

## بخش بندی مشتریان صادراتی میوه های خوراکی

دکتر مهدی غضنفری\*      سمیرا ملک محمدی\*\*      دکتر سمیه علیزاده\*\*\*  
دکتر مهدی فتح الله\*\*\*\*

پذیرش: ۸۸/۱۰/۹

دریافت: ۸۶/۱/۱۷

بخش بندی مشتریان / مدیریت ارتباط با مشتری / RFM / داده کاوی

### چکیده

داده کاوی یکی از تکنیک های جدید برای کاویدن الگوها و روندها با توجه به داده های مشتریان است که سبب بهبود ارتباط با مشتری می شود و از ابزارهای مطرح در مدیریت ارتباط با مشتری است. بخش بندی شیوه ای برای شناخت مشتری و شکستن کل جمعیت مشتریان به گروه های کوچکتر است. هدف این تحقیق ارائه مدلی برای بخش بندی کشورهای براساس تازگی (Recency)، تکرار (Frequency) و ارزش پولی (Monetary) صادرات میوه های خوراکی از کشور جمهوری اسلامی ایران به سایر کشورها در طول یازده سال (۱۳۷۴-۱۳۸۴) است. پس از محاسبه متغیرها، با استفاده الگوریتم های خوشه بندی (Kmeans و Fuzzy Kmeans) کشورها خوشه بندی شده اند و نتایج این خوشه بندی ها از نظر کیفیت بر اساس سه معیار مختلف سنجیده شده است. با استفاده از هرم ارزش مشتری و درخت تصمیم گیری خوشه ها تحلیل و به چهار گروه اصلی تقسیم شده اند: مشتریان وفادار، مشتریان فعال، مشتریان جدید و مشتریان غیر فعال. در پایان سیاست های مرتبط با مدیریت ارتباط با مشتری برای هر بخش مطرح شده است.

### طبقه بندی JEL: J41.

- \* دانشیار دانشگاه علم و صنعت ایران
  - \*\* کارشناس ارشد مهندسی فناوری اطلاعات و تجارت الکترونیک دانشگاه علم و صنعت
  - \*\*\* استادیار دانشگاه خواجه نصرالدین طوسی
  - \*\*\*\* دکتری مهندسی صنایع دانشگاه علم و صنعت
- سمیرا ملک محمدی، مسئول مکاتبات.

## مقدمه

در طی چند سال گذشته تعامل شرکت‌ها با مشتریان بطور قابل توجهی تغییر کرده است به طوری که تضمین بلندمدتی برای تداوم کسب و کار با مشتریان وجود ندارد. به همین دلیل سازمان‌ها باید مشتریان‌شان را به درستی شناخته، نیازها و خواسته‌های آنها را پیش‌بینی کنند و با استفاده از این نوع اطلاعات بهره‌وری خود را افزایش دهند. سازمان‌ها برای حفظ رقابت، نیاز به تدوین استراتژی‌های تمرکز بر مشتری، مشتری محوری، مشتری مداری دارند و مدیریت ارتباط با مشتری (CRM) راه‌حلی است که این تلاش‌ها را برای سازمان‌ها و همچنین مشتریان محقق می‌سازد. مدیریت ارتباط با مشتری به سازمان‌ها این اجازه را می‌دهد که مشتریان خود را بهتر بشناسند و تفاوت بین آنها را بهتر درک کنند، در نتیجه در تخصیص منابع مورد نیاز به مشتریان نیازمند کارآمدی بیشتری داشته باشند<sup>۱</sup>. CRM سازمان را قادر می‌سازد که گفتگوی یگانه، یکتا و مستمری را با هر یک از مشتریان برقرار نماید؛ تمامی تعامل‌های پیشین را به خاطر داشته و به آنها رجوع نماید و رفتار مشتریان را تحت تأثیر قرار دهد<sup>۲</sup>. بسیاری از سازمان‌ها داده‌های بسیار زیادی را درباره مشتریان، تامین کنندگان و شرکای تجاری جمع‌آوری و ذخیره می‌کنند ولی ناتوانی این سازمان‌ها برای کشف دانش پنهان و با ارزش موجود در این داده‌ها سبب می‌شود که این داده‌ها مورد استفاده قرار نگرفته و جمع‌آوری داده‌ها در عمل بیهوده باشد. صاحبان کسب و کارها میل به استخراج دانش ناشناخته، معتبر و قابل درک از بانک‌های اطلاعاتی عظیم خود و استفاده از این اطلاعات برای کسب سود بیشتر دارند. یکی از روش‌های شناخت مشتری، رویکرد بخش‌بندی مشتریان است. بخش‌بندی، مشتریان را به خوشه‌های همگنی تقسیم می‌کند که نیازها و خصوصیات مشتریان درون هر خوشه با یکدیگر مشابه است. به این ترتیب می‌توان از مشتریان شناخت بهتری پیدا کرد و گوناگونی نیازها و خواسته‌هایشان را برآورده ساخت و به بهترین نحو ارتباط با آنها را مدیریت کرد. هدف از بخش‌بندی مشتریان، گروه‌بندی مشتریان با نیازها و رفتار خرید مشابه برای حداکثر کردن پاسخ به

---

1. Berson et al., (2001).

2. Brown, (2000).

برنامه‌های بازاریابی هدفدار، برای گروه‌های مختلف است. بخش‌بندی مشتریان امکان ارائه محصول یا خدمت مناسب را به مشتریان خوشه هدف فراهم نموده و امکان ارتباط نزدیک با آنها را ایجاد می‌کند. بخش‌بندی یکی از مهمترین مقوله‌ها در رسیدن به بازاریابی مدرن و مدیریت موفق ارتباط با مشتری است.<sup>۱</sup> بخش‌بندی می‌تواند با استفاده از تجزیه و تحلیل خوشه‌ای (جمعیت‌شناختی، روان‌شناختی، تراکنشی و ترویجی) انجام شود و جمعیت زیادی از مشتریان را به بخش‌های مختلف تقسیم کند به طوری که مشتریان هر بخش به یکدیگر شبیه، و مشتریان بخش‌های مختلف با یکدیگر متفاوت باشند. یکی از این روش‌های بخش‌بندی، بخش‌بندی داده محور است که با استفاده از تکنیک‌های مختلف آماری و داده‌کاوی صورت می‌گیرد. داده‌کاوی یکی از ابزارهای بنیادی در آشکارسازی جمعیت‌شناختی مشتریان است که از تکنیک‌های آن می‌توان برای کسب دامنه وسیعی از اهداف صنایع مختلف استفاده کرد.<sup>۲</sup> عبارت داده‌کاوی مترادف با یکی از عبارات‌های استخراج دانش، برداشت اطلاعات، واری‌داده‌ها و حتی لایروبی کردن داده‌ها می‌باشد که در حقیقت کشف دانش در پایگاه داده<sup>۳</sup> را توصیف می‌کند. بنابراین ایده‌ای که مبنای داده‌کاوی است یک فرآیند با اهمیت از شناخت الگوهای بالقوه مفید، تازه و درنهایت قابل درک در داده‌ها می‌باشد.<sup>۴</sup> همه روش‌های داده‌کاوی برای بخش‌بندی مناسب نیستند و مدل‌های صحیحی را ایجاد نمی‌کنند و معمولاً روش‌های درخت تصمیم‌گیری و خوشه‌بندی از سایر روش‌ها برای بخش‌بندی مناسب‌تر است. اما این روش‌ها یک تفاوت اساسی از نظر یادگیری نظارتی یا غیرنظارتی دارند. یادگیری نظارتی، به تکنیک‌هایی اشاره دارد که قبل از بخش‌بندی، بخش‌ها با هدف خاصی که در ذهن است ایجاد شده است. به عنوان مثال بخش‌بندی مشتریان به مشتریان با ارزش کم و ارزش زیاد، اما در یادگیری غیرنظارتی هیچ بخش از پیش تعیین‌شده‌ای وجود ندارد و در واقع این تکنیک‌ها به ارائه یک تصویر کلی از همه مشتریان و سازمان‌دهی داده کمک

---

1. Berson et al., (2001).

2. Rygielski et al., (2002).

3. Knowledge Discovery of Database.

4. Ye, 2003.

می‌کنند.<sup>۱</sup> از دیگر مسائل پراهمیت در بخش‌بندی می‌توان به انتخاب نوع داده و متغیرها اشاره کرد که در موفقیت خوشه‌بندی بسیار تاثیرگذار است. در برخی مطالعات صورت گرفته در حوزه بخش‌بندی بازار خوشه‌بندی صرفاً بر روی داده‌های تراکنشی صورت گرفته است که از آن جمله می‌توان به بخش‌بندی مشتریان بازار بورس کشور کره و تبیین سیاست‌های بازاریابی برای حداکثر کردن سود بلندمدت اشاره کرد که با استفاده از روش‌های خوشه‌بندی مثل SOM، Kmeans و Fuzzy Kmeans بر روی داده‌های تراکنشی مشتریان صورت گرفته است.<sup>۲</sup> Chiu و Tsai در سال ۲۰۰۴ نیز به بسط متدولوژی جدید بخش‌بندی بازار پرداختند. در این روش بر مبنای متغیرهایی مانند اقلام کالاهای خریداری شده و میزان درآمد مربوط به آنها با توجه به تراکنش‌های قبلی مشتریان، بازار بخش‌بندی شده است. در برخی موارد داده‌های توصیفی<sup>۳</sup> و ترویجی<sup>۴</sup> برای بخش‌بندی مشتریان نیز مورد توجه قرار گرفته است. برای مثال Hwang و همکارانش سال ۲۰۰۴ با در نظر گرفتن سهم سود ایجاد شده، سود بالقوه و تعریف سودآوری مشتری، مدلی برای اندازه‌گیری طول عمر ارزش مشتری<sup>۵</sup> پیشنهاد کرده‌اند و بر اساس ارزش فعلی، ارزش بالقوه و وفاداری مشتریان را بخش‌بندی کرده‌اند. Kim و همکارانش در سال ۲۰۰۶ با در نظر گرفتن اهمیت شناخت مشتری در جهت ایجاد ارتباط بلندمدت، کسب وفاداری و سودآوری بیشتر مشتری، ارزش هر مشتری را تعیین نموده، سپس بر مبنای این ارزش مشتریان بخش‌بندی و استراتژی‌های مناسب برای هر بخش تبیین شده است.

یکی از مدل‌های مطرح در تحلیل ارزش مشتری، مدل RFM است که توسط Hughes در سال ۱۹۹۴ ارائه شده است و به بیان تفاوت مشتریان با استفاده از سه متغیر تازگی، تکرار و ارزش پولی می‌پردازد. پارامترهای مدل RFM به شرح زیر است:

تازگی آخرین خرید (R): R نشان دهنده تازگی است که مدت زمان بین آخرین تعامل تجاری با زمان حال را نشان می‌دهد، هرچه این مدت کمتر باشد، R بیشتر است.

- 
1. Hwang et al., (2004).
  2. Shina & Sohn, (2004).
  3. Descriptive data.
  4. Promotional data.
  5. Customer Life Time Value.

تکرار خرید (F): F نشان دهنده تکرار است و تعداد تراکنش‌ها را در یک بازه مشخص نشان می‌دهد، برای مثال دو بار در سال، دو بار در یک فصل یا دو بار در یک ماه. هر چه تکرار بیشتر باشد، F بزرگ‌تر است.

ارزش پولی خرید (M): M نشان دهنده ارزش پولی است و ارزش پولی تعاملات را در بازه‌ای خاص نشان می‌دهد. هر چه ارزش پولی بیشتر باشد، M بزرگ‌تر است.

تحقیقات Wu و Lin نشان می‌دهد که هر چه R و F بیشتر باشد احتمال آنکه تراکنش جدیدی با مشتری صورت بگیرد بیشتر است و همچنین اگر M نیز بزرگ‌تر باشد احتمال بازگشت مشتری برای خرید بیشتر است<sup>۱</sup>. متغیرهای مدل RFM برای بخش‌بندی مشتریان بسیار کارا هستند<sup>۲</sup>. از دیگر کاربردهای مدل RFM در بخش‌بندی مشتریان براساس متغیرهای تازگی، تکرار و ارزش پولی می‌توان به بخش‌بندی مشتریان به منظور تعیین سیاست‌های بهینه بازاریابی برای هر بخش اشاره کرد<sup>۳</sup>. در سال ۲۰۰۹ نیز Yeh و همکارانش برای انتخاب بازار هدف در بازاریابی مستقیم از متغیرهای مدل RFM به عنوان ورودی خوشه‌بندی محاسبه متغیرهای ورودی خوشه‌بندی استفاده کرد و با اضافه کردن دو پارامتر زمان اولین خرید و احتمال ریزش، مدل RFM را بسط داد<sup>۴</sup>.

همان‌طور که اشاره شد بخش‌بندی مشتریان به شیوه‌های گوناگونی صورت می‌گیرد. در این تحقیق از متغیرهای مدل RFM به عنوان متغیرهای ورودی به الگوریتم خوشه‌بندی استفاده شده است و با توجه به مورد مطالعاتی مورد بررسی، صادرات میوه‌های خوراکی از ایران به سایر کشورها، از داده‌های تراکنشی صادرات بانک اطلاعاتی سازمان توسعه تجارت ایران برای بخش‌بندی استفاده شده است که مناسب‌تر از داده‌های جمعیت‌شناختی و توصیفی است. در این مقاله مطالب به صورت زیر ساختاردهی شده است:

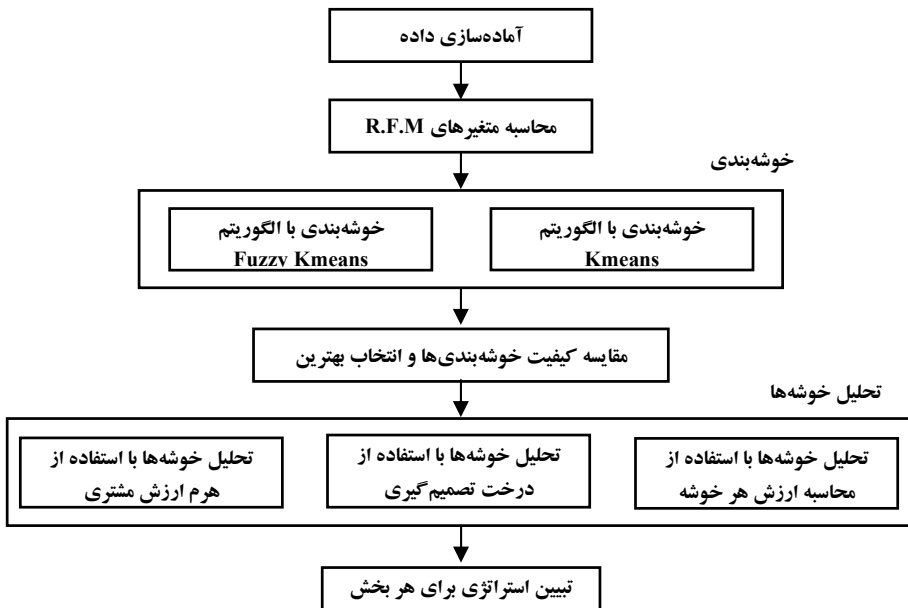
در بخش بعدی متدولوژی پیشنهاد شده به منظور بخش‌بندی مشتریان بازرگانی خارجی توضیح داده شده است. در بخش سوم مدل پیشنهادی بر روی داده‌های مشتریان بازرگانی خارجی ایران در بخش میوه و مرکبات پیاده‌سازی و نتایج حاصله مورد بررسی قرار گرفته

1. Wu and Lin, (2005).
2. Newell, (1997).
3. Jonkera et al., (2004).
4. Yeh et al., (2009).

است. بخش چهارم به تحلیل نتایج و خوشه تشکیل شده پرداخته است و راهکارهای اجرایی مناسبی در این خصوص پیشنهاد شده است. در بخش پایانی نیز پیشنهاداتی برای مطالعات آتی ارائه شده است.

## ۱. مدل پیشنهاد شده

مدل پیشنهاد شده شامل دو بخش اصلی است بخش بندی مشتریان و تحلیل بخش های تشکیل شده. پس از آماده سازی داده ها و استخراج متغیرهای تازگی (R) تکرار (F) ارزش پولی (M) از بانک اطلاعاتی، با استفاده از الگوریتم خوشه بندی K means و Fuzzy Kmeans کشورها خوشه بندی شده اند. عملکرد خوشه بندی به وسیله سه معیار سنجش کیفیت خوشه ارائه شده، مقایسه و در نهایت تعداد خوشه بهینه و بهترین خوشه بندی مشخص شده است. در بخش دوم با در نظر گرفتن مفاهیم هرم ارزش مشتری و تحلیل خوشه ها در فضای RFM به کمک درخت تصمیم گیری کشورها دسته بندی و استراتژی مناسب برای هر بخش تبیین شده است. چارچوب کلی مدل پیشنهاد شده در شکل (۱) نشان داده شده است.



شکل ۱- چارچوب مدل پیشنهاد شده

## ۱-۱. محاسبه متغیرهای تازگی، تکرار و ارزش پولی صادرات

از سه متغیر تازگی (R) تکرار (F) ارزش پولی (M) برای بخش‌بندی کشورها استفاده شده است. متغیر تازگی فاصله بین تازه‌ترین زمان تراکنش و زمان تحلیل را اندازه می‌گیرد، در واقع آخرین باری را که در طی بازه زمانی مورد نظر کشور ایران به کشورهای دیگر صادرات داشته است، مشخص می‌کند. متغیر تکرار، تکرار صادرات در بازه زمانی مورد نظر را اندازه می‌گیرد. ارزش پولی، کل ارزش دلاری صادرات در طی بازه مفروض را مشخص می‌کند. سه متغیر ذکر شده بعد از استخراج از بانک اطلاعاتی بصورت زیر نرمال می‌گردند:

$$R(c_i) = \frac{Q^R - Q_{Min}^R}{Q_{Max}^R - Q_{Min}^R} \quad (1)$$

$$F(c_i) = \frac{Q^F - Q_{Min}^F}{Q_{Max}^F - Q_{Min}^F} \quad (2)$$

$$M(c_i) = \frac{Q^M - Q_{Min}^M}{Q_{Max}^M - Q_{Min}^M} \quad (3)$$

که  $Q^R$ ،  $Q^F$  و  $Q^M$  مقدار اصلی امتیازات کشور  $c_i$  مطابق با تعریف  $R$ ،  $F$  و  $M$  است.  $Q_{Min}^R$ ،  $Q_{Min}^F$  و  $Q_{Min}^M$  حداقل و  $Q_{Max}^R$ ،  $Q_{Max}^F$  و  $Q_{Max}^M$  حداکثر مقدار  $R$ ،  $F$  و  $M$  برای همه کشورها را نشان می‌دهند.

## ۲-۱. خوشه‌بندی

خوشه‌بندی تقسیم یک گروه ناهمگن به چندین زیر گروه است. خوشه‌بندی به طور خودکار ویژگی‌های متمایزکننده زیر گروه‌ها را تعریف می‌کند و زیر گروه‌ها را شکل می‌دهد به طوری که پایگاه داده را به چندین بخش تقسیم می‌کند و گروه‌هایی از رکوردها را به وجود می‌آورد که نمایانگر یا صاحب صفت خاصی هستند. الگوهای به

دست آمده در ذات پایگاه داده نهاده است و نشانگر بعضی اطلاعات غیر منتظره و در عین حال ارزشمند است<sup>۱</sup>. در این تحقیق براساس سه متغیر تازگی صادرات به هر کشور  $R(c_i)$ ، تکرار صادرات به هر کشور  $F(c_i)$  و ارزش پولی صادرات صورت گرفته به هر کشور  $M(c_i)$ ، کشورها خوشه‌بندی شده‌اند. برای خوشه‌بندی از دو الگوریتم خوشه‌بندی Kmeans و Fuzzy Kmeans استفاده شده است.

#### ۱-۲-۱. الگوریتم خوشه‌بندی Kmeans

الگوریتم Kmeans ابتدا K عضو (که K تعداد خوشه‌هاست) را به صورت تصادفی از میان N عضو انتخاب می‌نماید و آن‌ها را به عنوان مراکز خوشه‌ها در نظر می‌گیرد. سپس N-K عضو باقیمانده به نزدیک‌ترین خوشه تخصیص می‌یابند. بعد از تخصیص همه اعضا مجدداً مراکز خوشه‌ها محاسبه شده و اعضا با توجه به مراکز جدید به خوشه‌ها تخصیص می‌یابند و این کار تا زمانی که مراکز خوشه‌ها ثابت بمانند، ادامه می‌یابد<sup>۲</sup>.

#### ۱-۲-۲. الگوریتم خوشه‌بندی Fuzzy Kmeans

در الگوریتم خوشه‌بندی Fuzzy Kmeans درجه عضویت هر نمونه به هر خوشه به طور تصادفی تعیین می‌گردد. با استفاده از درجه عضویت و مختصات مرکز خوشه‌ها مختصات جدید مراکز خوشه‌ها از رابطه زیر محاسبه می‌شود:

$$C_{ij} = \frac{\sum (\mu_{ik})^q X_{ik}}{\sum_{k=1}^n (\mu_{ik})^q} \quad (۴)$$

که در آن  $C_{ij}$  مقدار متغیر زام از مرکز خوشه  $i$ ام،  $\mu_{ik}$  درجه عضویت نمونه  $k$ ام به خوشه  $i$ ام و  $X_{ik}$  مقدار متغیر زام در نمونه  $k$ ام است.  $q$  معرف فازی شدن متغیر زام در نمونه  $k$ ام است. اگر این مقدار برابر واحد باشد الگوریتم تحلیل خوشه‌ای به نوع خوشه‌ای غیرفازی تبدیل می‌شود که در آن مراکز خوشه‌ها همان مراکز ثقل داده‌ها و درجه عضویت هر نمونه به این مرکز صفر یا یک است. اگر مقدار  $q$  به سمت بی‌نهایت میل کند همه

1. Ye, (2003).

2. Soman et al., (2006).



مراکز خوشه‌ها به مرکز کل داده‌ها همگرا شده و درجه عضویت همه نمونه‌ها به مراکز خوشه‌ها برابر  $\frac{1}{C}$  می‌گردد. پس از محاسبه مراکز جدید خوشه‌ها لازم است درجه عضویت هر نمونه به همه مراکز جدید خوشه‌های جدید محاسبه شود این فاصله از طریق فرمول فاصله اقلیدسی صورت می‌گیرد.

$$\mu_{ik} = \frac{(d_{ik}^2)^{\frac{-1}{q-1}}}{\sum_{k=1}^c (d_{ik}^2)^{\frac{-1}{q-1}}} \quad (5)$$

که در آن  $\mu_{ik}$  درجه عضویت نمونه  $k$ ام به خوشه  $i$ ام و  $d$  فاصله نمونه  $i$ ام تا مرکز خوشه  $i$ ام است که به طریق زیر محاسبه می‌شود:

$$d(i, j) = \sqrt{(|x_{i1} - x_{j1}|^2 + |x_{i2} - x_{j2}|^2 + \dots + |x_{ip} - x_{jp}|^2)} \quad (6)$$

محاسبه تابع هدف متغیر  $i$ ام در محیطی که به درجه  $q$  فازی شدگی دارد به صورت زیر تعریف می‌شود:

(7)

$$J_q = \sum_{i=1}^c \sum_{k=1}^n (\mu_{ik})^q (d_{ik}^2)^q$$

که در آن مقادیر  $\mu_{ik}$  و  $d_{ik}^2$  از روابط قبلی محاسبه می‌شود. تکرار محاسبات تا زمانی ادامه می‌یابد که اختلاف بین دو مرحله متوالی کمتر از دقت مورد نظر گردد.<sup>۱</sup>

### ۳-۱. مقایسه عملکرد خوشه‌بندی‌ها

برای سنجش همگنی درون خوشه‌ای و ناهمگنی بین خوشه‌ای تعاریف متفاوتی براساس فاصله اعضا وجود دارد که در زیر به تفصیل بررسی گردیده است. سپس با در نظر گرفتن این تعاریف معیارهایی برای سنجش کیفیت خوشه‌بندی تعریف شده است.

## ۱-۳-۱. اندازه‌گیری فاصله بین خوشه‌ها

برای محاسبه ناهمگنی بین خوشه‌ها معیارهای فاصله متفاوتی تعریف شده است که برخی از متداول‌ترین معیارهای اندازه‌گیری فاصله بین دو خوشه عبارتند از:

- حداقل فاصله ممکنه بین عناصر خوشه‌ها: اگر  $O^n$  مجموعه اعضای خوشه  $n$  و  $c_i$  اعضای خوشه  $n$  باشد و  $O^m$  مجموعه اعضای خوشه  $m$ ،  $c_j$  اعضای خوشه  $m$  باشد، فاصله بین خوشه  $n$  و  $m$  عبارت است از:

$$d(m, n) = \text{Min}(\text{Dist}(c_i, c_j)), c_i \in O^n, c_j \in O^m \quad (10)$$

- حداکثر فاصله ممکنه بین عناصر دو خوشه: فاصله بین خوشه  $n$  و  $m$  عبارتست از:

$$d(m, n) = \text{Max}(\text{Dist}(c_i, c_j)), c_i \in O^n, c_j \in O^m \quad (11)$$

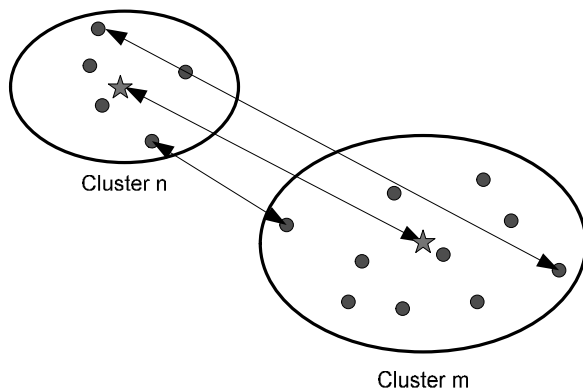
- مقادیر متوسط کلیه فاصله‌های ممکنه بین عناصر دو خوشه: فاصله بین خوشه  $n$  و  $m$  عبارت است از:

$$d(m, n) = \frac{1}{\|O^n\| \times \|O^m\|} \sum_{c_i \in O^n} \sum_{c_j \in O^m} \text{Dist}(c_i, c_j) \quad (12)$$

فاصله بین مراکز دو خوشه. اگر اعضای خوشه‌ها متغیرهای گسسته باشند از نزدیک‌ترین عضو به میانه خوشه برای سنجش فاصله استفاده شده و اگر این اعضا پیوسته باشند از میانه استفاده می‌شود. اگر  $c^n$  مرکز خوشه  $n$  و  $c^m$  مرکز خوشه  $m$  باشد، فاصله بین خوشه  $n$  و  $m$  عبارتست از:

$$d(m, n) = \text{Dist}(c^n, c^m) \quad (13)$$

همگنی درون خوشه‌ای را نیز می‌توان با توجه به بیشترین فاصله اعضای درون یک خوشه، کمترین فاصله اعضای درون یک خوشه، میانگین مجموع فواصل اعضای خوشه و یا فاصله اعضای خوشه با مرکز یا میانه خوشه سنجید<sup>۱</sup>.



شکل ۲- فواصل بین خوشه‌ای

## ۲-۳-۱. معیار سنجش کیفیت خوشه‌ها

همانطور که اشاره شد، خوشه‌بندی ای بهتر است که فاصله درون خوشه‌ها در آن کمترین و فاصله بین خوشه‌ها بیشترین باشد. برای مقایسه خوشه‌بندی‌های انجام شده به کمک الگوریتم‌های خوشه‌بندی، سه معیار سنجش کیفیت خوشه‌بندی مطرح شده‌است:

● معیار ۱: اگر  $O = \{c^n | n=1, \dots, K\}$  مجموعه مراکز خوشه‌ها و  $c^n$  مراکز خوشه‌ها باشد و  $O^n = \{c_i | i=1, \dots, \|T^c - O\|\}$  مجموعه باقیمانده کشورها که به عنوان مرکز خوشه انتخاب نشده‌اند و  $T^c$  مجموعه کلیه کشورهایی که خوشه‌بندی روی آنها صورت گرفته‌است، باشد، کیفیت نتایج خوشه‌بندی با  $K$  خوشه می‌تواند بصورت معادله ۱۴ تعریف شود:

$$\rho(K) = \frac{1}{K} \sum_{n=1}^K ( \text{Max}_{1 \leq m \leq K, m \neq n} \{ \frac{\eta_n + \eta_m}{\delta_{nm}} \} ) \quad (14)$$

$$\eta_n = \frac{1}{\|O^n\|} \sum_{c_i \in O^n} \text{Dist}(c_i, c^n) \quad (15)$$

$$\eta_m = \frac{1}{\|O^m\|} \sum_{c_j \in O^m} \text{Dist}(c_j, c^m) \quad (16)$$

$$\delta_{nm} = \text{Dist}(c^n, c^m) \quad (17)$$

معادله (۱۵)،  $\eta_n$  را به عنوان میانگین فاصله بین مرکز خوشه  $c^n$  و همه مشتریان خوشه  $O^n$  تعریف می کند. معادله (۱۶) توضیح می دهد که  $\eta_m$  میانگین فاصله بین مرکز خوشه  $c^m$  و همه مشتریان در خوشه  $O^m$  است. معادله (۱۷)  $\delta_{nm}$  را به عنوان فاصله  $c^m$  و  $c^n$  تعریف می کند. معیار فوق هرچه کمتر باشد تراکم درون خوشه‌ای بیشتر و فاصله بین خوشه‌ای بیشتر است، در نتیجه کیفیت خوشه‌بندی صورت گرفته بهتر است<sup>۱</sup>.

● **معیار ۲:** یکی از روش‌های ارزیابی عملکرد خوشه‌بندی، روش سنجش تراکم درون خوشه‌ای است. این معیار میزان تراکم خوشه را وقتی که تعداد خوشه‌ها ثابت است نشان می‌دهد. در این معیار باید متغیرها در یک محدوده باشند. با توجه به این که در این تحقیق داده‌ها گسسته هستند برای محاسبه تراکم یک خوشه، فاصله داده‌های درون یک خوشه باید با مرکز خوشه سنجیده شود. تراکم خوشه  $I_n$ ، بصورت زیر تعریف شده است:

$$I_n = \sum_{c_i \in O^n} \text{Dist}(c_i, c^n) \quad (17)$$

سپس  $F(K)$  برای  $K$  خوشه برابر است با:

$$F(K) = \frac{1}{K} \sum_{n=1}^K \sum_{c_i \in O^n} \text{Dist}(c_i, c^n) \quad (18)$$

در واقع  $F(K)$  میانگین مربعات فاصله اقلیدسی بین هر مشاهده و میانه خوشه مربوطه است. هر چه  $F(K)$  کمتر باشد نشان می‌دهد که خوشه‌ها متراکم‌تر بوده و خوشه‌بندی بهتر صورت گرفته است<sup>۲</sup>. در این تحقیق به جای مجموع مربعات فاصله، میانگین مربعات فاصله اعضای خوشه با مراکز خوشه در معادله (۱۳) در نظر گرفته شده است و معادله بصورت زیر درآمده است.

$$F(K) = \frac{1}{K} \sum_{n=1}^K \left( \frac{1}{\|O^n\|} \sum_{c_i \in O^n} \text{Dist}(c_i, c^n) \right) \quad (19)$$

1. Tsai and Chiu, (2004).  
2. Shina and Sohn (2004).

همچنین برای در نظر گرفتن ناهمگنی برون خوشه‌ای فاصله بین مراکز خوشه‌ها در نظر گرفته شده است.

$$F(K) = \frac{1}{K} \left( \frac{\sum_{n=1}^K \left( \frac{1}{\|O^n\|} \sum_{c_i \in O^n} \text{Dist}(c_i, c^n) \right)}{\sum_{n=1}^K \sum_{m < n}^K \text{Dist}(c^n, c^m)} \right) \quad (20)$$

هرچه معیار فوق کمتر باشد فاصله اعضای درون هر خوشه با مرکز آن خوشه کمتر بوده و فاصله بین مراکز خوشه‌ها نیز بیشتر است.

معیار ۳: این معیار بر اساس مفاهیم ارائه شده در بخش ۳-۲-۱ برای سنجش کیفیت خوشه‌بندی که همگنی درون خوشه‌ای و ناهمگنی برون خوشه‌ای بصورت همزمان تعریف شده است که به صورت زیر محاسبه می‌شود:

$$Q(K) = \frac{1}{K} \left( \frac{\text{IntraDist}}{\text{ExtraDist}} \right) \quad (21)$$

که IntraDist مجموع بیشترین فاصله اعضای درون هر خوشه است:

$$\text{IntraDist} = \sum_{n=1}^K \text{Max}(\text{Dist}(c_i, c_j)), c_i, c_j \in O^n \quad (22)$$

و ExtraDist مجموع کمترین فاصله بین خوشه‌هاست:

$$\text{ExtraDist} = \sum_{n=1}^K (\text{Min}(\text{Dist}(c_i, c_j)), c_i \in O^n, c_j \in O^m) \quad (23)$$

● معیار ۳: نسبت بیشترین فاصله اعضای درون هر خوشه را به کمترین فاصله مراکز خوشه‌ها با یکدیگر نشان می‌دهد و هرچه کمتر باشد خوشه‌بندی بهتری صورت گرفته است.

#### ۴-۱. تحلیل خوشه‌ها و تبیین استراتژی

در این مرحله براساس ارزش محاسبه شده برای هر خوشه و مفاهیم هرم ارزش مشتری، هر خوشه به کمک "درخت تصمیم" تحلیل شده است و استراتژی‌های مناسب جهت

ماندگاری و ارزش افزایی در بازار جهانی تبیین شده است.

#### ۱-۴-۱. محاسبه ارزش هر خوشه

ارزش هر مشتری را می‌توان براساس تازگی، تکرار و ارزش دلاری صادرات به هر کشور به صورت زیر مشخص کرد:

$$V(c_i) = W^R \times R(c_i) + W^F \times F(c_i) + W^M \times M(c_i) \quad (۲۴)$$

که  $R(c_i)$ ،  $F(c_i)$  و  $M(c_i)$  به ترتیب امتیازات مشتری  $c_i$  با توجه به معیارهای  $R$ ،  $F$  و  $M$  است.  $W^R$ ،  $W^F$  و  $W^M$  اهمیت وزنی برای معیارهای  $R$ ،  $F$  و  $M$  را به ترتیب نشان می‌دهد.

$$W^R + W^F + W^M = 1 \text{ بعلاوه داریم:}$$

سودآوری خوشه  $O^n$  با محاسبه میانگین ارزش همه کشورهای خوشه  $n$  حاصل می‌شود. می‌توان آن را از طریق معادله زیر تعریف کرد:

$$V(O^n) = W^R \times R(O^n) + W^F \times F(O^n) + W^M \times M(O^n) \quad (۲۵)$$

$$R(O^n) = \frac{\sum_{c_i \in O^n} R(c_i)}{\|O^n\|} \quad (۲۶)$$

$$F(O^n) = \frac{\sum_{c_i \in O^n} F(c_i)}{\|O^n\|} \quad (۲۷)$$

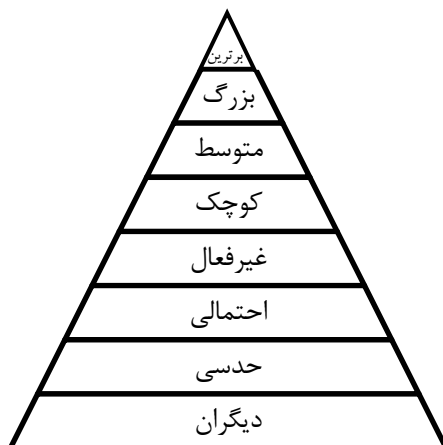
$$M(O^n) = \frac{\sum_{c_i \in O^n} M(c_i)}{\|O^n\|} \quad (۲۸)$$

که  $R(O^n)$ ،  $F(O^n)$  و  $M(O^n)$  امتیازات  $n$  امین خوشه با توجه به معیارهای  $R$ ،  $F$  و  $M$  است.

#### ۱-۴-۲. هرم ارزش مشتری

مفهوم بازاریابی مشتری و هرم مشتری به عنوان ابزاری برای شناخت و تجزیه و تحلیل

رفتار مشتری در سال ۱۹۸۹ معرفی شد. (شکل ۳)



شکل ۳ - هرم ارزش مشتری (بر اساس میزان درآمد و سودآوری)

با توجه به شکل (۳)، مشتریان در لایه‌های مختلف هرم را می‌توان به صورت زیر تعریف کرد:

**مشتریان فعال:** سازمان یا شرکت‌هایی که در یک محدوده زمانی مشخص، کالا یا خدمات خریداری کرده‌اند. مشتریان فعال به چهار بخش برترین، بزرگ، متوسط و کوچک تقسیم‌بندی می‌شوند.

**مشتریان غیرفعال:** شرکت‌هایی که قبلاً کالا یا خدمات خریداری می‌کردند ولی این کار را در محدوده زمانی مشخصی انجام نداده‌اند. مشتریان غیر فعال منبع مهمی برای درآمد بالقوه هستند. آنها همچنین منبع اطلاعات ارزشمندی هستند و از طریق این اطلاعات می‌توان از تبدیل مشتریان فعال به غیرفعال جلوگیری کرد.

**مشتریان احتمالی یا بالقوه:** شرکت‌هایی که به نوعی با سازمان ارتباط دارند، ولی هنوز کالا یا خدماتی خریداری نکرده‌اند. مشتریان احتمالی شرکت‌هایی هستند که انتظار داریم در آینده نزدیک به مشتریان فعال ارتقا یابند.

**مشتریان حدسی:** شرکت‌هایی هستند که می‌توان محصولات و یا خدماتی را به آنها سرویس داد ولی هنوز با آنها ارتباط برقرار نشده است. معمولاً تلاش می‌شود تا با مشتری

حدسی ارتباط برقرار شده و با این کار به هدف تبدیل آنان به مشتریان فعال در درازمدت نزدیک شد.

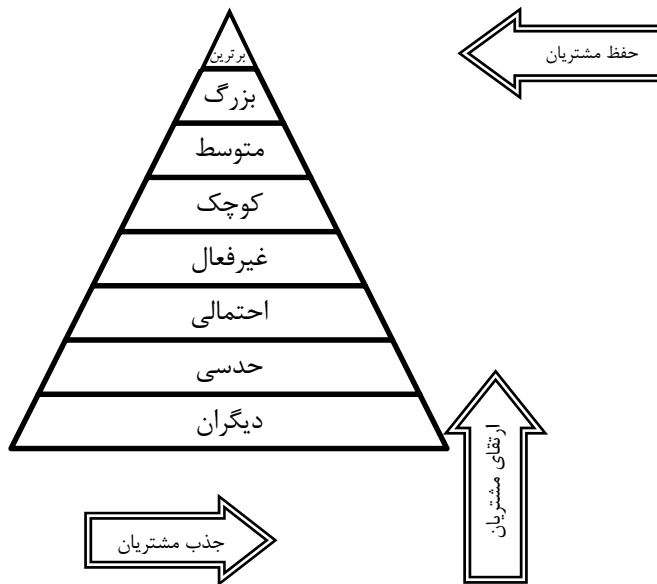
**سایرین:** شرکت‌هایی که خیلی ساده، نیاز یا علاقه‌ای به خرید یا استفاده از کالاها و خدمات ندارند. در حالی که از این گروه پولی بدست نخواهد آمد، ولی مهم است که شناخته شوند و میزان هزینه و زمان بازاریابی که برای برقراری ارتباط با این گروه افراد و شرکت‌ها که هرگز با سازمان معامله نخواهند کرد، مشخص شود.

مفهوم CRM را می‌توان از طریق هرم مشتری به سه بخش زیر مطرح نمود:

الف) جذب مشتریان جدید به هرم مشتری

ب) ارتقای مشتریان به سمت بالای هرم

ج) حفظ مشتریان در برابر ترک<sup>۱</sup> هرم



شکل ۴- مفهوم CRM از دیدگاه بخش بندی مشتری براساس ارزش



## ۳-۴-۱. دسته‌بندی

دسته‌بندی یکی دیگر از روش‌های داده کاوی است و به ارزشیابی ویژگی‌های مجموعه‌ای از داده‌ها می‌پردازد و سپس آنها را به مجموعه‌ای از گروه‌های از پیش تعریف شده اختصاص می‌دهد. این روش یکی از متداول‌ترین روش‌های داده کاوی است<sup>۱</sup>. برای تحلیل خوشه‌های تشکیل شده در این تحقیق از درخت تصمیم‌گیری C&R استفاده شده است<sup>۲</sup>. متغیرهای ورودی را برای یافتن بهترین تجزیه می‌آزمایید تا شاخص ناخالصی حاصل از تجزیه کمترین مقدار باشد. در تجزیه دوزیر گروه تعیین می‌شود و هر کدام در مرحله بعد نیز به دوزیر گروه دیگر تقسیم خواهند شد و این روند ادامه می‌یابد تا زمانی که یکی از معیارهای توقف برآورده شود<sup>۳</sup>. در این تحقیق از درخت تصمیم‌گیری برای دسته‌بندی و نمایش خوشه‌ها در فضای R، F و M استفاده شده است.

## ۲. مورد مطالعاتی

مورد مطالعاتی مورد بررسی در این تحقیق صادرات میوه‌های خوراکی از ایران به سایر کشورها است. داده‌های تراکنشی ذخیره شده در بانک اطلاعاتی سازمان توسعه تجارت ایران مبتنی بر سیستم کدینگ بین المللی HS است. بر همین اساس تعرفه‌های فصل هشت این تقسیم‌بندی مرتبط با میوه‌های خوراکی به عنوان مبنای تحقیق در نظر گرفته شده است. در گام نخست تحقیق، داده‌های صادرات میوه و مرکبات کشور جمهوری اسلامی ایران به ۱۴۸ کشور در طول ۱۱ سال یعنی از سال ۱۳۷۴ تا ۱۳۸۴ از بانک اطلاعاتی سازمان توسعه تجارت ایران استخراج شده است. این داده‌ها شامل ۱۲۱۵۸ رکورد و هر رکورد شامل اطلاعات نظیر: نام کشوری که صادرات به آن صورت گرفته، گروه تعرفه صادراتی، سال صادرات این تعرفه به کشور مربوطه و ارزش دلاری گروه تعرفه صادراتی است.

همانطور که پیشتر گفته شد، متغیر تازگی نشان دهنده آخرین باری است که به کشور مربوطه صادرات صورت گرفته است. با توجه به اینکه تحقیق بر روی صادرات بین سال‌های

1. Soman et al., (2006).

2. Classification and Regression Tree .

3. Ye, (2003).

۱۳۷۴ و ۱۳۸۴ صورت گرفته و بازه زمانی مورد بررسی ۱۱ سال است، مقدار این متغیر بین ۱۱ و ۱ است که ارزش امتیاز ۱۱ از همه بیشتر است و به این معنی می‌باشد که به تازگی (سال ۸۴) به کشور مربوطه صادرات صورت گرفته است. با توجه به فرمول (۱) مقدار  $R(C_i)$  برای هر کشور محاسبه شده است. تکرار صادرات، تعداد تکرارهای صادرات به هر کشور را در طی بازه مورد بررسی (۱۱ سال) نشان می‌دهد که مقدار این متغیر نیز بین ۱۱ و ۱ است و با استفاده از معادله (۲) مقدار  $F(C_i)$  برای هر کشور محاسبه شده است. برای محاسبه ارزش پولی هر کشور نیز کل حجم صادرات در طی بازه زمانی ۱۱ ساله گذشته به کشور مربوطه محاسبه شده است. با توجه به فرمول (۳) مقدار  $M(C_i)$  برای هر کشور محاسبه شده است.

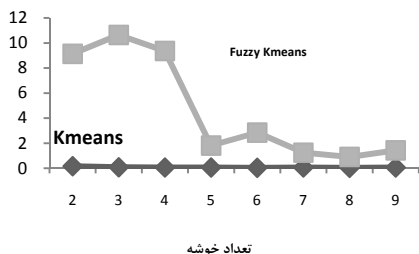
#### ۲-۱. بخش‌بندی کشورها

پس از محاسبه متغیرها، با استفاده از دو الگوریتم Kmeans و FuzzyKmeans کشورها خوشه‌بندی شده‌اند. برای مقایسه کیفیت خوشه‌بندی دو الگوریتم با یکدیگر و همچنین تعیین تعداد بهینه خوشه، سه معیار ذکر شده در بخش ۲-۴-۱ برای خوشه‌بندی بین ۲ تا ۹ خوشه محاسبه شده است. حال به بررسی نتایج حاصله در حالت‌های مختلف می‌پردازیم:

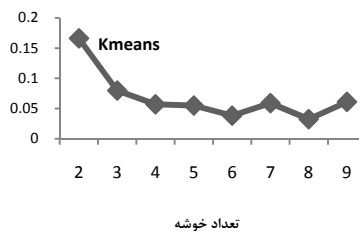
● **معیار ۱:** با توجه به معادله (۱۴)، کیفیت نتایج خوشه‌بندی با استفاده از الگوریتم‌های Kmeans و Fuzzy Kmeans در جدول (۱) نشان داده شده است. همانطور که پیشتر اشاره شده کمتر بودن هر یک از سه معیار ذکر شده نشان می‌دهد که خوشه‌های تشکیل شده کیفیت بهتری دارند. همانطور که در شکل ۵ (الف) مشاهده می‌شود، کیفیت خوشه‌های تشکیل شده با استفاده از الگوریتم Kmeans از الگوریتم Fuzzy Kmeans بهتر است. با توجه به تفاوت مقیاس معیار ۱ برای دو الگوریتم، شکل ۵ (ب) کیفیت خوشه‌های تشکیل شده با الگوریتم Kmeans را دقیق‌تر نشان می‌دهد. با توجه به نمودار شکل ۵ (ب)، خوشه‌بندی با هشت خوشه بهترین کیفیت را دارد.

جدول ۱- مقایسه الگوریتم‌های خوشه‌بندی بر اساس معیار ۱

معیار ۱		
تعداد خوشه	Fuzzy Kmeans	Kmeans
۲	۹/۱۳۳۷۷۵۲۹	۰/۱۶۶۳۱۲۰۳۹
۳	۱۰/۶۴۵۵۷۵۲۷	۰/۰۷۹۷۶۷۳۸۹
۴	۹/۳۶۳۰۵۷۹۱۸	۰/۰۵۶۹۸۷۸۴۵
۵	۱/۸۱۸۶۴۰۴۳۴	۰/۰۵۵۰۵۱۱۳۶
۶	۲/۸۴۵۹۳۱۳۹۶	۰/۰۳۸۲۳۹۰۳۷
۷	۱/۲۴۵۱۳۲۶۸۶	۰/۰۵۸۸۹۶۷۹۹
۸	۰/۹۰۵۴۹۴۱۲۵	۰/۰۳۲۳۹۴۲۲۵
۹	۱/۴۳۴۷۸۴۶۶۱	۰/۰۶۱۰۸۵۷۵۴



(الف)



(ب)

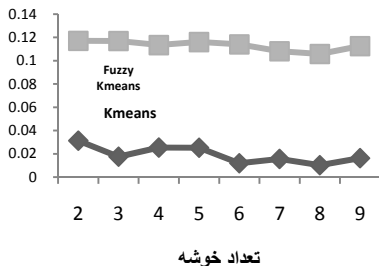
شکل ۵- کیفیت خوشه‌بندی الگوریتم‌ها با استفاده از معیار ۱ (الف) مقایسه الگوریتم

Kmeans و Fuzzy Kmeans، (ب) الگوریتم Kmeans

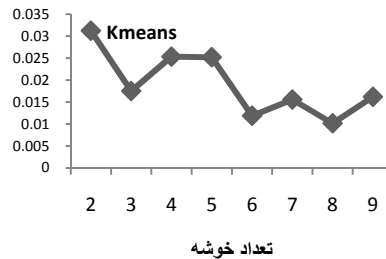
● معیار ۲: با توجه به معادله ۲۰، تراکم درون خوشه‌های تشکیل شده با استفاده از الگوریتم‌های Kmeans و Fuzzy Kmeans در جدول (۲) نشان داده شده است. همانطور که در شکل (۶) (الف) مشاهده می‌شود تراکم خوشه‌های تشکیل شده با استفاده از الگوریتم Kmeans از الگوریتم Fuzzy Kmeans بیشتر است. با توجه به نمودار شکل (۶) (ب)، خوشه‌بندی با هشت خوشه، خوشه‌های متراکم‌تری دارد.

## جدول ۲- مقایسه الگوریتم‌های خوشه‌بندی بر اساس معیار ۲

معیار ۲		
تعداد خوشه	Fuzzy Kmeans	Kmeans
۲	۰/۰۳۱۲۴۰۸۵۲	۰/۱۱۷۰۷۳۷۹۶
۳	۰/۰۱۷۵۰۲۷۴۲	۰/۱۱۶۹۹۴۶۲۹
۴	۰/۰۲۵۳۲۴۶۰۵	۰/۱۱۳۴۳۷۲۸۸
۵	۰/۰۲۵۲۰۷۰۷۴	۰/۱۱۶۰۶۳۱۵۸
۶	۰/۰۱۱۸۷۴۱۲۷	۰/۱۱۴۰۸۹۶۳۲
۷	۰/۰۱۵۵۶۹۶۴۷	۰/۱۰۸۱۲۵۵۵۵
۸	۰/۰۱۰۱۵۶۱۵	۰/۱۰۵۸۱۹۲۸۱
۹	۰/۰۱۶۱۹۰۷۱۵	۰/۱۱۲۵۲۸۵۳۱



(الف)



(ب)

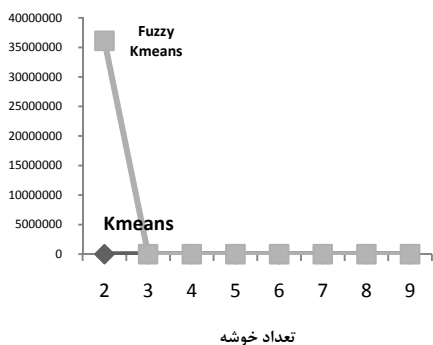
شکل ۶- تراکم خوشه‌بندی الگوریتم‌ها با استفاده از معیار ۲ (الف) مقایسه الگوریتم

Kmeans و Fuzzy Kmeans، (ب) الگوریتم Kmeans

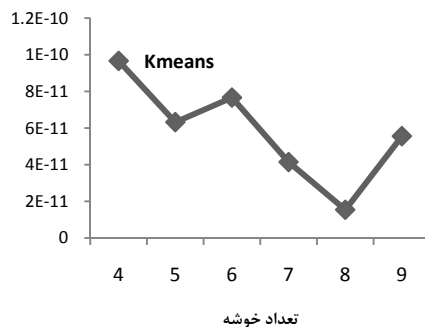
● معیار ۳: با توجه به معادله ۲۳، کیفیت خوشه‌های تشکیل شده با استفاده از الگوریتم‌های Kmeans و Fuzzy Kmeans در جدول (۳) نشان داده شده است. همانطور که در شکل (۷) (الف) مشاهده می‌شود همانند دو معیار قبل، کیفیت خوشه‌های تشکیل شده با استفاده از الگوریتم Kmeans از الگوریتم Fuzzy Kmeans بهتر است. با توجه به نمودار شکل (۷) (ب)، در خوشه‌بندی با ۸ خوشه، کیفیت خوشه‌ها بهتر است.

## جدول ۳- مقایسه الگوریتم‌های خوشه‌بندی بر اساس معیار ۳

معیار ۱		
تعداد خوشه	FuzzyKmeans	Kmeans
۲	۱۰/۳۱۳۲۲۵۹۸	۳۶۱۱۵۹۳۹/۸۲
۳	۰/۸۶۴۰۸۸۹۸	۸۱۷۴/۶۱۵۵۵۲
۴	۹/۶۶۵۷۹	۱۵/۱۱۷۹۸۱۰۶
۵	۶/۳۲۴۹۹	۲۹/۲۰۳۴۶۱۳۶
۶	۷/۶۶۶۲۴	۳/۳۱۸۶۹
۷	۴/۱۴۹۷۸	۹/۵۴۵۹۳
۸	۱/۵۴۴۳۴	۳/۵۶۷۲۶
۹	۵/۵۶۱۷۴	۴/۵۰۵۵



(الف)

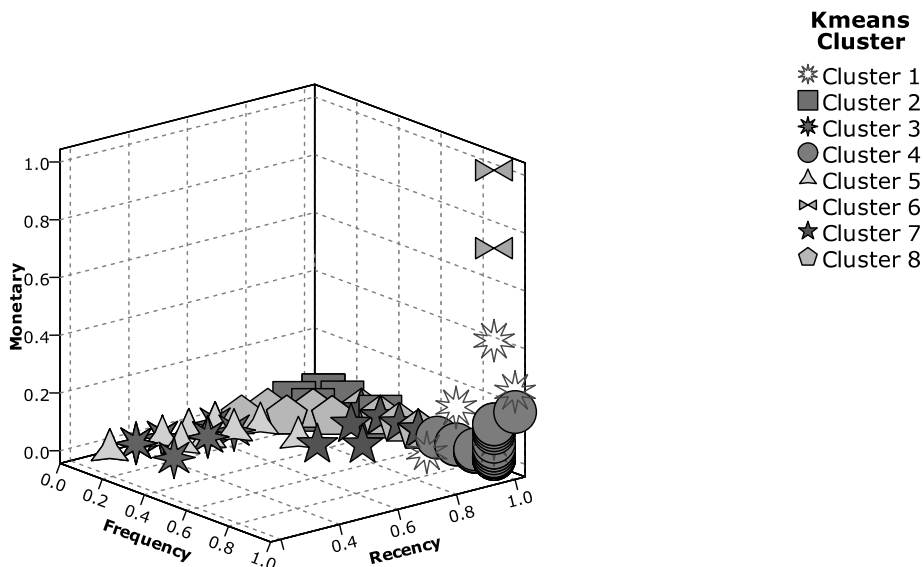


(ب)

شکل ۷- کیفیت خوشه‌بندی الگوریتم‌ها با استفاده از معیار ۳ (الف) مقایسه

## الگوریتم Kmeans و Fuzzy Kmeans، (ب) الگوریتم Kmeans

همانطور که نتایج حاصله از سه معیار ارزیابی کیفی نشان می‌دهد الگوریتم خوشه‌بندی Kmeans در این مورد مطالعاتی از الگوریتم Fuzzy Kmeans خوشه‌های با کیفیت‌تری را تشکیل داده است. همچنین بنابر محاسبات صورت گرفته تعداد بهینه خوشه نیز هشت خوشه است. در شکل (۸) کشورهای مورد مطالعه در فضای سه بعدی R,F,M تصویر شده است که اعضای هر خوشه به شکل واحدی نشان داده شده است.



شکل ۸- تصویر خوشه‌های تشکیل شده در فضای سه بعدی RFM

### ۳. تحلیل خوشه‌ها

در این بخش با استفاده از ارزش محاسبه شده برای خوشه‌ها بر اساس متغیرهای تازگی، تکرار و ارزش دلاری صادرات، هرم ارزش مشتری و درخت تصمیم به تحلیل بخش‌های تشکیل شده توسط الگوریتم خوشه‌بندی Kmeans پرداخته شده است.

ارزش خوشه‌ها با توجه به فرمول (۲۵) محاسبه و وزن متغیرها نیز براساس نظر خبرگان تعیین شده است. با توجه به اهمیت تکرار و پایداری صادرات وزن این متغیر از دو متغیر دیگر بیشتر است و برای آن وزن  $0/5$  در نظر گرفته شده است. برای متغیر M وزن  $0/3$  و متغیر R وزن  $0/2$  در نظر گرفته شده است. با توجه به معادلات (۲۶)، (۲۷) و (۲۸) تازگی، تکرار و ارزش پولی صادرات هر خوشه و با استفاده از معادله (۲۵) ارزش هر خوشه تعیین می‌گردد. نتایج به صورت جدول زیر است:

## جدول ۴- ارزش خوشه‌های تشکیل شده

شماره خوشه	میانگین ارزش پولی صادرات	میانگین تکرار صادرات	میانگین تازگی صادرات	ارزش هر خوشه	تعداد اعضای خوشه
۶	۰/۸۶۸۹۶	۰/۹۰۹	۱/۰۰۰	۰/۹۱۵	۲
۴	۰/۰۲۸۰۵	۰/۸۸۶	۱/۰۰۰	۰/۶۵۱	۴۳
۱	۰/۰۳۲۷۷	۰/۸۸۱	۰/۹۹۸	۰/۶۵۰	۴۲
۷	۰/۰۰۰۵۸	۰/۳۵۷	۰/۹۴۴	۰/۳۶۷	۱۳
۸	۰/۰۰۰۰۷	۰/۲۴۶	۰/۹۳۰	۰/۳۰۹	۱۷
۲	۰/۰۰۰۰۶	۰/۱۵۶	۰/۹۷۴	۰/۲۷۳	۷
۵	۰/۰۰۰۲۵	۰/۱۴۰	۰/۵۴۵	۰/۱۷۹	۱۳
۳	۰/۰۰۰۰۳	۰/۱۳۲	۰/۴۷۱	۰/۱۶۰	۱۱

در شکل (۹) مراکز خوشه‌های تشکیل شده نمایش داده شده است. همانطور که مشاهده می‌شود محور عمودی ارزش دلاری صادرات را نشان می‌دهد. بدیهی است از پایین به بالا ارزش دلاری خوشه‌ها افزایش می‌یابد برای مثال خوشه شش نسبت به سایرین از ارزش دلاری بیشتری برخوردار است. محور افقی میزان تازگی صادرات به کشورها را نشان می‌دهد. نزدیکی خوشه‌ها به ابتدای محور نشانگر غیر فعال شدن آنها است و نزدیکی به عدد یک نشانگر وجود رابطه تجاری اخیر با این کشورها است. تکرار صادرات نیز با رنگ نشان داده شده است. با توجه به نمایش خوشه‌ها بر اساس سه متغیر تازگی، تکرار و ارزش دلاری صادرات می‌توان خوشه‌ها را با یک خوشه‌بندی سلسه مراتبی متا خوشه‌بندی کرد و به چهار گروه اصلی تقسیم کرد:

گروه ۱: مشتریان وفادار (خوشه ۶)

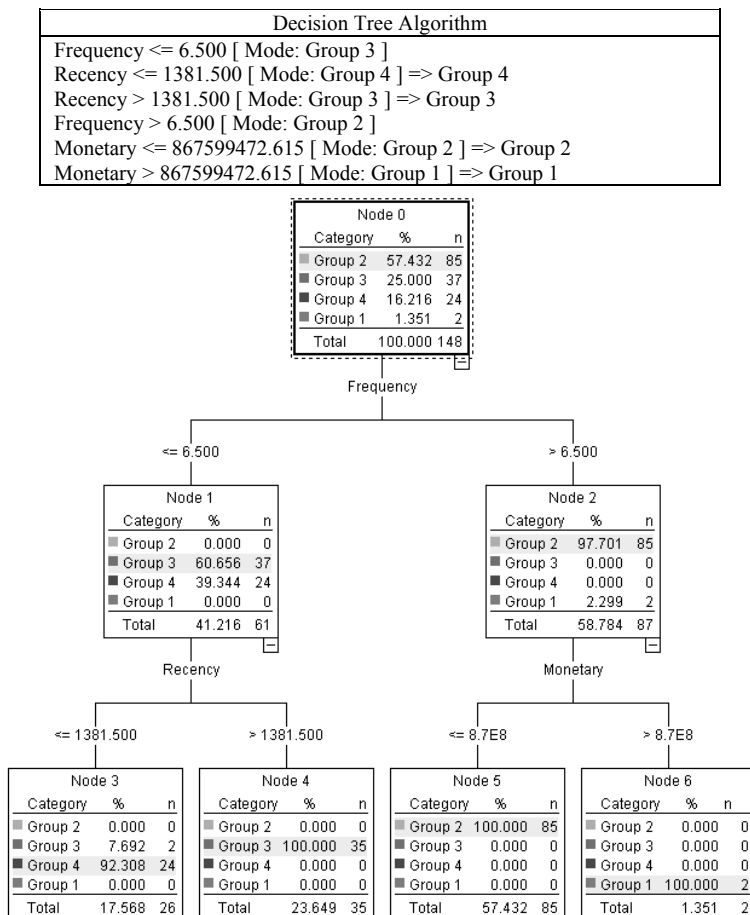
گروه ۲: مشتریان فعال (خوشه ۱ و ۴)

گروه ۳: مشتریان جدید (خوشه ۲ و ۷ و ۸)

گروه ۴: مشتریان رویگردان (خوشه ۳ و ۵)

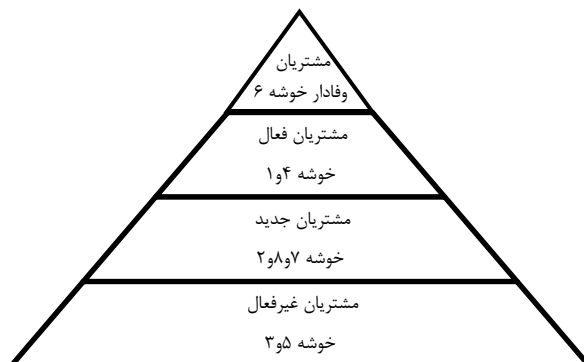






شکل ۱۰- دسته‌بندی مشتریان به کمک درخت تصمیم‌گیری

با توجه به دانش استخراج شده می‌توان هرم ارزش مشتری را براساس خوشه‌ها و گروه‌های تشکیل شده برای مشتریان صادراتی میوه‌های خوراکی ایران را ترسیم کرد (شکل ۱۱). همانطور که پیشتر اشاره شد برای داشتن ارتباط موفق و سودآور با کشورها باید کشورهای راس هرم که با ارزش‌ترین مشتریان هستند حفظ شوند. همچنین باید مشتریان فعال حفظ و ارتقا ارزش یابند، مشتریان جدید برای صادرات بیشتر مورد توجه قرار گیرند و علت ریزش مشتریان غیرفعال برای جلوگیری از رویگردانی‌های بعدی نیز مشخص شود.



شکل ۱۱- هرم ارزش مشتریان بازرگانی میوه‌های خوراکی  
سازمان توسعه تجارت ایران

نتیجه خروجی مدل پیشنهاد شده هرم ارزش مشتریان بازرگانی خارجی میوه‌های خوراکی است که می‌توان هر لایه هرم را به صورت زیر تحلیل نمود:

● **کشورهای گروه ۱:** مشتریان وفادار و با ارزش بالا در بازرگانی خارجی میوه‌های خوراکی هستند. صادرات به این کشورها مکرراً و به تازگی صورت گرفته است، ارزش دلاری صادرات صورت گرفته به این کشورها نیز قابل توجه است. این گروه از سایر کشورها در صادرات میوه با ارزش تر هستند، بنابراین باید حفظ شوند و خدمات ویژه به آنان ارائه شود. اعضای این گروه دو کشور آلمان و امارات متحده عربی هستند. کشور امارات متحده عربی یکی از همسایگان ایران است و حجم زیادی از تبادلات تجاری ایران به سایر کشورها به صورت غیرمستقیم توسط این کشور صورت می‌گیرد. در واقع این کشور مصرف‌کننده واقعی محصولات ایران نیست بلکه صرفاً به عنوان یک واسطه عمل می‌کند و کالاهای صادر شده به آن مجدداً به سایر کشورها صادرات مجدد می‌شود. بنابراین بررسی بازار مصرف امارات نمی‌تواند شناخت عمیق تری از نیازها و خواسته‌های بازار حاصل کند، لذا باید ابتدا شبکه صادرات میوه‌های خوراکی کشور امارات را بررسی کرد و با خوشه‌بندی کشورهایی که از کالاهای صادرات ایران در تجارت با امارات سهمیم هستند، مشتریان با ارزش را یافت و مستقیماً به بازارهای واقعی راه یافت. اما جایگاه آلمان در صادرات میوه‌های خوراکی با کشور امارات متفاوت است و امکان مشارکت

استراتژیک در صادرات میوه‌های خوراکی با این کشور وجود دارد. صادرات به کشور آلمان صادرات مستقیم است و حفظ بازار و ارتباط مستمر با این کشور بسیار حائز اهمیت است. این کشور دارای بازارهای آزاد بین‌المللی است لذا شناسایی رقبا، افزایش توان رقابتی، برندسازی ملی و شرکت در نمایشگاه‌های بین‌المللی در کنار رقبا می‌تواند در حفظ این بازار بسیار موثر باشد. از دیگر اقدامات می‌توان به دعوت برای بازدید کارخانه‌ها و مزارع با تکیه بر تولید ارگانیک محصولات، حضور در شبکه تامین فروشگاه زنجیره‌ای و امکان مشارکت در سرمایه‌گذاری برای گسترش بازار اشاره کرد.

● **کشورهای گروه ۲:** مشتریان وفادار با ارزش متوسط و کمتر از گروه ۱ است. در دوره مذکور صادرات پایدار و مداوم به این کشورها صورت گرفته است اما ارزش دلاری آن پایین است باید سعی شود علاوه بر حفظ و نگهداری این مشتریان فعال، ارزش آنها را افزایش داد تا در هرم ارزش به لایه بالایی ارتقا یابند. به این منظور می‌توان با این کشورها مشارکت جنبی در تولید و یا توزیع داشت. با توجه به اینکه تعداد اعضای این گروه (۸۵ عضو) بیشتر از سایر گروه‌ها است با تقسیم کار در سطح ملی بین بنگاه‌ها و اتحادیه‌های مربوطه می‌توان شناخت بیشتری از بازارهای این گروه حاصل کرد. همچنین ایجاد نمایندگی فروش، جذب سرمایه‌گذاری و برگزاری نمایشگاه می‌تواند در جذب بیشتر مشتریان و گسترش بیشتر در این بازارها موثر باشد.

● **کشورهای گروه ۳:** ۳۷ کشور عضو این گروه مشتریان جدیدی هستند که سابقه ارتباط قبلی با آنها وجود ندارد و باید سعی شود این کشورها جذب و پایه‌های ارتباط پایدار، بلندمدت و با ارزش بالا با آنها طرح‌ریزی شود تا در لایه‌های هرم ارتقا یابند. در این راستا با کاهش موقت قیمت‌ها، ارسال نمونه‌های برتر، پرداخت مشوق به مشتریان، دعوت به شرکت در نمایشگاه‌های ایران و بازدید از مرکز تولید و توزیع میوه می‌توان موثر باشد. همچنین با ارزیابی پتانسیل بازار فروش رقبا و حتی برقرار ارتباط با رقبا می‌توان به این بازارهای جدید راه یافت.

● **کشورهای گروه ۴:** ۲۴ عضو این گروه، مشتریانی هستند که غیر فعال و رویگردان شده‌اند و در حال ریزش هستند. با ارزیابی دقیق رقبا و نیازهای مشتریان، علل شکست و

امکان حضور مجدد در این بازارها قابل بررسی است. باید سعی شود کشورهای با ارزش در این گروه با دعوت از نمایندگان این کشورها برای بازدید از نمایشگاه‌های ایرانی مجدداً جذب شوند و یا سعی شود از طریق رقبا و استفاده از شبکه توزیع آنها در این بازارها حضور یافت.

### جمع‌بندی و ملاحظات

رویکرد بازاریابی انبوه پاسخگوی نیازهای مشتریان گوناگون امروزی نیست. بخش‌بندی مشتریان امکان برآورد نمودن نیازها و خواسته‌های مشتریان هر بخش و مدیریت ارتباط با مشتریان را به بهترین نحوه میسر می‌سازد. بخش‌بندی یکی از مهمترین مقوله‌ها در رسیدن به بازاریابی مدرن و مدیریت ارتباط با مشتری موفق است.

در این مقاله ابتدا متغیرهای تکرار، تازگی و ارزش پولی صادرات برای هر کشور محاسبه گردید، سپس با استفاده از الگوریتم‌های خوشه‌بندی Kmeans و Fuzzy Kmeans کشورها خوشه‌بندی شده‌اند. کیفیت خوشه‌بندی‌های صورت گرفته با معیارهای سنجش کیفیت که بر مبنای همگنی درون خوشه‌ای و ناهمگنی برون خوشه‌ای تعریف شده‌اند، سنجیده شده‌است. پس از تعیین تعداد بهینه خوشه توسط معیارهای مطرح شده خوشه‌ها بر اساس متغیرهای تازگی، تکرار و ارزش پولی به چهار گروه اصلی تقسیم شده‌اند: مشتریان وفادار، مشتریان فعال، مشتریان جدید و مشتریان غیرفعال. سپس به کمک مفاهیم هرم ارزش مشتری، درخت تصمیم‌گیری هر گروه تحلیل و سیاست‌های متناسب با هر گروه تبیین شده‌است.

در مطالعات آتی می‌توان بازه‌های سالانه بر روی بازه‌های ماهانه متغیرهای تکرار، تازگی و ارزش پولی صادرات را بررسی نمود. همچنین می‌توان از سایر الگوریتم‌ها خوشه‌بندی نیز استفاده نمود و عملکرد آن را با الگوریتم‌های بکار گرفته شده مقایسه نمود.

## منابع

- Berson, A., Smith, S. and Thearling, K. (2001); *Bulding Data Mining Application for CRM*, McGraw-Hill.
- Brown, S. A.(2002); "Performance Driven CRM: How to Make Your Customer Relationship Management Vision a Reality.", *John Wiley & Sons*.
- Dibb, S. and Simkin, L. (1996); *The Market Segmentation Workbook: Target Marketing for Marketing Managers*, London: Routledge.
- Helsen, K., Jedidi, K. and DeSarbo, W.S. (1993); 'A New Approach to Country Segmentation Utilizing Multinational Diffusion Patterns', *Journal of Marketing*, pp.63–71.
- Hruschka, H. (1986); 'Market Definition and Segmentation Using Fuzzy Clustering Methods', *International Journal of Research in Marketing*, Vol. 3, pp.117–134.
- Hwang, H., Jung, T. and Suh, E. (2004); 'An LTV Model and Customer Segmentation Based on Customer Value: A Case Study on the Wireless Telecommunication Industry', *Expert Systems with Applications*, Vol. 26, pp.181–188.
- Hughes, A. M. (1994); 'Strategic Database Marketing'. Chicago: Probus Publishing Company.
- Jonkera, J-J., Piersmab, N. and Van den Poelc, D. (2004); 'Joint Optimization of Customer Segmentation and Marketing Policy to Maximize Long-Term Profitability', *Expert Systems with Applications*, Vol. 27, pp.159–168.
- Kamakura, W.A., Wedel, M., de Rosa, F. and Mazzon, J.A. (2003); 'Cross-Selling Through Database Marketing: A Mixed Data Factor

- Analyzer for Data Augmentation and Prediction’, *Intern J. of Research in Marketing*, Vol. 20, pp.45–65.
- Kim, S-Y., Jung, T-S., Suh, E-H. and Hwang, H-S. (2006); ‘Customer Segmentation and Strategy Development Based on Customer Lifetime Value: A Case Study’, *Expert Systems with Applications*, Vol. 31, pp.101–107.
- Michaud, P. (1997); ‘Clustering Techniques’, *Future Generation Computer System*, Vol. 13, No. 2, pp.135–147.
- Newell, F. (1997); *The New Rules of Marketing: How to Use One-to-one Relationship Marketing to Be the Leader in Your Industry*, New York: McGraw-Hills Companies Inc.
- Shina, H.W. and Sohn, S.Y. (2004); ‘Segmentation of Stock Trading Customers According to Potential Value’, *Expert Systems with Applications*, Vol. 27, pp.27–33.
- Rygielski, C., Wang, J-C. and Yen, D.C. (2002); ‘Data Mining Techniques for Customer Relationship Management’, *Technology in Society*, Vol. 24, pp.483–502.
- Soman, K.P., Diwakar, S. and Ajay, V. (2006); *Insight into Data Mining (Theory and Practice)*, India: Prentice-Hall.
- Tsai, C-Y. and Chiu, C-C. (2004); ‘A Purchase-based Market Segmentation Methodology’, *Expert Systems with Applications*, Vol. 27, pp.265–276.
- Wang, H-F. and Hong, W-K. (2005); *Managing Customer Profitability in a Competitive Market by Continuous Data Mining*, Department of Industrial Engineering and Engineering Management, National Tsing Hua University, Hsinchu, Taiwan, ROC.
- Weber, R. (1996); ‘Customer Segmentation for Banks and Insurance Groups with Fuzzy Clustering Techniques’, in J.F. Baldwin (Ed.) *Fuzzy Logic*, New York: Wiley.

- Wilcox, P.A. and Gurau, C. (2003); 'Business Modeling with UML: the Implementation of CRM System for Online Retailing', *Journal of Retailing and Consumer Services*, Vol. 10, pp.181–191.
- Wu, J., & Lin, Z. (2005); Research on Customer Segmentation Model by Clustering. *ACM International Conference Proceeding Series*, 113.
- Ye, N. (2003); *The Handbook of Data Mining*, Mahwah, NJ/London: Lawrence Erlbaum Associates, Publishers.
- Yeh, C., Yang, K. and Ting, T. (2009); 'Knowledge Discovery on RFM Model Using Bernoulli Sequence', *Expert Systems with Applications*, Vol 36, pp5866–5871
- Zeithaml, V.A., Rust, R.T. and Lemon, K.N. (2001); 'The Customer Pyramid: Creating and Serving Profitable Customers', *California Management Review*.