

پیش‌بینی ارزش مشتریان جدید بانک بر مبنای مدل آر.اف.ام با استفاده از درخت تصمیم بهبودیافته در راستای کاهش حداکثر حافظه مورد نیاز

محمدرضا غلامیان *

عظیمه مظفری **

چکیده

یکی از مهم‌ترین فاکتورهای بانکداری در راستای کاهش هزینه‌ها و افزایش سودآوری، مدیریت و ارزیابی مشتریان با ارزش است. در دهه‌های اخیر محققان بسیاری به تجزیه و تحلیل ویژگی‌های مشتریان به منظور تعیین ارزش آن‌ها با استفاده از تکنیک‌های داده‌کاوی پرداخته‌اند و درخت تصمیم یکی از پرکاربردترین الگوریتم‌های داده‌کاوی در این زمینه است. از آنجایی که این الگوریتم برای ساخت درخت، تنها یک ویژگی را در یک زمان برای آزمون در هر گره در نظر گرفته و وابستگی بین ویژگی‌ها را نادیده می‌گیرد، بنابراین این مسئله باعث افزایش ماکزیمم حافظه مورد نیاز می‌شود. به منظور برطرف نمودن این مشکل، در این پژوهش روشی برای بهبود درخت تصمیم با استفاده از شبکه عصبی برای کشف وابستگی بین ویژگی‌ها با رویکرد کاهش ماکزیمم حافظه

* دکترای مهندسی صنایع، عضو هیئت‌علمی دانشگاه علم و صنعت، تهران.

** کارشناس ارشد مهندسی صنایع گرایش مدیریت سیستم و بهره‌وری، دانشگاه علم و صنعت، تهران. (نویسنده مسئول)،
azime.mozafari@yahoo.com

تاریخ پذیرش: ۱۳۹۵/۸/۱۶

تاریخ دریافت: ۱۳۹۴/۹/۲۰

۹۴ مطالعات مدیریت فناوری اطلاعات، سال پنجم، شماره ۱۷، پاییز ۹۵

مورد نیاز پیشنهاد شده که در کنار مدل آر.اف.ام برای پیش‌بینی ارزش مشتریان جدید استفاده می‌شود. نتایج آزمایش‌ها نشان می‌دهد که روش پیشنهادی با استفاده از وابستگی بین ویژگی‌ها می‌تواند ارزش مشتریان جدید را با ماکزیمم حافظه مورد نیاز کم‌تری نسبت به روش پایه پیش‌بینی کند.

کلیدواژگان: ارزش مشتری، خوشه‌بندی، درخت تصمیم، شبکه عصبی، مدل آر.اف.ام

مقدمه

بانکداری یکی از نهادهای مالی تأثیرگذار در اقتصاد هر کشور است و یکی از مهم‌ترین مسائل مورد بررسی در آن، عدم شناسایی رفتار مشتریان و اتخاذ تصمیم‌های سلیقه‌ای در برخورد با آنها است. به این ترتیب یکی از عوامل اصلی در توسعه بانکداری، مدیریت صحیح ارتباط با مشتری^۱ است که با ورود رقبای جدید و مسائلی از قبیل خصوصی‌سازی، اهمیت آن بیش از پیش نمایان شده است. از آنجاکه مشتریان مهم‌ترین عامل بقا در دنیای رقابت می‌باشند، ارتباط با مشتریان و شناخت ویژگی‌های رفتاری آنها یک موضوع قابل توجه در فعالیتهای بانکداری است.

شناسایی مشتریان مستلزم تجزیه و تحلیل ویژگی‌ها و طبقه‌بندی^۲ کردن آنها است که منجر به یافتن گروه‌هایی از مشتریان سودآور بر اساس ویژگی‌های آنها می‌شود. یکی از حوزه‌های در حال رشد برای محقق شدن این هدف، داده‌کاوی^۳ است. کشف دانش در پایگاه داده^۴ که داده‌کاوی خوانده می‌شود، عبارت است از فرایندی که با استفاده از تکنیک‌های هوشمند دانش را در مجموعه‌ای بزرگ از داده‌ها استخراج می‌کند. داده‌کاوی امکان استخراج دانش با ارزش از داده‌های قدیمی و پیش‌بینی نتایج حاصل از موقعیت‌های آینده را می‌دهد، هم‌چنین به تصمیم‌گیری بهینه جهت بهبود کارایی کمک می‌کند. خوشه‌بندی^۵ از مفیدترین کارکردهای داده‌کاوی است و به فرایند گروه‌بندی داده‌ها به خوشه یا کلاس گفته می‌شود به طوری که اشیا در یک کلاس دارای بیشترین شباهت با هم و بیشترین تفاوت با اشیا در دیگر کلاس‌ها باشند (هان و همکاران^۶، ۲۰۰۶)؛ بنابراین بانک‌ها می‌توانند با استفاده از تکنیک‌های داده‌کاوی چارچوبی برای

-
1. Customer Relationship Management: CRM
 2. Classification
 3. Data Mining
 4. Knowledge Discovery in Data base: KDD
 5. Clustering
 6. Han et al.

تجزیه و تحلیل ویژگی‌های مشتریان و شناسایی و پیش‌بینی ارزش آن‌ها داشته باشند. یکی از ابزارهای پیش‌بینی ارزش مشتریان، الگوریتم درخت تصمیم^۱ است که یکی از پرکاربردترین روش‌های تجزیه و تحلیل داده‌ها است. درخت تصمیم یک ساختار درختی شبیه نمودار جریان است که در آن هر گره داخلی^۲ بیانگر یک آزمون روی ویژگی برای ساخت درخت^۳ است (هان و همکاران، ۲۰۰۶). با توجه به این‌که اغلب پیاده‌سازی‌ها از درخت تصمیم تنها یک ویژگی برای آزمون در هر گره به منظور ساخت درخت را در نظر گرفته و وابستگی بین ویژگی‌ها را نادیده می‌گیرند، این مسئله باعث می‌شود که این الگوریتم با مشکلاتی از جمله بزرگی اندازه درخت^۴ روبرو شود و در نتیجه ماکزیمم حافظه مورد نیاز^۵ افزایش یابد. به این ترتیب در این پژوهش یک روش پیشنهادی به منظور بهبود درخت تصمیم با استفاده از الگوریتم شبکه عصبی^۶ به عنوان یک تکنیک تجزیه و تحلیل داده ارائه می‌شود. شبکه عصبی کاربردهای متفاوتی دارد و با توجه به این‌که برای بهبود وابستگی بین ویژگی‌ها در درخت تصمیم عملکرد بهتری دارد (لی و همکاران^۷، ۲۰۱۳؛ لی و چن^۸، ۲۰۱۰)، در این پژوهش برای کشف وابستگی بین ویژگی‌ها استفاده شده است (ین و لی^۹، ۲۰۱۱). لازم به ذکر است که در این پژوهش داده‌های مشتریان بر مبنای مدل آ.ر.اف.ام^{۱۰} استخراج شده‌اند که سه فاکتور تازگی^{۱۱} (R)، تعداد دفعات^{۱۲} (F) و ارزش پولی^{۱۳} (M) را برای هر رکورد یا مشتری در نظر

-
1. Decision Tree algorithm: DT
 2. Internal node
 3. Generate tree
 4. Tree size
 5. Maximum required memory
 6. Neural Network algorithm: NN
 7. Li et al.
 8. Li & Chan
 9. Yen & Lee
 10. Recently, Frequency, Monetary (RFM)
 11. Recency (R)
 12. Frequency (F)
 13. Monetary (M)

پیش‌بینی ارزش مشتریان جدید بانک بر مبنای ... ۹۷

می‌گیرد. (رزمی و قنبری، ۱۳۸۸؛ هوقز^۱، ۲۰۰۵؛ چن و چن^۲، ۲۰۰۹؛ بوتل^۳، ۲۰۰۹) در واقع در این مدل، فرض بر این است که مشتریان دارای ارزش بالای هر یک از شاخص‌های مدل، بهترین مشتریان هستند، البته تا زمانی که در آینده همانند گذشته رفتار نمایند (کینیقام و همکاران^۴، ۲۰۰۶).

دلیل استفاده از مدل آ.راف.ام آن است که در این مدل فقط نگرش‌های مالی مطرح نمی‌باشند و گرایش اصلی آن در تحلیل ویژگی‌ها، به سمت مسائل غیرمالی است (حسینی، ملکی و غلامیان، ۲۰۱۰).

بنابراین در این پژوهش هدف این است که برای کاهش میزان ماکزیمم حافظه مورد نیاز در پیش‌بینی ارزش مشتریان جدید بر مبنای مدل آ.راف.ام، از الگوریتم درخت تصمیم بهبودیافته به‌وسیله شبکه عصبی استفاده گردد.

پیشینه پژوهش

پژوهش‌های متنوعی در زمینه خوشه‌بندی و پیش‌بینی ارزش مشتریان بر مبنای مدل آ.راف.ام صورت گرفته است؛ از جمله سهرابی و خانلری (۲۰۰۷) در پژوهش خود به محاسبه ارزش دوره عمر مشتریان یک بانک خصوصی بر اساس مدل آ.راف.ام پرداختند و مشتریان را با استفاده از رویکرد خوشه‌بندی ک-میانگین^۵ بخش‌بندی نمودند که در پایان مشتریان با ارزش و سودآور بر اساس ارزش دوره عمر به ۸ خوشه تفکیک شده و ویژگی‌های آن‌ها مورد تحلیل قرار گرفت. فراجین و محمدی (۲۰۱۰) با استفاده از روش ک-میانگین و الگوریتم قوانین انجمنی به تجزیه و تحلیل رفتار مشتریان در یک سیستم بانکی بر مبنای مدل آ.راف.ام پرداختند. در این پژوهش مشتریان به سه گروه سودآور

1. Hughes
2. Cheng & Chen
3. Buttle
4. Keiningham et al.
5. K-Means

تقسیم شدند و قوانین رابطه‌ای از الگوریتم قوانین انجمنی برای تشخیص گروه‌های مشتریان با ایجاد پروفایل‌های مشتری مورد استفاده قرار گرفت. نامور، غلامیان و خاک‌آبی (۲۰۱۰) در پژوهش خود برای تحلیل مشتریان، از داده‌های آ.راف.ام و برای محاسبه ارزش دوره عمر مشتریان از الگوریتم کامیانگین استفاده نمودند. سیدحسینی، ملکی و غلامیان (۲۰۱۰) با استفاده از الگوریتم کامیانگین و مدل توسعه‌یافته آ.راف.ام به داده‌کاوی در پایگاه داده یک شرکت طراحی مهندسی و تأمین قطعات خودرو پرداختند. با توجه به حساس بودن کامیانگین به تعداد خوشه‌ها از شاخص دیویس بولدین برای تعیین تعداد خوشه و از روش درون خوشه^۱ برای محاسبه کیفیت خوشه‌بندی استفاده نمودند. کفاش‌پور، توکلی و علیزاده زوارم (۱۳۹۱) نیز در پژوهش خود با استفاده از روش کامیانگین مشتریان شرکت بازرگانی طوس شرق را بر مبنای مدل آ.راف.ام و ارزش دوره عمر مشتریان بخش‌بندی کردند که مشتریان در ۸ خوشه اصلی گنجانده شدند. خواجه‌وند و تارخ (۲۰۱۱) برای تخمین ارزش آینده مشتری از مدل آ.راف.ام، الگوریتم خوشه‌بندی کامیانگین و محاسبه ارزش دوره عمر مشتری استفاده کردند و سپس با استفاده از روش سری‌های زمانی به پیشگویی ارزش آینده هر بخش در هر خوشه پرداختند. زین‌العابدینی، مهدوی و خان‌بابایی (۱۳۹۱) در پژوهشی به شناسایی و تحلیل رفتار مشتریان بانکی پرداختند. آن‌ها در این پژوهش با استفاده از تکنیک‌های داده‌کاوی و مدل تحلیل آ.راف.ام به بخش‌بندی، شناسایی و تحلیل رفتار مشتریان بانکداری الکترونیکی و تعیین میزان ارزش هر یک از آن‌ها پرداختند. همچنین با استفاده از الگوریتم قوانین وابستگی به یافتن ارتباط بین ویژگی‌های مشتریان خدمات بانکداری الکترونیکی پرداختند. الوندی، فضل‌ی و عبدلی (۲۰۱۲) از الگوریتم خوشه‌بندی کامیانگین و مدل آ.راف.ام برای طبقه‌بندی مشتریان به منظور تبیین استراتژی‌های بازاریابی در یک بانکی دولتی استفاده کردند. نتایج خوشه‌بندی مشتریان را

در نه گروه بخش‌بندی کرد و از مدل آ.راف.ام برای تعیین ارزش طول عمر مشتری استفاده شد. رضایی‌نیا، کرامتی و البدوی (۲۰۱۲) از الگوریتم خوشه‌بندی کا-میانگین و مدل آ.راف.ام برای شناسایی و حفظ مشتریان با ارزش استفاده نمودند. در این پژوهش الگوریتم کا-میانگین مشتریان را به گروه‌های مختلف تقسیم کرد و روش تحلیل سلسله‌مراتبی برای محاسبه وزن معیارهای مدل آ.راف.ام استفاده شد. لیو و شیه^۱ (۲۰۰۵) در پژوهش خود متدولوژی ارائه دادند که از ترکیب تکنیک‌های تصمیم‌گیری گروهی و داده‌کاوی تشکیل شده بود. در این روش از مفهوم ارزش دوره عمر مشتری استفاده شده و برای خوشه‌بندی مشتریان بر اساس مدل آ.راف.ام از تکنیک خوشه‌بندی کا-میانگین استفاده شد. هو و جینگ^۲ (۲۰۰۸) به بررسی قابلیت مدل آ.راف.ام در بخش‌بندی مشتریان در شرکت‌های خدمات پس از فروش خودرو پرداختند. در این تحقیق مشتریان بر اساس روش خوشه‌بندی K-میانگین به هشت خوشه تفکیک شدند. در نهایت، پس از تحلیل ویژگی‌های مشتریان، ارزش دوره عمر آن‌ها در هر خوشه تعیین گردید. وو و همکاران^۳ (۲۰۰۹) با استفاده از مدل آ.راف.ام و روش خوشه‌بندی کا-میانگین به تحلیل ارزش مشتریان یک شرکت ساخت تجهیزات صنعتی پرداختند. پس از آماده‌سازی داده‌ها، مشتریان بر اساس شاخص‌های آ.راف.ام در شش گروه خوشه‌بندی شدند و ویژگی‌های مشتریان در قالب خوشه‌ها و با استفاده از تحلیل ارزش دوره عمر مشتری مورد تحلیل قرار گرفت. چنگ و چن^۴ (۲۰۰۹) مدلی بر مبنای ترکیب ارزش عددی متغیرهای آ.راف.ام و الگوریتم کا-میانگین با تئوری مجموعه‌های سخت^۵ پیشنهاد کردند. بر اساس این مدل پیشنهادی، وفاداری مشتریان با توجه به شاخص‌های آ.راف.ام و با تعیین تعداد خوشه‌ها به سه، پنج و هفت کلاس درجه‌بندی شد، سپس با کشف و

1. Liu & Shih
2. Hu & Jing
3. Wu et al.
4. Cheng & Chen
5. Rough set theory

توصیف ویژگی‌های مشتریان هر خوشه و با استخراج قوانین تصمیم‌گیری قابل فهم به ارزیابی و پیاده‌سازی مدیریت ارتباط با مشتری پرداخته شد. لی و همکاران^۱ (۲۰۱۱) با استفاده از یک روش خوشه‌بندی دو مرحله‌ای به تحلیل ویژگی‌های مشتریان یک کارخانه بافندگی پرداختند. مبنای تحلیل خوشه‌ای در این پژوهش، مدل آر.اف.ام بود و مشتریان با الگوریتم کا-میانگین به ۵ خوشه به منظور تعیین استراتژی‌های بازاریابی متمایز تفکیک شدند و تحلیل ویژگی‌های هر خوشه بر اساس مدل امتیازدهی آر.اف.ام انجام گرفت. وو و همکاران^۲ (۲۰۱۳) از الگوریتم خوشه‌بندی و طبقه‌بندی برای پیش‌بینی رفتار بیمه‌گذاران در صنعت بیمه عمر استفاده کردند. آن‌ها از الگوریتم کا-میانگین برای بخش‌بندی بیمه‌گذاران بر اساس شاخص‌های موردنظر و از مدل آر.اف.ام برای محاسبه ارزش بیمه‌گذاران استفاده کردند و در نهایت الگوریتم‌های درخت تصمیم C5 و مجموعه سخت را برای تعیین عوامل و فاکتورهای مؤثر در ارزش‌ها مقایسه کردند.

پژوهش‌هایی نیز در زمینه پیش‌بینی رفتار مشتریان با استفاده از روش‌های داده‌کاوی و از جمله درخت تصمیم انجام شده است که از آن جمله می‌توان موارد زیر را نام برد. خالصی و شکوهی (۱۳۸۹) برای پیش‌بینی و اعتبارسنجی مشتریان بانکی، تکنیک‌های داده‌کاوی درخت تصمیم، ماشین بردار پشتیبان و الگوریتم ژنتیک را مقایسه کردند که نتایج نشان داد درخت تصمیم عملکرد بهتری دارد و برای اعتبارسنجی دقیق‌تر مشتریان به سیستم بانکی ارائه گردید. اسدپور و شاهرودی (۱۳۹۲) برای جداسازی و رتبه‌بندی مشتریان خوش حساب بانک رفاه از داده‌کاوی استفاده نمودند. آن‌ها با استفاده از روش‌های درخت تصمیم و شبکه عصبی به تجزیه و تحلیل داده‌ها پرداخته و در پایان با مقایسه سه روش، مدل نهایی را انتخاب نموده و رده مشتریان جدید را بر اساس مدل نهایی پیشگویی نمودند. در پایان نیز با استفاده از درخت تصمیم نحوه ارتباط بانک با

1. Li et al.
2. Wu et al.

پیش‌بینی ارزش مشتریان جدید بانک بر مبنای ... ۱۰۱

مشتریان را تعیین کردند. لیانگ^۱ (۲۰۱۰) به بخش‌بندی مشتریان در صنعت تعمیر و نگهداری خودرو در تایوان پرداخت. در این مقاله ابتدا مشتریان با استفاده از الگوریتم شبکه عصبی و خوشه‌بندی کا-میانگین به سه گروه مشتریان با ارزش بالا، پایین و متوسط گروه‌بندی شدند. سپس الگوریتم درخت تصمیم برای استخراج خصوصیات هر بخش از مشتریان به کار رفت. لی و همکاران^۲ (۲۰۱۰) با استفاده از الگوریتم‌های داده‌کاوی به بخش‌بندی و پیش‌بینی ارزش مشتریان کارت اعتباری بانک‌ها در چین پرداختند. در این مدل ابتدا مشتریان با استفاده از الگوریتم کا-میانگین به چهار گروه دسته‌بندی شدند و سپس مدل‌های پیش‌بینی ارزش مشتریان برای بانک به‌طور جداگانه توسط چهار الگوریتم داده‌کاوی درخت تصمیم C5، CHAID، CART و شبکه عصبی ایجاد شدند. نتایج نشان داد که اطلاعات مفیدی از قوانین درخت تصمیم به دست آمده که بهترین مدل برای پیش‌بینی بوده است. بلاجی و سریویستا^۳ (۲۰۱۲) از الگوریتم خوشه‌بندی کا-میانگین برای تجزیه و تحلیل اولویت مشتری روی محصولات بیمه عمر و بهبود استراتژی فروش محصول استفاده کردند. سپس درخت تصمیم برای توضیح وابستگی بین متغیرهای مختلف از جمله وضعیت مالی و اولویت محصول استفاده شد. ایزدپرست و همکاران (۱۳۹۱) دو الگوریتم داده‌کاوی کا-میانگین و درخت تصمیم را برای دسته‌بندی بیمه‌گذاران و پیش‌بینی سطح خسارت آن‌ها در بیمه بدنه خودرو ایران مقایسه و ارزیابی کردند. الگوریتم کا-میانگین بیمه‌گذاران را در ۱۱ خوشه بخش‌بندی کرد و با تحلیل خوشه‌ها از روی ویژگی‌ها و فراوانی هر کدام از آن‌ها، بیمه‌گذاران را در سه سطح ریسک کم، متوسط و بالا گروه‌بندی کرد. سپس از الگوریتم درخت تصمیم به‌منظور پیش‌بینی ریسک بیمه‌گذاران استفاده نمودند. سووینی و رودپیش^۴ (۲۰۱۲) به ارزیابی روش‌های داده‌کاوی به‌منظور ارائه روش بهینه برای پیش‌بینی رفتار بیمه‌گذاران

1. Liang

2. Li et al.

3. Balaji & Srivatsa

4. Soeini & Rodpysch

در بیمه خودرو ایران پرداختند. در مرحله اول ابتدا بیمه‌گذاران بر اساس ویژگی‌های آن‌ها با استفاده از الگوریتم ک-میانگین به ۲، ۳، ۴، ۵، ۶ بخش خوشه‌بندی شدند. تعداد مناسب خوشه بر اساس شاخص مجموع مربعات خطا^۱ محاسبه شد و با توجه به نتایج این شاخص تعداد ۴ خوشه انتخاب شد. سپس در مرحله دوم از الگوریتم‌های درخت تصمیم C5، CART، CHAID، QUEST و الگوریتم شبکه عصبی و شبکه بیزین برای پیش‌بینی رفتار بیمه‌گذاران استفاده شد.

در این پژوهش روشی معرفی می‌گردد که برای پیش‌بینی ارزش مشتریان بر مبنای مدل آر.اف.ام از طریق بهبود درخت تصمیم‌گیری با استفاده از شبکه عصبی برای کشف وابستگی بین ویژگی‌ها از حافظه کمتری بهره گرفته و نتایج دقیقی نیز ارائه می‌کند.

روش پژوهش

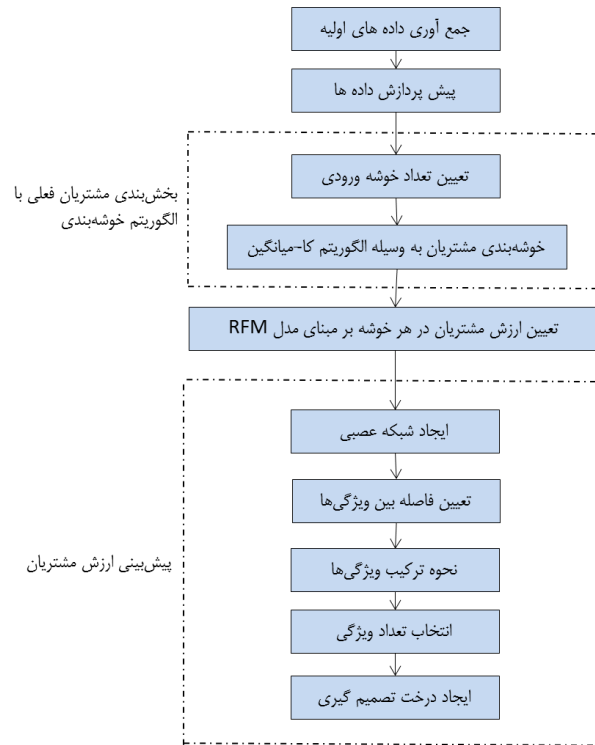
به‌منظور پیش‌بینی ارزش مشتریان روش‌های متعددی وجود دارند که با تجزیه و تحلیل ویژگی‌های در نظر گرفته شده از مجموعه داده، شناسایی ارزش هر یک از مشتریان و سودآوری آن‌ها را امکان‌پذیر می‌سازند. بدیهی است استفاده از یک الگوریتم نامناسب برای مجموعه داده مورد بررسی، ماکزیمم حافظه مورد نیاز نتایج خروجی را تحت تأثیر قرار می‌دهد. به همین دلیل در این پژوهش روشی پیشنهاد شده است که ضمن بهره‌گیری از قدرت شبکه عصبی، میزان ماکزیمم حافظه مورد نیاز را کاهش دهد.

برای انجام پژوهش حاضر ابتدا داده‌های اولیه مربوطه به مشتریان یک بانک خصوصی در شیراز بر مبنای مدل آر.اف.ام از پایگاه داده آن جمع‌آوری شدند. معمولاً بانک‌های اطلاعاتی به علت حجم زیاد داده و ارتباط با منابع اطلاعاتی مختلف ممکن است در معرض وجود داده مفقود یا گم‌شده، مغشوش و ناسازگار باشند، لذا داده‌های مشتریان در مرحله بعد پیش‌پردازش شدند تا برای انجام مراحل بعدی آماده باشند. پس از آن

1. Sum of Squares due to Error: SSE

پیش‌بینی ارزش مشتریان جدید بانک بر مبنای ... ۱۰۳

مشتریان با استفاده از الگوریتم کاسمیانگین بخش‌بندی شدند. الگوریتم خوشه‌بندی کاسمیانگین یک رویکرد بسیار محبوب در روش‌های بخش‌بندی به دلیل سهولت پیاده‌سازی و اجرای سریع آن است و به طور گسترده‌ای در تقسیم‌بندی بازار، شناخت الگو و غیره استفاده می‌شود (هان و همکاران، ۲۰۰۶). از آنجاکه برای استفاده از این الگوریتم کاسمیانگین باید تعداد خوشه‌ها از قبل مشخص باشند، لذا تعداد خوشه‌ها با استفاده از الگوریتم دومرحله‌ای^۱ تعیین شد. سپس ارزش مشتریان بر مبنای مدل آراف.ام تعیین گردید. در مرحله بعد برای پیش‌بینی ارزش مشتریان با استفاده از روش پیشنهادی، ابتدا ویژگی‌های ورودی و خروجی مجموعه داده مورد بررسی، آماده‌سازی شده و سپس شبکه عصبی موردنظر با تعیین پارامترهای اولیه برای این مجموعه داده آموزش داده شد. در ادامه با توجه به وزن‌های بین ویژگی‌های ورودی و اولین لایه پنهان به دست آمده از نتایج آموزش شبکه، ویژگی‌های ورودی مجموعه داده گروه‌بندی شدند. پس از آن با انتخاب تعداد ویژگی مناسب بنا بر گروه‌بندی تعیین‌شده از ویژگی‌ها، این مجموعه داده به‌عنوان یک مجموعه داده جدید که وابستگی بین ویژگی‌ها مشخص شده، برای ساخت درخت و مقایسه نتایج با دیگر ویژگی‌ها بر اساس گروه‌بندی مورد نظر، به‌منظور انتخاب بهترین نحوه گروه‌بندی از ویژگی‌ها و مقایسه و ارزیابی نتایج آن در نظر گرفته شد (ین و لی، ۲۰۱۱). مراحل انجام پژوهش در شکل ۱ نشان داده شده است.



شکل ۱. مراحل انجام پژوهش

قدم اول. جمع آوری داده های اولیه

داده های مورد استفاده در این پژوهش، از پایگاه داده یک بانک خصوصی در شیراز در بازه زمانی دو ساله مربوط به ۱۳۹۰/۱/۱ تا ۱۳۹۲/۱۲/۲۹ جمع آوری شده است. همانطور که پیش تر نیز گفته شد، داده های مربوط به شاخص های مدل آر.اف.ام برای انجام عملیات خوشه بندی و اطلاعات مربوط به مشتریان برای تحلیل خوشه ها در نظر گرفته شده است و دلیل انتخاب این مدل نیز آن است که ویژگی هایی از مشتری در نظر گرفته شوند که به خوبی نشان دهنده ارزش مشتری باشند. فایل اولیه داده ها حاوی ۳۰۹۶ رکورد (سطر) و ۱۳ ویژگی (ستون) بوده که هر رکورد بیانگر اطلاعات مربوط به یک

پیش‌بینی ارزش مشتریان جدید بانک بر مبنای ... ۱۰۵

مشتری است. ویژگی‌های مشتریان حاوی دو نوع اطلاعات می‌باشند که عبارت‌اند از:

۱. اطلاعات مربوط به مشخصات مشتریان شامل: شماره مشتری، نام و نام خانوادگی، جنسیت، تاریخ تولد، تحصیلات، وضعیت تأهل، مالکیت منزل، شغل و محل زندگی.
۲. اطلاعات مربوط به مدل آ.ر.اف.ام شامل: تازگی مبادله، تعداد مبادله و ارزش پولی مبادله.

قدم دوم. پیش‌پردازش داده‌ها

مجموعه داده جمع‌آوری شده مشتریان بانک از کیفیت لازم برای انجام عملیات داده‌کاوی برخوردار نبودند و شاید به این دلیل باشد که این اطلاعات فقط در راستای ذخیره سوابق مشتریان برای دسترسی به مشخصات آن‌ها کاربرد دارند و اصولاً برای کاربردهای داده‌کاوی جمع‌آوری نشده‌اند. به همین دلیل بعد از شناخت نوع و ابعاد داده‌ها به منظور استخراج دانش مفید، پیش‌پردازش شدند. به این ترتیب از میان ویژگی‌های اولیه تهیه شده فقط هشت ویژگی جنسیت، تاریخ تولد، تحصیلات، وضعیت تأهل، مالکیت خودرو، تازگی مبادله، تعداد مبادله و ارزش پولی مبادله در نظر گرفته شدند. در واقع باید ویژگی‌هایی در نظر گرفته شوند که هم قابلیت تفکیک مشتریان به گروه‌های مختلف را داشته باشند و هم بتوانند آن‌ها را با توجه به ارزشی که برای بانک دارند، شناسایی کنند. به این ترتیب ویژگی‌هایی مانند شماره مشتری، نام و نام خانوادگی، شغل و محل زندگی حذف گردید. در ادامه به منظور بهبود کیفیت داده‌ها برای عملیات داده‌کاوی، مقادیر پرت از ویژگی‌ها حذف شده و ویژگی‌هایی با مقادیر صفر و یا نامعلوم حذف شدند. همچنین ویژگی سن مشتری از روی ویژگی تاریخ تولد محاسبه گردید. در پایان تعداد رکوردهای پایگاه داده از ۳۰۹۶ رکورد به ۳۰۱۲ رکورد کاهش یافت. محدوده ویژگی‌های مورد استفاده در جدول ۱ نشان داده شده‌اند.

جدول ۱. معرفی شاخصه‌های مربوط به مشتریان

نوع ویژگی	شاخصه	محدوده تغییرات
مشخصات مشتریان	جنسیت	زن و مرد
	سن	از ۱۸ تا ۷۹ سال
	تحصیلات	از بی‌سواد تا دکتر
	وضعیت تأهل	مجرد، متأهل و غیره
	مالکیت خودرو	مالک و غیر مالک
مدل آر.اف.ام	تازگی مبادله	از ۲ تا ۶۵۹ روز
	تعداد مبادله	از ۲ تا ۲۳ بار
	ارزش پولی مبادله	از ۱۴۴۰۰۰ تا ۲۴۲۴۰۰۰ تومان

با توجه به این‌که نوع داده در نظر گرفته شده برای ایجاد روش پیشنهادی و مقایسه با روش پایه اهمیت دارد بنابراین در این پژوهش ویژگی‌های مشتریان به صورت طبقه‌بندی و گسسته در نظر گرفته شدند. در جدول ۲ چگونگی تبدیل ویژگی‌های مجموعه داده‌ها نشان داده شده است.

جدول ۲. چگونگی تبدیل ویژگی‌های انتخاب شده

ردیف	ویژگی	مقدار	شماره گروه	ردیف	ویژگی	مقدار	شماره گروه
۱	گروه جنسیت	مرد	۱	۵	گروه مالکیت منزل	مالک	۱
		زن	۲			غیر مالک	۲
۲	گروه سنی	۱۸-۲۴	۱	۶	گروه متغیر R	از ۲ تا ۱۳۳	۱
		۲۵-۳۴	۲			از ۱۳۴ تا ۲۶۴	۲
		۳۵-۴۴	۳			از ۲۶۵ تا ۳۹۶	۳

پیش‌بینی ارزش مشتریان جدید بانک بر مبنای ... ۱۰۷

ردیف	ویژگی	مقدار	شماره گروه	ردیف	ویژگی	مقدار	شماره گروه
			۷	۴	۴۵-۵۴	از ۳۹۷ تا ۵۲۷	۴
				۵	۵۵-۶۴	از ۵۲۷ تا ۶۵۹	۵
				۶	۶۴ به بالا	از ۲ تا ۵	۱
				۱	بی سواد	از ۶ تا ۹	۲
				۲	سیکل	از ۱۰ تا ۱۳	۳
۳	گروه تحصیلات		۸	۳	دیپلم	از ۱۴ تا ۱۷	۴
				۴	فوق دیپلم	از ۱۸ تا ۲۱	۵
				۵	کارشناسی	از ۱۴۴۰۰۰ تا ۴۵۷۶۰۰	۱
				۶	کارشناسی ارشد	از ۴۵۷۷۰۰ تا ۹۱۳۳۰۰	۲
				۷	دکتر و بالاتر	از ۹۱۳۴۰۰ تا ۱۳۶۹۰۰	۳
۴	گروه وضعیت تأهل		۸	۱	مجرد	از ۱۳۶۹۱۰ تا ۱۸۲۴۷۰۰	۴
				۲	متاهل	از ۱۸۲۴۸۰۰ تا ۲۲۸۰۴۰۰	۵

قدم سوم. بخش‌بندی مشتریان فعلی

تعیین تعداد خوشه ورودی:

در الگوریتم کا-میانگین، انتخاب تعداد مناسب خوشه برای تقسیم مشتریان به بخش‌های مختلف، بر روی نتایج نهایی تأثیر مستقیم دارد؛ بنابراین به منظور افزایش دقت خوشه‌بندی این الگوریتم، تعداد خوشه ورودی با استفاده از الگوریتم دومرحله‌ای تعیین

۱۰۸ مطالعات مدیریت فناوری اطلاعات، سال پنجم، شماره ۱۷، پاییز ۹۵

شده و تعداد خوشه ۷ به عنوان خروجی به دست آمد.

خوشه بندی مشتریان به وسیله الگوریتم کا-میانگین:

پس از تعیین تعداد خوشه بهینه با استفاده از الگوریتم دو مرحله ای، در این مرحله الگوریتم کا-میانگین با توجه به تعداد ۷ خوشه پیاده سازی شد. درصد فروانی و ویژگی های مشتریان هر ۷ خوشه در جدول ۳ نشان داده شده است.

جدول ۳. ویژگی های مشتریان در هر خوشه

ویژگی خوشه	تعداد اعضای خوشه	جنسیت	میانگین سنی	تحصیلات	تأهل	مالکیت منزل
خوشه اول	۶۳۵ (۲۱ درصد)	۵۲ درصد مرد ۴۸ درصد زن	۴۸ سال	۱۵ درصد بی سواد ۴۰ درصد سیکل ۱۷ درصد دیپلم ۱۸ درصد فوق دیپلم ۹ درصد کارشناسی ۱ درصد دکترا و بالاتر	۱۰۰ درصد مجرد ۰ درصد متأهل	۱۰۰ درصد مالک ۰ درصد غیر مالک
خوشه دوم	۲۶۶ (۹ درصد)	۱۰۰ درصد مرد ۰ درصد زن	۴۶ سال	۲۶ درصد بی سواد ۳۹ درصد سیکل ۱۵ درصد دیپلم ۱۲ درصد فوق دیپلم ۶ درصد کارشناسی	۰ درصد مجرد ۱۰۰ درصد متأهل	۰ درصد مالک ۱۰۰ درصد غیر مالک

پیش‌بینی ارزش مشتریان جدید بانک بر مبنای ... ۱۰۹

ویژگی خوشه	تعداد اعضای خوشه	جنسیت	میانگین سنی	تحصیلات	تأهل	مالکیت منزل
				۱ درصد کارشناسی ارشد ۱ درصد دکترا و بالاتر		
خوشه سوم	۲۵۹ (۸ درصد)	۱۰۰ درصد مرد ۰ درصد زن	۴۸ سال	۱۳ درصد بی‌سواد ۳۵ درصد سیکل ۲۲ درصد دیپلم ۲۱ درصد فوق‌دیپلم ۷ درصد کارشناسی ۱ درصد کارشناسی ارشد ۱ درصد دکترا و بالاتر	۰ درصد مجرد ۱۰۰ درصد متأهل	۰ درصد مالک ۱۰۰ درصد غیر مالک
خوشه چهارم	۲۰۸ (۷ درصد)	۱۰۰ درصد مرد ۰ درصد زن	۵۳ سال	۲۵ درصد بی‌سواد ۳۵ درصد سیکل ۱۷ درصد دیپلم ۱۷ درصد فوق‌دیپلم ۴ درصد کارشناسی ۲ درصد کارشناسی ارشد	۰ درصد مجرد ۱۰۰ درصد متأهل	۱۰۰ درصد مالک ۰ درصد غیر مالک
خوشه پنجم	۲۸۷ (۱۰ درصد)	۰ درصد مرد	۴۷ سال	۱۸ درصد بی‌سواد	۰ درصد مجرد	۱۰۰ درصد مالک

۱۱۰ مطالعات مدیریت فناوری اطلاعات، سال پنجم، شماره ۱۷، پاییز ۹۵

ویژگی خوشه	تعداد اعضای خوشه	جنسیت	میانگین سنی	تحصیلات	تأهل	مالکیت منزل
		۱۰۰ درصد زن		۲۹ درصد سیکل ۲۰ درصد دیپلم ۲۷ درصد فوق دیپلم ۴ درصد کارشناسی ۱ درصد کارشناسی ارشد ۱ درصد دکترا و بالا تر	۱۰۰ درصد متأهل	۰ درصد غیر مالک
خوشه ششم	۵۰۸ (۱۷ درصد)	۰ درصد مرد ۱۰۰ درصد زن	۴۴ سال	۱۶ درصد بی سواد ۲۷ درصد سیکل ۲۷ درصد دیپلم ۲۲ درصد فوق دیپلم ۶ درصد کارشناسی ۲ درصد کارشناسی ارشد	۰ درصد مجرد ۱۰۰ درصد متأهل	۰ درصد مالک ۱۰۰ درصد غیر مالک
خوشه هفتم	۸۴۹ (۲۸ درصد)	۴۸ درصد مرد ۵۲ درصد زن	۴۸ سال	۱۹ درصد بی سواد ۲۸ درصد سیکل ۲۰ درصد دیپلم ۲۶ درصد فوق دیپلم ۵ درصد کارشناسی	۱۰۰ درصد مجرد ۰ درصد متأهل	۰ درصد مالک ۱۰۰ درصد غیر مالک

ویژگی خوشه	تعداد اعضای خوشه	جنسیت	میانگین سنی	تحصیلات	تأهل	مالکیت منزل
				۱ درصد کارشناسی ارشد ۱ درصد دکترا و بالاتر		

قدم چهارم. تعیین ارزش مشتریان در هر خوشه

بعد از انجام عملیات خوشه‌بندی، ارزش مشتریان هر خوشه بر مبنای شاخص‌های مدل آر.اف.ام محاسبه شد. به این ترتیب که ارزش هر مشتری برابر است با میانگین ارزش متغیرهای آر.اف.ام. بعد از محاسبه ارزش مشتریان، خوشه‌ها با توجه به میانگین شاخص ارزش آن‌ها در سه سطح مشتریان با ارزش کم، متوسط و زیاد گروه‌بندی شدند. بدین معنی که خوشه‌هایی که میانگین شاخص ارزش نزدیک به هم دارند در کلاس ارزش یکسان قرار گرفتند (حنفی‌زاده و پایدار، ۲۰۱۳). ارزش مشتریان هر خوشه و نحوه کلاس‌بندی خوشه‌ها در جدول ۴ قابل مشاهده است.

جدول ۴. میانگین شاخص ارزش مشتریان هر خوشه

خوشه	میانگین شاخص ارزش	کلاس ارزشی
خوشه اول	۲/۵۷	زیاد
خوشه دوم	۲/۰۷	کم
خوشه سوم	۲/۶۴	زیاد
خوشه چهارم	۲/۳۳	متوسط
خوشه پنجم	۲/۳۵	متوسط
خوشه ششم	۲/۰۹	کم
خوشه هفتم	۲/۳۶	متوسط

به این ترتیب طبق جدول ۴ خوشه‌های اول و سوم با ارزش‌ترین مشتریان را دارند که به عنوان کلاس ۱ نام‌گذاری شدند. در واقع این کلاس ۲۹ درصد از مشتریان بانک را شامل می‌شود و در برگیرنده مشتریان کلیدی است که باید در حفظ و نگهداری آن‌ها تلاش شود. خوشه‌های چهارم، پنجم و هفتم دارای مشتریانی با کلاس سطح ارزش متوسط هستند، در واقع این مشتریان نسبتاً سودآور هستند و باید روی آن‌ها شناخت بیشتری صورت گیرد تا با اعمال استراتژی‌های مناسب سعی در جذب آن‌ها گردد و یا این‌که سعی شود هزینه‌های مربوط به آن‌ها کاهش یابد. خوشه‌های دوم و ششم دارای مشتریان کم ارزش می‌باشند که به عنوان زنگ خطری برای بانک می‌باشند که باید با اتخاذ استراتژی مناسب در مقابل آن‌ها سعی در بهبود رفتارشان داشت.

قدم پنجم. پیش‌بینی ارزش مشتریان

الگوریتم درخت تصمیم یکی از کاربردی‌ترین الگوریتم‌های طبقه‌بندی و پیش‌بینی است که محبوبیت آن به طور فزاینده‌ای در حال افزایش است. یک مسئله مهم در این الگوریتم عدم توجه به وابستگی بین ویژگی‌ها است که باعث افزایش حداکثر میزان حافظه مورد نیاز می‌گردد لذا در این پژوهش از شبکه عصبی به منظور کشف وابستگی بین ویژگی‌ها برای بهبود درخت تصمیم در پیش‌بینی ارزش مشتریان استفاده می‌شود.

ایجاد شبکه عصبی:

یکی از مهم‌ترین مسائل در هنگام استفاده از شبکه عصبی پیش‌پردازش داده‌های ورودی و خروجی آن است. با انجام این مرحله داده‌های ورودی و خروجی به شکلی که برای استفاده در شبکه مناسب باشند، تبدیل می‌شوند لذا فرایند نرمال‌سازی^۱ به عنوان تابع پیش‌پردازش داده‌ها در نظر گرفته می‌شود (جایا لاکشمی، ۲۰۱۱)^۲

1. Normalization
2. Jayalakshmi & Santhakumaran, 2011

با توجه به این‌که ویژگی‌های ورودی به طور گسسته می‌باشند، از روش نرمال‌سازی معیار 'z طبق رابطه زیر استفاده شد.

$$V' = \frac{V - \bar{A}}{\sigma_A}$$

که در آن \bar{A} میانگین و σ_A انحراف معیار صفت A است.

علاوه بر ویژگی‌های ورودی، ویژگی خروجی که همان ارزش مشتریان است، نیز پیش‌پردازش می‌شود. برای نرمال کردن داده خروجی از نرمال‌سازی کمینه - بیشینه با انتخاب تصادفی بازه ۰/۲ تا ۰/۸ استفاده شد. انتخاب این بازه تصادفی به این علت است که تابع فعال‌سازی سیگموئید^۲ که در حین آموزش شبکه استفاده می‌شود، به کران بالای ۰ و ۱ نزدیک می‌شود ولی هیچ‌گاه نمی‌تواند به طور دقیق ارزش ۰ و ۱ را برای ویژگی هدف به دست آورد (دموس و بیل^۳، ۱۹۹۳).

با فرض Min_A و Max_A به‌عنوان مقادیر حداقل و حداکثر یک ویژگی مانند A ، این نرمال‌سازی یک مقدار V از A را به دامنه $[new_MinA, new_MaxA]$ با محاسبه رابطه زیر نگاشت می‌کند:

$$V' = \frac{V - Min_A}{Max_A - Min_A} (new_MaxA - new_MinA) + new_MinA$$

سپس کل مجموعه داده به دو زیرمجموعه آموزشی (۷۰٪) و آزمایشی (۳۰٪) تقسیم شد که به ترتیب ۲۱۰۸ و ۹۰۴ تعداد رکورد را شامل می‌شوند. با توجه به وجود سه سطح ارزشی مشتریان در مجموعه داده، به جای انتخاب تصادفی از میان همه کلاس‌ها، زیرمجموعه‌ها طوری انتخاب شدند که از هر کلاس تعداد رکورد به نسبت مساوی در

1. Z-score or zero-mean normalization
2. Sigmoid activation function
3. Demuth & Beale

زیرمجموعه‌ها قرار گیرند. شبکه ساخته شده در این پژوهش، شامل یک لایه ورودی، یک لایه پنهان و یک لایه خروجی است. شبکه عصبی مصنوعی پرسپترون چندلایه پیشرو از نوع پس انتشار با ضریب همبستگی $0/944$ (R) و مقدار خطای $0/00227$ (MSE) مناسب‌ترین شبکه است. تعداد نرون لایه ورودی و پنهان برابر با ۵ نرون (تعیین‌کننده تعداد ویژگی‌های مشتریان) و تعداد نرون لایه خروجی برابر با ۳ نرون (تعیین‌کننده تعداد کلاس‌های ارزش مشتری) است. همچنین از تابع فعال‌سازی سیگموئید برای لایه پنهان و تابع فعال‌سازی خطی برای لایه خروجی استفاده شد. بعد از آموزش شبکه، به منظور ارزیابی صحت شبکه بر اساس زیرمجموعه داده آزمایشی و اطمینان از نتایج آن ماتریس متقاطع محاسبه شده که در شکل ۲ نشان داده شده است.

Confusion Matrix

	1	2	3	
1	0 0.0%	0 0.0%	0 0.0%	NaN% NaN%
2	0 0.0%	746 82.5%	0 0.0%	100% 0.0%
3	78 8.6%	0 0.0%	80 8.8%	50.6% 49.4%
	0.0% 100%	100% 0.0%	100% 0.0%	91.4% 8.6%
	1	2	3	

Target Class

شکل ۲. نتایج صحت آموزش شبکه در تعداد دور ۶۰ هزار

نتایج نشان داد که صحت شبکه $91/4$ درصد بوده و آموزش مناسبی داشته است. در این مرحله وزن بین ویژگی‌های ورودی و اولین لایه پنهان محاسبه شد.

تعیین فاصله بین ویژگی‌ها:

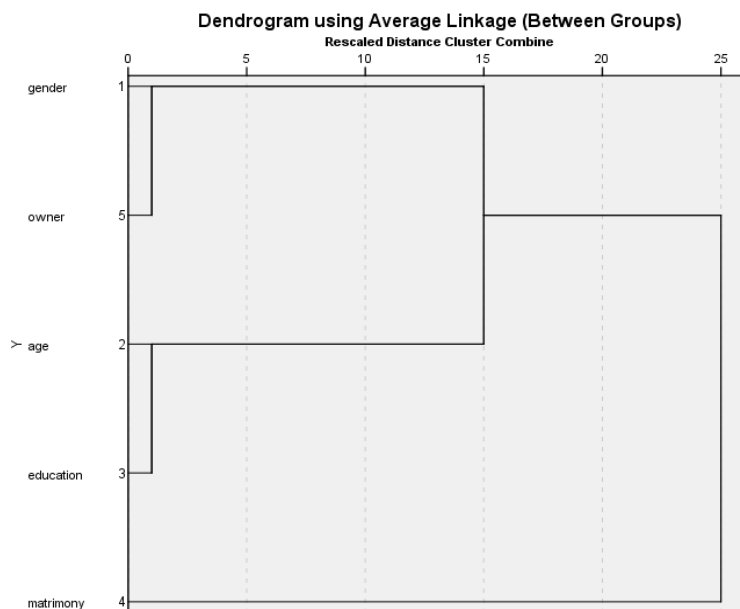
برای تعیین این‌که کدام یک از ویژگی‌های مجموعه داده مشتریان دارای وابستگی با

پیش‌بینی ارزش مشتریان جدید بانک بر مبنای ... ۱۱۵

یکدیگر می‌باشند، با توجه به ویژگی‌های موجود و وزن‌های به دست آمده در مرحله قبل، فاصله بین هر ویژگی با دیگر ویژگی‌ها، به وسیله معیار اندازه‌گیری فاصله اقلیدسی^۱ محاسبه شد.

نحوه ترکیب ویژگی‌ها:

در این مرحله نحوه ترکیب ویژگی‌ها در هر سطح، با استفاده از الگوریتم خوشه‌بندی میانگین اتصال^۲ مشخص شد. دیاگرام ایجاد شده از این روش با توجه به فاصله محاسبه شده در مرحله قبل، بر اساس پنج ویژگی ورودی جنسیت، تاریخ تولد، تحصیلات، وضعیت تأهل و مالکیت خودرو بنا بر مجموعه داده مشتریان بانک در شکل ۳ نشان داده شده است.



1. Euclidean distance
2. Average linkage

شکل ۳. وابستگی بین ویژگی‌ها در تعداد دور ۶۰ هزار

با توجه به شکل ۳، ابتدا در سطح اول ویژگی‌های جنسیت و مالکیت منزل، سن و تحصیلات ترکیب می‌شوند و سپس طبق سطوح مشخص شده این ترکیب ویژگی‌ها ادامه پیدا می‌کند. به این ترتیب، ویژگی‌ها تا دو سطح ترکیب شده و ۳ ویژگی تعیین شده و با داشتن سه ویژگی وارد مرحله بعد می‌شویم.

ایجاد درخت تصمیم‌گیری:

با توجه به این که درخت تصمیم به منظور پیش‌بینی ارزش مشتریان بر اساس زیرمجموعه آموزشی ساخته می‌شود، ابتدا مجموعه داده اولیه به دو زیرمجموعه داده آموزشی و آزمایشی به ترتیب با ۷۰٪ و ۳۰٪ داده و انتخاب به نسبت مساوی از هر کلاس تقسیم می‌شود. سپس الگوریتم درخت تصمیم C5 برای ساخت درخت پیش‌بینی ارزش مشتریان بر اساس مجموعه ویژگی‌های به دست آمده از نتایج خروجی شبکه که وابستگی بین آن‌ها تعیین شده، استفاده شد.

به منظور ارزیابی ماکزیمم حافظه مورد نیاز برای پیش‌بینی ارزش مشتریان، ابتدا درخت تصمیم C5 بر اساس نحوه ترکیب ویژگی‌ها با تعداد ۳ ویژگی ایجاد شد و سپس برای اطمینان از این که تعداد ویژگی انتخاب شده بهترین بهبود را در تعداد قوانین و عمق و سایز درخت ویژگی داشته است، درخت تصمیم با ۵ ویژگی نیز ایجاد شد. نتایج خروجی ساخت درخت تصمیم در جدول ۵ نشان داده شده است.

جدول ۵. مقایسه وابستگی بین ویژگی‌ها

ویژگی درخت				عمق	تعداد ویژگی مشتریان
صحت	تعداد قوانین (تعداد برگ)	اندازه (تعداد گره)	تعداد ویژگی مشتریان		
۹۰/۰۷ درصد	۱۰	۹	۸	۵	تعداد ویژگی مشتریان
۹۳/۳۸ درصد	۴	۳	۴	۳	

طبق نتایج جدول ۱۱، تعداد سه ویژگی باعث بهبود در اندازه درخت شده است. اندازه درخت بر مبنای عمق، تعداد گره و تعداد قوانین تولید شده سنجدیده می‌شود. به این ترتیب با روش جدید و پیشنهادی در این پژوهش، اندازه درخت کاهش یافته و لذا حداکثر حافظه مصرفی مورد نیاز کاهش می‌یابد.

نتیجه‌گیری

با گسترش رقابت و اهمیت بانکداری در اقتصاد کشور، تجزیه و تحلیل ویژگی‌های مشتریان بانک برای شناخت رفتار آن‌ها و پیش‌بینی ارزش مشتریان جدید، از اهمیت بالایی برخوردار است. امروزه به علت کند بودن و هزینه‌بر بودن تجزیه و تحلیل دستی داده‌ها، یکی از پرکاربردترین روش‌ها برای پیش‌بینی ارزش مشتریان، تکنیک‌های داده‌کاوی هستند که در این پژوهش مورد استفاده قرار گرفته‌اند. در پژوهش حاضر با استفاده از مدل آر.اف.ام و روش خوشه‌بندی به پیش‌بینی ارزش مشتریان پرداخته شد که از این نظر با پژوهش‌های لیو و شیبه (۲۰۰۵)، سهرابی و خانلری (۲۰۰۷)، هو و جینگ (۲۰۰۸)، وو، چنگ و لو (۲۰۰۹)، چنگ و چن (۲۰۰۹)، فراجین و محمدی (۲۰۱۰)، نامور، غلامیان و خاک‌آبی (۲۰۱۰)، سیدحسینی، ملکی و غلامیان (۲۰۱۰)، کفاش‌پور، توکلی و علیزاده زوارم (۱۳۹۱)، زین‌العابدینی، مهدوی و خان‌بابایی (۱۳۹۱)، لی، دای و تسنگ (۲۰۱۱)، خواجه‌وند و تارخ (۲۰۱۱)، الوندی، فضلی و عبدلی (۲۰۱۲)، رضایی‌نیا، کرامتی و البدوی (۲۰۱۲)، وو، وانگ و شینگ (۲۰۱۳) همسو است. ابزار مورد استفاده به منظور پیش‌بینی رفتار مشتریان در این پژوهش، درخت تصمیم است که از این نظر با پژوهش‌های خالصی و شکوهی (۱۳۸۹)، ایزدپرست و همکاران (۱۳۹۱)، اسدپور و شاهرودی (۱۳۹۲)، لیانگ (۲۰۱۰)، لی، وو، سان و ژانگ (۲۰۱۰)، بلاچی و سریویستا (۲۰۱۲)، سووینی و رودپیش (۲۰۱۲) هم‌جهت است.

پس از مطالعه پژوهش‌های پیشین و در نظر گرفتن معایب آن‌ها، در این پژوهش برای پیش‌بینی ارزش مشتریان یک بانک خصوصی در شیراز، روشی بر اساس مفهوم

وابستگی بین ویژگی‌ها پیشنهاد گردید. در روش پیشنهادی بر اساس وزن بین ویژگی‌های ورودی و اولین لایه پنهان به دست آمده از آموزش شبکه عصبی، فاصله بین ویژگی‌ها محاسبه شد. با در نظر گرفتن این فاصله، نحوه ترکیب ویژگی‌ها در هر سطح با توجه به وابستگی بین آن‌ها مشخص شد. سپس با ایجاد درخت تصمیم بنا بر تعداد ویژگی‌های مختلف، نتایج خروجی مورد مقایسه قرار گرفته و از نظر ماکزیمم حافظه مورد نیاز ارزیابی شد. با توجه به نتایج ارزیابی روش پیشنهادی مشخص شد که روش ارائه شده برای تعیین وابستگی بین ویژگی‌ها با رویکرد کاهش ماکزیمم حافظه مورد نیاز، می‌تواند نتایج بهتری داشته باشد، به این معنی که می‌توان بر اساس وابستگی بین ویژگی‌ها در پیش‌بینی ارزش مشتریان با رویکرد کاهش ماکزیمم حافظه مورد نیاز، به نتایجی بهتر از نتایج روش پایه دست یافت.

منابع

اسدپور، نوید؛ و شاهرودی، کامبیز. (۱۳۹۲). جداسازی و رتبه‌بندی مشتریان خوش‌حساب بانک رفاه با رویکرد داده‌کاوی. دومین همایش ملی بررسی راهکارهای ارتقاء مباحث مدیریت، حسابداری و مهندسی صنایع در سازمان‌ها.

ایزدپرست، سید محمود؛ و فتح‌نژاد، فرامرز. (۱۳۹۱). ارائه چارچوب برای پیش‌بینی سطح خسارت مشتریان بیمه بدنه اتومبیل با استفاده از راهکار داده‌کاوی، تازه‌های جهان بیمه، ۱۵۶، ۳۲-۱۵.

خالصی، نرگس؛ و شکوهی، امیرحسین. (۱۳۸۹). ارائه روشی جدید برای اعتبارسنجی مشتریان بانکی با استفاده از تکنیک‌های داده‌کاوی. چهارمین کنفرانس داده‌کاوی ایران. دانشگاه صنعتی شریف ۱۳۸۹. تهران.

رزمی، جعفر؛ و قنبری، آرش. (۱۳۸۸). ارائه مدلی نوین جهت محاسبه ارزش دوره عمر مشتری، نشریه مدیریت فناوری و اطلاعات، ۱(۱)، ۳۵-۵۰.

زین‌العابدینی، سیده فاطمه. مهدوی، مهرگان؛ و خان بابایی، محمد. به‌کارگیری تکنیک‌های داده‌کاوی و بازاریابی در بخش‌بندی و تحلیل رفتار مشتریان بانکداری الکترونیکی. کنفرانس ملی فناوری اطلاعات و جهاد اقتصادی.

کفاش‌پور، آذر. توکلی، احمد؛ و علیزاده زوارم، علی. (۱۳۹۱). بخش‌بندی مشتریان بر اساس ارزش دوره عمر آنها با استفاده از داده‌کاوی بر مبنای مدل آر.اف.ام. (RFM). پژوهش‌های مدیریت عمومی. (۱۵)، ۶۳-۸۴.

Alvandi, M. Fazli, S. & Abdoli, F. S. (2012). K-Mean clustering method for analysis customer lifetime value with LRFM relationship model in banking services. *International Research Journal of Applied and Basic Sciences*, 3(11), 2294-2302.

Balaji, S. & Srivatsa, S. K. (2012). Decision Tree induction based classification for mining Life Insurance Data bases. *International Journal of Computer Science and Information Technology & Security (IJCSITS)*, ISSN, 2249-9555.

Buttle, F. (2009). *Customer Relationship Management. Concepts and Tools*, Butterworthe.

Cheng, C. H. & Chen, Y. S. (2009). Classifying the segmentation of customer value via RFM model and RS theory. *Expert systems with*

- applications*, 36(3), 4176-4184.
- Demuth, H. & Beale, M. (1993). *Neural network toolbox for use with MATLAB*.
- Farajian, M. A. & Mohammadi, S. (2010). Mining the banking customer behavior using clustering and association rules methods. *International Journal of Industrial Engineering*, 21(4).
- Han, J. Kamber, M. & Pei, J. (2011). *Data mining: concepts and techniques: concepts and techniques*. Elsevier.
- Hanafizadeh, P. & Paydar, N. R. (2013). A data mining model for risk assessment and customer segmentation in the insurance industry. *International Journal of Strategic Decision Sciences*, 4(1), 52-78.
- Hosseini, S. M. S. Maleki, A. & Gholamian, M. R. (2010). Cluster analysis using data mining approach to develop CRM methodology to assess the customer loyalty. *Expert Systems with Applications*, 37(7), 5259-5264.
- Hu, W. & Jing, Z. (2008). *Study of segmentation for auto services companies based on RFM model*, [online], <http://www.pucsp.br/icim/ingles/downloads/pdf/proceeding_2008/66.pdf>.
- Hughes, A. M. (2005). *Strategic database marketing*. McGraw-Hill Pub. Co.
- Jayalakshmi, T. & Santhakumaran, A. (2011). Statistical normalization and back propagation for classification. *International Journal of Computer Theory and Engineering*, 3(1), 1793-8201.
- Keiningham, T. L. Aksoy, L. & Bejou, D. (2006). Approaches to the measurement and management of customer value. *Journal of Relationship Marketing*, 5(2-3), 37-54.
- Khajivand, M. & Tarokh, M. J. (2011). Estimating customer future value of different customer segments based on adapted RFM model in retail banking context. *Procedia Computer Science*, 3, 1327-1332.
- Li, D. C. Dai, W. L. & Tseng, W. T. (2011). A two-stage clustering method to analyze customer characteristics to build discriminative customer management: A case of textile manufacturing business. *Expert Systems with Applications*, 38(6), 7186-7191.
- Li, W. Wu, X. Sun, Y. & Zhang, Q. (2010, December). Credit card customer segmentation and target marketing based on data mining. *In Computational Intelligence and Security (CIS), 2010 International Conference on* (pp. 73-76). IEEE.
- Li, X. & Chan, C. W. (2010). Application of an enhanced decision tree learning approach for prediction of petroleum production. *Engineering Applications of Artificial Intelligence*, 23(1), 102-109.
- Li, X. Chan, C. W. & Nguyen, H. H. (2013). Application of the Neural Decision Tree approach for prediction of petroleum production. *Journal of Petroleum Science and Engineering*, 104, 11-16.
- Liang, Y. H. (2010). Integration of data mining technologies to analyze

- customer value for the automotive maintenance industry. *Expert Systems with Applications*, 37(12), 7489-7496.
- Liu, D. R. & Shih, Y. Y. (2005). Hybrid approaches to product recommendation based on customer lifetime value and purchase preferences. *Journal of Systems and Software*, 77(2), 181-191.
- Namvar, M. Gholamian, M. R. & KhakAbi, S. (2010, January). A two phase clustering method for intelligent customer segmentation. In *Intelligent Systems, Modelling and Simulation (ISMS), 2010 International Conference on* (pp. 215-219). IEEE.
- Rezaeinia, S. M. Keramati, A. & Albadvi, A. (2012). An integrated AHP-RFM method to banking customer segmentation. *International Journal of Electronic Customer Relationship Management*, 6(2), 153-168.
- Soeini, R. A. & Rodpysh, K. V. (2012). Evaluations of Data Mining Methods in Order to Provide the Optimum Method for Customer Churn Prediction: Case Study Insurance Industry. In *2012 International Conference on Information and Computer Applications (ICICA 2012) IPCSI* (Vol. 24).
- Sohrabi, B. & Khanlari, A. (2007). Customer lifetime value (CLV) measurement based on RFM model. *Iranian Accounting & Auditing Review*, 14(47), 7-20.
- Wu, H. H. Chang, E. C. & Lo, C. F. (2009). Applying RFM model and K-means method in customer value analysis of an outfitter. In *Global Perspective for Competitive Enterprise, Economy and Ecology* (pp. 665-672). Springer London.
- Wu, K. S. Wang, F. K. & Shyng, J. Y. (2013). Applied data mining techniques in insurance company: A comparative study of rough sets and decision tree. *African Journal of Business Management*, 7(24), 2309-2318.
- Yen, S. J. & Lee, Y. S. (2011). A neural network approach to discover attribute dependency for improving the performance of classification. *Expert Systems with Applications*, 38(10), 12328-12338.