ارزیابی تحلیلی و مدیریت اطلاعات فروش محصولات لبنی با استفاده از داده‌کاوی

فصلنامه علمی- پژوهشی
مدیریت اطلاعات
دوره ۲، شماره ۵ و ۶
زمستان ۹۵ و بهار ۹۶

بابک سهرابی
استاد مدیریت فناوری اطلاعات، دانشکده مدیریت، دانشگاه تهران [1]
ایمان رئیسی وانانی
استادیار مدیریت صنعتی، دانشکده مدیریت و حسابداری، دانشگاه علامه طباطبائی
مهتاب امامی
فارغ التحصیل مدیریت فناوری اطلاعات، دانشکده مدیریت، دانشگاه تهران

چکیده: پیش‌بینی دقیق میزان فروش نقش مهمی در مدیریت مؤثر تولید و منابع انسانی و درنهایت بهبود سودآوری شرکت‌های تولیدی دارد. پژوهش حاضر، از روش‌های داده‌کاوی برای تجزیه‌وتحلیل و استخراج خوشه‌های فروش در یک شرکت تولید محصولات لبنی استفاده می‌نماید و درنهایت رفتار مشتریان این شرکت را در مناطق مختلف ایران توصیف می‌کند. مسئله پیش‌بینی فروش در این شرکت ازاین‌جهت حائز اهمت است که اقلام لبنی باید به‌صورت روزانه تولید، توزیع و مصرف شوند چراکه به‌سرعت فاسد می‌گردند. درنتیجه، نوسانات غیرمنتظره در تقاضای مصرف‌کنندگان ممکن است مشکلات جدی برای توزیع این محصولات ایجاد نماید. مدیریت ارشد این شرکت براین باور است که شرایط اقلیمی مناطق مختلف ایران و نیز تغییرات جوی در فصول مختلف سال به‌شدت بر میزان فروش مؤثر است؛ اما تابه امروز، این متغیر مستقل بسیار مهم در پیش‌بینی فروش دخیل نبوده است. تحقیق حاضر علاوه بر طراحی سیستم تحلیل و پیش‌بینی فروش بر آن است تا تأثیر این متغیرهای جوی را بر فروش انواع مختلف محصول در سطح کشور تعیین نماید و دستورالعمل‌های جدیدی را به تولیدکنندگان محصولات لبنی ارائه کند. در این مقاله از الگوریتم‌های خوشه‌بندی و طبقه‌بندی برای تمایز رفتار مشتریان در سراسر ایران استفاده‌شده است. این فرآیند در طی شش مرحله‌ی درک کسب‌وکار، درک داده‌ها، تهیه داده‌ها و پیش‌پردازش، مدل‌سازی، ارزیابی و پیش‌بینی انجام‌شده است. مراحل جمع‌آوری داده‌ها و مدل‌سازی با پیش‌بینی‌های تحلیلی و نتیجه‌گیری تکمیل می‌گردد. با توجه به محدودیت دسترسی به پایگاه‌های داده‌ی مرکزی شرکت لبنی موردنظر، ضروری بود که محققان بخشی از داده‌های محصولات لبنی را که مربوط به بازه زمانی محدود و محصولات منتخب بود، برای تحقیق حاضر انتخاب نمایند. یافته‌های پژوهش شامل یک تحلیل عملی از سیاست‌های درست فروش برای هر دسته از رفتار مشتریان، باهدف بهبود سودآوری است. این پژوهش بر حفظ تعادل بین نیازهای مشتری و کنترل هزینه‌های موجودی انبار تمرکز دارد و نیاز مبرم شرکت‌هایی که به دنبال فروش خوب و سودآوری از طریق پیش‌بینی دقیق تحلیلی هستند را برآورده می‌سازد.

کلید واژه‌ها: *محصولات لبنی، داده‌کاوی، تحلیل اطلاعات، پیش‌بینی فروش.*

---

[1] bsohrabi@ut.ac.ir

مقدمه

پیش‌بینی فروش در خرده‌فروشی، تخمین تقاضای آینده محصول است که برای برنامه‌ریزی صحیح کسب‌وکار ضروری است و نقش مهمی را در زنجیره تأمین دائمی در مواجهه با رقابت شدید جهانی بازی می‌کند. پیش‌بینی فروش پایه‌ای است که در آن طرح‌های شرکت ساخته‌شده و برای بهبود ظرفیت رقابت شرکت‌های خرده‌فروشی نیز امری ضروری است (Guo, Wong and Li 2013). پیش‌بینی فروش، فرآیندی است که با سازمان‌دهی و تجزیه‌وتحلیل اطلاعات، امکان برآورد فروش را به وجود می‌آورد (Xiaodan Yu 2013). سیستم پیش‌بینی فروش موفق می‌تواند به خاصه برای صنایع غذایی بسیار مفید باشد، چراکه کوتاه بودن عمر مفید بسیاری از محصولات غذایی و اهمیت بالای کیفیت آن‌ها که با سلامت انسان‌ها مرتبط است، چالش‌های پیچیده‌ای را برای صاحبان این صنایع به همراه دارد (Doganis, et al. 2006). با توجه به محیط رقابتی، جهانی‌شدن، زمان زیاد بین شروع و تکمیل یک فرآیند[۲] و عدم اطمینان از تقاضای مشتری، پیش‌بینی فروش عامل اصلی موفقیت در بهینه‌سازی زنجیره تأمین شرکت‌های تولیدی به شمار می‌رود (Thomassey and Happiette,2007). مزایای استفاده از تکنیک‌های پیشرفته پیش‌بینی در سطوح مختلف را می‌توان به شرح زیر دانست (S. Thomassey 2010):

- کاهش اثر شلاقی[۳] بدون تغییر عمده در سازمان‌دهی زنجیره تأمین. سیستم پیش‌بینی پیشرفته، یک روش کارآمد برای کاهش این پدیده منفی است.
- فراهم آمدن امکان تسریع تولید برای تأمین‌کننده، بهینه‌سازی منابع، کاهش هزینه‌ها و بهبود کارایی در استراتژی فروش خرده‌فروشان.
- کاهش فروش ازدست‌رفته، تنزل قیمت و درنتیجه افزایش حاشیه سود.

بااین‌حال، پیش‌بینی‌کنندگان با محدودیت‌های خاص صنایع لبنی مواجه هستند که آن را از دیگر صنایع فرآوری محصولات غذایی متمایز می‌کند. به‌عنوان‌مثال تقاضای ناپایدار، تقاضای فصلی فروش، وجود طیف وسیعی از اقلام، عمر مفید کوتاه محصولات لبنی و شرایط خاص انبارداری و حمل. در این صنعت، خرید و نگهداری شیر، به‌عنوان ماده‌ی اولیه اصلی کلیه محصولات لبنی، بسیار حائز اهمیت است. چراکه شیر به‌صورت روزانه و با هزینه‌ی حمل‌ونقل بالایی از دامداران جمع‌آوری می‌شود و برای جلوگیری از فساد سریع آن، باید در شرایط ایده‌آل نگهداری گردد و به‌سرعت وارد خط تولید شود (Francesconi, Heerink and D'Haese 2010). درنتیجه، جمع‌آوری بیش‌ازاندازه‌ی شیر از دامداران و تولید و توزیع بیش‌ازاندازه محصولات لبنی، هزینه‌های قابل‌توجهی را به شرکت تحمیل

[۲] lead times
[۳] bullwhip effect

خواهد کرد. از سوی دیگر، با توجه به محیط رقابتی محصولات لبنی در ایران، از دست دادن فروش ناشی از دست‌کم گرفتن تقاضای بازار نیز قابل‌قبول نیست؛ بنابراین، به‌منظور افزایش سودآوری، حفظ رقابت و ایجاد یک زنجیره تأمین مؤثر، شرکت در درجه اول نیازمند تکیه‌بر یک سیستم دقیق پیش‌بینی فروش است (S. Thomassey 2010).

بررسی ادبیات موضوعی

برای مدت طولانی بهبود عملکرد پیش‌بینی فروش، نگرانی اصلی مدیران بوده و تحقیقات بر آن تمرکز داشته است. برای این منظور، محققان تکنیک‌های پیش‌بینی پیچیده‌تری را منتشر کرده‌اند و معتقدند که یک مدل دقیق‌تر، شرایط پیچیده بازار را بهتر مدل‌سازی می‌کند. بااین‌حال، تکنیک‌های پیش‌بینی پیشرفته تنها در صورت استفاده در فرایندهای تصمیم‌گیری و برنامه‌ریزی سازمان‌ها مفید هستند (Davis and Mentzer 2007). تاریخچه پیش‌بینی فروش به بیش از ۵۰ سال قبل بازمی‌گردد. از آن زمان به بعد مقالات زیادی در مورد پیش‌بینی فروش منتشرشده‌اند که گویای کاربرد وسیع آن در صنایعی مانند الکترونیک، غذا و پوشاک است (Guo, Wong and Li 2013). مطالعه ولیدی و همکاران، (۲۰۱۴) اهمیت پیش‌بینی دقیق فروش را در مدیریت زنجیره تأمین صنایع غذایی، به‌ویژه لبنیات، موردبحث قرار می‌دهد. (Validi, Bhattacharya and Byrne 2014)

اهمیت پیش‌بینی فروش در دنیای واقعی، تحقیقات زیادی را در این زمینه برانگیخته است. طبق نظر شادون، روش پیش‌بینی فروش می‌تواند به سه حوزه تحقیق، یعنی تحقیقات در عملیات، آمار و داده‌کاوی تقسیم شود (Xiaodan Yu 2013). مدل‌های آماری سنتی، دارای محدودیت‌های زیادی هستند چراکه پیش‌بینی آینده را صرفاً از طریق بررسی توابع خطی گذشته انجام می‌دهند؛ بنابراین مدل‌های غیرخطی مختلفی به‌منظور غلبه بر این محدودیت‌ها ارائه‌شده است (Khashei and Bijari ۲۰۱۲). در طی چند دهه گذشته، به خاصه برای داده‌هایی که حاوی روند و الگوهای فصلی هستند، روش‌های متعددی برای پیش‌بینی فروش ارائه‌شده است. عدم توجه به الگوهای فصلی حاضر در داده‌ها، منجر به ایجاد مدل‌های پیش‌بینی ضعیفی می‌گردد.

مدل‌های آماری که تابه‌حال در تحقیقات پیش‌بینی فروش استفاده‌شده‌اند عبارت‌اند از روش میانگین‌گیری متغیر وزن‌دار وینتر[۴] (Tratar 2015)، مدل بوکس-جنکین آریما[۵] (Park, Park and Lee 1991)، رگرسیون خطی تکی و چندگانه (Papalexopoulos and Hesterberg ۱۹۹۰)، مدل‌های خودگردان (Huang, Fildes and Soopramanien 2014) و میانگین

[۴] Winters exponential smoothing
[۵] Box-Jenkins ARIMA model

متحرک (Hassan, Shalaby and Gazarin 1988). این روش‌ها در چند دهه گذشته بارها پیشنهادشده و مورداستفاده قرارگرفته‌اند.

مقالات و پژوهش‌های متعددی در ایران و در سطح بین‌المللی، به استفاده از داده‌کاوی به‌منظور توصیف و پیش‌بینی وقایع مختلف پرداخته‌اند (قانعی استاد و دیگران ۱۳۹۶؛ رضایی نور و شیخ بهایی ۱۳۹۶). «داده‌کاوی» به‌عنوان فرآیند تجزیه‌وتحلیل و پیدا کردن روابط معنی‌دار یا پیدا کردن قوانین در میان داده‌ها به‌صورت خودکار یا نیمه‌خودکار (Chen and Huang 2011) تعریف می‌شود. تکنیک‌های داده‌کاوی مانند خوشه‌بندی[۶] طبقه‌بندی[۷] شبکه‌های عصبی، الگوریتم‌های تکامل‌یافته و رویکردهای مبتنی بر فازی به‌طور گسترده‌ای برای اهداف پیش‌بینی در صنایع مختلف مانند فروشگاه‌های خرده‌فروشی ((Guo, Wong and Li 2013)؛ (Alon, Qi and Sadowski 2001))، صنعت پوشاک (S. Thomassey 2010)، بازار عمده‌فروشی کامپیوتر (Lu, Lee and Lian 2012)، صنایع خرده‌فروشی مواد غذایی تازه (Chen, et al. 2010)؛ (Chen and Ou 2009)) تولید محصولات غذای با عمر مفید کوتاه (Doganis, et al. 2006)، صنایع تولید بوردهای وسایل الکترونیکی (Chang, Wang and Liu 2007) (Chang, Wang and Liu 2007)، بازار نیمه‌هادی[۸] (Navarro-Barrientos, et al. 2014) تولید مواد شیمیایی (Shah 2012) مورداستفاده قرارگرفته است.

در میان روش‌های داده‌کاوی فوق، پژوهش حاضر، روش ترکیبی خوشه‌بندی و طبقه‌بندی را برای انجام تحقیق انتخاب کرده است. روش‌های خوشه‌بندی و طبقه‌بندی به‌صورت جداگانه در موارد مختلفی ازجمله شناخت الگو، ارزیابی اعتبار و وام، پیش‌بینی ترافیک وب، تشخیص نفوذ تقلب و شبکه[۹] تشخیص پزشکی (X.-B. Li 2005)، پیش‌بینی تولید (Li and Chan 2010) مدیریت ارتباط با مشتری (Ngai, Xiu and Chau 2009) حسابرسی و کنترل اعتبار (Chen and Huang 2011) و بررسی روند و پیش‌بینی تقاضا (Ma, Kwak and Kim 2014) کاربرد فراوان دارد. همچنین مقالات متعددی نیز از روش ترکیبی خوشه‌بندی و طبقه‌بندی در صنایع مختلفی ازجمله بازار نساجی و پوشاک (Thomassey and Happiette, 2007) (Fiordaliso and Thomassey 2006) منتشرشده‌اند. این روش ترکیبی، در زمینه‌های دیگری همچون شناسایی و پیش‌بینی تقلب (Farvaresh and Sepehri 2011)و بهبود کیفیت (Köksal, Batmaz and Caner Testik ۲۰۱۱)نیز مورداستفاده قرارگرفته است؛ اما تنها چند تحقیق درباره پیش‌بینی فروش بر اساس رویکرد

[۶] Clustering
[۷] Classification
[۸] semiconductor market
[۹] fraud and network intrusion detection

هیبریدی در صنعت غذایی فاسدشدنی و به‌ویژه محصولات لبنی موردبحث قرارگرفته است. به‌علاوه، اغلب تحقیقات، متغیرهای قیمت و تبلیغات را در مدل‌های پیش‌بینی فروش انتخاب می‌کنند؛ بنابراین، پیش‌بینی فروش بر اساس پیش‌بینی شرایط اقلیمی و جوی یک موضوع جدید در این زمینه است.

روش تحقیق

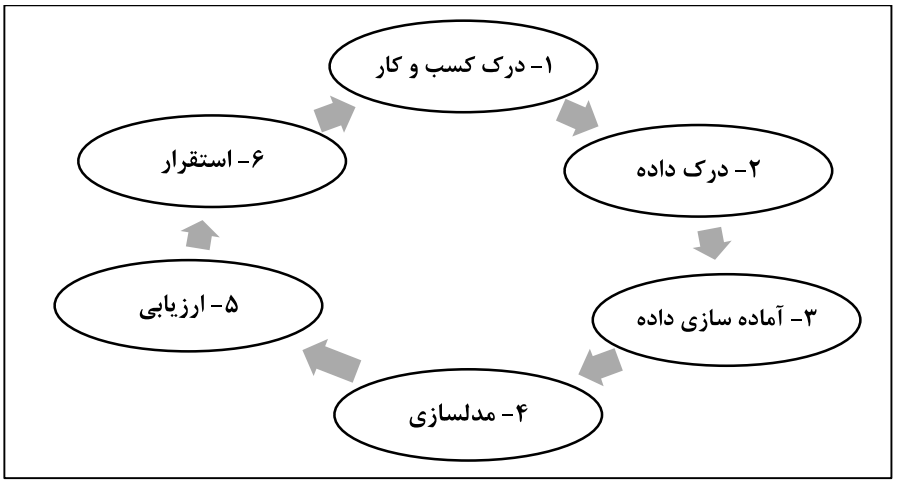
تحقیق حاضر یک پژوهش کاربردی[1] بر روی‌داده‌های از نوع کمی است که بر اساس مطالعه موردی شرکت فرآورده‌های لبنی منتخب صورت گرفته است. تصمیم‌گیری داده محور یکی از کلیدهای اصلی ایجاد موفقیت در سازمان‌ها و تدوین راهبردهای دقیق و کارآمد است و شامل داده‌کاوی و تحلیل پیشرفته حجم زیادی از داده‌ها و یافتن الگوهای منظم و پنهان در آن‌ها می‌شود. با توجه به پیچیدگی و گستردگی حوزه تحلیل‌های پیشرفته و داده‌کاوی، نیاز به تدوین روش‌شناسی اثربخش در این حوزه شناسایی‌شده است. به همین جهت روش‌های مختلفی پیشنهاد شدند که یکی از آن‌ها با عنوان CRISP-DM[2] از مقبولیت و کارآمدی فراوانی برخوردار شده است (Sharma, Osei-Bryson and Kasper 2012؛ Kurgan and Musilek 2006).

به‌طورکلی دلیل انتخاب این روش‌شناسی در وهله اول، محبوبیت این چارچوب و گستردگی استفاده از آن توسط سازمان‌ها است. دلیل دوم، دقیق‌تر بودن آن نسبت به دیگر متدولوژی‌های داده‌کاوی و امکان ارائه رهنمودهای جامع است. دلیل سوم این است که استفاده از CRISP-DM محدود به صنایع خاص نیست و در کلیه زمینه‌ها قابل‌استفاده و پیاده‌سازی است (Chalaris, et al. 2014). همان‌طور که در شکل ۱ نشان داده‌شده است این فرآیند شامل شش مرحله یا گام است. اولین گام CRISP-DM، درک کسب‌وکار سازمان است. این مرحله تمرکز بر درک اهداف و الزامات پروژه از منظر تجاری و یا سازمان دارد. گام دوم درک داده است که در آن داده‌های اولیه جمع‌آوری می‌شوند، سپس مشکلات کیفیت داده و یا فرضیه‌های وجود اطلاعات پنهان شناسایی می‌شوند. مرحله سوم آماده‌سازی داده‌ها است. در این مرحله اموری مانند مرتب کردن داده‌ها، تبدیل داده‌ها و انتخاب داده‌ها به‌منظور ساخت مجموعه داده نهایی انجام می‌شود. در گام چهارم، تکنیک‌های مختلف مدل‌سازی انتخاب می‌شود و مورداستفاده قرار می‌گیرند. در گام ارزیابی، ارزشمند بودن مدل و میزان برآورده شدن اهداف کسب‌وکار بررسی و تعیین می‌شود. گام نهایی استقرار است که مشخص‌کننده اقداماتی است که باید برای استفاده از مدل‌های توسعه‌یافته صورت گیرد (Chalaris, et al. 2014).

---

[1] Applied Research

[2] Cross-Industry Standard Process for Data Mining

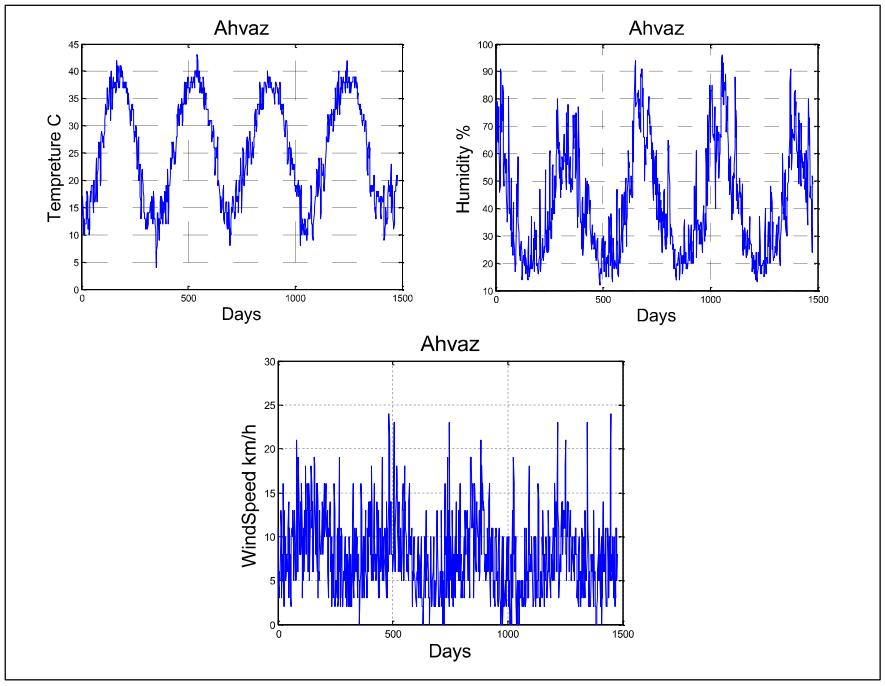


شکل ۱. روش CRISP-DM

پژوهشگر داده‌های فروش مربوط به ۲۰ محصول لبنی را برای یک دوره ۴ ساله از شرکت موردپژوهش دریافت نموده است. داده‌های تحقیق متعلق به فاصله زمانی دی‌ماه ۱۳۸۹ تا اسفندماه ۱۳۹۳ است. محصولات از ۴ گروه محصولات مختلف لبنی شامل ۵ نوع خامه (طعم‌های ساده، شکلاتی، شکلاتی مغز دار، عسلی و انجیری)، ۵ نوع دوغ (گازدار/بدون گاز و با سبزی‌های معطر متنوع)، ۵ نوع شیر پاستوریزه (با درصد چربی، سایز و طعم‌های متفاوت) و ۵ نوع ماست (با درصد چربی و سایزهای مختلف) انتخاب‌شده‌اند. اطلاعات فروش ۶ شهر بزرگ ایران به دلیل اهمیت بازار آن‌ها برای شرکت انتخاب‌شده است. شهرهای انتخابی عبارت‌اند از: اهواز، آمل، اصفهان، مشهد، شیراز و تهران. اطلاعات تاریخی شرایط جوی این شهرها نیز از سایت معتبر https://weatherspark.com دانلود شده است. داده‌های شرایط جوی شامل درجه حرارت ($^{\circ}C$)، رطوبت ($\%$) و سرعت باد (Km / H) برای هرروز در طول دوره تحقیق است.

تجزیه‌وتحلیل توصیفی داده‌های فروش

قبل از شروع عملیات مرتب‌سازی داده‌ها و فرایندهای داده‌کاوی، اول ویژگی‌های اصلی داده‌ها به‌صورت بصری بررسی می‌شود. توصیفی از اطلاعات تاریخی اقلیمی در شکل ۲ برای شهر اهواز نشان داده‌شده است. برای این شهر وضعیت آب و هوایی بر اساس سه متغیر نمایش داده‌شده است. انتخاب این متغیرها، به درخواست شرکت لبنی منتخب و به دلیل اطمینان از داده‌های قابل‌اتکا و در دسترس بوده است:



شکل ۲. شرایط آب و هوایی در شهرستان اهواز طی دوره تحقیقاتی

بدیهی است که در شکل ۲، چهار دوره کامل که متعلق به ۴ سال مختلف است و همچنین تغییرات روند آب و هوایی مشابه در هرسال نشان داده‌شده است. به‌عنوان‌مثال، ماکزیمم‌های نسبی نمودار دما، در طی دوره تحقیق به ۴ تابستان اشاره دارد. به‌راحتی می‌توان گذشت فصول را در نمودار مشخص نمود.

در گام بعد روند حجم فروش ۲۰ محصول در شهر اهواز بررسی شد. نتایج این بررسی، نوسانات زیاد و سریع فروش هر یک از محصولات را نشان می‌دهد که کار مدل‌سازی و تجزیه‌وتحلیل را بیشتر به چالش می‌کشد. نقطه قابل‌توجه دیگر در این بررسی، پایداری در متوسط فروش برخی محصولات در طول دوره تحقیق است. این موارد بیشتر محصولات اصلی در سبد غذایی هستند که به قیمت و یا هوا خیلی حساس نیستند (به‌عنوان‌مثال گروه محصول شیر پاستوریزه).

در گام بعد، روند نوسان قیمت هریک از محصولات در شهر اهواز بررسی شد. بررسی این روند به‌صراحت نشان می‌دهد که قیمت‌ها در یک دوره نسبتاً طولانی ثابت بوده است. این در حالی است که حجم فروش، ارتعاشات زیاد و سریعی را در فواصل زمانی مشابه نشان می‌دهد. با نگاه دقیق به حجم فروش در دو روز متوالی، باقیمت‌های ثابت، تفاوت‌های زیادی در میزان فروش پدیدار می‌شود.

در مرحله بعد، میانگین فروش روزانه در شهرهای مختلف مورد تجزیه‌وتحلیل قرار می‌گیرد. ازآنجاکه جمعیت هر شهر متغیر اصلی در تعیین حجم فروش روزانه در آن شهر است، لازم است قبل از تجزیه‌وتحلیل داده‌ها، اثر این متغیر بر میزان تقاضای شهر را خنثی گردد. به این منظور اطلاعات جمعیتی هر شهر بر اساس داده‌های مرکز آمار ایران (وب‌سایت www.amar.org.ir قسمت سرشماری نفوس و مسکن) جمع‌آوری گردید.

جدول شماره ۱ نشان‌دهنده تقاضا به ازای ۱۰۰۰ نفر در هر شهر است. بالاترین مقدار فروش روزانه در هر ردیف محصول مشخص‌شده است.

جدول ۱. تقاضا به ازای هر ۱۰۰۰ نفر در هر شهر

	اهواز	آمل	اصفهان	مشهد	شیراز	تهران
جمعیت	۱،۳۹۵	۳۷۱	۴،۸۷۹	۳،۰۷۰	۱،۷۰۱	۱۲،۱۸۳
خامه	۰,۰۱	۰,۱۳	۰,۰۱	۰,۰۲	۰,۰۱	۰,۰۱
خامه ۱	۰,۰۰	۰,۰۹	۰,۰۰	۰,۰۱	۰,۰۰	۰,۰۰
خامه ۲	۰,۰۱	۰,۳۵	۰,۰۱	۰,۰۵	۰,۰۲	۰,۰۲
خامه ۳	۰,۰۰	۰,۰۲	۰,۰۰	۰,۰۱	۰,۰۰	۰,۰۱
خامه ۴	۰,۰۰	۰,۰۱	۰,۰۰	۰,۰۱	۰,۰۰	۰,۰۱
خامه ۵	۰,۰۲	۰,۱۸	۰,۰۱	۰,۰۱	۰,۰۴	۰,۰۰
دوغ	۰,۸۱	۰,۸۴	۰,۱۵	۰,۲۹	۰,۲۱	۰,۵۰
دوغ ۱	۱,۶۸	۰,۷۸	۰,۳۶	۰,۸۶	۰,۶۶	۰,۹۹
دوغ ۲	۰,۶۸	۱,۰۰	۰,۰۹	۰,۱۲	۰,۱۱	۰,۲۴
دوغ ۳	۰,۷۳	۰,۸۲	۰,۱۰	۰,۰۹	۰,۰۷	۰,۵۵
دوغ ۴	۰,۷۶	۰,۱۵	۰,۱۰	۰,۱۴	۰,۱۰	۰,۲۸
دوغ ۵	۰,۱۷	۱,۴۳	۰,۱۲	۰,۲۳	۰,۱۳	۰,۴۴
شیر	۰,۵۵	۱,۰۵	۰,۱۱	۰,۳۶	۰,۲۵	۰,۱۸
شیر ۱	۰,۳۴	۰,۸۴	۰,۰۸	۰,۲۰	۰,۲۰	۰,۳۲
شیر ۲	۰,۸۶	۰,۴۷	۰,۰۶	۰,۲۲	۰,۲۰	۰,۱۳
شیر ۳	۰,۴۶	۰,۸۶	۰,۱۱	۰,۳۰	۰,۲۳	۰,۰۸
شیر ۴	۰,۸۹	۲,۲۸	۰,۱۸	۰,۹۰	۰,۴۳	۰,۱۸
شیر ۵	۰,۲۲	۰,۸۲	۰,۱۰	۰,۱۹	۰,۱۸	۰,۱۸
ماست	۰,۳۹	۰,۵۱	۰,۰۶	۰,۲۶	۰,۲۸	۰,۵۱
ماست ۱	۰,۰۹	۰,۲۵	۰,۰۰	۰,۰۱	۰,۱۲	۰,۲۲
ماست ۲	۰,۱۶	۰,۶۶	۰,۰۱	۰,۰۴	۰,۰۳	۰,۰۹
ماست ۳	۰,۱۱	۰,۳۱	۰,۰۳	۰,۰۵	۰,۰۴	۰,۱۶
ماست ۴	۰,۰۰	۰,۰۴	۰,۰۱	۰,۰۲	۰,۱۰	۰,۴۹
ماست ۵	۱,۵۹	۱,۲۸	۰,۲۴	۱,۱۸	۱,۰۹	۱,۶۰

در جدول شماره ۱ شهر آمل به‌عنوان طرفدار بزرگ شرکت لبنیاتی موردتحقیق است که تعجب‌آور نیست زیرا کارخانه تولید در این شهر واقع‌شده است. این جدول نشان می‌دهد که پس از آمل، این محصولات لبنی در اهواز نیز محبوبیت زیادی دارند؛ اما فقط ۲ نوع ماست در بین مردم پایتخت (تهران) محبوب هستند.

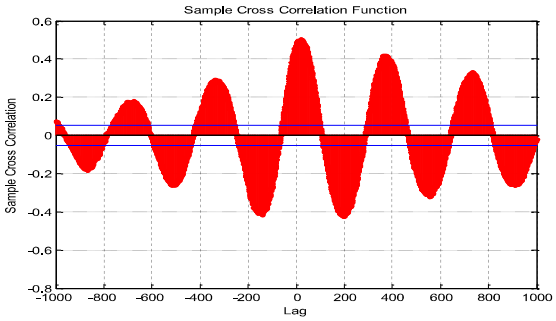
عوامل مختلفی می‌تواند باعث ایجاد تفاوت در تقاضای محصولات لبنی شود. بررسی پیشینه در کشورهای درحال‌توسعه بر وجود مجموعه‌ای از روابط کلیدی میان ویژگی‌های اجتماعی-اقتصادی خانوار و مصرف لبنیات تأکید دارد. تحقیقات نشان می‌دهد که مصرف شیر در هر خانواده تابع درآمد سرانه، اندازه خانواده، تعداد فرزندان، سطح تحصیلات و قیمت خرده‌فروشی است (Francesconi, Heerink and D'Haese 2010).

در مرحله بعدی، مجموع تقاضای کلیه محصولات در روزهای مختلف هفته مورد تجزیه‌وتحلیل و بررسی قرار می‌گیرد. مشاهده می‌شود که روند غالب، یک روند کاهشی از بالاترین مقدار تقاضا در اولین روز کاری (شنبه) است که به‌تدریج کاهش می‌یابد تا زمانی که به پایین‌ترین مقدار خود در آخر هفته (پنجشنبه و جمعه) برسد.

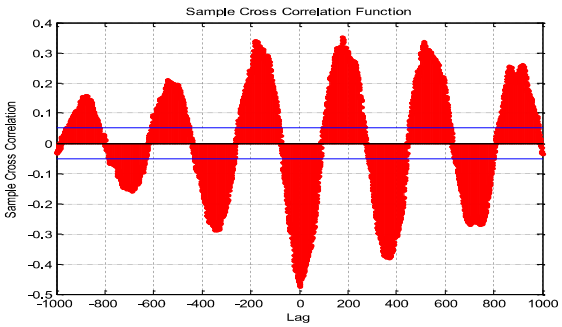
جدول۲. میانگین مجموع تقاضا (به ازای هر هزار نفر) در روزهای مختلف هفته

گروه محصول	شنبه	یکشنبه	دوشنبه	سه‌شنبه	چهارشنبه	پنجشنبه	جمعه	مجموع
خامه	۵۳	۵۲	۴۷	۴۶	۴۸	۳۴	۱۹	۴۳
دوغ	۱,۹۱۸	۱,۹۲۹	۱,۷۶۰	۱,۷۱۲	۱,۷۵۵	۱,۱۸۱	۷۳۷	۱,۵۸۸
شیر	۱,۰۳۸	۱,۰۳۵	۱,۰۲۵	۹۷۴	۱,۰۱۹	۷۷۶	۲۷۰	۸۹۳
ماست	۱,۷۲۹	۱,۶۶۳	۱,۵۰۳	۱,۵۸۲	۱,۵۷۳	۱,۰۶۱	۷۲۳	۱,۴۱۹
مجموع	۱,۱۸۵	۱,۱۷۱	۱,۰۸۵	۱,۰۷۹	۱,۱۰۰	۷۶۳	۴۴۴	۹۸۸

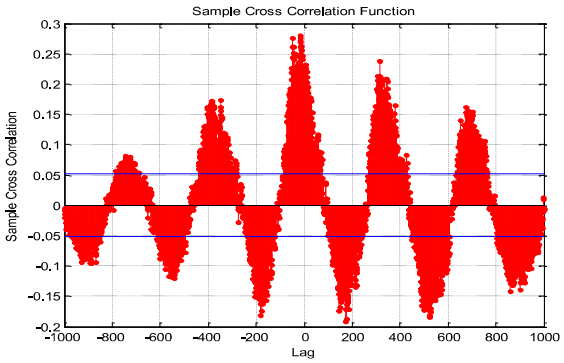
سپس، رابطه متغیر وابسته‌ی حجم فروش با متغیرهای مستقل اقلیمی موردبررسی قرار می‌گیرد. بدین منظور، از تابع همبستگی متقابل با تعداد ۱۰۰۰ نمونه استفاده می‌شود؛ که نتیجه آن در شکل‌های ۳ تا ۵ نشان داده‌شده است:



شکل ۳. نمودار همبستگی متقابل متغیرهای «حجم فروش» و «دما»



شکل ۴. نمودار همبستگی متقابل دو متغیر «حجم فروش» و «رطوبت»



شکل ۵. نمودار همبستگی متقابل دو متغیر «حجم فروش» و «سرعت باد»

همان‌طور که در نمودارهای بالا نشان داده‌شده است، سیگنال فروش به شرایط آب و هوایی وابستگی زیادی دارد.

پیش‌پردازش داده‌ها

داده‌های فروش ۲۰ محصول از ۴ گروه عمده محصول (خامه، دوغ، شیر پاستوریزه و ماست) برای این تحقیق انتخاب‌شده‌اند. این داده‌ها حجم فروش روزانه در ۶ شهر بزرگ ایران (اهواز، آمل، اصفهان، مشهد، شیراز و تهران) را طی سال‌های ۲۰۱۱ تا ۲۰۱۵ در مدت ۱۴۹۴ روز نشان می‌دهند. تاریخچه شرایط هواشناسی دوره موردتحقیق نیز موجود است. این داده‌ها شامل رطوبت روزانه، دما و سرعت باد است. داده‌ها به‌صورت یک جدول جمع‌آوری‌شده است.

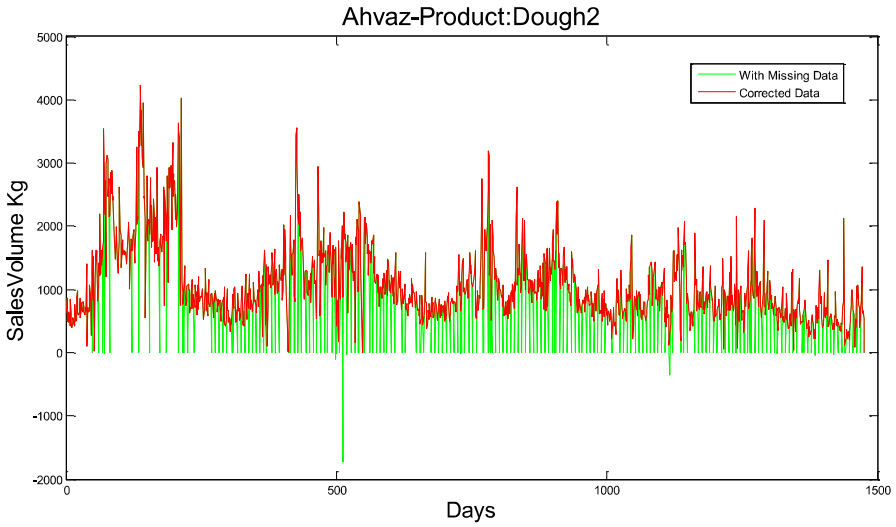
کلیه مراحل داده‌کاوی با استفاده از نرم‌افزار MATLAB انجام‌شده‌اند. برای بارگیری داده‌ها در این نرم‌افزار، مجموعه داده به یک آرایه ساختاری ۶ سلولی تبدیل‌شده، در هر سلول برای یک شهر موارد زیر نمایش داده‌شده است:

- تاریخ
- حجم (کیلوگرم)
- قیمت (ریال/ کیلوگرم)
- دما (سانتی‌گراد)
- سرعت باد (کیلومتر/ ساعت)
- نام محصولات

در اصل مجموعه داده شامل ۱۷۸،۵۶۰ مورد بود. به‌عنوان اولین مرحله قبل از پردازش اطلاعات، سوابق داده‌ای روزها بدون اطلاعات فروش کامل از تمام محصولات حذف شد که منجر به باقی ماندن ۱۷۷،۱۲۰ نمونه بدون ارزش‌های ازدست‌رفته شد که مربوط به ۱۴۷۶ روز از اطلاعات فروش است.

مقدار فروش در مجموعه داده دارای اعداد اعشاری زیادی است که نتیجه تبدیل واحد به کیلوگرم است. در مرحله بعد، تعداد فروش به نزدیک‌ترین عدد صحیح گرد می‌شوند تا از اشتباه در فرایند داده‌کاوی جلوگیری شود.

نکته مهم در مورد داده‌ها، تعداد بالای روزهای فروش صفر یا منفی است. راه‌حل مناسب این است که میانگین فروش روز قبل و روز بعد برای هر ردیف فروش صفر/منفی محاسبه شود. سپس مقدار این میانگین، جایگزین داده‌ی صفر/منفی شود. نتیجه بدین‌صورت می‌شود:



شکل ۶. نمودار مقدار فروش قبل و بعد از حذف پرونده صفر / منفی

مورد بعدی مربوط به داده‌های موجود در روزهایی است که میزان فروش در آن به‌طور شگفت‌انگیزی پایین‌تر یا بالاتر نسبت به‌روزهای قبل یا بعدشان است. این نوع سوابق قطعاً خطای سیستم پیش‌بینی را به میزان غیرقابل قبولی افزایش می‌دهد و در مراحل بعدی باعث سردرگمی می‌شود. برای تشخیص این داده‌ها، به مدیران ارشد فروش و توزیع شرکت مراجعه شد و رایزنی صورت گرفت. طبق نظر کارشناسان مذکور، این ناپایداری‌ها مربوط به بازگشت غیرمعمول فروش و یا دور انداختن محصول ناشی از مسائل کیفیت یا تعمیر و نگهداری بد در طی روند توزیع است؛ اما طبق نظر مدیران ارشد فروش، به‌طورمعمول نوسانات حجم فروش در روزهای متوالی بیش از ۵۰٪ نیست. درنتیجه، مقدار فروش در هرروز با مقدار فروش در روزهای قبل و بعد، مقایسه می‌شود. اگر واریانس بیش از ۵۰٪ باشد، مبلغ فروش آن روز با مقدار متوسط ذکرشده جایگزین می‌شود. این فرآیند "۱۶۸۶۶" رکورد را که ۱۰ درصد کل رکوردها در مجموعه داده‌ها بود را شناسایی و اصلاح نمود. این نتیجه توسط کارشناسان موردبحث و تائید قرار گرفت.

مدل‌سازی داده

لازم به ذکر است که قیمت‌های محصولات لبنی اغلب از سوی مراجع دولتی تعیین می‌گردد و نوسانات کمی دارد. درنتیجه نوسانات زیاد تقاضا باوجود قیمت‌های ثابت، نشان می‌دهد که نوسانات به متغیرهای اقلیمی و جوی وابسته هستند. در این بخش، روش خوشه‌بندی برای تحلیل رفتار تقاضا استفاده می‌شود. روش خوشه‌بندی می‌تواند بر اساس روش کلاسیک و روش مستقیم k-means و یا بر اساس

تکنیک‌های پیشرفته‌تر مبتنی بر تکنیک‌های عصبی مانند نقشه خودسازمانی[۲](SOM) باشد (S. Thomassey 2010). بنابراین، از سه متغیر آب‌وهوا برای خوشه کردن روزها با ۳ روش مختلف استفاده می‌شود. روش‌های انتخاب‌شده عبارت‌اند از Kmeans، FCM و SOM.

۱- روشK-Means

اولین روش منتخب پژوهشگر برای خوشه‌بندی آب‌وهوا در شهرهای مختلف، روش K-means است. این روش، ماتریس چندبعدی از داده را به تعداد K خوشه تقسیم‌بندی می‌کند. اساس این تقسیم‌بندی حداقل کردن فاصله‌ی نقاط موجود در هر خوشه از مرکز[۳]آن است. K-means یک بردار شامل شماره خوشه هریک از نقاط ایجاد می‌کند (توماسی، ۲۰۱۰).

۲- روشFCM

دومین روش منتخب برای اجرای خوشه‌بندی روی‌داده‌های هواشناسی، FCM است. FCM یک روش خوشه‌بندی داده است که هریک از نقطه‌های داده تا میزان خاصی که با درجه‌ی عضویت[۴]اندازه‌گیری می‌شود عضو هریک از خوشه‌ها هستند.

۳- روش شبکه‌ی خودسازمان‌ده(SOM)

سومین تکنیک خوشه‌سازی که در این تحقیق استفاده می‌شود، شبکه‌ی خود سازمان‌ده است. در این تکنیک، از روش یادگیری رقابتی برای آموزش استفاده می‌شود و مبتنی بر مشخصه‌های خاصی از مغز انسان توسعه‌یافته است. سلول‌ها در مغز انسان در نواحی مختلف طوری سازمان‌دهی شده‌اند که در نواحی حسی مختلف، با نقشه‌های محاسباتی مرتب و معنی‌دار ارائه می‌شوند. برای نمونه، ورودی‌های حسی لامسه –شنوایی و ... با یک ترتیب هندسی معنی‌دار به نواحی مختلف مرتبط هستند (S. Thomassey 2010).

---

[۱] self-organizing map (SOM)
[۲] Cluster Centroid
[۳] Membership Grade

در داده‌کاوی، ترکیبی از روش‌های خوشه‌بندی و طبقه‌بندی، اغلب هنگامی‌که تعداد داده‌ها زیاد است و داده‌ها مختل شده‌اند، با موفقیت مورداستفاده قرار می‌گیرند. درواقع، روش خوشه‌بندی که در گروهی از موارد مشابه انجام می‌شود، سیستم را قادر می‌سازد که نویز و پیچیدگی را کاهش دهد و بنابراین مرحله بعدی، طبقه‌بندی را تسهیل می‌کند. هدف مرحله طبقه‌بندی اختصاص دادن هر آیتم به خوشه‌ای است که متعلق به آن است (S. Thomassey 2010). به‌عبارت‌دیگر، درخت تصمیم‌گیری پیوندهای قابل‌فهم میان این خوشه‌ها و معیارهای توصیفی را پیدا می‌کند (Thomassey and Fiordaliso, 2006). درخت تصمیم‌گیری به‌عنوان یکی از محبوب‌ترین روش‌های طبقه‌بندی، مدل‌ها را مبتنی بر تفکیک مجدد داده‌ها می‌سازد. به‌طورمعمول یک الگوریتم درخت تصمیم‌گیری با تمام مجموعه داده‌ها شروع می‌شود، داده‌ها را به دو یا چند زیرمجموعه بر اساس یک یا چند ویژگی تقسیم می‌کند و سپس هر زیرمجموعه را به‌طور مکرر به زیرمجموعه‌های ریزتر تقسیم می‌کند تا اندازه هر زیرمجموعه به یک سطح مناسب برسد. کل فرآیند مدل‌سازی می‌تواند در یک ساختار درختی نشان داده شود و مدل تولیدشده می‌تواند به‌صورت مجموعه‌ای از قوانین «if-then» خلاصه شود. ویژگی درختان تصمیم‌گیری بدین شرح است: آسان در تفسیر، ارزان در محاسبات و قادر به مقابله با داده‌های پر نوسان و ناپایدار (X.-B. Li 2005).

در این بخش از مطالعه، پس از اعمال روش‌های خوشه‌بندی برای تقسیم داده‌ها بر اساس متغیرهای آب و هوایی، از روش درخت تصمیم‌گیری برای پیش‌بینی مقدار فروش بر اساس پیش‌بینی آب‌وهوا استفاده می‌شود. داده‌های ورودی درخت طبقه‌بندی، شاخص ماتریس حاصل از هر روش خوشه‌بندی است. درخت طبقه‌بندی در هر قسمت نشان داده‌شده است.

- خوشه‌بندی K-means

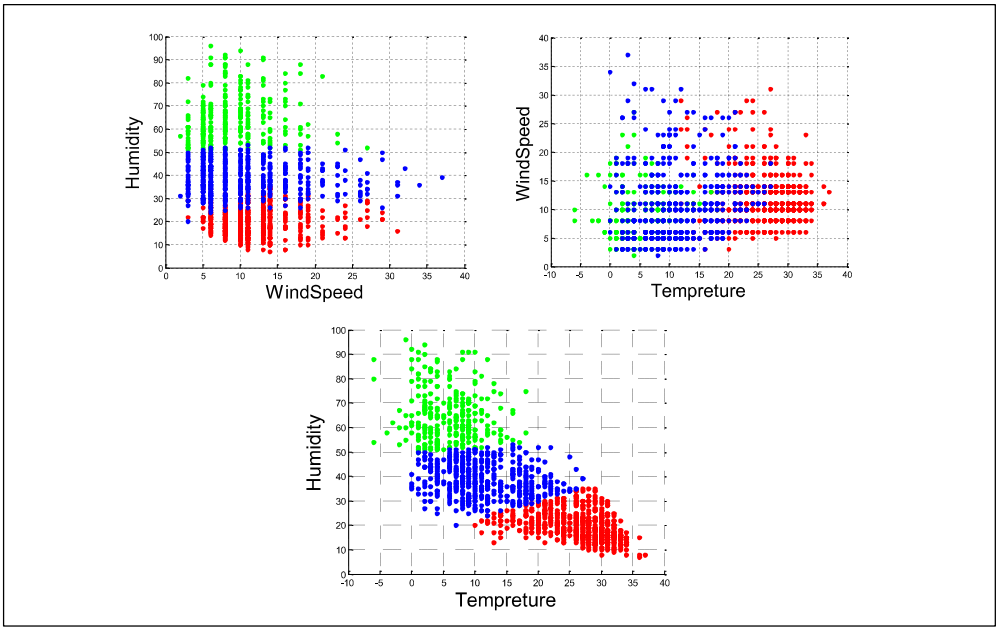
K-means اولین روش انتخاب‌شده برای خوشه‌بندی داده‌های آب و هوایی است.K-means ماتریس داده چندبعدی را به خوشه‌های K تفکیک می‌کند. تکرار این تفکیک موجب به حداقل رساندن فاصله مجموع کل خوشه‌های درون خوشه‌ای نسبت به میانگین خوشه[۱۵]می‌شود. K-means یک بردار حاوی شاخص خوشه‌ای هر نقطه است. برای پیدا کردن بهترین تعداد خوشه‌ها، از شاخص "Silhouette" استفاده می‌شود. طرح Silhouette نشان می‌دهد که زمانی که تعداد خوشه برابر ۳ شود ضریب silhouette حداکثر می‌شود و این نشان می‌دهد که تعداد خوشه‌ها بهینه است.

[۱۵] point-to-cluster-centroid distances

جدول ۳. شاخص میانگین Silhouette بر اساس تعداد خوشه‌های مختلف

تعداد خوشه	شاخص میانگین Silhouette
۳	۰,۶۷۲۲
۴	۰,۵۸۷۵
۵	۰,۵۲۷۴
۶	۰,۴۸۷۷
۷	۰,۴۲۵۹
۸	۰,۴۰۲۷
۹	۰,۴۳۶۸
۱۰	۰,۴۰۵۰
۱۱	۰,۳۶۰۹
۱۲	۰,۳۵۱۴
۱۳	۰,۳۳۶۶
۱۴	۰,۳۵۴۳
۱۵	۰,۳۴۱۰
۱۶	۰,۳۲۲۵

نتایج برای شهر «تهران» نشان داده‌شده است.



شکل ۷. نتایج خوشه‌بندی K-means

خوشه‌های شرایط آب و هوایی در جدول ۴ شرح داده‌شده است:

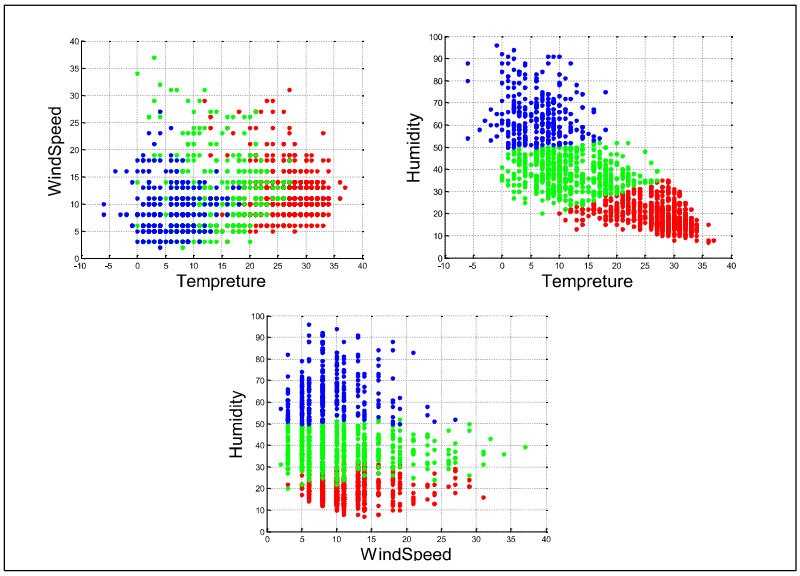
جدول ۴. نتایج خوشه‌بندی K-means

K-means	تعداد خوشه‌بندی		
	۱	۲	۳
متوسط دما	۲۶,۴	۶,۰	۱۱,۶
SD دما	۵,۰	۴,۳	۵,۹
متوسط رطوبت	۱۹,۸	۶۴,۷	۳۸,۶
SD رطوبت	۵,۶	۱۰,۳	۶,۷
متوسط سرعت باد	۱۱,۶	۸,۷	۱۰,۵
SD سرعت باد	۴,۱	۴,۲	۶,۲

- خوشه‌بندی فازی C-Means (FCM)

در بخش دوم، از روش FUZZY C-means برای خوشه‌بندی داده‌های آب و هوایی استفاده می‌شود. FCM یک تکنیک خوشه‌بندی داده است که در آن هرکدام از داده‌ها متعلق به یک خوشه است که با درجه عضویت تعیین می‌شود.

نمودار شماره ۱۲ نتایج خوشه‌بندی در شهر «تهران» را نشان می‌دهد. خوشه‌بندی در ۳ بعد انجام می‌شود. نمودارها به‌صورت جفت‌جفت ارائه می‌شوند.



شکل ۸. نتایج خوشه‌بندی FCM

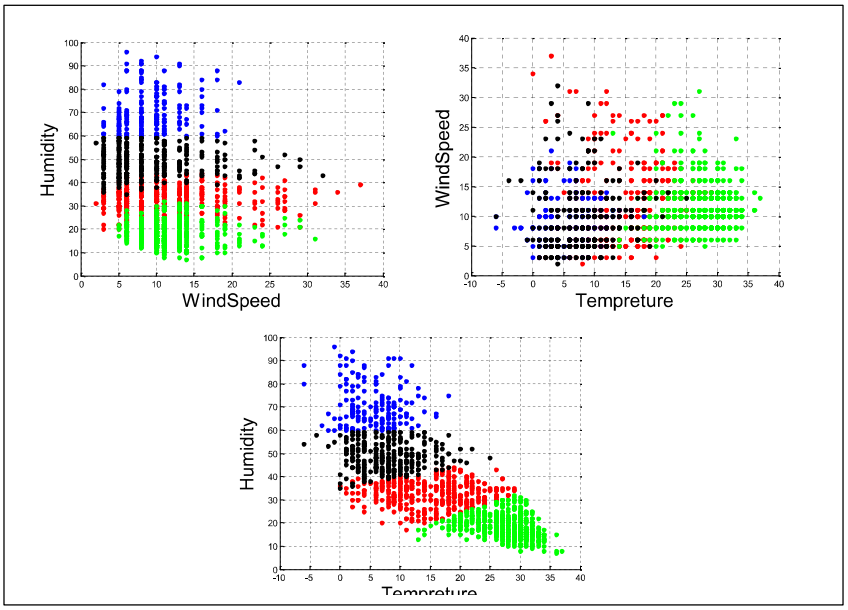
جدول ۵ میانگین و انحراف معیار مقادیر مختلف اقلیمی را برای هر خوشه نشان می‌دهد:

جدول ۵. نتایج خوشه‌بندی FCM

FCM	تعداد خوشه‌ها		
	۱	۲	۳
متوسط دما	۲۶,۶	۱۱,۹	۶,۰
SD دما	۴,۸	۶,۰۰	۴,۳
متوسط رطوبت	۱۹,۶	۳۸,۰۰	۶۴,۳
SD رطوبت	۵,۵	۶,۸	۱۰,۴
متوسط سرعت باد	۱۱,۵	۱۰,۵	۸,۸
SD سرعت باد	۴,۰	۶,۲	۴,۳

- خوشه‌بندی نقشه خودسازمان‌دهی (SOM)

در این بخش، از یک الگوریتم نقشه خودسازمان‌دهی برای خوشه‌بندی داده‌های آب و هوایی استفاده می‌شود. نقشه‌های خودسازمان‌دهی یاد می‌گیرند تا داده‌ها را بر اساس شباهت، توپولوژی، با اولویت (اما بدون تضمین) اختصاص یکسان تعداد نمونه به هر طبقه خوشه‌بندی کنند. [۲ ۲] برای سایز ابعاد انتخاب‌شده است، بنابراین به نظر می‌رسد که "۴" مناسب‌ترین تعداد خوشه باشد. نتیجه SOM برای شهر «تهران» در شکل ۹ نشان داده‌شده است.



شکل ۹. نتایج خوشه‌بندی SOM

خوشه‌ها در جدول ۶ شرح داده‌شده است:

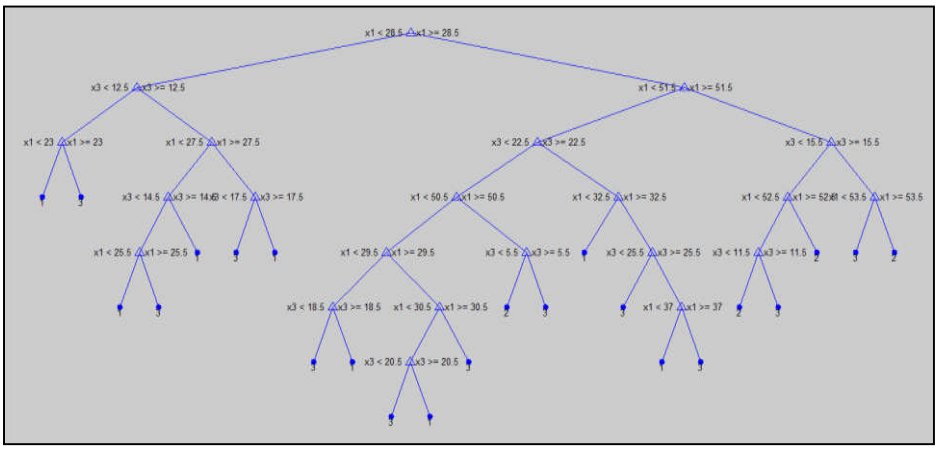
جدول ۶. نتایج خوشه‌بندی SOM

SOM	تعداد خوشه			
	۱	۲	۳	۴
متوسط دما	۱۵,۲	۲۷,۵	۵,۷	۷,۸
SD دما	۶,۳	۴,۲	۴,۱	۴,۷
متوسط رطوبت	۳۲,۶	۱۸,۵	۷۰,۵	۴۸,۹
SD رطوبت	۵,۲	۴,۷	۸,۹	۵,۸
متوسط سرعت باد	۱۱,۵	۱۱,۴	۹,۰	۹,۰
SD سرعت باد	۶,۳	۳,۷	۳,۷	۵,۴

حالا که تعداد محدودی از موارد به گروه‌های از پیش تعریف‌شده (Clusters) اختصاص داده‌شده است، می‌توانیم از روش‌های طبقه‌بندی برای به دست آوردن یک سیستم پیشنهادکننده استفاده کنیم. برای این منظور، خروجی خوشه‌بندی که یک ماتریس شاخص است به‌عنوان ورودی درخت طبقه‌بندی

مورداستفاده قرار می‌گیرد. مقایسه نتایج حاصل از سه روش خوشه‌بندی بر اساس شاخص Silhouette نشان می‌دهد که K-means و FCM تقریباً با دقت و اعتبار یکسان عمل می‌کنند؛ بنابراین در بخش بعدی، درخت طبقه‌بندی بر اساس خروجی هر دو روش خوشه‌بندی ساخته می‌شود؛ و سپس بهترین سیستم پیش‌بینی بر اساس معیار «ضرر جایگزینی مجدد»[۱] انتخاب خواهد شد.

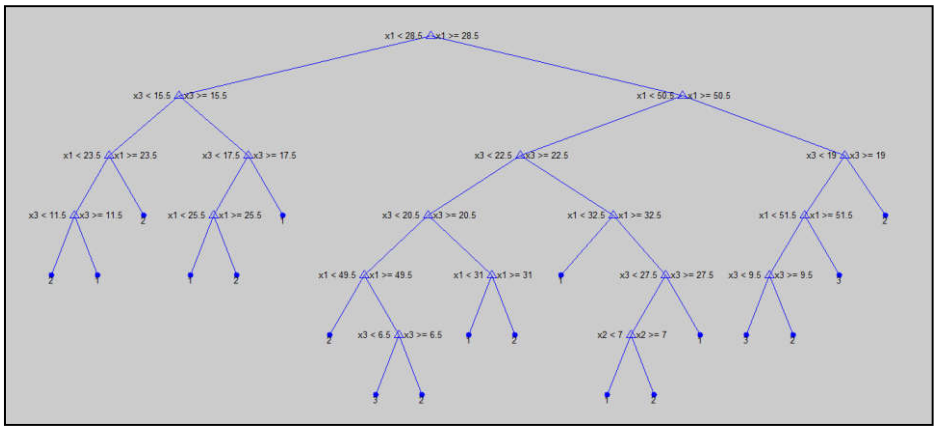
اول، ماتریس خروجی شاخص‌های خوشه‌بندی K-means به‌عنوان یک ورودی برای درخت طبقه‌بندی استفاده می‌شود. نتیجه در شکل ۱۰ نشان داده‌شده است. ضرر جایگزینی مجدد برابر با ۰,۰۰۴۱ است.



شکل ۱۰. درخت طبقه‌بندی بر اساس نتایج K-means

دوم، از نتیجه FCM برای ایجاد یک درخت طبقه‌بندی استفاده می‌شود. درخت طبقه‌بندی خارج‌شده نشان می‌دهد که زیان جایگزینی مجدد ۰,۰۰۶۱ است که نشان‌دهنده ضرر بیشتر نسبت به طبقه‌بندی مرحله اول است.

[۱] Re-substitution loss



شکل ۱۱. درخت طبقه‌بندی بر اساس نتایج FCM

ازآنجاکه روش خوشه‌بندی K-means، سطح بالایی از دقت و قابلیت اطمینان نسبت به FCM را نشان می‌دهد، K-means به‌عنوان سیستم پیش‌بینی دقیق‌تر برای شرکت لبنی ایرانی انتخاب می‌شود. سیستم پیش‌بینی انتخاب‌شده قادر است که با دریافت شرایط آب و هوایی، خوشه تقاضا را مشخص کند. پس از مشخص کردن خوشه، متوسط فروش پیش‌بینی‌شده کلیه محصولات، به تحلیلگر فروش داده خواهد شد. بر اساس این سیستم پیش‌بینی، رفتار فروش را می‌توان در ۳ خوشه دسته‌بندی کرد. برای طراحی سودمندترین سیستم پیشنهاد استراتژی‌های فروش، لازم است که هر خوشه در هر شهر به‌طور کامل تجزیه‌وتحلیل و درک شود، زیرا خوشه‌ها اشاره به ماه‌های مختلف سال در شهرهای مختلف دارد. برای این منظور، ماتریس شاخص‌های خوشه‌ای حاصل از K-means برای مشخص کردن اینکه ماه‌ها در دوره موردمطالعه در کدام خوشه قرار می‌گیرند، استفاده می‌شود. نتیجه در جدول ۷ نشان داده‌شده است:

جدول ۷. توزیع هر خوشه K-means در ماه های مختلف سال

اهواز (K-means)	۱	۲	۳	۴	۵	۶	۷	۸	۹	۱۰	۱۱	۱۲	مجموع
۱			۱۶	۶۵	۱۰۷	۱۱۷	۱۲۰	۱۱۰	۱۰۶	۵۰	۳		۶۹۴
۲	۴۸	۷۵	۹۸	۴۸	۱۰			۸	۱۰	۶۷	۵۰	۳۳	۴۴۷
۳	۱۰۲	۶۲	۱۴	۱	۴					۱	۶۲	۸۹	۳۳۵
مجموع	**۱۵۰**	**۱۳۷**	**۱۲۸**	**۱۱۴**	**۱۲۱**	**۱۱۷**	**۱۲۰**	**۱۱۸**	**۱۱۶**	**۱۱۸**	**۱۱۵**	**۱۲۲**	**۱۴۷۶**

آمل (K-means)	۱	۲	۳	۴	۵	۶	۷	۸	۹	۱۰	۱۱	۱۲	مجموع
۱	۲۲	۶	۲۲	۲۷	۴	۶۱	۶۱	۵۲	۲۳	۲۲	۷	۴	۳۵۰
۲	۱۲۸	۱۳۱	۱۰۴	۵۱	۸				۱	۳۲	۹۷	۱۱۶	۶۶۸
۳			۲	۳۶	۷۰	۵۶	۵۹	۶۶	۹۲	۶۴	۱۱	۲	۴۵۸
مجموع	**۱۵۰**	**۱۳۷**	**۱۲۸**	**۱۱۴**	**۱۲۱**	**۱۱۷**	**۱۲۰**	**۱۱۸**	**۱۱۶**	**۱۱۸**	**۱۱۵**	**۱۲۲**	**۱۴۷۶**

اصفهان (K-means)	۱	۲	۳	۴	۵	۶	۷	۸	۹	۱۰	۱۱	۱۲	مجموع
۱	۷۹	۹۹	۹۲	۴۹	۲۸	۲		۵	۶	۵۷	۵۰	۴۱	۵۰۸
۲		۲	۱۳	۴۴	۸۴	۱۱۵	۱۲۰	۱۱۳	۱۱۰	۵۴			۶۵۵
۳	۷۱	۳۶	۲۳	۲۱	۹					۷	۶۵	۸۱	۳۱۳
مجموع	**۱۵۰**	**۱۳۷**	**۱۲۸**	**۱۱۴**	**۱۲۱**	**۱۱۷**	**۱۲۰**	**۱۱۸**	**۱۱۶**	**۱۱۸**	**۱۱۵**	**۱۲۲**	**۱۴۷۶**

مشهد (K-means)	۱	۲	۳	۴	۵	۶	۷	۸	۹	۱۰	۱۱	۱۲	مجموع
۱	۲	۳	۶	۳۵	۶۸	۱۰۴	۱۲۰	۱۱۵	۱۰۱	۴۱			۵۹۵
۲	۹۱	۷۶	۵۶	۲۰	۵					۲۰	۷۰	۷۶	۴۱۴
۳	۵۷	۵۸	۶۶	۵۹	۴۸	۱۳		۳	۱۵	۵۷	۴۵	۴۶	۴۶۷
مجموع	**۱۵۰**	**۱۳۷**	**۱۲۸**	**۱۱۴**	**۱۲۱**	**۱۱۷**	**۱۲۰**	**۱۱۸**	**۱۱۶**	**۱۱۸**	**۱۱۵**	**۱۲۲**	**۱۴۷۶**

شیراز (K-means)	۱	۲	۳	۴	۵	۶	۷	۸	۹	۱۰	۱۱	۱۲	مجموع
۱	۶۶	۴۹	۱۷	۸	۴					۴	۵۷	۳۶	۲۴۱
۲	۲	۱	۲۰	۴۸	۱۰۱	۱۱۷	۱۲۰	۱۱۳	۱۱۳	۸۳			۷۱۸
۳	۸۲	۸۷	۹۱	۵۸	۱۶			۵	۳	۳۱	۵۸	۸۶	۵۱۷
مجموع	**۱۵۰**	**۱۳۷**	**۱۲۸**	**۱۱۴**	**۱۲۱**	**۱۱۷**	**۱۲۰**	**۱۱۸**	**۱۱۶**	**۱۱۸**	**۱۱۵**	**۱۲۲**	**۱۴۷۶**

تهران (K-means)	۱	۲	۳	۴	۵	۶	۷	۸	۹	۱۰	۱۱	۱۲	مجموع
۱		۱	۳۸	۵۹	۸۹	۱۱۴	۱۱۹	۱۱۰	۱۰۶	۴۶			۶۸۲
۲	۶۴	۴۷	۲۳	۹	۱			۱		۱۰	۶۳	۶۹	۲۸۷
۳	۸۶	۸۹	۶۷	۴۶	۳۱	۳	۱	۷	۱۰	۶۲	۵۲	۵۳	۵۰۷
مجموع	**۱۵۰**	**۱۳۷**	**۱۲۸**	**۱۱۴**	**۱۲۱**	**۱۱۷**	**۱۲۰**	**۱۱۸**	**۱۱۶**	**۱۱۸**	**۱۱۵**	**۱۲۲**	**۱۴۷۶**

ازآنجاکه خوشه‌ها به‌صورت فصلی دریافت می‌شوند، اکنون زمان آن است که روند تقاضای خوشه‌ها را

تحلیل کنیم. بدین منظور، متوسط فروش در هر خوشه محاسبه می‌شود تا به رفتار مشتری در شرایط مختلف آب و هوایی رسیدگی شود. نتیجه روش K-Means در جدول ۸ نشان داده‌شده و خوشه با حداکثر تقاضا در هر شهر برجسته شده است:

جدول ۸. میانگین تقاضای روزانه محصولات در هر خوشه K-means

تعداد خوشه	اهواز			آمل			اصفهان		
	۱	۲	۳	۱	۲	۳	۱	۲	۳
خامه ۱	-	۰,۲	۰,۰	۳۳,۹	۳۴,۵	۳۶,۹	۴,۵	۱۲,۱	۳,۸
خامه ۲	۱۳,۵	۱۵,۴	۱۶,۶	۱۳۲,۳	۱۲۳,۱	۱۴۲,۸	۶۷,۳	۷۶,۴	۷۲,۹
خامه ۳	۰,۳	۰,۴	۱,۵	۶,۵	۷,۸	۷,۹	۲۱,۵	۱۷,۶	۱۸,۹
خامه ۴	۰,۰	۰,۰	۰,۰	۲,۱	۲,۵	۲,۶	۲۰,۷	۱۸,۴	۲۵,۱
خامه ۵	۲۰,۹	۲۳,۶	۲۷,۷	۶۴,۵	۶۷,۵	۶۷,۶	۵۰,۲	۴۴,۰	۴۹,۰
دوغ ۱	۲۷۸۰,۰	۲۱۹۲,۰	۱۶۹۸,۰	۳۵۴,۰	۲۰۵,۷	۳۶۲,۶	۱۶۸۹,۳	۱۹۹۹,۸	۱۳۵۴,۱
دوغ ۲	۱۱۳۵,۱	۸۸۹,۳	۶۳۹,۵	۴۶۶,۶	۲۶۴,۰	۴۵۷,۶	۴۳۶,۰	۵۳۰,۱	۳۴۳,۴
دوغ ۳	۱۱۳۹,۵	۱۰۴۴,۴	۷۶۹,۷	۳۷۱,۸	۲۲۵,۱	۳۷۲,۳	۴۷۰,۴	۵۰۳,۷	۳۸۵,۵
دوغ ۴	۱۴۱۸,۴	۹۱۹,۵	۵۶۳,۲	۷۰,۹	۳۶,۹	۷۵,۶	۴۶۶,۶	۶۰۷,۳	۳۷۸,۸
دوغ ۵	۳۱۳,۹	۲۱۴,۸	۱۲۵,۵	۶۵۲,۸	۳۶۶,۲	۶۸۷,۹	۵۴۱,۱	۶۷۰,۹	۴۵۹,۷
شیر ۱	۴۸۰,۶	۴۲۶,۶	۴۷۹,۲	۳۱۹,۶	۲۶۹,۴	۳۶۰,۷	۳۸۳,۱	۴۰۴,۷	۳۰۹,۶
شیر ۲	۱۳۲۷,۴	۹۸۵,۸	۱۲۵۳,۲	۲۱۳,۵	۱۵۱,۴	۱۷۹,۰	۳۳۵,۷	۲۳۷,۳	۲۳۴,۱
شیر ۳	۶۴۱,۳	۶۰۴,۹	۵۲۲,۴	۲۹۴,۸	۳۰۰,۹	۳۶۰,۵	۵۰۷,۰	۵۵۱,۲	۵۲۴,۹
شیر ۴	۱۳۵۲,۳	۱۳۲۶,۲	۸۹۳,۰	۸۴۷,۰	۷۵۹,۵	۹۷۳,۸	۸۴۶,۶	۹۰۵,۳	۹۷۰,۵
شیر ۵	۲۵۹,۷	۳۳۹,۸	۳۳۲,۵	۲۹۱,۹	۳۰۲,۷	۳۲۰,۵	۵۴۳,۸	۴۳۸,۹	۵۲۳,۲
ماست ۱	۸۰,۳	۱۵۰,۶	۱۶۳,۱	۶۵,۴	۱۲۴,۱	۶۷,۲	۲۴,۶	۱۶,۱	۱۳,۱
ماست ۲	۲۲۶,۱	۲۲۸,۲	۲۲۳,۹	۲۵۶,۹	۲۰۷,۹	۲۸۰,۸	۵۱,۲	۴۱,۲	۴۰,۶
ماست ۳	۱۶۰,۷	۱۶۳,۵	۱۵۰,۷	۱۲۵,۴	۹۵,۷	۱۳۶,۶	۱۳۲,۶	۱۳۶,۸	۱۲۰,۹
ماست ۴	۴,۸	۱,۰	۳,۷	۱۷,۳	۱۵,۴	۱۷,۵	۵۴,۲	۶۳,۴	۴۹,۱
ماست ۵	۲۳۰۷,۷	۲۲۱۹,۵	۲۰۱۵,۷	۵۱۲,۱	۴۰۷,۷	۵۴۱,۹	۱۱۲۸,۱	۱۱۸۴,۶	۱۱۱۹,۳

ادامه جدول ۸. میانگین تقاضای روزانه محصولات در هر خوشه K-means

تعداد خوشه	مشهد			شیراز			تهران		
	۱	۲	۳	۱	۲	۳	۱	۲	۳
خامه ۱	۱۵,۳	۱۶,۰	۱۷,۳	۱,۶	۱,۶	۱,۳	۲۰,۴	۲۳,۲	۲۳,۴
خامه ۲	۱۵۳,۳	۱۵۷,۳	۱۷۰,۶	۳۰,۷	۳۰,۹	۲۸,۹	۳۰۲,۲	۲۹۹,۵	۲۹۵,۵
خامه ۳	۲۵,۰	۲۸,۵	۲۸,۷	۴,۷	۳,۲	۴,۷	۶۴,۶	۶۹,۵	۸۱,۵
خامه ۴	۲۳,۹	۲۳,۱	۲۵,۰	۵,۲	۵,۶	۴,۶	۸۵,۲	۸۰,۰	۸۰,۱
خامه ۵	۱۸,۱	۲۰,۴	۲۰,۴	۵۹,۹	۶۷,۶	۶۳,۶	۳۴,۹	۲۹,۴	۳۸,۴
دوغ ۱	۳۰۲۳,۸	۲۲۵۵,۴	۲۵۵۲,۴	۹۲۵,۹	۱۲۵۰,۸	۱۰۳۰,۶	۱۳۰۶۷,۰	۹۷۵۲,۰	۱۲۱۰۲,۰
دوغ ۲	۴۴۴,۷	۲۸۹,۹	۳۵۱,۳	۱۴۲,۵	۲۰۶,۳	۱۶۷,۳	۳۴۶۶,۰	۲۱۹۷,۰	۲۵۷۱,۰
دوغ ۳	۲۹۳,۲	۲۴۷,۱	۲۷۶,۸	۹۰,۳	۱۴۰,۴	۱۲۳,۳	۷۳۷۰,۰	۵۶۳۸,۰	۶۳۷۶,۰
دوغ ۴	۴۵۹,۰	۴۰۹,۱	۴۵۰,۶	۱۱۴,۷	۱۹۷,۲	۱۵۴,۷	۳۸۷۵,۰	۲۸۲۷,۰	۳۲۰۰,۰
دوغ ۵	۸۱۰,۲	۵۸۱,۳	۶۵۷,۹	۱۵۵,۳	۲۶۳,۰	۱۹۸,۰	۶۳۱۱,۰	۴۱۵۹,۰	۴۷۷۲,۰
شیر ۱	۵۷۷,۲	۵۵۷,۸	۶۶۲,۲	۳۰۱,۰	۳۳۸,۹	۳۸۳,۲	۳۵۴۹,۱	۳۹۳۴,۹	۴۳۴۴,۵
شیر ۲	۷۵۸,۸	۴۴۲,۴	۷۳۲,۷	۲۹۲,۸	۳۱۶,۳	۳۶۸,۸	۱۶۹۱,۳	۱۱۴۰,۶	۱۶۳۱,۷
شیر ۳	۹۸۰,۴	۸۵۵,۹	۸۷۵,۵	۴۱۲,۱	۳۷۲,۹	۴۱۲,۹	۹۰۴,۷	۱۱۶۷,۵	۸۷۳,۸
شیر ۴	۲۷۷۴,۴	۲۶۹۲,۲	۲۷۸۳,۴	۷۶۱,۶	۶۸۷,۱	۷۶۳,۳	۱۸۵۶,۵	۳۳۷۸,۰	۲۰۸۹,۳
شیر ۵	۵۴۵,۹	۵۹۳,۰	۶۳۰,۰	۳۲۲,۴	۲۸۴,۶	۳۲۹,۵	۱۸۷۲,۰	۲۳۹۳,۲	۲۴۳۳,۳
ماست ۱	۲۸,۶	۷۲,۵	۳۳,۸	۲۱۵,۲	۱۹۲,۶	۲۲۹,۳	۲۶۴۴,۰	۲۵۲۰,۰	۲۸۶۷,۰
ماست ۲	۱۳۱,۹	۱۱۸,۲	۱۲۱,۲	۳۵,۶	۵۷,۰	۵۳,۳	۱۱۵۹,۰	۸۳۷,۰	۱۲۵۸,۰
ماست ۳	۱۷۰,۶	۱۴۸,۵	۱۵۷,۵	۵۸,۰	۷۱,۲	۶۶,۱	۱۹۷۱,۰	۱۷۳۴,۰	۱۹۰۳,۰
ماست ۴	۸۵,۲	۶۰,۲	۶۴,۸	۱۴۱,۴	۱۶۸,۴	۱۶۶,۸	۶۱۶۹,۰	۴۹۱۸,۰	۵۹۸۸,۰
ماست ۵	۳۷۶۲,۷	۳۴۹۳,۹	۳۵۵۰,۵	۱۷۴۰,۶	۱۸۹۱,۲	۱۸۳۱,۱	۱۹۳۵۸,۰	۱۹۱۴۲,۰	۱۹۷۱۰,۰

با تجزیه‌وتحلیل نتایج جدول ۸، این تحقیق مدل رفتار مشتری را بر اساس تغییرات اقلیمی معرفی می‌کند. به‌عنوان‌مثال، در آب‌وهوای گرم خوشه اول شهر اهواز، تقاضای نوشیدنی دوغ حداکثر است که سرنخ‌هایی را به نفع تولید بیشتر این محصول ارائه می‌دهد.

بنابراین بر اساس پیش‌بینی آب‌وهوا، درخت طبقه‌بندی خوشه آب‌وهوا، تقاضا را در شهرهای مختلف تعیین می‌کند. برنامه‌ریزی مناسبی باید با توجه به تقاضای پیش‌بینی‌شده اتخاذ شود. ابزارهای تبلیغاتی مختلف می‌تواند در بازار مورداستفاده قرار گیرد، مانند نمونه‌های رایگان، کوپن، جایزه، بسته‌بندی و تبلیغات متقابل. یانگ و لای (۲۰۰۶) به معرفی و مقایسه راهبردهای مختلف بسته‌بندی مبتنی بر داده‌های فروش آنلاین پرداخته‌اند. بسته‌بندی محصول، یک استراتژی فروش فراگیر در بازار است. این استراتژی یک شیوه رایج است که شامل ترکیب دو یا چند محصول یا خدمات و فروش آن‌ها باقیمت یک مجموعه می‌شود. چالش این است که چگونه محصول مناسب را انتخاب کنید تا به عملکرد مورد انتظار دست‌یابید، مانند ایجاد بازارهای جدید یا افزایش وفاداری مشتری، فروش یا سود (Yang and Lai 2006).

نتیجه‌گیری و پیشنهاد برای تحقیقات آتی

ازآنجاکه محصولات لبنی سریع فاسد می‌شوند و روزانه تولید می‌شوند، نوسانات غیرمنتظره در تقاضای مصرف‌کننده، مشکلات مهمی را برای تجاری‌سازی محصولات لبنی ایجاد می‌کند. به‌طورکلی شرکت‌ها امیدوار هستند که مشتریانشان در صورت نبود محصول موردنظرشان در فروشگاه، موارد جایگزین بخرند یا خریدشان را به بعد موکول کنند؛ اما در محیط رقابتی امروز بیشتر احتمال دارد که مشتریان از تولیدکننده‌ی دیگری خرید کنند و هرگز به تولیدکننده‌ی قبلی بازنگردند. برای جلوگیری از کمبود محصول، خرده‌فروشان عملاً می‌توانند ذخیره ایمنی را افزایش دهند که این امر سود را به‌طور هنگفتی کاهش می‌دهد. در چنین شرایطی، خرده‌فروشان با یک دوراهی مواجه می‌شوند: آن‌ها باید بین از دست دادن مشتری و هزینه‌های ذخیره تعادل برقرار کنند (Huang, Fildes and Soopramanien ۲۰۱۴).

بنابراین، سازمان‌هایی که به دنبال فروش خوب هستند اغلب نیاز به حفظ تعادل بین تقاضای مشتری و کنترل هزینه‌های موجودی دارند. موجودی بیشتر انبار، تقاضای مشتری را در تمام مراحل برآورده می‌سازد، اما ذخیره بیش‌ازحد می‌تواند منجر به مشکلاتی ازجمله از بین رفتن سرمایه، فاسدشدن موجودی انبار و کاهش سود شود. در مقابل، سطح موجودی پایین‌تر اگرچه هزینه‌های موجودی را کاهش می‌دهد، اما هزینه فرصت‌های فروش ازدست‌رفته را افزایش می‌دهد و کاهش رضایت مشتری و سایر مشکلات را به همراه دارد. پیش‌بینی فروش می‌تواند برای تعیین سطح موجودی متناسب و اجتناب

از مشکل ذخیره بیشتر یا کمتر از نیاز مورداستفاده قرار گیرد. علاوه بر این، پیش‌بینی فروش می‌تواند پیامدهایی برای برنامه‌ریزی مالی، بازاریابی، مدیریت مشتری و دیگر زمینه‌های کسب‌وکار داشته باشد. بدین ترتیب، ارتقای دقت پیش‌بینی‌های فروش، تبدیل به یک جنبه مهم در مدیریت کسب‌وکار شده است (Lu, Lee and Lian 2012).

شرکت‌های مواد غذایی به دلیل ویژگی‌های خاص خود، از قبیل کوتاه بودن عمر محصولاتشان، نیاز به حفظ کیفیت بالای محصول و نوسانات خواسته‌های مصرف‌کننده، بیشتر به پیش‌بینی‌های فروش توجه دارند. درحالی‌که محصولات فقط برای مدت محدودی فروخته می‌شوند، کمبود و مازاد کالا می‌تواند منجر به از دست رفتن درآمد برای شرکت شود (Doganis, et al. 2006). در مطالعه حاضر سعی شده است که به این موارد با تجزیه‌وتحلیل و استخراج داده‌های فروش ۲۰ محصول از ۴ گروه عمده لبنیات (خامه، دوغ، شیر پاستوریزه و ماست) که مربوط به یک تولیدکننده بود جامه عمل پوشانده شود. برای ایجاد یک سیستم پیش‌بینی فروش و همچنین یک سیستم پیشنهادی استراتژی فروش، از روش‌های مختلف داده‌کاوی استفاده شد. ورودی اصلی تمام مراحل داده‌کاوی وضعیت هوا بود که به نظر می‌رسد تأثیر قابل‌توجهی بر تقاضای محصولات لبنی داشته است. در ضمن، این مطالعه ثابت کرد که تقاضای مشتری به نوسانات قیمت بسیار حساس نیست. روش‌های داده‌کاوی در ۲ مرحله اعمال شد. در مرحله اول داده‌های تاریخی آب‌وهوا با استفاده از ۳ روش K-means، FCM و SOM خوشه‌بندی شد. نتیجه هر روش خوشه‌بندی با استفاده از معیارهای شبیه‌سازی silhouette مورد ارزیابی قرار گرفت که FCM و K-means به‌عنوان روش‌هایی شناخته شدند که میزان دقت مشابهی دارند؛ بنابراین خروجی هر دو روش خوشه‌بندی وارد مرحله دوم داده‌کاوی شد. در مرحله دوم، درخت طبقه‌بندی بر اساس ماتریس شاخص هر دو روش انتخاب‌شده، طراحی شد. با مقایسه زیان جایگزینی مجدد دو درخت طبقه‌بندی، K-means به‌عنوان روشی دقیق‌تر معرفی شد؛ بنابراین، مناسب‌ترین روش برای پیش‌بینی فروش در شرکت لبنی ایرانی در این تحقیق، روشی ترکیبی از خوشه‌بندی K-means با ۳ خوشه و درخت تصمیم‌گیری است. این سیستم پیش‌بینی، شرایط آب و هوایی را به‌عنوان ورودی دریافت می‌کند و خوشه فروش را به‌عنوان خروجی ارائه می‌دهد. خوشه‌های فروش به‌طور کامل در این مقاله برای هر شهر شرح داده‌شده و بر این اساس، استراتژی فروش و بازاریابی مناسب پیشنهادشده است.

مدل پیش‌بینی نهایی نشان می‌دهد که تقاضا در بازار لبنیات نسبت به تغییرات شرایط آب و هوایی بسیار حساس است. این در حالی است که تقاضا واکنش سریعی به تغییرات قیمت نشان نمی‌دهد.

برای تحقیق در آینده، تجزیه‌وتحلیل سبد بازار برای کشف الگوهای خرید مشتری همراه با کشف وابستگی‌ها یا همپوشانی پایگاه داده‌های فروش و توزیع جالب به نظر می‌رسد. اطلاعات به‌دست‌آمده از

تجزیه‌وتحلیل می‌تواند در تشکیل استراتژی‌های بازاریابی و فروش مانند تبلیغات بسته‌بندی مؤثر واقع شود.

لازم به ذکر است که شرکت موردپژوهش دارای مدل‌های پیش‌بینی فروشی است که بر اساس متغیرهای مستقل شناخته‌شده‌ی شرکت طراحی‌شده است و تغییرات اقلیمی و شرایطی جوی و فصلی در آن در نظر گرفته نشده است. بر این اساس پیشنهاد می‌گردد که سیستم جامعی طراحی شود که مدل تحقیق حاضر را با مدل‌های قدیمی پیش‌بینی فروش شرکت ترکیب نموده و کلیه متغیرهای مستقل مهم و تعیین‌کننده را بکار ببرد.

ضمناً تحقیق حاضر بر اساس داده‌های منتخب از شرکت بزرگ لبنیاتی ایرانی صورت گرفته است. توصیه می‌شود داده‌های کامل لبنیات برای طیف گسترده‌ای از محصولات به‌منظور اعتبارسنجی جامع‌تر و بررسی نتایج در سراسر این صنعت جمع‌آوری گردد.

فهرست منابع

رضایی نور, جلال, و محمدرضا شیخ بهایی. ۱۳۹۶. "کاربردهای دادهکاوی متنی در حوزه مدیریت دانش زنجیره خدمات دولت الکترونیکی." مجله مدیریت فناوری اطلاعات، دانشکده مدیریت دانشگاه تهران ۹ (۱): ۳۹-۶۰.

قانعی استاد, محمد،, حسین خسروی مهموئی, و مجید، عبدالرزاق نژاد. ۱۳۹۶. "کاهش ویژگی سیستمهای اطلاعاتی ناقص بر مبنای تئوری مجموعه راف با استفاده از الگوریتم رقابت استعماری فازی." مجله مدیریت فناوری اطلاعات، دانشکده مدیریت دانشگاه تهران ۹ (۱): ۱۲۳-۱۴۲.

Alon, Ilan, Min Qi, and Robert J. Sadowski. 2001. "Forecasting aggregate retail sales: A comparison of artificial neural networks and traditional methods." *Journal of Retailing and Consumer Services* 147-156.

Chalaris, Manolis, Stefanos Gritzalis, Manolis Maragoudakis, Cleo Sgouropoulou, and Anastasios Tsolakidis. 2014. "Improving Quality of Educational Processes Providing New Knowledge using Data Mining Techniques." *Procedia - Social and Behavioral Sciences* 390 – 397.

Chang, Pei-Chann, Chen-Hao Liu, and Yen-Wen Wang. 2006. "A hybrid model by clustering and evolving fuzzy rules for sales decision supports in printed circuit board industry." *Decision Support Systems* ۴۲ (۳): ۱۲۵۴۱۲۶۹.

Chang, Pei-Chann, Yen-Wen Wang, and Chen-Hao Liu. 2007. "The development of a weighted evolving fuzzy neural network for PCB sales forecasting." *Expert Systems with Applications* 32 (1): 86–96.

Chen, Chen-Yuan, Wan-I Lee, Hui-Ming Kuo, Cheng-Wu Chen, and Kung-Hsing Chen. 2010. "The study of a forecasting sales model for fresh food." *Expert Systems with Applications* 37 (12): 7696–۷۷۰۲.

Chen, F.L., and T.Y. Ou. 2009. "Gray relation analysis and multilayer functional link network sales forecasting model for perishable food in convenience store." *Expert Systems with Applications* 36 (3): 7054–۷۰۶۳.

Chen, S., and M. Huang. 2011. "Constructing credit auditing and control & management model with data mining technique." *Expert Systems with Applications* 5359–5365.

Davis, Donna F., and John T. Mentzer. 2007. "Organizational factors in sales forecasting management." *International Journal of Forecasting* ۴۷۵۴۹۵.

Doganis, Philip, Alex Alexandridis, Panagiotis Patrinos, and Haralambos Sarimveis. 2006. "Time series sales forecasting for short shelf-life food products based on artificial neural networks and evolutionary computing." *Journal of Food Engineering* 196–204.

Farvaresh, Hamid, and MohammadMehdi Sepehri. 2011. "A data mining frame work for detecting subscription fraud in telecommunication." *Engineering Applications of Artificial Intelligence* 24 (1): 182–194.

Francesconi, G. N., N. Heerink, and M. D'Haese. 2010. "Evolution and challenges of dairy supply chains: Evidence from supermarkets, industries and consumers in Ethiopia." *Food Policy* 35 (1): 60–68.

Guo, Z.X., W.K. Wong, and Min Li. 2013. "A multivariate intelligent decision-making model for retail sales forecasting." *Decision Support Systems* 247–255.

Hassan, M.F., M.A. Shalaby, and I.A. Gazarin. 1988. "A comparative evaluation of some non seasonal time series forecasting methods." *Current Advances in Mechanical Design & Production IV* Dec: 637–۶۴۳.

Huang, T., R. Fildes, and D. Soopramanien. 2014. "The value of competitive information in forecasting FMCG retail product sales and the variable selection problem." *European Journal of Operational Research* 237 (2): 738–748.

Khashei, Mehdi, and Mehdi Bijari. 2012. "A new class of hybrid models for time series forecasting." *Expert Systems with Applications* 4344–۴۳۵۷.

Köksal, Gülser, Inci Batmaz, and Murat Caner Testik. 2011. "A review of data mining applications for quality improvement in manufacturing industry." *Expert Systems with Applications* 38 (10): 13448–13467.

Kurgan, L. A., and P. Musilek. 2006. "A survey of knowledge discovery and data mining process models." *The Knowledge Engineering Review* ۲۱ (۱): ۱–۲۴.

Li, Xiao-Bai. 2005. "A scalable decision tree system and its application in pattern recognition and intrusion detection." *Decision Support Systems* ۴۱ (۱): ۱۱۲–۱۳۰.

Li, Xiongmin, and ChristineW. Chan. 2010. "Application of an enhanced decision tree learning approach for prediction of petroleum production." *Engineering Applications of Artificial Intelligence* 23 (1): ۱۰۲–۱۰۹.

Lu, Chi-Jie , Tian-Shyug Lee, and Chia-Mei Lian. 2012. "Sales forecasting for computer wholesalers: A comparison of multivariate adaptive regression splines and artificial neural networks." *Decision Support Systems* 584–596.

Ma, Jungmok, Minjung Kwak, and Harrison M. Kim. 2014. "Demand Trend Mining for Predictive Life Cycle Design." *Journal of Cleaner Production* 68: 189-199.

Navarro-Barrientos, J.-Emeterio, Dieter Armbruster, Hongmin Li, Morgan Dempsey, and Karl G. Kempf. 2014. "Characterization and analysis of sales data for the semiconductor market: An expert system approach." *Expert Systems with Applications* 41 (3): 893–903.

Ngai, E.W.T., Li Xiu, and D.C.K. Chau. 2009. "Application of data mining techniques in customer relationship management: A literature review and classification." *Expert Systems with Applications* 36 (2): 2592–۲۶۰۲.

Papalexopoulos, A.D., and T.C. Hesterberg. 1990. "A regression-based approach to short-term system load forecasting." *IEEE Transactions on Power Systems* 5 (4): 1535–1547.

Park, J.H., Y.M. Park, and K.Y Lee. 1991. "Composite modeling for adaptive short-term load forecasting." *IEEE Transactions on Power Systems* 6 (2): 450– 457.

Shah, Mrinalini. 2012. "Fuzzy based trend mapping and forecasting for time series data." *Expert Systems with Applications* 39 (7): 6351–6358.

Sharma, S., K.M. Osei-Bryson, and G.M. Kasper. 2012. "Evaluation of an integrated Knowledge Discovery and Data Mining process model." *Expert Systems with Applications* 39: 11335-11348.

Thomassey, Se´ bastien. 2010. "Sales forecasts in clothing industry: The key success factor of the supply chain management." *International Journal of Production Economics* 128 (2): 470–483.

Thomassey, Se´bastien, and Antonio Fiordaliso. 2006. "A hybrid sales forecasting system based on clustering and decision trees." *Decision Support Systems* 408–421.

Thomassey, Se´bastien, and Michel Happiette. 2007. "A neural clustering and classification system for sales forecasting of new apparel items." *Applied Soft Computing* 1177–1187.

Tratar, Liljana Ferbar. 2015. "Forecasting method for noisy demand." *International Journal of Production Economics* 161: 64–73.

Validi, Sahar, Arijit Bhattacharya, and P.J. Byrne. 2014. "A case analysis of a sustainable food supply chain distribution system—A multi-objective approach." *Int. J.ProductionEconomics.*

Xiaodan Yu, Zhiquan Qi, Yuanmeng Zhao. 2013. "Support Vector Regression for Newspaper/Magazine Sales Forecasting." *Information Technology and Quantitative Management* 1055 – 1062.

Yang, Tzyy-Ching, and Hsiangchu Lai. 2006. "Comparison of product bundling strategies on different online shopping behaviors." *Electronic Commerce Research and Applications* 295–304.

Analytical Assessment of Retail Sales of Dairy Products Using Data Mining

Abstract

Accurate forecasting of sales amount plays an important role in effective production and sourcing management of manufacturing companies which ultimately leads to high profitability. The paper proposes an effective use of data mining approaches to analyze and extract clusters of purchases that best characterizing the purchasing behavior of the dairy customers in different regions of Iran in a dominant dairy production company. The paper applies clustering and classification algorithms for differentiating the purchasing behavior across Iran. It employs six phases of business understanding, data understanding, data preparation and pre-processing, modeling, evaluation, and explanation of predictions for the managers of selected company. The data gathering and modeling phases are complemented by analytical predictions and conclusion. The research proves that for the dairy products that are highly perishable and should be produced on a daily basis, unexpected fluctuations in consumer demand may pose significant problems to the commercialization and sales of such products. It also categorizes the products sales behavior in different seasons and provides new directions for practitioners on the dairy products that should be selected for each season and city based on seasonal fluctuations. Due to the limitations in accessing the central databases of dairy companies and products, the researchers were required to gather data from daily databases for a limited period of time and for specific products. It is recommended to gather a full dairy data set for a broader range of products for the purpose of a more comprehensive validation and examination of results throughout the industry. The paper includes a practical analysis of appropriate sales policies for each category of sales behavior for the purpose of high profitability. The final forecasting solution demonstrates that demand in dairy market is very sensitive to changes in weather condition while it does not show a quick reaction to price changes. The paper concentrates on the need to maintain a balance between meeting customer demands and controlling inventory costs and also fulfills a major longtime need for enterprises seeking good sales performance and profitability through accurate analytical predictions.

Keywords: *Dairy Products, Retail Sales, Data Mining, Information Analysis, Prediction*