهدوی‌راد (۱۳۹۰)، «طراحی مدل اعتبارسنجی و پیش‌بینی ریسک اعتباری مشتریان تسهیلات لیزینگ(مورد مطالعه : شرکت لیزینگ ایران خودرو)»، *پژوهشنامه اقتصادی*، ۱۲، ۴۴، ۲۳۴-۲۱۳.

عیسی‌زاده، سعید و بهرام عریانی (۱۳۹۲)، «رتبه‌بندی مشتریان حقوقی بانک‌ها برحسب ریسک اعتباری به روش تحلیل پوشی داده‌ها: مطالعه موردی شعب بانک کشاورزی»، *فصلنامه پژوهش‌ها و سیاست‌های اقتصادی*، سال هجدهم، شماره ۵۵، صفحات ۸۶-۵۹.

مهدوی، غلامحسین و کاظم گودرزی (۱۳۸۹)، «ارائه یک شبکه عصبی مصنوعی جهت پیش بینی ریسک سیستماتیک با استفاده از متغیرهای کلان اقتصادی (مطالعه موردی: شرکت سایپا)»، *پژوهشنامه اقتصادی*، ۱۱، ۴۳، ۲۳۷-۲۱۹.

مدیریت تلفیق برنامه و فناوری اطلاعات (۱۳۹۵)، *تبیین عملکرد صندوق کارآفرینی امید*، تهران: صندوق کارآفرینی امید کشور.

هاشمیان، مژده، محمدتقی عیسایی، فتاح میکائیلی و محسن طباطبائی (۱۳۹۱)، «عوامل موثر بر پذیرش ابزارهای بانکداری الکترونیک از سوی مشتریان (پیمایشی درباره بانک سامان)»، *مدیریت فناوری اطلاعات*، ۴ (۱۱)، ۱۵۵-۱۷۴.

ب- انگلیسی

Amin, H., Abdul Hamdi, M.R., Lada, S., and Baba, R., (2013), "Cluster Analysis for Bank Customers' Selection of Islamic Mortgages in Eastern Malaysia an Empirical Investigation", *International Journal of Islamic and Middle Eastern Finance and Management*, No. 82, 277-294.

Bošnjak, Z., and Grljevic, O. (2011), Credit User's Segmentation for Improved Customer Relationship Management in Banking, In Applied Computational Intelligence and Informatics. Proceedings (pp. 379-384). IEEE.

www.SID.ir

Archive of SID

Chen, F. L., and Li, F. C. (2010), "Combination of Feature Selection Approaches with SVM in Credit Scoring", *Expert Systems with Applications*, *37*(7), 4902-4909.

Chiraphadhanakul, S., Dangprasert, P., and Avatchanakorn, V. (2013), *Genetic Forecasting Algorithm with Financial Applications,* InIntelligent Information Systems, IIS'97. Proceedings (pp. 174-178). IEEE.

Fang, B., and Ma, S. (2009), Data Mining Technology and its Application in CRM of Commercial Banks, In 2009 First International Workshop on Database Technology and Applications (pp. 243-246). IEEE.

Hendalianpour, A., Razmi, J., and Gheitasi, M. (2017), "Comparing Clustering Models in Bank Customers: Based on Fuzzy Relational Clustering Approach", *Accounting*, *3*(2), 81-94.

Herrera-Restrepo, O., Triantis, K., Seaver, W. L., Paradi, J. C., and Zhu, H. (2016), "Bank Branch Operational Performance: A Robust Multivariate and Clustering Approach", *Expert Systems with Applications*, *50*, 107-119.

Haenlein, M., Kaplan, A. M., and Beeser, A. J. (2015), "A Model to Determine Customer Lifetime Value in a Retail Banking Context", *European Management Journal*, 25(3), 221-234.

Hubert, M., Rousseeuw, P. J., and Vanden Branden, K. (2005), "ROBPCA: A New Approach to Robust Principal Component Analysis", *Technometrics*, *47*(1), 64-79.

Kumar, V., Shah, D., and Venkatesan, R. (2014), "Managing Retailer Profitability-one Customer at a Time!", *Journal of Retailing*, 82(4), 277-294.

Lee, B., Cho, H., Chae, M., And Shim, S. (2010), "Empirical Analysis of Online Auction Fraud: Credit Card Phantom Transactions", *Expert Systems with Applications*, *37*(4), 2991-2999.

Lin, C. S., Tzeng, G. H., and Chin, Y. C. (2011), "Combined Rough set Theory and Flow Network Graph to Predict Customer Churn in Credit card Accounts", *Expert Systems with Applications*, 38(1), 8-15.

Ngai, E. W., Xiu, L., and Chau, D. C. (2009), "Application of Data Mining Techniques in Customer Relationship Management: A Literature Review and Classification", *Expert Systems with Applications*, 36(2), 2592-2602.

www.SID.ir

Archive of SID

Purba, I. Permanasari, A.E. Setiawan, N.A.(2014), *Optimization of Neural Using Genetic Algorithm in Forecasting Third Party Funds Bank*, Electrical Engineering and Informatics (MICEEI), Makassar International Conference on.IEEE.pp184-188.

Saar, D. Yagil, Y. (2015), "Forecasting Sectorial Profitability and Credit Spreads Using Bond Yields", *International Review of Economics and Finance*, vol (38), pp29-43.

Xie, Y., Li, X., Ngai, E. W. T., and Ying, W. (2009), "Customer Churn Prediction Using Improved Balanced Random Forests", *Expert Systems with Applications*, 36(3), 5445-5449.

www.SID.ir