نشریـه علمـی
دانش مالی تحلیل اوراق بهادار
سال چهاردهم، شماره پنجاه‌ام
تابستان ۱۴۰۰

ارائه‌ی یک مدل دو مرحله‌ی مبتنی بر داده کاوی جهت ارزیابی عملکرد مشتریان مشترک بانک و بیمه

حمیدرضا امیرحسنخانی[۱]
عباس طلوعی اشلقی[۲]
رضا رادفر[۳]
علیرضا پورابراهیمی[۴]

تاریخ دریافت: ۱۴۰۰/۰۲/۰۷ تاریخ پذیرش: ۱۴۰۰/۰۴/۱۹

چکیده

اکتشاف دانش از پایگاه داده و داده کاوی یکی از مهم ترین ابزار های مدیریت ارتباط با مشتری است که می-تواند به سازمان با تاکید بر یافتن اطلاعات مفید یا دانش مورد علاقه آن کمک کند. امروزه بانک‌ها و بیمه‌ها از پایگاه‌های اطلاعاتی متعدد و گسترده‌ای برخوردار هستند که حاوی اطلاعات مربوط به مبادلات و جزئیات دیگر مربوط به مشتریانشان هستند. اطلاعات ارزشمند تجاری از درون همین ذخایر داده‌ای قابل بازیابی است. اما پشتیبانی از این دست تحلیل‌ها و تصمیم‌گیری با استفاده از زبان‌های گزارش‌گیری سنتی امکان‌پذیر نخواهد بود. بنابراین با توجه به اهمیت اطلاعات مشتریان مشترک بانک و بیمه باید آنها را هرچه دقیق تر تحلیل نمود. در این تحقیق با گردآوری و تجزیه و تحلیل اطلاعات مشتریان مشترک بانک و بیمه یک متدولوژی مبتنی بر داده‌کاوی جهت ارزیابی مشتریان بر اساس شاخص های عملکردی آنها در حوزه بانکی و بیمه‌ای ارائه می شود. همچنین با تحلیل رفتار مشتریان با استفاده از یه رویکرد دو مرحله‌ای مبتنی بر یادگیری بدون نظارت و یادگیری با نظارت اقدام به پیش بینی رفتار مشتریان جدیدالورود خواهیم نمود.

واژه‌های کلیدی: مدیریت ارتباط با مشتری، داده کاوی، بانک، بیمه.

۱- دانشجوی دکتری، گروه مدیریت فناوری اطلاعات، واحد امارات، دانشگاه آزاد اسلامی، دبی، امارات متحده عربی
۲- استاد گروه مدیریت، گروه مدیریت صنعتی، دانشکده مدیریت و اقتصاد، واحد علوم و تحقیقات، دانشگاه آزاد اسلامی، تهران، ایران(عهده دار مکاتبات) (toloie@gmail.com)
۳- استاد گروه مدیریت، گروه مدیریت صنعتی، دانشکده مدیریت و اقتصاد، واحد علوم و تحقیقات، دانشگاه آزاد اسلامی، تهران، ایران
۴- استادیار گروه مدیریت، دانشکده مدیریت، واحد کرج، دانشگاه آزاد اسلامی، کرج، ایران

۱- مقدمه

بنا به تحقیق مجله بررسی بازرگانی هاروارد برخی شرکت ها تنها با پنج درصد تلاش بیشتر در حفظ مشتری های موجود سود خود را تا صد در صد افزایش می دهند. [۱]. اکتشاف دانش از پایگاه داده و داده کاوی یکی از مهم ترین ابزار های مدیریت ارتباط با مشتری است که می‌تواند به سازمان با تاکید بر یافتن اطلاعات مفید یا دانش مورد علاقه آن کمک کند[۳] [۲]. در [۴] محققان از روش‌هایی از جمله قوانین وابستگی، برای امتیازدهدی و تحلیل رفتاری مشتریان بانک استفاده کرده‌اند. به طور کلی تحقیقات زیادی در زمینه بخش‌بندی مشتریان در صنعت بانکداری انجام گرفته است. به عنوان مثال در [۵] به استفاده از تکنیک‌های داده‌کاوی در بخش‌بندی مشتریان اشاره شده است. همچنین می‌توان به تجربه موفق یکی از بانک‌های اروپایی در زمینه استفاده از داده‌کاوی در مدیریت ارتباط با مشتریان اشاره کرد [۶]. همچنین تحقیقات زیادی در مدلسازی فرآیند رویگردانی مشتریان به چاپ رسیده است. برای مثال در [۷] از دو الگوریتم داده‌کاوی برای ساخت مدلی به منظور پیش-بینی رویگردانی مشتریانی که از کارت‌های اعتباری استفاده می‌کنند، استفاده شده است. مثال دیگری که در این مورد می‌توان اشاره کرد مدلسازی رویگردانی مشتریان در ZBANK است [۸]. تحقیقات زیادی در زمینه استفاده از روش‌های داده‌کاوی در رتبه‌بندی اعتبار مشتریان بانک‌ها انجام شده است که در ادامه به مهمترین آنها اشاره می کنیم. یکی از این مطالعات، مقاله هانگ و همکاران است. در این تحقیق مشخص شده که متعادل نبودن تعداد نمونه‌های موجود در هر کلاس، می‌تواند منجر به انطباق بیش از حد مدل با داده‌ها شود [۹]. جمشیدی و همکاران [۱۰] یک رویکرد چند هدفه مبتنی بر سیستم استنتاج نورو دیفیوژن سازگار برای شناسایی پولشویی بانکی و مبادله ارز ارائه دادند. مگومدوو و همکاران [۱۱]، دورفیو و همکاران [۱۲] و پلاکسی و همکاران [۱۳] از روشهای یادگیری ماشین مبتنی بر هوش مصنوعی جهت طراحی و نظارت بر سیستمهای ضدپولشویی بهره برده‌اند. لیت و همکاران [۱۴]، ریدر [۱۵] و تیاواری و همکاران [۱۶] در مقالات مروری مجموعه‌ای غنی از تحقیقات صورت گرفته مبتنی بر داده‌کاوی و هوش مصنوعی جهت مقابله با پولشویی و سایر جرائم بانکی را گردآوری کرده‌اند. از جمله پژوهشهایی که در سالهای اخیر می توان به آنها اشاره کرد مطالعه دیپالی و همکاران در سال ۲۰۱۹ است که کاربرد تکنیکهای داده کاوی در سیستمهای بانکی را با هدف کشف تقلب و مدیریت ارتباط با مشتری مورد بررسی قرار دادند [۱۷]. همچنین در سال ۲۰۲۰ کانچانا و همکاران با استفاده از تکنیکهای طبقه بندی در داده-کاوی به بررسی حملات فیشینگ به سایتهای بانکی پرداختند [۱۸]. شهبازی در مقاله‌ی دیگری با استفاده از الگوریتم‌های طبقه بندی مبتنی بر درخت تصمیم یک مدل رتبه بندی اعتباری برای مشتریان بانک ارائه دادند [۱۹]. دینسر وهمکاران یک روش مبتنی بر داده کاوی جهت ارزیابی رضایت مشتریان از نرم‌افزارهای موبایلی در صنعت بانکداری ارائه دادند [۲۰]. کور و همکاران نیز از داده کاوی برای پیشبینی ریسک اعتباری مشتریان بانک استفاده کردند [۲۱]. مطالعه دیگری از حسنی و همکارن منتشر شده است که در آن دیجیتال سازی و داده کاوی در سیستم های بانکی مورد مطالعه قرار گرفته است [۲۲].

در این تحقیق قصد داریم یک متدولوژی کارا مبتنی بر داده‌کاوی جهت بخشبندی، ارزیابی عملکرد و همچنین مدلسازی پیش‌بینی عملکرد مشتریان جدیدالورود برای مشتریان مشترک بانک و بیمه ارائه دهیم. هدف این تحقیق توانمندسازی هلدینگ‌های سرمایه گذاری که سهامداران توام بانک و بیمه هستند در بالاترین سطح سفارشی‌سازی در تصمیم گیری برای مشتریان و اتخاذ تصمیمات متنوع و کارا طبق خصوصیات مشتریان شان و قوی‌تر کردن تعاملات با مشتری، رفع بهتر نیازهای مشتری و ارتقای سطح رضایت مشتریان و در نتیجه وفادار کردن آنهاست. بر این اساس این هلدینگ‌ها می‌توانند از طریق تقویت

پایگاه داده ها، لینک های ارتباطی شرکت های اطلاعاتی و افزایش دقت در ورود و ثبت اطلاعات اولیه و با اتکا به روش های داده کاوی، در هریک از زمینه های فوق الذکر به نتایج قابل توجه دست یابد.

۲- بیان مسئله

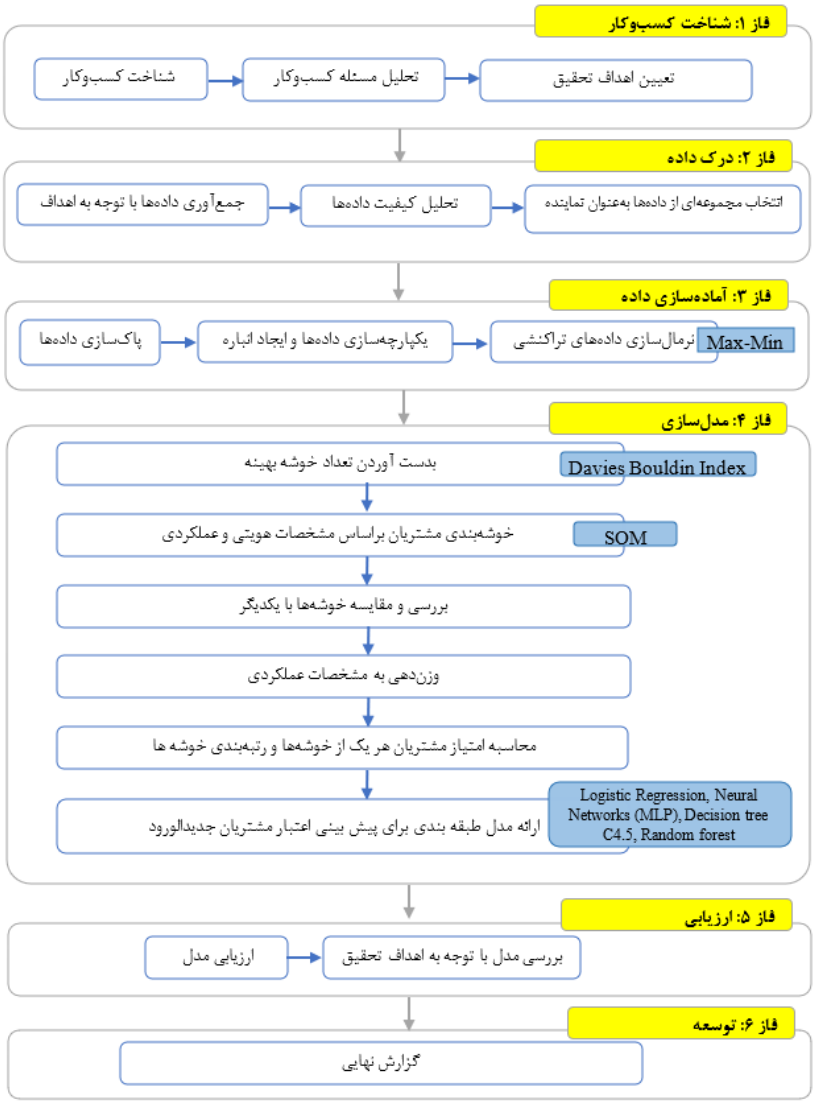
مسئله‌ی اصلی که در این تحقیق بر روی آن تمرکز شده است، ارزیابی و رتبه بندی مشتریان مشترک بیمه آرمان و بانک ایران زمین است. با توجه به اینکه بانک ایران زمین یکی از سهامدارن اصلی بیمه آرمان، می باشد، لذا اساسا جذب مشتریان مشترک بانک و بیمه به منظور ارائه خدمات بانک-بیمه‌ای به آنها در راستای یکی از اهداف سهامداران این موسسات قرار گرفته است. مع‌الاسف یکی از مسائلی که در زمینه جذب مشتریان مشترک برای این دو موسسه با آن مواجه و در این تحقیق به آن پرداخته شده است، مسئله ارزیابی و رتبه‌بندی مشتریان مشترک این دو می‌باشد. ایجاد سیستم بخش‌بندی و رتبه‌بندی مشتریان بر اساس عوامل مؤثر بر میزان ارزش مشتریان می‌تواند راهگشای حفظ مشتریان سودآور، جذب مشتریان مشترک جدید و تبیین برخی فعالیت‌های ارتباط با مشتری باشد. در این تحقیق از اطلاعات هویتی و شاخص های عملکردی مشتریان مشترک در بانک و بیمه به صورت توام به‌عنوان عامل تأثیرگذار بر بخش‌بندی مشتریان استفاده خواهد شد.

معرفی مورد مطالعاتی

در این تحقیق بر روی اطلاعات مستخرج از بانک ایران زمین و بیمه آرمان متمرکز می‌شویم. با توجه به وجود سهامداران مشترک این موسسات تحلیل اطلاعات مشتریان مشترک این موسسات در راستای هدف تحقیق می تواند نتایج کاربردی مدونی را ایجاد کند.

۳- متدولوژی پیشنهادی تحقیق

متدولوژی‌های مختلفی جهت پیاده‌سازی پروژه‌های داده‌کاوی ارائه شده‌اند، این متدولوژی‌ها به صورت یک الگوی فرآیندی، چرخه حیات یک پروژه داده‌کاوی را در بر می‌گیرد. متدولوژی پیشنهادی در این تحقیق به طور کلی شش مرحله عملیاتی را به طور کامل طی می‌کند. این شش مرحله شامل شناخت کسب‌وکار، شناخت داده، پیش‌پردازش، مدل‌سازی، ارزیابی مدل و توسعه مدل می‌باشد. شکل ۱ متدولوژی پیشنهادی در این تحقیق را توصیف می‌نماید.



شکل ۱: چارچوب تحقیق

۴- مدل‌سازی

این فاز شامل عملیات مدل‌سازی می‌باشد که بر روی داده‌های پردازش شده صورت می‌گیرد. این فاز شامل شش گام است:

۱) به کمک روش دیوس بولدین[1] تعداد خوشه‌های بهینه را تعیین می‌کنیم.

۲) مشتریان بر اساس اطلاعات هویتی و شاخص های عملکردی تطبیق یافته با استفاده از نقشه های خودسازمانده خوشه‌بندی می‌شوند.

۳) بررسی و مقایسه خوشه‌ها با یکدیگر

۴) وزن شاخص های عملکردی مؤثر در خوشه‌بندی را بر اساس نظر خبرگان محاسبه می‌کنیم.

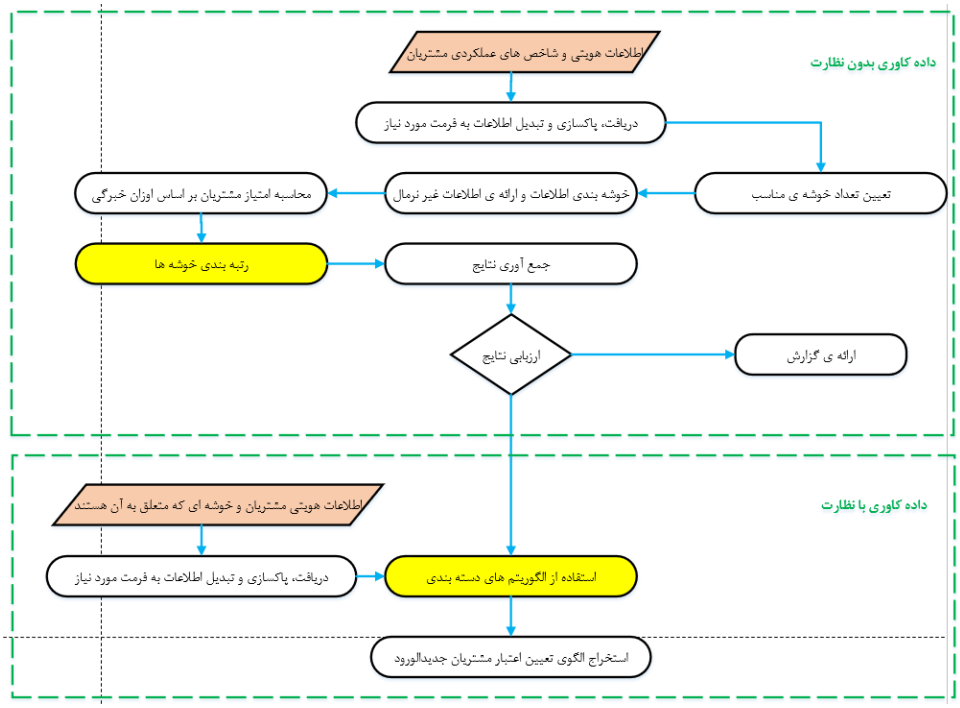
۵) محاسبه امتیاز مشتریان برای هر یک از خوشه‌های موجود و رتبه‌بندی خوشه‌ها بر اساس ارزش مشتریان آن‌ها.

۶) مدلی را جهت پیش‌بینی ارزش مشتریان جدیدالورود بر اساس اطلاعات هویتی آنها مبتنی بر طبقه‌بندی طراحی می‌کنیم. برای این منظور از روشهای رگرسیون لجستیک،

درخت تصمیم، جنگل تصادفی و شبکه‌های عصبی چندلایه بهره می بریم.

شکل ۲ روال کلی مدلسازی در این تحقیق بر مبنای ارتباط کلی داده کاوی بدون نظارت و داده‌کاوی با نظارت طی دو فاز متوالی نشان می دهد.

در ادامه این بخش مفاهیم کاربردی و نحوه‌ی استفاده از الگوریتم‌های استفاده‌شده در این تحقیق، ارائه می‌گردد. این الگوریتم ها در فاز مدلسازی متدولوژی پیشنهادی مورد استفاده قرار می‌گیرد.



شکل ۲: روال کلی مدلسازی بر مبنای ارتباط کلی داده کاوی بدون نظارت و داده‌کاوی با نظارت طی دو فاز متوالی

۴-۱- شاخص دیویس بولدین

شاخص دیویس بولدین نخستین بار در سال ۱۹۷۹ مطرح شد. این شاخص برای ارزیابی عدم تشابه میان خوشه‌ها، درواقع میانگین شباهت هر خوشه را با شبیه‌ترین خوشه محاسبه می‌کند، بنابراین هرچقدر این شاخص کمتر باشد، خوشه‌های بهتری تولیدشده است.

۴-۲- خوشه بندی بر اساس نقشه های خود سازمانده

نگاشت‌های خود سازمانده در حالت کلی یک ساختار دو لایه با یک لایه ورودی و یک لایه خروجی دارند. نرونهای لایه ورودی وظیفه انتقال داده‌ها به شبکه را برعهده داشته و در حالت کلی تعداد آنها با بعد بردارها در فضای ورودی برابر است.

۴-۳- رگرسیون لجستیک

رگرسیون لجستیک نوعی از رگرسیون است که متغیرهای مستقل می‌تواند هم در مقیاس کمی و هم در مقیاس کیفی باشد، ولی متغیر وابسته کیفی دوسطحی است. رگرسیون لجستیک می‌تواند یک مورد خاص از مدل خطی عمومی و رگرسیون خطی دیده شود.

۴-۴- درخت تصمیم

درختان تصمیم، درختانی هستند که نمونه‌ها را توسط دسته‌ی آن‌ها براساس مقادیر ویژگی‌هایشان

دسته‌بندی می‌کنند. هر گره در درخت تصمیم، ویژگی یک نمونه برای دسته‌بندی شدن را نشان می‌دهد، و هر شاخه مقداری را که گره می‌تواند اختیار کند ارائه می‌دهد. نمونه‌های دسته‌بندی شده از گره اصلی شروع می‌شوند و براساس مقادیر ویژگی‌شان دسته‌بندی می‌شوند. درختان تصمیم شیوه‌ای برای نمایش یک سری از قوانین می‌باشند که منتهی به یک رده یا مقدار می‌شوند.

۴-۵- جنگل تصافی

جنگل تصادفی یک الگوریتم یادگیری نظارت شده محسوب می‌شود. همانطور که از نام آن مشهود است، این الگوریتم جنگلی را به طور تصادفی می‌سازد. جنگل ساخته شده، در واقع گروهی از درخت‌های تصمیم است. به بیان ساده، جنگل تصادفی چندین درخت تصمیم ساخته و آن‌ها را با یکدیگر ادغام می‌کند تا پیش‌بینی‌های صحیح‌تر و پایدارتری حاصل شوند.

۴-۶- شبکه‌های عصبی مصنوعی چند لایه

شبکه‌هایی که بر اساس فرایندهای یادگیری، الگوهای موجود در داده‌ها را شناسایی می کنند در اصطلاح، شبکه عصبی مصنوعی می‌نامند. یک نرون مصنوعی در حقیقت مدلی محاسباتی است که از نرون‌های عصبی واقعی انسان، الهام گرفته است. به هنگام مدل کردن اعصاب، از پیچیدگی‌های آن‌ها صرف نظر می‌شود و تنها به مفاهیم پایه ای بها داده می‌شود، چرا که در غیر این صورت رویکرد مدلسازی بسیار دشوار خواهد شد. در یک نگاه ساده، مدل یک عصب باید شامل ورودی‌هایی باشد که در نقش سیناپس انجام وظیفه کنند. این ورودی‌ها در وزنهایی ضرب می‌شوند تا قدرت سیگنال را تعیین کنند. نهایتاً یک عملگر ریاضی تصمیم‌گیری می‌کند که آیا نرون فعال شود یا خیر و اگر جواب مثبت باشد، میزان خروجی را مشخص می‌سازد. بنابراین شبکه عصبی مصنوعی با استفاده از مدل ساده شده عصب واقعی به پردازش اطلاعات می‌پردازد.

۵- پیاده سازی و تجزیه و تحلیل نتایج

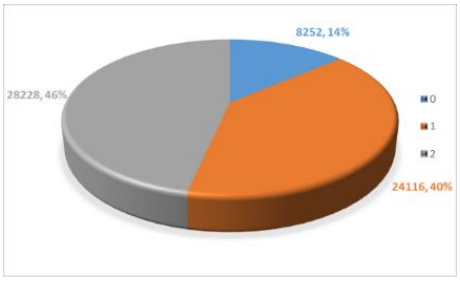
۵-۱- اطلاعات هویتی و شاخص‌های عملکری مشتریان

اطلاعات مشتریان مشترک بیمه آرمان و بانک ایران زمین در سه سرفصل کلی اطلاعات هویتی، شاخص‌های عملکردی در حوزه بانک و شاخص های عملکردی در حوزه بیمه گردآوری شده است. این اطلاعات در ارتباط با ۶۰۵۹۶ مشتری بوده است. پس از بررسی های اولیه و پیش پردازش اطلاعات در نهایت جدول نهایی اطلاعات مشتریان مشترک به صورت جدول ۵ قابل ارائه است. در این جدول سرفصل‌های سه گانه اطلاعت مشتریان طبقه بندی و نحوه‌ی کدینگ آنها در فرایند پردازش اطلاعات ارائه می شود. در ادامه توزیع برخی از ویژگی های مرتبط با اطلاعات هویتی مشتریان مشترک بانک و بیمه در مجموعه داده‌ی مورد مطالعه، به صورت نمودارهای دایره‌ای مصور شده است.

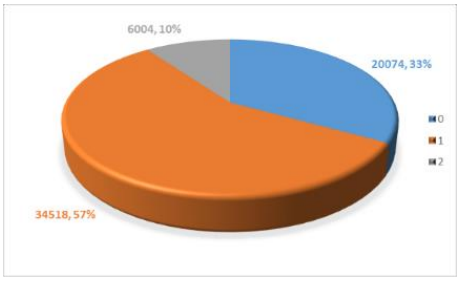
جدول ۵- شاخص های سه گانه اطلاعات مشتریان مشترک بانک و بیمه

سر فصل	نام شاخص در پایگاه داده	نام فارسی شاخص	نحوه ی کدینگ
اطلاعات هویتی	Cust_Brith_Date	سن	
	Sex_Desc	جنسیت	زن =۰، مرد = ۱، نامشخص = ۲
	Cust_Married_Desc	وضعیت تاهل	مجرد = ۰، متاهل = ۱، نامشخص = ۲
	Cust_Grad_Desc	وضعیت تحصیل	بیسواد = ۰، زیردیپلم =۱، دیپلم =۲، فوق دیپلم = ۳، لیسانس = ۴، فوق لیسانس = ۵،، دکتری = ۶، نامشخص = ۷

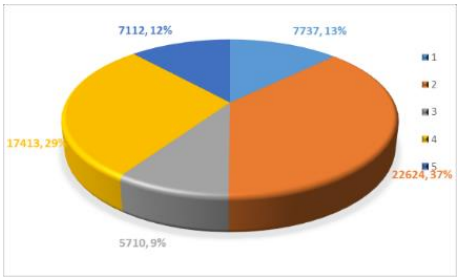
سر فصل	نام شاخص در پایگاه داده	نام فارسی شاخص	نحوه ی کدینگ
	Cust_Group_Desc	حقیقی یا حقوقی بودن	حقوقی = ۱، حقیقی = ۰
	count_card	تعداد کارت های مشتری	
	has_longterm_deposit	افتتاح سپرده بلند مدت	ندارد = ۰، دارد = ۱
	Dpst_Group_Desc	نوع سپرده	سپرده سرمایه گذاری بلند مدت = ۱، سپرده سرمایه گذاری کوتاه مدت = ۲، سپرده قرض الحسنه پس انداز = ۳، سپرده قرض الحسنه جاری = ۴، غیره =۵
	Count_Acc	تعداد حساب در بانک	
شاخص های عملکردی در حوزه بانک	m_totalmnt	مجموع مبلغ واریز	
	n_totalamnt	مجموع برداشت	
	abm_count	تعداد واریز	
	abn_count	تعداد برداشت	
	CHQ_total	تعداد چک های برگشتی در طول دوره	
	Avg_Amount_Trans	میانگین مبالغ تراکنش های مشتری	
شاخص های عملکردی در حوزه بیمه	haghebimeh	حق بیمه پرداختی	
	tedad_h	تعداد بیمه نامه صادره	
	khesarat	خسارت دریافتی	
	tedad_k	تعداد خسارت	



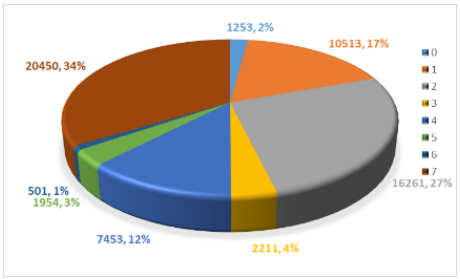
شکل ۴: نمودار دایره ای از توزیع وضعیت تاهل مشتریان مشترک بانک و بیمه در مجموعه داده‌ی مورد مطالعه



شکل ۳: نمودار دایره ای از توزیع جنسیت مشتریان مشترک بانک و بیمه در مجموعه داده‌ی مورد مطالعه



شکل ۶: نمودار دایره ای از توزیع نوع سپرده مشتریان مشترک بانک و بیمه در مجموعه داده‌ی مورد مطالعه



شکل ۵: نمودار دایره ای از توزیع وضعیت تحصیل مشتریان مشترک بانک و بیمه در مجموعه داده‌ی مورد مطالعه

۵-۲- نرمال سازی داده‌ها جهت نگاشت به بازه‌ی بین صفر و یک

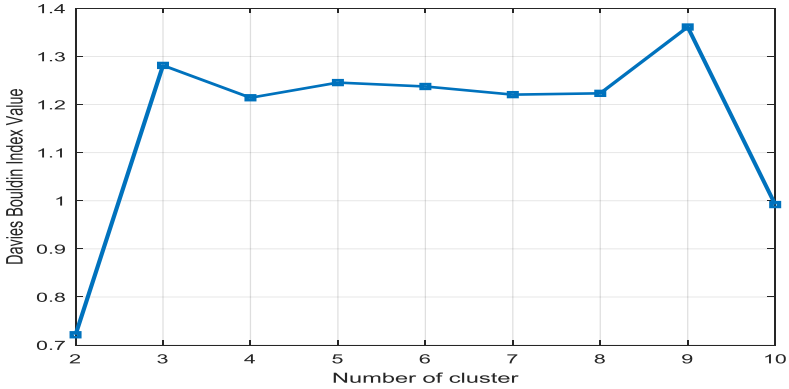
در این تحقیق از نرمال‌سازی Max-Min که یک انتقال خطی روی داده‌های اصلی ایجاد می‌کند استفاده شده است. به دلیل تفاوت در واحد هر یک از شاخص‌ها، لازم است تا مقادیر این شاخص‌ها بر اساس یک واحد یکسان، نرمال‌سازی گردند. در فرایند نرمال سازی شاخص های عملکردی منفی ضمن نگاشت در بازه بین صفر و یک جنبه‌ی مثبت پیدا می کنند. این شاخص ها شامل مجموع برداشت، تعداد چک های برگشتی در طول دوره، خسارت دریافتی توسط مشتری و تعداد خسارت می شوند.

۵-۳- تعیین تعداد خوشه مناسب

قبل از خوشه‌بندی داده‌ها ابتدا تعداد خوشه‌های بهینه را به‌دست می‌آوریم. در این تحقیق برای به‌دست آوردن تعداد خوشه‌های بهینه از شاخص دیویس بولدین استفاده می‌کنیم. برای این منظور تعداد۶۰۵۹۶ عضو را با K=2 تا K=10 خوشه، با استفاده از نگاشت های خود سازمانده (SOM) بخش‌بندی کردیم. در حقیقت این شاخص، میانگین حداکثر نسبت پراکندگی درون خوشه‌ای به پراکندگی بین خوشه‌ها را محاسبه می‌کند. هر چه مقدار شاخص دیویس بولدین کمتر باشد، عمل خوشه‌بندی بهتر صورت گرفته است. نمودار ۷ بهترین تعداد خوشه را نشان می‌دهد. در شکل ۷ محور افقی نشان‌دهنده تعداد خوشه‌ها و محور عمودی بیانگر شاخص دیویس بولدین می‌باشند. همچنین در این نمودار مشاهده می‌شود که خوشه‌بندی با دو خوشه دارای کیفیت بیشتری نسبت به تعداد خوشه‌های دیگر است.

۵-۴- خوشه‌بندی مشتریان با استفاده از نقشه‌های خودسازمانده (SOM)

همان‌طور که در قبل اشاره شد ابزار داده‌کاوی به‌منظور بخش‌بندی مشتریان بر اساس پارامترهای ارزش مشتری استفاده می‌شود. پس از تعیین تعداد خوشه‌های بهینه، جهت خوشه‌بندی مشتریان به منظور تفکیک آن‌ها در گروه‌های همگن بر اساس ارزش شاخص‌های عملکردی در حوزه های بانک و بیمه، از روش خوشه‌بندی نقشه‌های خود سازمانده استفاده شده است. از آنجائیکه این تحقیق به دنبال خوشه‌بندی مشتریان مشترک بیمه و بانک می‌باشد تا اعتبار مشتریان را براساس خوشه‌ای که به آن تعلق دارند بسنجد باید در هریک از خوشه‌های بدست آمده نماینده‌ای از آن خوشه وجود داشته باشد تا معرف آن خوشه باشد. از این رو روش‌های خوشه‌بندی براساس نقشه‌های خودسازمانده نیز از آن دسته است که می‌تواند مفید باشد. در جدول ۶ مرکز خوشه‌ها بر اساس ۱۹ بعد معرفی شده نشان داده شده است.



شکل ۷: مقدار شاخص دیویس بولدین برای تعداد خوشه های مختلف

جدول ۶- مرکز خوشه‌ها بر اساس ۱۹ بعد نرمال شده

نوع اطلاعات	شاخص	خوشه ۱	خوشه ۲
	میانگین نرمال شده سن	0.36717	0.292674
	میانگین نرمال شده جنسیت	0.30733	0.560768
	میانگین نرمال شده وضعیت تاهل	0.52821	0.977828
	میانگین نرمال شده وضعیت تحصیل	0.37073	0.978619
اطلاعات هویتی	میانگین نرمال شده حقیقی یا حقوقی بودن	0.00066	0.318330
	میانگین نرمال شده تعداد کارت های مشتری	0.00000	0.000243
	میانگین نرمال شده افتتاح سپرده بلند مدت	0.00084	0.216576
	میانگین نرمال شده نوع سپرده	0.42488	0.585904
	میانگین نرمال شده تعداد حساب در بانک	0.00012	0.000395
	میانگین نرمال شده مجموع مبلغ واریز	0.00000	0.000248
	میانگین نرمال شده مجموع برداشت	1.00000	0.999752
شاخص های عملکردی در	میانگین نرمال شده تعداد واریز	0.00002	0.000270
حوزه بانکی	میانگین نرمال شده تعداد برداشت	0.99973	0.999549
	میانگین نرمال شده تعداد چک های برگشتی در طول دوره	0.99925	0.999144
	میانگین نرمال شده میانگین مبالغ تراکنش های مشتری	0.00011	0.000726
	میانگین نرمال شده حق بیمه پرداختی	0.00000	0.000469
شاخص های عملکردی در	میانگین نرمال شده تعداد بیمه نامه صادره	0.00001	0.000099
حوزه بیمه‌ای	میانگین نرمال شده خسارت دریافتی	0.99999	0.999753
	میانگین نرمال شده تعداد خسارت	1.00000	0.999959

با توجه به قابلیت الگوریتم خوشه بندی مورد استفاده در این تحقیق در شناسایی الگوهای پیچیده نسبت به الگوهای خطی که با کمک شاخص‌ها استخراج می‌گردد می‌توان گفت که این روش اجماع خوبی روی داده‌ها ایجاد خواهد نمود. با توجه به اینکه داده واقعی در مورد برچسب مشتریان خوب و بد در دسترس نیست از این خروجی‌ها به عنوان برچسب برای فاز بعدی که الگوریتم‌های نظارت‌شده است استفاده می‌کنیم.

۵-۵- گامهای محاسبه امتیاز مشتریان و ارزیابی خوشه‌ها

برای امتیاز دهی به مشتریان موسسات مالی عمدتا از روشهای وزن‌دهی استفاده می شود. اختلافی که در موسسات رتبه بندی مشاهده می شود به گزینش معیارهای آن ها، بویژه اهمیت زیادی که آن ها به (Soft Factors) یا عوامل کیفی می‌دهند، باز می‌گردد. این عوامل شامل ارزیابی کیفیت مدیریت یا پیش‌بینی کلی دورنمای شرکت در بازار است. هرچند در این مورد، که کلیت فرآیندی که براساس آن موسسات رتبه بندی، رتبه نهایی خود را از معیارهای مورد اشاره استخراج می‌نمایند، مشخص نیست ولی ما فرض را بر این قرارمی دهیم که شروط پذیرفته شده برای رتبه بندی می تواند برای موسسات رتبه بندی و مدل های رتبه بندی داخلی بانک ها بطور یکسان اعمال گردد.

در این تحقیق امتیاز مشتریان بر اساس ۱۰ شاخص که شامل ۶ شاخص عملکردی در حوزه بانکی و ۴ شاخص عملکردی در حوزه بیمه‌ای است محاسبه می شود. در این راستا اوزان هر یک از این شاخص ها بر اساس نظرسنجی از خبرگان شرکت های بانک و بیمه تعیین می گردد. همچنین امتیاز هر یک از خوشه های بدست آمده از فرایند خوشه بندی مشتریان بر اساس میانگین امتیاز مشتریان در هر

خوشه قابل محاسبه است. بدیهی است خوشه با امتیاز بزرگتر حاوی مشتریانی با ارزش بیشتر می باشد. جدول ۷ نظر ۹ نفر از خبرگان حوزه‌ی بانک و بیمه به شاخص های عملکردی مشتریان مشترک بانک و بیمه را نشان می دهد. در واقع امتیاز مشتریان مشترک بانک ایران زمین و بیمه آرمان بر اساس اوزان بدست آمده در سطر آخر این جدول که برایند نظر خبرگان است محاسبه می‌شود.

با توجه به اوزان بدست آمده در جدول ۷ امتیاز مشتریان قابل محاسبه است. همانطور که در جدول ۸ مشاهده می‌شود. دو خوشه بدست آمده از فرایند خوشه بندی به ترتیب دارای حدودا ۷۰ و ۳۰ درصد مجموعه مشتریان مورد مطالعه هستند. بر اساس میانگین امتیازات بدست آمده از مشتریان هر خوشه مشخص گردید مشتریان قرار گرفته در خوشه ۲ دارای ارزش بیشتری بر اساس مجموع شاخص های عملکردی در حوزه بانک و بیمه هستند. این نتیجه از گزاراشات مندرج در جدول ۶ نیز تایید می شود. زیرا جدول ۶ که حاوی اطلاعات مرکز خوشه‌ها بر اساس ۱۹ بعد معرفی شده پس از نرمال سازی است، نشان می دهد در تمامی شاخص‌های عملکردی مشتریان خوشه ۲ وضعیت بهتری نسبت به مشتریان خوشه ۱ دارند.

جدول ۷- امتیاز خبرگان به هر یک از شاخص‌های عملکردی

خبره	شاخص‌های عملکردی در حوزه بانکی						شاخص‌های عملکردی در حوزه بیمه‌ای			
	مجموع مبلغ واریز	مجموع برداشت	تعداد واریز	تعداد برداشت	تعداد چک های برگشتی در طول دوره	میانگین مبالغ تراکنش های مشتری	حق بیمه پرداختی	تعداد بیمه نامه صادره	خسارت دریافتی	تعداد خسارت
مجموع امتیازات	45	42	15	15	45	30	45	31	45	31
مجموع نرمال شده (وزن)	0.1308	0.1221	0.0436	0.0436	0.1308	0.0872	0.1308	0.0901	0.1308	0.0901

جدول ۸- رتبه بندی مشتریان بر اساس امتیاز محاسبه شده برای مشتریان هر خوشه

خوشه	تعداد اعضای هر خوشه	درصد اعضای هر خوشه	رتبه خوشه
1	42216	69.67%	2
2	18380	30.33%	1

۵-۶- ارائه مدل‌های طبقه بندی برای پیش بینی اعتبار مشتریان جدیدالورود

در این مرحله به وسیله خروجی بدست آمده از فاز قبل (داده‌کاوی بدون نظارت) به عنوان برچسب، الگوریتم‌های رگرسیون لجستیک، درخت تصمیم، جنگل تصادفی و شبکه عصبی را آموزش داده و در پایگاه دانش ذخیره می‌نمائیم. هدف از اجرای این روش‌ها در متدولوژی پیشنهادی، پیش‌بینی کلاس یا طبقه مشتریان بر مبنای فرایند خوشه‌بندی است. در این مرحله به منظور طبقه‌بندی مشتریان داده‌های موجود به دو مجموعه آموزش (۹۰ درصد داده‌ها) و آزمایش با (۱۰ درصد داده‌ها) تقسیم شدند. از مجموعه داده‌های آموزش جهت یادگیری و مدل‌سازی و از مجموعه آزمایش برای ارزیابی مدل و دقت پیش‌بینی استفاده گردید. در این قسمت رگرسیون لجستیک، درخت تصمیم، جنگل تصادفی و شبکه عصبی که زیرمجموعه‌ای از روش‌های هوش‌مصنوعی به شمار می‌آیند بکار گرفته شده است. از میان انواع الگوریتم‌های درخت تصمیم، از الگوریتم C۴/۵ برای ساخت مدل‌های طبقه‌بندی استفاده شده است.

در این بخش طبقه بندی کننده‌های رگرسیون لجستیک، درخت تصمیم، جنگل تصادفی و شبکه های

عصبی با استفاده از نرم‌افزار MATLAB پیاده‌سازی شده است. خطای فاز آموزش با رگرسیون لجستیک ۰,۰۰۱۶، درهت تصمیم ۰,۰۰۰۱، جنگل تصادفی ۰,۰۰۰۰۹ و شبکه های عصبی ۰,۰۰۰۰۵ می‌باشد. در جدول ۹ تا ۱۲، دقت و خطای پیش‌بینی طبقه بندی مشتریان توسط این چها روش در قالب ماتریس درهم-ریختگی، به تفکیک خوشه‌ها مشاهده می‌شود.

جدول ۹- دقت و خطای پیش‌بینی خوشه مشتریان توسط رگرسیون لجستیک در داده‌های فاز آزمایش

پیش‌بینی \ واقعی	1	2
1	4235	5
2	10	1810

جدول ۱۰- دقت و خطای پیش‌بینی خوشه مشتریان توسط درهت تصمیم در داده‌های فاز آزمایش

پیش‌بینی \ واقعی	1	2
1	4242	4
2	3	1811

جدول ۱۱- دقت و خطای پیش‌بینی خوشه مشتریان توسط جنگل تصادفی در داده‌های فاز آزمایش

پیش‌بینی \ واقعی	1	2
1	4243	2
2	2	1813

جدول ۱۲- دقت و خطای پیش‌بینی خوشه مشتریان توسط شبکه عصبی چندلایه در داده‌های فاز آزمایش

پیش‌بینی \ واقعی	1	2
1	4244	1
2	1	1814

۶- ارزیابی مدل پیش بینی طبقه مشتریان جدیدالورود

پیش از برشمردن معیارهای ارزیابی باید مفهوم ماتریس درهم‌ریختگی روشن شود. این ماتریس چگونگی عملکرد الگوریتم رده‌بندی را با توجه به مجموعه داده ورودی به تفکیک انواع رده‌های مسئله نشان می‌دهد. مفاهیم ماتریس درهم‌ریختگی به شرح زیر تعریف می‌شوند:

- ✓ تعداد منفی‌های صحیح[۲] (TN): تعداد رکوردهایی که رده واقعی آن‌ها منفی بوده و الگوریتم رده‌بندی نیز آن‌ها را به‌درستی منفی تشخیص داده است.
- ✓ تعداد مثبت‌های ناصحیح[۳] (FP): تعداد رکوردهایی که رده واقعی آن‌ها منفی بوده ولی الگوریتم رده‌بندی آن‌ها را به‌اشتباه مثبت تشخیص داده است.
- ✓ تعداد منفی‌های ناصحیح[۴] (FN): تعداد رکوردهایی که رده واقعی آن‌ها مثبت بوده ولی الگوریتم رده‌بندی آن‌ها را به‌اشتباه منفی تشخیص داده است.
- ✓ تعداد مثبت‌های صحیح[۵] (TP): تعداد رکوردهایی که رده واقعی آن‌ها مثبت بوده ولی الگوریتم رده‌بندی آن‌ها را به‌درستی مثبت تشخیص داده است.

براساس مفاهیم فوق معیارهای زیر برای ارزیابی نتایج طبقه بندی استفاده می‌شود:

$$\text{Accuracy}=\frac{tp+tn}{tp+tn+fp+fn} \quad (۱)$$

$$\text{Recall}=\frac{tp}{tp+fn} \quad (۲)$$

$$\text{Precision}=\frac{tp}{tp+fp} \quad (۳)$$

نتایج جدول ۱۳ نشان می‌دهد که شبکه عصبی طراحی‌شده برای طبقه‌بندی مشتریان، ۹۹,۹۷ درصد داده‌های آزمایشی را به درستی تشخیص داده است، و با توجه به بالا بودن درصد سه معیار دقت، بازخوانی و صحت روش شبکه عصبی نتیجه می‌گیریم این روش نسبت به سه روش دیگر کارامدتر می‌باشد. برای تحلیل رفتار تراکنش مشتری که عوامل تأثیرگذار در ارزش مشتری در آن دخیل است، الگوریتم‌های دسته‌بندی به‌منظور بیان دانش و استخراج قوانین در ارتباط با هر بخش از مشتریان به‌کار می‌روند. در این

تحقیق الگوریتم‌های متفاوتی مورداستفاده قرارگرفته است و الگوریتم شبکه عصبی چندلایه به دلیل ارائه نتایج ارزیابی با دقت، صحت و بازخوانی بهتر از باقی الگوریتم‌ها بوده است.

جدول ۱۳: متوسط دقت، یادآوری و صحت مدل‌ها طبقه بندی کننده مشتریان

الگوریتم طبقه‌بندی	دقت	بازخوانی	صحت
رگرسیون لجستیک	99.75%	99.74%	99.67%
درخت تصمیم	99.88%	99.85%	99.87%
جنگل تصادفی	99.93%	99.92%	99.92%
شبکه‌عصبی	99.97%	99.96%	99.96%

هدف از انجام تحقیق حاضر ارائه یک مدل داده محور جهت ارزیابی مشتریان مشترک بانک و بیمه بود. در این راستا با توجه به عدم label گذاری مشتریان در پایگاه داده‌های بانک و بیمه از یک رویکرد دو مرحله-ای مبتنی داده کاوی بهره برده شد. تحلیل رفتار مشتریان و طبقه بندی آنها و نیز به کارگیری از یک روش ارزیابی بر اساس شاخص های عملکردی در حوزه‌ی بانک و بیمه تصویری از مشتریان پر ریسک و مشتریان کم ریسک بدست داد. در این تحقیق امتیاز مشتریان بر اساس ۱۰ شاخص که شامل ۶ شاخص عملکردی در حوزه بانکی و ۴ شاخص عملکردی در حوزه بیمه‌ای است محاسبه می شود. در این راستا اوزان هر یک از این شاخص ها بر اساس نظرسنجی از خبرگان شرکت‌های بانک و بیمه تعیین می گردد. همچنین امتیاز هر یک از خوشه های بدست آمده از فرایند خوشه بندی مشتریان بر اساس میانگین امتیاز مشتریان در هر خوشه قابل محاسبه است. بدیهی است خوشه با امتیاز بزرگتر حاوی مشتریانی با ارزش بیشتر می باشد. همانطور که در جدول ۸ مشاهده می‌شود. دو خوشه بدست آمده از فرایند خوشه بندی به ترتیب دارای حدودا ۷۰ و ۳۰ درصد مجموعه مشتریان مورد مطالعه هستند. بر اساس میانگین امتیازات بدست آمده از مشتریان هر خوشه مشخص گردید مشتریان قرار گرفته در خوشه ۲ دارای ارزش بیشتر یا ریسک کمتری بر اساس مجموع شاخص های عملکردی در حوزه بانک و بیمه هستند. این نتیجه از گزاراشات مندرج در جدول ۶ نیز تایید می شود. زیرا جدول ۶ که حاوی اطلاعات مرکز خوشه‌ها بر اساس ۱۹ بعد معرفی شده پس از نرمال سازی است، نشان می دهد در تمامی شاخص‌های عملکردی مشتریان خوشه ۲ وضعیت بهتری نسبت به مشتریان خوشه ۱ دارند.

۷- نتیجه‌گیری

در این تحقیق با ارائه خوشه‌های مشتریان مشترک بانک و بیمه و رتبه‌بندی آنان می‌توان استراتژیهای مؤثر در جهت افزایش سوددهی و بازاریابی برای هر یک از خوشه‌ها را تعیین و طرح‌های مناسب برای بازاریابی و رفتار با مشتریان بالقوه و یا متوسط ارائه نمود. در این تحقیق از یک مدل موزون جهت محاسبه ارزش هر مشتری استفاده‌شده است. بعد از دسترسی به داده‌ها، آن‌ها را نرمال‌سازی کرده سپس تعداد خوشه‌های بهینه به‌دست‌آمده و خوشه‌بندی مشتریان صورت گرفت. بر اساس نتایج بدست آمده از شاخص دیویس بولدین ۲ خوشه بهینه به‌دست آمد. سپس به کمک الگوریتم نگاشت خودسازمانده فرایند خوشه‌بندی پیاده سازی شد. در مرحله بعد شاخص‌های عملکردی مشتریان در حوزه‌ی بانک و بیمه توسط افراد خبره وزن دهی گردید و سپس به محاسبه امتیاز مشتریان پرداخته شد و خوشه‌ها رتبه‌بندی شده و با توجه به ویژگی‌ها و خصوصیات هر خوشه از مشتریان تجزیه و تحلیل‌های مرتبط ارائه گردید. در تحلیل رتبه‌بندی خوشه‌ها، خوشه دو از نظر ارزش مشتریان رتبه بالاتری را دریافت کرده است و درصد اعضای آن حدود ۳۰ درصد کل مشتریان تحت مطالعه بوده است. در انتها با هدف تعیین طبقه مشتریان جدیدالورود، داده‌ها توسط چهار الگوریتم طبقه‌بندی (رگرسیون لجستیک، درخت تصمیم C4.5، جنگل تصادفی و شبکه عصبی چندلایه) مدل‌سازی و با ارزیابی عملکرد هر الگوریتم، شبکه عصبی چندلایه با دقت بالاتر (۹۹,۹۷٪) نسبت به سه

الگوریتم دیگر به‌عنوان الگوریتم برتر مشخص گردید. علاوه بر شاخص‌های در نظر گرفته شده در این تحقیق برای خوشه‌بندی مشتریان پیشنهاد می‌شود شاخص‌های دیگری نظیر حوزه کاری مشتری، میزان خوش حسابی آنها در بازپرداخت وام‌های گرفته شده و غیره در نظر گرفته شود تا بدین ترتیب تجزیه‌وتحلیل بهتری بر روی رفتار مشتریان بر اساس مشخصات آن‌ها صورت گیرد.

فهرست منابع

* [4]. مینایی، بهروز؛ اصغری، فاطمه؛ "بکارگیری داده کاوی برای کشف مدل امتیازبندی و تحلیل رفتاری مشتریان بانک"، دومین کنفرانس داده کاوی ایران، دانشگاه صنعتی امیرکبیر، ۱۳۸۷.
* [1]. Filippidou, D., Keane, J.A., Svinterikou, S. and Murray, J., Data Mining for Business Improvement: Applying the HyperBank Approach, PADD'98 - 2nd Int.Conf., on the Practical Application of Knowledge Discoveryand Data Mining, Practical Application Company, 1998, pp.1-10.
* [2]. Fayyad, U., Shapiro, G.S. and Smyth, P., From Data Mining to Knowledge Discovery in Databases, AI Magazine, 1996, pp.37-54.
* [3]. Wiz, K., The Hidden World of Data Mining, Think Analytics Corporation, 2002, pp.1-21.
* [5]. Antonios Chorianopoulos and Konstantinos Tsiptsis, Data Mining Techniques in CRM: John Wiley and Sons, 2009.
* [6]. Lindgreen Adam and Antioco Michael, "Customer relaitonship management: the case of a European bank, "Marketing Intelligence & Planning. Vol. 23, no. 2, pp. 136-154, 2005.
* [7]. Nie G, Rowe. W, Zhang. L, Tian. Y, and Shi. Y, "Credit card churm forcasting by logistic regression and decision tree, "Expert Systems with Applications, 2011.
* [8]. Koh Hian Chye and Chan Kin Leong Gerry, Data mining and customer relationship marketing in the banking industry.: Singapore Management Review, 2000.
* [9]. Hung. C, Jiau. H Huang. Y, "Evaluation of neural networks and data mining methods on credit assessment task for class imbalance problem," Nonlinear Analysis: Real World Application, pp. pp 720-747, 2006.
* [10]. Jamshidi, M.B.; Gorjiankhanzad, M.; Lalbakhsh, A.; Roshani, S. A Novel Multiobjective Approach for Detecting Money Laundering with a Neuro-Fuzzy Technique. In Proceedings of the 2019 IEEE 16th International Conference on Networking, Sensing and Control (ICNSC), Ban_, AB, Canada, 9–11 May 2019; pp. 454–458.
* [11]. Magomedov, G.S.; Dobrotvorsky, A.S.; Khrestina, M.P.; Pavelyev, S.A.; Yusubaliev, T.R. Application of Artificial Intelligence Technologies for the Monitoring of Transactions in AML-Systems Using the Example of the Developed Classification Algorithm. Int. J. Eng. Technol. 2018, 7, 76–79.
* [12]. Dorofeev, D.; Khrestina, M.; Usubaliev, T.; Dobrotvorskiy, A.; Filatov, S. Application of Machine Analysis Algorithms to Automate Implementation of Tasks of Combating Criminal Money Laundering. In Digital Transformation and Global Society, DTGS 2018, Communications in Computer and Information Science; Alexandrov, D., Boukhanovsky, A., Chugunov, A., Kabanov, Y., Koltsova, O., Eds.; Springer: Cham, Switzerland, 2018.
* [13]. Plaksiy, K.; Nikiforov, A.; Miloslavskaya, N. Applying Big Data Technologies to Detect Cases of Money Laundering and Counter Financing of Terrorism. In Proceedings of the 2018 6th International Conference on Future Internet of Things and Cloud Workshops (FiCloudW), Barcelona, Spain, 6–8 August 2018; pp. 70–77.
* [14]. Sobreira Leite, Gleidson, Adriano Bessa Albuquerque, and Plácido Rogerio Pinheiro. "Application of Technological Solutions in the Fight Against Money Laundering—A Systematic Literature Review." Applied Sciences 9.22 (2019): 4800.
* [15]. Ryder, Nicholas. "Money laundering: A review of the UKs anti-money laundering policy." (2020).
* [16]. Tiwari, Milind, Adrian Gepp, and Kuldeep Kumar. "A review of money laundering literature: the state of research in key areas." Pacific Accounting Review (2020).

* [17]. Kamath, D., Pavithra, K., & Pujari, K. (2019). Data mining techniques applied in banking sector-A review. International Journal of Social and Economic Research, 9(3), 358-365.
* [18]. Kanchana, M., Chavan, P., & Johari, A. (2020). Detecting Banking Phishing Websites Using Data Mining Classifiers (No. 2855). Easy Chair.
* [19]. Shahbazi, F. (2020). Using Decision Tree Classification Algorithm to Design and Construct the Credit Rating Model for Banking Customers. IOSR Journal of Business and Management, 21(3, Series 2), 24-28.
* [20]. Dinçer, H., Yüksel, S., Canbolat, Z. N., & Pınarbaşı, F. (2020). Data Mining-Based Evaluating the Customer Satisfaction for the Mobile Applications: An Analysis on Turkish Banking Sector by Using IT2 Fuzzy DEMATEL. In Tools and Techniques for Implementing International E-Trading Tactics for Competitive Advantage (pp. 320-339). IGI Global.
* [21]. Kaur, M., Bhaddal, P., & Singh, G. (2019). Calculation of client credit risk prediction in banking sector using data mining.
* [22]. Hassani, H., Huang, X., & Silva, E. (2018). Digitalization and big data mining in banking. Big Data and Cognitive Computing, 2(3), 18.

یادداشت‌ها

[1] Davies Bouldin Index
[2] True Negative
[3] False Positive
[4] False Negative
[5] True Positive