Journal of Water and Wastewater Science and Engineering (JWWSE)
Vol. 8, No. 2, PP. 16-24, Summer 2023
DOI: 10.22112/JWWSE.2022.343238.1316

نشریه علمی
علوم و مهندسی آب و فاضلاب
سال هشتم، شماره ۲، صفحات ۱۶-۲۴، تابستان ۱۴۰۲

Research Paper

Simultaneous Analysis of Monthly Water Clustering Shiraz Water Subscribers Based on the Value of Their Life Cycle and RFM Model

Azime Mozafari[1*] and Razie Ranjbar[2]
1- Master of Industrial Engineering, Iran University of Science and Technology, Shiraz Water and Wastewater Company, Shraz, Iran.
2- Master of Industrial Engineering, Islamic Azad Shiraz University, Shraz, Iran.
*Corresponding Author, Email: azime.mozafari@yahoo.com

Received: 10/07/2022
Revised: 14/09/2022
Accepted: 26/11/2022
© IWWA

Abstract

Due to the importance of identifying low water consumption subscribers, in this study, using data mining methods, segmentation and identification of water subscribers based on the life cycle value and RFM model are discussed. First, the values of RFM model indices for Shiraz water subscribers were extracted from the subscribers' database and pre-processed. After weighting these three indicators using hierarchical analysis, subscribers were segmented using self-organizing neural network; Then the lifetime value pyramid was drawn to identify key subscribers and clusters of key and valuable subscribers were identified. The results show that the water supply index is the most important and the recent payment index is the least important in determining the value of subscribers and based on this, subscribers were divided into six clusters. According to the life cycle value pyramid, the most valuable subscribers are located in cluster 3, which are mainly residential, commercial and industrial uses, are mainly residents of Zone 2, most of these subscribers have paid their bills on time and their water's amount is mostly low. By identifying valuable and key subscribers, suggestions were made to improve the service to subscribers by considering their value.

Keywords: Value of subscribers' lifespan, Hierarchical analysis, Self-organizing neural network, Shiraz Water and WasteWater Company.

مقاله پژوهشی

بخش‌بندی مشترکین آب شیراز بر اساس ارزش دوره عمر آن‌ها و مدل RFM

عظیمه مظفری[1*] و راضیه رنجبر[2]
۱- کارشناسی ارشد مهندسی صنایع، دانشگاه علم و صنعت ایران، شرکت آب و فاضلاب شیراز، شیراز، ایران.
۲- کارشناسی ارشد مهندسی صنایع، دانشگاه آزاد اسلامی واحد شیراز، شیراز، ایران.
*نویسنده مسئول، ایمیل: azime.mozafari@yahoo.com

تاریخ دریافت: ۱۴۰۱/۰۴/۱۹
تاریخ اصلاح: ۱۴۰۱/۰۶/۲۳
تاریخ پذیرش: ۱۴۰۱/۰۹/۰۵
© انجمن آب و فاضلاب ایران

چکیده

با توجه به اهمیت شناسایی مشترکین کم‌مصرف آب، در این پژوهش با استفاده از روش‌های داده‌کاوی، بخش‌بندی و شناسایی مشترکین آب براساس ارزش دوره عمر و مدل RFM انجام می‌شود. ابتدا مقادیر شاخص‌های مدل RFM برای مشترکین آب شیراز از پایگاه اطلاعات مشترکین استخراج شده و پیش‌پردازش شد. پس از وزن‌دهی به این سه شاخص با استفاده از روش تحلیل سلسله مراتبی، مشترکین با استفاده از شبکه عصبی خودسازمان‌ده بخش‌بندی شدند. سپس هرم ارزش دوره عمر برای شناسایی مشترکین کلیدی رسم شده و خوشه‌های مشترکین کلیدی و با ارزش شناسایی شدند. نتایج نشان می‌دهند که شاخص مبلغ آب‌بها دارای بیشترین اهمیت و شاخص تازگی پرداخت دارای کمترین اهمیت در تعیین ارزش مشترکین بوده و بر این اساس مشترکین به شش خوشه تقسیم شدند. طبق هرم ارزش دوره عمر، با ارزش‌ترین مشترکین در خوشه ۳ قرار گرفته‌اند که عمدتا کاربری مسکونی و تجاری و صنعتی دارند، به‌طور عمده ساکن منطقه ۲ هستند، اکثر این مشترکین قبوض خود را به‌موقع پرداخت نموده‌اند و مبلغ آب‌بهای آن‌ها نیز عمدتا کم است. با شناسایی مشترکین با ارزش و کلیدی، پیشنهادهایی برای بهبود امر خدمت‌رسانی به مشترکین با در نظر گرفتن ارزشی که دارند، ارائه شد.

کلمات کلیدی: ارزش دوره عمر مشترکین، تحلیل سلسله مراتبی، شبکه عصبی خودسازمان‌ده، شرکت آب و فاضلاب شیراز.

۱- مقدمه

مصرف آب در ایران بیشتر از مصرف سرانه آب در مکان‌هایی است که از نظر آب و هوایی و زندگی اجتماعی و اقتصادی مشابه ایران هستند و لذا این موضوع نیازمند توجه ویژه به مدیریت مصرف آب است. رفتار مردم در مصرف آب به نگرش و میزان آگاهی آنان بستگی دارد (Foster and Beattie, 1981). لذا با توجه به شرایط کنونی بحران آب و افزایش تقاضا برای این ماده حیاتی و کالای ارزشمند، نگاه فراگیر به آب و مدیریت آن اهمیت خاصی دارد (امینی و همکاران، ۱۳۹۷-الف). برای اجرای موفقیت‌آمیز مدیریت مصرف آب، لازم است رفتار مشترکین یا مصرف‌کنندگان بررسی و درصورت لزوم اصلاح شود. از این‌رو امروزه بررسی الگوها و نحوه مصرف آن از اهمیت بالایی برای شرکت‌های آب و فاضلاب برخوردار است و ارائه هرگونه راه‌کاری برای کاهش مصرف و جلوگیری از هدررفت آن بسیار با ارزش است. شرکت‌های آب و فاضلاب همواره برای جلب رضایت مشترکین، تامین و توزیع آب با قیمت مناسب و حداقل نمودن هزینه‌ها تلاش کرده‌اند. این اهداف هنگامی میسر خواهد شد که این شرکت‌ها شناخت و درک صحیحی نسبت به رفتار مشترکین داشته باشند (کجوری و فریدونیان، ۱۳۹۴).

امروزه با توجه به تولید حجم انبوهی از داده‌ها و ذخیره آن‌ها، بازیابی اطلاعات از چنین پایگاه‌های بزرگی، می‌تواند یکی از نقاط قوت یک سازمان محسوب شود. رویکردهای متعددی برای گردآوری، ذخیره، سازمان‌دهی و مدیریت کارآمد داده‌های موجود و رسیدن به نتایج معنی‌دار به‌کار گرفته شده‌اند که داده‌کاوی[۱] یکی از رویکردهای اخیر در این زمینه است که از طریق آن داده‌ها می‌توانند به‌صورت موثر ذخیره و استخراج شوند (Monika and Amarpreet, 2018). داده‌کاوی فرآیندی است که از ابزارهای تحلیلی گوناگونی برای کشف الگوها و روابط بین داده‌ها استفاده می‌کند (Berry and Linoff, 2016). ازاین‌رو فرآیندی است کاملا علمی برای کشف دانش‌های پنهان و روابط ناشناخته بین داده‌ها که با نگرشی نو، به مسئله استخراج داده‌ها می‌پردازد (طباطبایی، ۱۳۹۸). داده‌کاوی مجموعه‌ای از روش‌هایی است که به شخص امکان می‌دهد تا ورای داده‌پردازی معمولی حرکت کند و به استخراج اطلاعاتی که در انبوه داده‌ها مخفی و یا پنهان است، کمک می‌کند. داده‌کاوی همچنین می‌تواند به عنوان بخشی از فرایند بزرگتر کشف دانش در بانک‌های اطلاعاتی در محیط‌های مختلف درنظر گرفته شود (غضنفری، ۱۳۹۹).

یکی از کاربردی‌ترین تکنیک‌های داده‌کاوی، الگوریتم خوشه‌بندی[۲] است که به‌منظور استخراج و شناسایی الگو از بین حجم انبوه داده‌ها استفاده می‌شود. این تکنیک داده‌های مشابه را در یک خوشه قرارداده و الگویی را به‌عنوان نماینده ارائه می‌کند به‌طوری‌که این نماینده بیانگر رفتار داده‌های آن خوشه است (آموزگار، ۱۳۹۵). در بحث خوشه‌بندی، هدف این است که داده‌هایی که در یک خوشه قرار دارند بیشترین تشابه را با یکدیگر و کمترین تشابه را با اعضای خوشه‌های دیگر داشته باشند (کجوری و فریدونیان، ۱۳۹۴). ازجمله تکنیک‌هایی که برای این منظور استفاده می‌شود، شبکه عصبی مصنوعی[۳] خودسازمان‌ده[۴] است که بر پایه اتصال به‌هم پیوسته چندین واحد پردازشی به‌نام نرون[۵]، ساخته می‌شود که در لایه‌های مشخصی آرایش یافته‌اند (Rosenblatt, 1962). مزایای این روش ازجمله تحمل بالای داده‌های مغشوش[۶]، کارکرد موازی، قابلیت استفاده زمانی که دانش بسیار کمی در مورد مسئله وجود دارد، استفاده از آن‌را در بسیاری از کاربردها افزایش داده است (Han et al., 2016). آموزش شبکه خودسازمان‌ده بر مبنای الگوریتم یادگیری بدون ناظر و فاصله اقلیدسی[۷] بنا شده است. پس از پایان مرحله آموزش، نقشه‌ای از نرون‌ها به‌دست می‌آید که در واقع چکیده‌ای از فضای مورد تحلیل شبکه است (Saggaf et al., 2003; Kreidler, 2018).

داده‌های موجود در بانک اطلاعاتی مشترکین شرکت آب و فاضلاب شیراز همواره در حال افزایش هستند، در حالی‌که در عمق این حجم داده‌ها، الگوها و روابط بسیار جالبی میان پارامترهای مختلف به‌صورت پنهان باقی می‌ماند که به شناسایی مشترکین و تعیین ارزش آن‌ها کمک می‌کند. این پژوهش درصدد است تا با استفاده از داده‌های موجود در بانک اطلاعاتی مشترکین شرکت آب و فاضلاب شیراز، به بخش‌بندی آن‌ها بپردازد. شناسایی مشترکین مستلزم تحلیل مشترکین هدف و دسته‌بندی کردن آن‌ها است که به یافتن گروه‌هایی از مشترکین کلیدی براساس ویژگی‌های آن‌ها منجر می‌شود. به این ترتیب با استفاده از خوشه‌بندی می‌توان مشترکین مشابه را در گروه‌های یکسان دسته‌بندی نمود. اما سؤال اصلی این است که چگونه می‌توان مشترکین کلیدی را شناسایی کرد؟ ارزش دوره عمر مفهومی است که می‌تواند در این راستا کمک فراوانی نماید. بر این اساس، مدل پیشنهادی بر مدل RFM[۸] است که برای استخراج الگوهای رفتار مشتری استفاده می‌شود و سه فاکتور تازگی (R)[۹]، تعداد دفعات (F)[۱۰] و ارزش پولی (M)[۱۱] را برای هر رکورد در نظر می‌گیرد (رزمی و قنبری، ۱۳۸۸؛ Buttle, 2018; Hughes, 1994; Cheng and Chen, 2009). در واقع در این مدل، فرض بر این است که مشترکین دارای ارزش بالای هریک

از شاخص‌های مدل، بهترین مشترکین هستند، البته تا زمانی‌که در آینده همانند گذشته رفتار نمایند (Keiningham et al., 2006). دلیل استفاده از مدل RFM آن است که در این مدل فقط نگرش‌های مالی مطرح نیست و گرایش اصلی آن در تحلیل ویژگی‌ها، به‌سمت مسائل غیرمالی است (SeyedHosseini et al., 2011).

نظرات مختلفی پیرامون اهمیت شاخص‌های مدل RFM وجود دارد (Bin et al., 2008) که در این پژوهش به‌منظور رتبه‌بندی این شاخص‌ها از روش تحلیل سلسله مراتبی (AHP)[۱۲] استفاده شده که یکی از تکنیک‌های تصمیم‌گیری چندمعیاره (MCDM)[۱۳] است. این روش که براساس مقایسات زوجی[۱۴] بنا شده، وزن هر یک از فاکتورها را در راستای گزینه‌های رقیب مشخص می‌سازد و در نهایت یک الگوریتم ریاضی به‌گونه‌ای ماتریس‌های حاصل از مقایسات زوجی را با همدیگر تلفیق می‌سازد که تصمیم بهینه حاصل شود (مهرگان، ۱۳۹۷). به این ترتیب تحقیق حاضر در پی پاسخگویی به سه سوال زیر است:

سوال اول: اولویت و اهمیت شاخص‌های RFM در تعیین ارزش مشترکین شرکت آب و فاضلاب شیراز چگونه است؟

سوال دوم: مشترکین شرکت آب و فاضلاب شیراز را می‌توان به چند بخش یا خوشه تقسیم نمود و مشترکین هر خوشه چه ویژگی‌هایی دارند؟

سوال سوم: کدام مشترکین برای شرکت آب و فاضلاب شیراز ارزش بیشتری دارند؟

۲- روش تحقیق

شرکت‌های آب و فاضلاب داده‌های زیادی راجع به مشترکین خود ذخیره و نگهداری می‌کنند. این درحالی است که در کشف دانش یا ارزش نهفته در این داده‌ها کم‌توان هستند. پایگاه داده مشترکین یکی از مهم‌ترین دارایی‌های این شرکت‌ها است که می‌توانند از آن برای توسعه استراتژی در درک بهتر رفتار مشترکین و مصرف‌کنندگان با استفاده از تجزیه و تحلیل مشخصات و رفتار مصرف هر مشترک استفاده نمایند. با توجه به اهمیت این موضوع، تحقیق حاضر با هدف بخش‌بندی مشترکین آب شیراز براساس مدل RFM انجام شده که برای این منظور از روش‌های تصمیم‌گیری چندمعیاره و داده‌کاوی در کنار مفهوم مهمی به‌نام هرم ارزش مصرف مشترکین استفاده شده است.

از این‌رو تحقیق حاضر به‌دلیل انجام در یک مجموعه مشخص (شرکت آب و فاضلاب شیراز) براساس طبقه‌بندی بر مبنای هدف، از نوع تحقیقات کاربردی است، زیرا نتایج حاصل از آن می‌تواند مورد استفاده سازمان‌ها و شرکت‌های آب و فاضلاب در اتخاذ سیاست‌های مناسب قرار گیرد. تحقیق حاضر با روش پیمایشی - تحلیلی به اجرا درآمده که در آن با ترکیب روش‌های تصمیم‌گیری چندمعیاره و داده‌کاوی چارچوبی برای خوشه‌بندی و پیش‌بینی رفتار مشترکین شرکت آب و فاضلاب شیراز ارائه شده است تا زمینه را برای شناسایی مشترکین کلیدی و انتخاب استراتژی‌های مناسب با توجه به ویژگی‌های مشترکین هر بخش فراهم نماید. داده‌های مورد استفاده از بانک اطلاعاتی مشترکین شرکت آب و فاضلاب شیراز استخراج شده‌اند و پژوهش در قالب دوره زمانی مقطعی و در بازه زمانی مربوط به ۱۴۰۰/۰۱/۰۱ تا ۱۴۰۰/۱۲/۲۹ انجام شده است و بانک اطلاعاتی مشترکین مورد استفاده در پژوهش شامل ۸۱۳،۶۸۵ رکورد یا مشترک است. مراحل انجام پژوهش در شکل ۱ نشان داده شده است.

گام اول	• بیان مسئله و تعیین هدف
گام دوم	• جمع‌آوری، پیش‌پردازش و آماده‌سازی داده‌های مشترکین شرکت آب و فاضلاب شیراز
گام سوم	• نرمال‌سازی داده‌های تحقیق
گام چهارم	• وزن‌دهی به شاخص‌های مدل RFM
گام پنجم	• تعیین ارزش شاخص‌های مدل RFM
گام ششم	• خوشه‌بندی مشترکین
گام هفتم	• تعیین متوسط ارزش شاخص‌های مدل در هر خوشه
گام هشتم	• تحلیل خوشه‌ای
گام نهم	• محاسبه ارزش دوره عمر مشترکین هر خوشه
گام دهم	• تحلیل مشترکین براساس ارزش دوره عمر آن‌ها

شکل ۱- گام‌های انجام پژوهش

۳- یافته‌ها

۳-۱- جمع‌آوری و آماده‌سازی داده‌های تحقیق

یکی از مهم‌ترین مراحل فرایند استخراج دانش، جمع‌آوری داده‌های خام و شناسایی اطلاعات مفید برای حل مسئله است. پس از شناخت مسئله باید در پی شناسایی اطلاعات مفید برای حل مسئله و انتخاب منابع داده مناسب بود. به‌طور کلی برای پیاده‌سازی تکنیک‌های داده‌کاوی باید معیارهایی درنظر گرفته شوند که بتوان براساس آن‌ها مشترکین را به گروه‌های مختلف خوشه‌بندی نمود. داده‌های مناسب از پایگاه داده مشترکین شرکت آب و فاضلاب در سال ۱۴۰۰ استخراج شد که به‌شرح جدول ۱ است.

جدول ۱- متغیرهای تحقیق

ردیف	شاخص	واحد	نوع داده	نوع متغیر	محدوده تغییر
۱	تعداد مشترکین	فقره	عددی	---	از ۱ تا ۸۱۳،۶۸۵
۲	نوع کاربری	---	اسمی	---	مسکونی، تجاری و صنعتی، عمومی و اداری، اماکن مذهبی و آموزشی، نظامی، غیردولتی
۳	محل سکونت	---	اسمی	---	از منطقه ۱ تا ۵
۴	آخرین پرداخت آب‌بها (R)	روز	تاریخ	معکوس	از ۰ تا ۱۰۳ روز
۵	تعداد دفعات پرداخت به‌موقع (F)	دفعه	عددی	مستقیم	از ۱ تا ۱۰ بار
۶	مبلغ هزینه‌های آب‌بها (M)	ریال	عددی	معکوس	از ۰ تا ۱۰،۵۲۱،۰۹۱ ریال

پس از شناخت نوع و ابعاد داده‌ها به منظور استخراج دانش مفید از میان آن‌ها باید داده‌ها را برای استفاده آماده کرد. به‌وسیله تکنیک‌های آماده‌سازی داده می‌توان کیفیت داده‌ها و در نتیجه کیفیت نتایج خروجی را افزایش داد. آماده‌سازی داده‌ها که تحت‌عنوان پیش‌پردازش شناخته می‌شود، بیشترین زمان را به‌خود اختصاص می‌دهد و طبق ارزیابی‌های انجام شده، یکی از مهم‌ترین مراحل کشف دانش است. به‌منظور آماده‌سازی داده‌های تحقیق حاضر نرم‌افزار Modeler به‌کار گرفته‌شده و اقدامات زیر انجام شده است:

- با توجه به این‌که برخی داده‌های عددی معکوس هستند، نیاز به نرمال‌سازی داده‌ها است. منظور از داده‌های معکوس عددی، داده‌هایی است که کمتر بودن آن‌ها نشان‌گر بالا بودن ارزش شاخص است. به‌عنوان مثال می‌توان از شاخص آخرین پرداخت نام برد.
- شاخص «تعداد مشترکین» صرفا بیان‌کننده تعداد رکوردهای موجود است و برای عملیات داده‌کاوی مورد استفاده قرار نمی‌گیرد.

همان‌طور که در جدول ۱ بیان شده، شاخص آخرین پرداخت (R) مربوط به فاصله زمانی بین آخرین پرداخت آب‌بهای صورت‌گرفته توسط مشترک تا پایان دوره مورد بررسی است که کمتر بودن این فاصله نشان‌گر بالا بودن ارزش این شاخص در مدل است. لازم‌به ذکر است که دوره بررسی به‌صورت سالیانه بوده و منظور از پایان دوره مورد بررسی، پایان سال ۱۴۰۰ است.

شاخص تعداد پرداخت به‌موقع (F) بیان‌گر تعداد دفعاتی است که مشترک در طول دوره زمانی یک ساله ۱۴۰۰، مبلغ آب‌بهای خود را به‌موقع پرداخت نموده و بیشتر بودن آن نشان‌گر بالابودن ارزش این شاخص در مدل است. لازم‌به ذکر است که تعداد پرداخت توسط تعداد دوره‌های قرائت مشترکین در طول سال ۱۴۰۰ تعیین می‌شود. شاخص ارزش پولی (M) بیان‌گر مبلغ هزینه آب‌بها است که کمتر بودن آن نشان‌گر بالابودن ارزش این شاخص در مدل است. به این معنی که مشترک، کم‌مصرف بوده و آب‌بهای آن نیز کم شده است و لذا یک مشترک ارزشمند با مصرف کم محسوب می‌شود.

۳-۲- نرمال‌سازی داده‌های تحقیق

با توجه به این‌که آموزش شبکه خودسازمان‌ده بر مبنای فاصله اقلیدسی است، نرمال بودن داده‌های آموزش شبکه می‌تواند نقشه‌های خروجی را به‌شدت تحت‌تاثیر قراردهد که با استفاده از روابط (۱) تا (۳) انجام می‌شود.

$$R' = \frac{R_{max} - R}{R_{max} - R_{min}} \quad (۱)$$

$$F' = \frac{F - F_{min}}{F_{max} - F_{min}} \quad (۲)$$

$$M' = \frac{M - M_{min}}{M_{max} - M_{min}} \quad (۳)$$

که R، F و M: مقادیر اصلی شاخص‌ها، R_{max}، F_{max} و M_{max}: بیشترین مقادیر شاخص‌ها، R_{min}، F_{min} و M_{min}: کمترین مقدار شاخص‌ها در طول سال ۱۴۰۰ و R'، F' و M' نیز مقادیر نرمال‌شده شاخص‌ها هستند.

۳-۳- وزن‌دهی به شاخص‌های مدل RFM

وزن نسبی شاخص‌ها، با استفاده از روش تحلیل سلسله مراتبی محاسبه شده و با W_M، W_F و W_R نشان داده می شود. با استفاده از نظرات ۱۰ متخصص صنعت آب و فاضلاب، ماتریس مقایسات زوجی برای معیارها و گزینه‌ها نسبت به شاخص‌ها تشکیل و با کمک نرم افزار Expert Choice وزن گزینه‌ها محاسبه شد. طبق خروجی این روش، وزن‌های نسبی شاخص تازگی پرداخت معادل ۰/۳۰، تعداد پرداخت به‌موقع معادل ۰/۳۴ و مبلغ آب‌بها معادل ۰/۳۶ تعیین شد. هم‌چنین در پایان برای اطمینان از درستی جواب، نرخ سازگاری محاسبه شد. نرخ سازگاری شاخصی است که میزان صحت و درستی قضاوت‌ها را نشان می‌دهد و بدین منظور باید مقدار آن کمتر از ۰/۱ باشد. اگر این مقدار بیشتر از ۰/۱ باشد قضاوت‌ها متضاد و تصادفی درنظر گرفته می‌شود که باید در آن‌ها تجدیدنظر به‌عمل آید. در این پژوهش نرخ سازگاری برابر با ۰/۰۷ محاسبه شد که کمتر از ۰/۱ است. وزن‌های به‌دست آمده میزان اهمیت شاخص‌ها را نشان می‌دهند، به این ترتیب در این قسمت به سوال اول تحقیق پاسخ داده شده و شاخص مبلغ آب‌بها (M) دارای بیشترین اهمیت و شاخص تازگی پرداخت (R) دارای کمترین اهمیت در تعیین ارزش مشترکین است.

۴-۳- تعیین ارزش شاخص‌های مدل RFM

ارزش هر شاخص از ضرب مقدار نرمال شده آن در وزن آن طبق روابط (۴) تا (۶) تعیین شده و به‌ترتیب با R''، F'' و M'' نشان داده می‌شوند.

$$R'' = R' \times W_R \quad (۴)$$

$$F'' = F' \times W_F \quad (۵)$$

$$M'' = M' \times W_M \quad (۶)$$

سپس متوسط ارزش هر یک از این شاخص‌ها با تقسیم مجموع ارزش آن شاخص در همه رکوردها به تعداد کل مشترکین طبق فرمول های (۷) تا (۹) تعیین می‌شود.

$$\bar{R}'' = \frac{\sum R''}{n} \quad (۷)$$

$$\bar{F}'' = \frac{\sum F''}{n} \quad (۸)$$

$$\bar{M}'' = \frac{\sum M''}{n} \quad (۹)$$

پس از تعیین ارزش شاخص‌های مدل، مقدار متوسط ارزش شاخص‌ها طبق دستورالعمل گفته شده، محاسبه شد که برای فاکتور تازگی پرداخت معادل ۰/۱۱۴۲، تعداد پرداخت به‌موقع معادل ۰/۱۲۲۰ و مبلغ آب‌بها معادل ۰/۲۰۸۱ به‌دست آمد.

۵-۳- خوشه‌بندی مشترکین

برای خوشه‌بندی از شبکه عصبی خودسازمان‌ده استفاده شد و نتایج حاصل از خروجی شبکه در قالب نقشه‌های گرافیکی تحلیل می‌شود. این عملیات در نرم‌افزار MATLAB انجام شده است. برای انجام خوشه‌بندی، داده‌های آموزش شبکه از ۸۱۳،۶۸۵ بردار سه‌بعدی شاخص‌های مدل تشکیل شده که هر بردار نماینده یک رکورد از رکوردهای پایگاه داده مشترکین شرکت آب و فاضلاب شیراز است. لازم‌به ذکر است که کلیه مشترکین انتخاب شده، دارای قرائت یکسان در بازه زمانی سال ۱۴۰۰ بوده‌اند.

در این قسمت به سوال دوم پاسخ داده شده و پس از انجام مراحل در پایان، ۶ بخش یا خوشه اطلاعاتی به‌دست آمد و هریک از مشترکین در یکی از این خوشه‌ها قرار می‌گیرند. نتایج خوشه‌بندی در جدول ۲ ذکر شده است.

۶-۳- تعیین متوسط ارزش شاخص‌های مدل در هر خوشه

متوسط ارزش هر یک از شاخص‌ها در هر خوشه از تقسیم مجموع ارزش شاخص در آن خوشه به تعداد مشترکین آن خوشه تعیین می‌شود که به‌ترتیب با $M_{M''}$، $M_{F''}$ و $M_{R''}$ نمایش داده شده و در جدول ۳ قابل مشاهده‌اند.

۷-۳- تحلیل خوشه‌ای

این تحلیل از طریق مقایسه متوسط ارزش شاخص‌ها در هر خوشه با متوسط ارزش شاخص‌ها در کل داده‌ها و هم‌چنین با مقایسه رتبه‌های خوشه‌ها در هریک شاخص‌ها صورت می‌گیرد. این مقایسه مشخص می‌کند که متوسط ارزش هریک از شاخص‌های مدل در هر خوشه نسبت به متوسط ارزش این شاخص‌ها در کل داده‌ها در چه وضعیتی قرار دارند. برای هریک از شاخص‌ها، وضعیت مطلوب، وضعیتی است که متوسط ارزش شاخص در یک خوشه بهتر از متوسط ارزش آن شاخص در کل داده‌ها باشد و با علامت (↑) نشان داده می‌شود. وضعیت نامطلوب یعنی وضعیتی که متوسط ارزش شاخص در یک خوشه بهتر از متوسط ارزش آن در کل داده‌ها نباشد و با علامت (↓) نشان داده می‌شود. نتایج حاصل در جدول ۴ نشان داده شده است.

جدول ۲- نتایج خوشه‌بندی مشترکین شهر شیراز در سال ۱۴۰۰

خوشه‌ها		۱	۲	۳	۴	۵	۶
تعداد مشترکین (فقره)		۴۰۱۵۲	۱۴۶۲۰۳	۲۷۱۱۰۲	۹۹۰۲۵	۲۰۱۰۲۵	۵۶۱۷۸
درصد مشترکین		۵	۱۸	۳۳	۱۲	۲۵	۷
نوع کاربری (درصد)	مسکونی و خانگی	۹۰	۹۱	۵۱	۸۰	۵۶	۶۸
	تجاری و صنعتی	۵	۵	۳۴	۱۴	۲۶	۱۰
	عمومی و اداری	۲	۱	۵	۲	۱۰	۶
	اماکن مذهبی و آموزشی	۱	۱	۱	۱	۳	۵
	نظامی	۱	۱	۱	۱	۱	۳
	غیردولتی	۱	۱	۸	۲	۴	۸
محل سکونت (درصد)	منطقه ۱	۱۲	۵	۹	۵	۶۹	۱۴
	منطقه ۲	۴	۸	۶۵	۵	۱۴	۱۰
	منطقه ۳	۹	۶	۱۲	۷۲	۵	۶
	منطقه ۴	۶۹	۱۰	۶	۱۰	۵	۵۵
	منطقه ۵	۶	۷۱	۸	۸	۷	۱۵
آخرین پرداخت آب‌بها (درصد)	کمتر از ۱۵ روز (مناسب)	۴۴	۷۷	۸۹	۳۲	۸۰	۲۲
	بیشتر از ۱۶ روز (نامناسب)	۵۶	۲۳	۱۱	۶۸	۲۰	۷۸
دفعات پرداخت به‌موقع (درصد)	از ۰ تا ۱ بار (کم)	۱۸	۸	۵	۲۵	۱۰	۵۹
	از ۲ تا ۵ (متوسط)	۳۲	۲۴	۱۴	۳۳	۲۱	۱۶
	بیشتر از ۶ بار (زیاد)	۵۰	۶۸	۸۱	۴۲	۶۹	۲۵
مبلغ هزینه‌های آب‌بها (درصد)	از ۰ تا ۲۰۰،۰۰۰ ریال (مصرف کم)	۲۷	۵۴	۵۹	۱۰	۵۸	۵۵
	از ۲۰۰،۰۰۰ تا ۱،۰۰۰،۰۰۰ ریال (مصرف متوسط)	۵۰	۳۸	۳۵	۷۹	۲۹	۲۸
	بیشتر از ۱،۰۰۰،۰۰۰ ریال (مصرف زیاد)	۲۳	۸	۶	۱۱	۱۳	۱۷

جدول ۳- متوسط ارزش شاخص‌ها در خوشه‌های مشترکین شهر شیراز در سال ۱۴۰۰

عنوان	مقدار
$M_{R''}$	۰/۲۵۸۷
$M_{F''}$	۰/۱۸۱۷
$M_{M''}$	۰/۰۸۵۲

جدول ۴- تحلیل خوشه‌های مشترکین شهر شیراز در سال ۱۴۰۰

خوشه‌ها		۱	۲	۳	۴	۵	۶
آخرین پرداخت ($M_{R''}$)	متوسط ارزش	۰/۲۳۵۴	۰/۲۶۳۲	۰/۲۷۸۵	۰/۲۱۰۴	۰/۲۸۷۵	۰/۲۲۴۱
	رتبه	چهارم	سوم	دوم	ششم	اول	پنجم
دفعات پرداخت‌به موقع ($M_{F''}$)	متوسط ارزش	۰/۱۷۹۴	۰/۲۰۱۵	۰/۲۱۲۰	۰/۱۸۰۰	۰/۱۹۱۱	۰/۱۲۲۶
	رتبه	پنجم	دوم	اول	چهارم	سوم	ششم
مبلغ آب‌بها ($M_{M''}$)	متوسط ارزش	۰/۰۶۸۸	۰/۰۸۵۵	۰/۰۹۷۵	۰/۰۷۱۴	۰/۰۸۵۸	۰/۰۷۲۱
	رتبه	ششم	سوم	اول	پنجم	دوم	چهارم
وضعیت متوسط ارزش شاخص‌ها ($M_{R''}$، $M_{F''}$، $M_{M''}$)		(↓،↓،↓)	(↓،↑،↑)	(↑،↑،↑)	(↓،↓،↓)	(↑،↑،↑)	(↑،↓،↓)

۳-۸- محاسبه ارزش دوره عمر مشترکین هر خوشه

ارزش دوره عمر مشترکین هر خوشه از مجموع وزین متوسط ارزش شاخص‌های RFM در آن خوشه طبق فرمول (۱۰) محاسبه می‌شود.

$$CLV = W_R \times M_{R''} + W_F \times M_{F''} + W_{M''} \times M_{M''} \quad (۱۰)$$

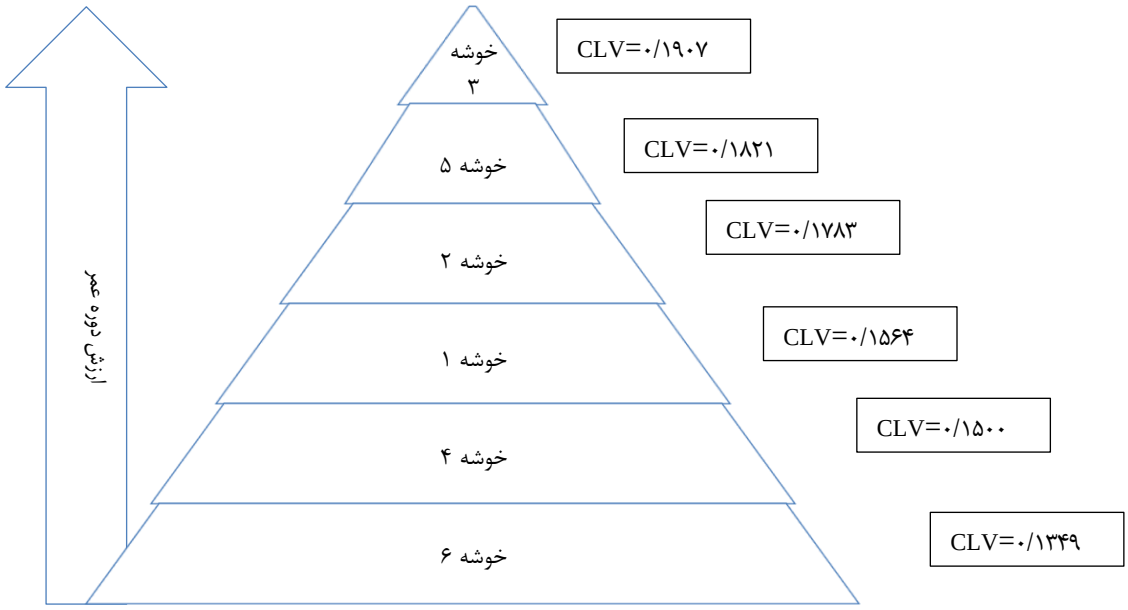
مقادیر ارزش دوره عمر در جدول ۵ قابل مشاهده است.

جدول ۵- مقادیر ارزش دوره عمر خوشه‌ها

خوشه‌ها	۱	۲	۳	۴	۵	۶
ارزش دوره عمر	۰/۱۵۶۴	۰/۱۷۸۳	۰/۱۹۰۷	۰/۱۵۰۰	۰/۱۸۲۱	۰/۱۳۴۹
	چهارم	سوم	اول	پنجم	دوم	ششم

۳-۹- تحلیل مشترکین بر اساس ارزش دوره عمر آن‌ها

در این گام هرکدام از خوشه‌ها براساس ارزش دوره عمر در قالب هرم ارزش عمر مشخص می‌شوند. ه چه از سطح پایین این هرم به سطوح بالای آن حرکت شود، خوشه‌های با ارزش دوره عمر بیشتر قرار گرفته‌اند. در این قسمت به سوال سوم پاسخ داده شده و خوشه‌ها براساس ارزش دوره عمر مشترکین آن‌ها، در قالب هرم ارزش عمر که در شکل ۱ نشان داده شده، مشخص شدند. در این هرم، به ارزش دوره عمر مشترکین در شش خوشه اصلی اشاره شده است. هر چه از سطح پایین این هرم به سطوح بالای آن حرکت شود، خوشه‌های با ارزش دوره عمر بیشتر قرارگرفته‌اند. به‌عبارتی، با حرکت به سمت رأس این هرم، مشترکینی قراردارند که ارزش و اهمیت بیشتری دارند.



شکل ۱- هرم ارزش دوره عمر مشترکین شرکت آب و فاضلاب شیراز

به این ترتیب طبق هرم ارزش دوره عمر، مشترکین خوشه ۳ دارای بیشترین ارزش عمر هستند. مشترکین این خوشه ۳۳ درصد از کل مشترکین را تشکیل می‌دهند، عمدتا کاربری مسکونی و تجاری و صنعتی دارند، به‌طور عمده ساکن منطقه ۲ هستند، به تازگی (کمتر از ۱۵ روز) قبوض خود را پرداخت نموده‌اند، تعداد دفعات پرداخت به‌موقع آن‌ها در این دوره یک‌ساله بیشتر از ۶ بار بوده و طبیعی است که مبلغ آب‌بهای آن‌ها نیز عمدتا کمتر از ۲۰۰،۰۰۰ ریال است.

بعد از خوشه ۳ مشترکین خوشه ۵ از اهمیت بالایی برخوردارند. مشترکین این خوشه ۲۵ درصد از کل مشترکین را تشکیل می‌دهند، عمدتا کاربری مسکونی و تجاری و صنعتی دارند، به‌طور عمده ساکن منطقه ۱ هستند، به تازگی (کمتر از ۱۵ روز) قبوض خود را پرداخت نموده‌اند، تعداد دفعات پرداخت به‌موقع آن‌ها در این دوره یک‌ساله بیشتر از ۶ بار بوده و مبلغ آب‌بهاء آن‌ها نیز عمدتا کمتر از ۲۰۰،۰۰۰ ریال است.

اولویت بعدی با خوشه ۲ است که حدود ۱۸ درصد از کل مشترکین را شامل می‌شود که عمدتا کاربری مسکونی دارند، به‌طور عمده ساکن منطقه ۵ هستند، این مشترکین به تازگی (کمتر از ۱۵ روز) قبوض خود را پرداخت نموده‌اند، تعداد دفعات پرداخت به‌موقع آن‌ها در این دوره یک‌ساله بیشتر از ۲ بار بوده و مبلغ آب‌بهای آن‌ها نیز عمدتا کمتر از ۱،۰۰۰،۰۰۰ ریال است.

خوشه ۱ در اولویت بعدی است که ۵ درصد مشترکین را شامل می‌شود که عمدتا کاربری مسکونی دارند، به‌طور عمده ساکن منطقه ۴ هستند، تعداد دفعات پرداخت به‌موقع آن‌ها در این دوره یک‌ساله بیشتر از ۲ بار بوده و مبلغ آب‌بهای آن‌ها نیز عمدتا کمتر از ۱،۰۰۰،۰۰۰ ریال است.

اولویت بعدی با خوشه ۴ است که حدود ۱۲ درصد از کل مشترکین را شامل می‌شود که عمدتا کاربری مسکونی دارند، به‌طور عمده ساکن منطقه ۳ هستند، عمدتا بیشتر از ۱۶ روز قبوض خود را پرداخت نموده‌اند، تعداد دفعات پرداخت به‌موقع آن‌ها در این دوره یک‌ساله بیشتر از ۲ بار بوده و مبلغ آب‌بهای آن‌ها نیز عمدتا بین ۲۰۰،۰۰۰ تا ۱،۰۰۰،۰۰۰ ریال است.

خوشه ۶ معرف مشترکینی است که کمترین ارزش را دارند و تنها ۷ درصد از کل مشترکین را شامل شده، عمدتا کاربری مسکونی دارند، به‌طور عمده ساکن منطقه ۴ هستند، عمدتا بیشتر از ۱۶ روز قبوض خود را پرداخت نموده‌اند، تعداد دفعات پرداخت به‌موقع آن‌ها در این دوره یک‌ساله کمتر از ۲ بار بوده و مبلغ آب‌بهای آن‌ها نیز عمدتا بین ۲۰۰،۰۰۰ تا ۱،۰۰۰،۰۰۰ ریال است.

۴- بحث و جمع‌بندی

با توجه به آن‌چه درمورد تکنیک‌های داده‌کاوی گفته شد و ضمن درنظر گرفتن گسترش فناوری اطلاعات و ارتباطات در سازمان‌ها، با به‌کارگیری بانک‌های اطلاعاتی موجود و استفاده از ابزارها و الگوریتم‌های مفیدی مانند داده‌کاوی، می‌توان ماهیت پیچیده داده‌های مرتبط با رفتار مصرفی مشترکین و روابط نامحسوس میان این داده‌ها را مدل کرده و الگوهای مصرف را شناسایی کرد (امینی و همکاران، ۱۳۹۷-ب). با استفاده از الگوریتم‌های داده‌کاوی می‌توان ماهیت پیچیده رفتار مصرفی مشترکین را در قالب الگوهایی شناسایی کرد که کمک شایسته‌ای به مدیریت مصرف می‌نماید (امینی، ۱۳۹۹). با توجه به اهمیت این موضوع، در پژوهش حاضر با استفاده از تکنیک‌های داده‌کاوی و تصمیم گیری چندمعیاره، مشترکین آب شیراز بخش‌بندی شده است. لازم‌به ذکر است که برای این منظور از مدل RFM و مفهوم هرم ارزش مصرف مشترکین نیز استفاده شده که تاکنون تحقیقی در این راستا و در صنعت آب و فاضلاب انجام نشده است. ارائه تحلیل خوشه‌ای به‌منظور بررسی دقیق‌تر ویژگی‌های مشترکین در خوشه‌ها و هم‌چنین، بخش‌بندی نهایی مشترکین براساس ارزش دوره عمر آن‌ها در قالب هرم ارزش دوره عمر، از مواردی است که در این مطالعه از آن‌ها بهره گرفته‌شده است. یکی از ویژگی‌های خاص این پژوهش استفاده از روش شبکه عصبی خودسازمان‌ده است که یکی از روش‌های دقیق در حوزه خوشه‌بندی و داده‌کاوی است.

براساس یافته‌های پژوهش، مشترکین آب شیراز به شش خوشه تفکیک شدند که با توجه به ارزش دوره عمر، خوشه‌های ۳، ۵، ۲، ۱، ۴ و ۶ به‌ترتیب در رتبه‌های اول تا ششم قرارگرفتند. طبق هرم ارزش، شرکت آب و فاضلاب شیراز باید برای جلب رضایت و تشویق مشترکین با ارزش خود (دارای بیشترین ارزش دوره عمر) در خوشه‌های ۵، ۳ و ۲ تلاش نماید و اهمیت ویژه‌ای برای مشترکین این خوشه‌ها قائل‌شود. از آن‌جا که مشترکین خوشه ۶ دارای کمترین ارزش دوره عمر هستند، لذا بهتر است که مطالعه دقیق‌تری پیرامون علل کم‌بودن مقادیر هر یک از شاخص‌های مدل در این مشترکین انجام گیرد.

لازم‌به ذکر است که رفتار مشترکین ثابت نیست و بانک اطلاعاتی آن‌ها تغییر می‌کند. مجموعه قوانینی که از بانک اطلاعاتی برای یک زمان خاص دریافت شده فقط برای یک زمان مشخص دارای اعتبار است و سپس منسوخ می‌شود. بنابراین پیشنهاد می‌شود که در مطالعات آتی، از مجموعه داده وسیع‌تری به لحاظ قلمروی زمانی استفاده شود که به‌طور قطع منتج به ایجاد نتایج قوی‌تر و گستره وسیع‌تری از دانش کاربردی پیرامون ویژگی‌های رفتاری مشترکین خواهد شد. ازطرفی، برش‌های زمانی متناوب از پایگاه داده شرکت، امکان اجرای پویای داده کاوی بر مبنای مدل RFM برای تعیین ارزش دوره عمر مشترکین را فراهم خواهد آورد که نتایج آن از این لحاظ که روند تغییرات در رفتار مشترکین را منعکس می‌کند، می‌تواند نقش شایانی در ارتباط با آن‌ها ایفا نماید.

۵- پی‌نوشت‌ها

1- Data Mining
2- Clustering
3- Artificial Neural Network
4- Self Organization Map (SOM)
5- Neuron
6- Noisy data
7- Euclidean distance
8- Recency, Frequency, Monetary (RFM)
9- Recency (R)
10- Frequency (F)
11- Monetary (M)
12- Analitical Hierarchy Process
13- Multi Criteria Decsion Making
14- Pair Wise Comparisons

۶- مراجع

آموزگار، م.، (۱۳۹۵)، "ارائه راه‌کار دو مرحله‌ای برای شناسایی الگوی مصرف برق"، *نشریه علمی پژوهشی کیفیت و بهره‌وری صنعت برق ایران*، ۵(۹)، ۴۸-۵۷.

امینی، ق.، فرمانی انتظام، ح.، جان صادق‌پور، ع.، و داودآبادی، ع.، (۱۳۹۷-الف)، "کاربرد داده‌کاوی در شناسایی مشترکین با مصارف غیرمجاز آب (مطالعه موردی: شرکت آب و فاضلاب قم)"، *دومین کنگره علوم و مهندسی آب و فاضلاب ایران*، اصفهان.

امینی، ق.، فرمانی انتظام، ح.، جان صادق‌پور، ع.، و داودآبادی، ع.، (۱۳۹۷-ب)، "شناسایی و استخراج الگوی مصرف آب به‌روش داده‌کاوی (مطالعه موردی: شرکت آب و فاضلاب قم)"، *دومین کنگره علوم و مهندسی آب و فاضلاب ایران*، اصفهان.

امینی، ق.، (۱۳۹۹)، "مدل‌سازی تشخیص مصرف غیرمجاز آب (مطالعه موردی: شهر قم)"، *مجله آب و فاضلاب*، ۳۱(۴)، ۱۸۴-۱۹۳.

رزمی، ج.، و قنبری، آ.، (۱۳۸۸)، "ارائه مدلی نوین برای محاسبه ارزش دوره عمر مشتری"، *نشریه مدیریت فناوری و اطلاعات*، ۱(۱)، ۳۵-۵۰.

طباطبایی، ا.، (۱۳۹۸)، *نگرشی بر داده‌کاوی*، دانشگاه آزاد قزوین، ۱۰-۱۵.

غضنفری، م.، (۱۳۹۹)، *داده کاوی و کشف دانش*. تهران، دانشگاه علم و صنعت ایران.

کاظمی، ز.، (۱۳۹۴)، "به‌کارگیری فرآیندکاوی برای بهبود فرآیندهای مدیریت دانش در مراکز تماس (مطالعه موردی: مرکز تماس ۱۲۲ سازمان آب و فاضلاب استان تهران)"، پایان‌نامه کارشناسی ارشد، دانشکده فنی، دانشگاه تربیت مدرس.

کجوری نفت‌چالی، م. و، فریدونیان، ع.، (۱۳۹۴)، "شناسایی الگوی مصرف انرژی الکتریکی با داده‌کاوی"، *سی‌امین کنفرانس بین‌المللی برق*، تهران.

مهرگان، م. ر.، (۱۳۹۷)، *پژوهش عملیاتی پیشرفته*، تهران، انتشارات کتاب دانشگاهی، چاپ اول.

Berry, M.J., and Linoff, G.S., (2016), *Data mining techniques: for marketing, sales, and customer relationship management*, John Wiley and Sons, https://doi.org/10.4324/9781351016551.

Bin, D., Peiji, S., and Dan, Z., (2008), "Data mining for needy student identify based on improved RFM model: A case study of university", *International Conference on Information Management, Innovation Management and Industrial Engineering*, New York, https://doi.org/10.1109/ICIII.2008.128.

Buttle, F., (2018), *Customer Relationship Management: Concepts and Tools*, Elsevier Butterworth Heinemann, https://doi.org/10.4324/9781351016551.

Cheng, C.H., and Chen, Y.S., (2009), "Classifying the segmentation of customer value via RFM model and RS theory", *Expert Systems with Applications*, 36, 4176-4184, https://doi.org/10.1016/j.eswa.2008.04.003.

Foster, H.S., and Beattie, B.R., (1981), "On the specification of price in studies of consumer demand under block price scheduling", *Land Economics*, 57(4), 624-629, https://doi.org/10.2307/3145676.

Han, J., Pei, J., and Kamber, M., (2016), *Data mining: Concepts and techniques*, Elsevier, https://doi.org/10.4324/9781351016551.

Hughes, A.M., (1994), *Strategic database marketing*, Chicago: Probus Publishing, https://doi.org/10.1287/ijoc.2021.1121.

Keiningham, T.L., Aksoy, A., and Bejou, D., (2006), "Approches to measurement and management of customer value", *Journal of Relationship Marketing*, 5(2), 37-54, https://doi.org/10.1300/J366v05n02_03.

Kreidler, M., (2018), Guide to auto insurance, Washington State Office of the Insurance Commissioner, viwed 2010/4/24, www.insurance.wa.gov.

Monika, C., and Amarpreet, K., (2018), "A comparative study of classification techniques for fraud detection", *Journal on Future Revolution in Computer Science and Communication Engineering*, 4, 19-23.

Rosenblatt, F., (1962). *Principles of neurodynamics: Perceptrons and the theory of brain mechanisms*, (Vol. 55), Washington, DC: Spartan books.

Saggaf, M.M., Toksoz, M.N., and Marhoon, M.I., (2003), "Seismic facies classification and dentification by competitive neural networks", *Geophysics*, (44), 1041-1063, https://doi.org/10.1190/1.1635052.

Seyedhosseini, S.M., Gholamian, M.R., and Maleki, A., (2011), "A methodology based on RFM using data mining approach to assess the customer loyalty", *International Journal of Industrial Engineering*, 22(2), 171-179, https://doi.org/10.1300/J366v05n02_03.

This article is an open-access article distributed under the terms and conditions of the Creative Commons Attribution (CC-BY) license.