

## سپاسگزاری

کلمات برای قدردانی از آنان که صبورانه در انجام این تحقیق یاریم کردند، قاصر از بیان عمق سپاسم هستند. من که دانسته‌هایم بس اندک بود و نیازمند حوصله و راهنمایی استادان. با این همه گزیری نیست جز آنکه به همان چارچوب دیرین اکتفا کرده، بگویم سپاس و امتنان برای:

جناب آقای دکتر امیر مسعود رحمانی، استاد گرانقدری که بر من منت گذارده، مسئولیت راهنمایی مرا در طول انجام این پروژه بر عهده گرفتند و به حق با راهنمایی‌های عالمانه و دلسوزانه خویش چه مهربانانه و متواضعانه مشوق و راهنمای من در این امر مهم بودند، اطمینان دارم که اگر راهنمایی‌ها، حمایت‌ها و دلگرمی‌های ایشان نبود اکنون این پروژه هنوز در ابتدای راه خود بود.

و در ادامه کمال سپاس خود را نسبت به جناب آقای دکتر احمد فراهی، که افتخار استاد مشاور را به من دادند ابراز می‌دارم.

اکنون بر این باورم که با کمک آن‌ها گام‌هایم استوار گردید و انگیزه‌ام قدرتی شگرف یافت. باشد که خداوند نگهدار این دو عزیز باشد.

سپاسگزارم از همسر مهربانم

که تقدیر کوچک‌ترین کلام است برای پاسخ به الطافش

و تعظیم و تکریم می‌طلبد این همگامی بردبارانه

## تقدیم به

مادر عزیزم

که بی‌شک دعای خیرش در تمام مراحل  
زندگی بدرقه راهم بوده است.

## چکیده

داده‌کاوی به عمل استخراج اطلاعات سودمند و دانش پنهان از پایگاه داده گفته می‌شود. کاوش قوانین انجمنی یکی از مهم‌ترین تکنیک‌های داده‌کاوی است که به استخراج الگوهای تکرارشونده از لابلای مجموعه ارقام پایگاه داده‌های تراکنشی با حجم بالا می‌پردازد و ریشه اصلی آن به تجزیه و تحلیل تعاملات نقطه فروش برمی‌گردد. اکثر الگوریتم‌های مرسوم کاوش قوانین انجمنی بر روی داده‌های دودویی کار می‌کنند، در حالی که پایگاه داده‌های استفاده شده در اکثر موارد تجاری و علمی حاوی داده‌های کمی هستند. این تحقیق برای کار با داده‌های کمی، مفاهیم منطق فازی و الگوریتم‌های کلاسیک کاوش قوانین انجمنی را مجتمع می‌کند تا قوانین انجمنی فازی مفید استخراج کند.

/

:

## فهرست مطالب

صفحه	عنوان
ث	چکیده .....
۴	فهرست جداول .....
۶	فهرست اشکال .....
	<b>فصل اول: مقدمه</b>
۸	۱-۱ مقدمه .....
۹	۲-۱ تعریف مسأله و بیان سؤالهای اصلی تحقیق .....
۱۱	۳-۱ سابقه و ضرورت انجام تحقیق .....
۱۲	۴-۱ فرضیه‌ها .....
۱۲	۵-۱ اهداف .....
۱۲	۶-۱ کاربردهای تحقیق .....
۱۳	۷-۱ جنبه جدید بودن و نوآوری طرح .....
۱۳	۸-۱ روش انجام تحقیق .....
۱۳	۹-۱ ساختار پایان نامه .....
	<b>فصل دوم: داده‌کاوی و کاوش قوانین انجمنی</b>
۱۴	۱-۲ داده‌کاوی و مراحل آن .....
۱۶	۲-۲ تکنیک‌های داده‌کاوی .....
۱۸	۳-۲ انواع داده‌کاوی .....
۱۹	۴-۲ کاوش قوانین انجمنی .....
۲۱	۵-۲ تعریف مسئله کاوش قوانین انجمنی .....
۲۲	۶-۲ الگوریتم‌های مطرح شده برای کاوش اقلام تکرارشونده .....
۲۲	۱-۶-۲ الگوریتم‌های پیروی .....
۲۶	۷-۲ قوانین انجمنی چند سطحی .....

## فصل سوم: خوشه‌بندی

- ۲۷ ..... ۱-۳ تعریف خوشه‌بندی
- ۲۷ ..... ۲-۳ کاربردهای خوشه‌بندی
- ۲۸ ..... ۳-۳ گام‌های اساسی در انجام خوشه‌بندی
- ۳۰ ..... ۴-۳ معیارهای شباهت - فاصله مورد استفاده در انجام خوشه‌بندی
- ۳۰ ..... ۱-۴-۳ چگونگی اندازه‌گیری شباهت - فاصله
- ۳۲ ..... ۲-۴-۳ اندازه‌گیری شباهت - فاصله بین متغیرهای گسسته
- ۳۴ ..... ۳-۴-۳ اندازه‌گیری شباهت - فاصله بین متغیرهای پیوسته
- ۳۶ ..... ۵-۳ روش‌های خوشه‌بندی
- ۴۰ ..... ۶-۳ اندازه‌گیری اعتبار خوشه‌ها

## فصل چهارم: منطق فازی

- ۴۲ ..... ۱-۴ مقدمه‌ای بر منطق فازی
- ۴۵ ..... ۲-۴ مفاهیم مورد استفاده در منطق فازی
- ۴۵ ..... ۳-۴ عملیات روی مجموعه‌های فازی
- ۴۷ ..... ۴-۴ فازی کردن و غیر فازی کردن

## فصل پنجم: کارهای مرتبط

- ۴۹ ..... ۱-۵ قوانین انجمنی کمی
- ۴۹ ..... ۱-۱-۵ نگاشت مسئله کاوش قوانین انجمنی کمی به کاوش قوانین انجمنی دودویی
- ۵۱ ..... ۲-۱-۵ استفاده از خوشه‌بندی برای تقسیم دامنه خصیصه‌ها
- ۵۴ ..... ۳-۱-۵ استفاده از مفاهیم منطق فازی جهت کاوش قوانین انجمنی کمی
- ۵۵ ..... ۲-۵ استفاده از خوشه‌بندی در کنار کاوش قوانین انجمنی
- ۵۵ ..... ۱-۲-۵ خوشه‌بندی تراکنش‌ها با استفاده از نتایج اجرای الگوریتم‌های کاوش قوانین انجمنی بر روی داده‌ها
- ۵۶ ..... ۲-۲-۵ کاوش قوانین انجمنی به کمک خوشه‌بندی
- ۵۸ ..... ۳-۵ روش‌های افزایش کارایی الگوریتم‌های کاوش قوانین انجمنی

## فصل ششم: الگوریتم پیشنهادی

- ۶۰ ..... ۱-۶ توضیح الگوریتم پیشنهادی
- ۶۱ ..... ۲-۶ گام‌های الگوریتم
- ۷۰ ..... ۳-۶ پیاده‌سازی الگوریتم پیشنهادی
- ۷۰ ..... ۱-۳-۶ مشخصات کلی پایگاه داده استفاده شده
- ۷۱ ..... ۲-۳-۶ محیط آزمایش
- ۷۲ ..... ۴-۶ نتایج اجرای الگوریتم بر روی پایگاه داده فروش فرآورده‌های نفتی

## فصل هفتم: ارزیابی الگوریتم پیشنهادی

- ۷۴ ..... ۱-۷ بررسی کارایی الگوریتم
- ۷۴ ..... ۱-۱-۷ نتایج پیاده‌سازی بر روی زمان اجرا
- ۷۷ ..... ۲-۱-۷ نتایج پیاده‌سازی بر روی دقت و صحت قوانین انجمنی استخراج شده
- ۸۰ ..... ۲-۷ مزایا و معایب الگوریتم پیشنهادی
- ۸۰ ..... ۱-۲-۷ مزایای الگوریتم پیشنهادی
- ۸۱ ..... ۲-۲-۷ معایب الگوریتم پیشنهادی

## فصل هشتم: نتیجه‌گیری و پیشنهادها

- ۸۲ ..... ۱-۸ نتیجه‌گیری
- ۸۳ ..... ۲-۸ پیشنهاد کارهای آتی

## فهرست جداول

صفحه	عنوان
<b>فصل دوم</b>	
۲۰	جدول ۱-۲: یک پایگاه داده نمونه .....
۲۴	جدول ۲-۲: نمادهای مهم مربوط به الگوریتم ای‌پریوری .....
<b>فصل سوم</b>	
۳۵	جدول ۱-۳: معیارهای اندازه‌گیری شباهت و کاربرد آن‌ها .....
<b>فصل پنجم</b>	
۵۰	جدول ۱-۵: یک جدول نمونه با داده‌های کمی .....
۵۰	جدول ۲-۵: جدول حاصل از تقسیم دامنه خصیصه‌های کمی جدول ۱-۵ .....
<b>فصل ششم</b>	
۶۴	جدول ۱-۶: نمونه‌ای از داده‌های میزان فروش فرآورده‌های نفتی .....
۶۶	جدول ۲-۶: مجموعه‌های فازی تشکیل شده از داده‌های جدول ۱-۶ .....
۶۷	جدول ۳-۶: عدد اصلی (میزان پشتیبانی) نواحی فازی جدول ۲-۶ .....
۶۷	جدول ۴-۶: مجموعه $L_1$ ناشی از جدول ۲-۶ .....
۶۸	جدول ۵-۶: مقادیر عضویت برای $(GH, LPGL)$ .....
۶۸	جدول ۶-۶: پشتیبانی تمام اعضای $C_2$ .....
۶۹	جدول ۷-۶: مجموعه $L_2$ و پشتیبانی اعضای آن .....
۶۹	جدول ۸-۶: قوانین فازی نهایی تشکیل شده بر روی داده‌های جدول ۱-۶ .....
۷۰	جدول ۹-۶: مشخصات کلی جداول مورد استفاده در تحقیق .....
۷۱	جدول ۱۰-۶: مشخصات کامپیوتر مورد استفاده در تحقیق .....
۷۲	جدول ۱۱-۶: خصوصیات خوشه‌های تشکیل شده بر روی داده‌های فروش فرآورده‌های نفتی .....



## فصل هفتم

- جدول ۱-۷: تعداد رکوردها در بزرگ‌ترین خوشه در انواع خوشه‌بندی ..... ۷۴
- جدول ۲-۷ الف: قوانین کشف شده از روی داده‌های سال‌های ۱۳۷۳ الی ۱۳۸۰ ..... ۷۵
- جدول ۲-۷ ب: قوانین کشف شده از روی داده‌های سال‌های ۱۳۷۳ الی ۱۳۸۸ ..... ۷۵

## فهرست اشکال

صفحه	عنوان
<b>فصل دوم</b>	
۱۵	شکل ۱-۲: کشف دانش در پایگاه داده
۲۳	شکل ۲-۲: الگوریتم ای‌پریوری
۲۴	شکل ۳-۲: عملیات الحاق در ای‌پریوری
۲۵	شکل ۴-۲: عملیات حذف در ای‌پریوری
۲۵	شکل ۵-۲: عملکرد الگوریتم ای‌پریوری
۲۶	شکل ۶-۲: نمونه ای از سلسله مراتب مفهومی در قوانین انجمنی چند سطحی
<b>فصل سوم</b>	
۲۹	شکل ۱-۳: مراحل خوشه‌بندی
۳۶	شکل ۲-۳: یک رده بندی از روش‌های خوشه‌بندی
<b>فصل چهارم</b>	
۴۳	شکل ۱-۴: انواع توابع عضویت در منطق فازی
۴۴	شکل ۲-۴: نمایش قد افراد با مجموعه‌های قاطع و مجموعه‌های فازی
۴۵	شکل ۳-۴: متمم یک مجموعه فازی
۴۶	شکل ۴-۴: اشتراک دو مجموعه فازی
۴۶	شکل ۵-۴: اجتماع دو مجموعه فازی
۴۷	شکل ۶-۴: فازی کردن به صورت مثلثی
۴۸	شکل ۷-۴: فازی کردن به صورت دوزنقه ای
<b>فصل پنجم</b>	
۵۳	شکل ۱-۵: تصویر کردن خوشه‌ها روی دامنه‌ای از خصیصه‌های دودویی
<b>فصل ششم</b>	
۶۵	شکل ۱-۶: نتیجه اجرای تابع Silhouette بر روی خوشه‌بندی با سه خوشه
۶۵	شکل ۲-۶: نتیجه اجرای تابع Silhouette بر روی خوشه‌بندی با چهار خوشه

شکل ۶-۳: نتیجه اجرای تابع Silhouette بر روی خوشه‌بندی با هشت خوشه ..... ۶۵

## فصل هفتم

شکل ۷-۱: مقایسه زمان کشف قوانین در دو حالت با خوشه‌بندی و بدون خوشه‌بندی ..... ۷۶

شکل ۷-۲ الف: مقایسه حداقل اطمینان با تعداد قوانین کشف‌شده در حالت بدون استفاده از خوشه‌بندی ..... ۷۸

شکل ۷-۲ ب: مقایسه حداقل اطمینان با تعداد قوانین کشف‌شده در حالت خوشه‌بندی با سه خوشه ..... ۷۸

شکل ۷-۲ پ: مقایسه حداقل اطمینان با تعداد قوانین کشف‌شده در حالت خوشه‌بندی با هفت خوشه ..... ۷۸

شکل ۷-۳: مقایسه حداقل پشتیبانی لازم برای کاوش قوانین در حالت با استفاده از خوشه‌بندی و بدون استفاده از خوشه‌بندی ..... ۷۹

## ۱-۱ مقدمه

در دو دهه قبل با رشد فزاینده داده‌ها در سیستم‌های مختلف و همچنین نیاز روزافزون سازمان‌ها به کسب دانش مفید و اطلاعات ارزشمند از داده‌های خام جهت انجام فعالیت‌های تصمیم‌گیری و پیش‌بینی، داده‌کاوی به عنوان روشی مهم و پرکاربرد برای استخراج اطلاعات و برآوردن نیازهای فوق مطرح شده است. در یک تعریف غیر رسمی داده‌کاوی فرآیندی است خودکار، برای استخراج الگوهایی که دانش را بازنمایی می‌کنند، که این دانش به صورت ضمنی در پایگاه داده‌های عظیم، انباره داده<sup>۱</sup> و دیگر مخازن بزرگ اطلاعات، ذخیره شده است.

داده‌کاوی موضوعی وابسته به کاربرد است و کاربردهای مختلف نیازمند روش‌ها و تکنیک‌های داده‌کاوی مختلفی هستند. کاوش قوانین انجمنی<sup>۲</sup>، دسته‌بندی<sup>۳</sup>، خوشه‌بندی<sup>۴</sup>، پیش‌بینی<sup>۵</sup> و تحلیل سری‌های زمانی<sup>۶</sup> از جمله مهم‌ترین کارهایی هستند که می‌توان با داده‌کاوی انجام داد.

کاوش قوانین انجمنی در پایگاه داده‌های رابطه‌ای<sup>۷</sup> یا تراکنشی<sup>۸</sup>، یکی از راهکارهای داده‌کاوی می‌باشد که اخیراً جذابیت زیادی را در انجمن‌های مربوط به پایگاه‌های داده به وجود آورده است، در این راهکار داده‌کاوی وابستگی‌ها و ارتباطات بین داده‌های موجود در یک پایگاه داده بدست می‌آیند. نتیجه این عملیات داده‌کاوی دسته‌ای از قوانین است که به آن‌ها قوانین انجمنی گفته می‌شود. اگرچه قوانین انجمنی را در موارد متعددی می‌توان بکار برد، ولی ریشه آن‌ها به تجزیه و تحلیل تعاملات نقطه فروش برمی‌گردد.

لازم به ذکر است که پایگاه داده تراکنشی، یک سیستم مدیریت پایگاه داده است که در آن عملیات کار با داده‌ها در صورت تکمیل نشدن صحیح، برگشت داده می‌شود و پایگاه داده رابطه‌ای مجموعه‌ای از اقلام داده‌هاست که در قالب جدول‌های مرتبط با هم نمایش داده می‌شوند.

---

<sup>۱</sup> Data warehouses

<sup>۲</sup> Association Rules Mining

<sup>۳</sup> Classification

<sup>۴</sup> Clustering

<sup>۵</sup> Prediction

<sup>۶</sup> Time Series Analysis

<sup>۷</sup> Relational Database

<sup>۸</sup> Transactional Database

پایگاه‌های استفاده شده در اکثر موارد تجاری و علمی، دارای انواع گوناگونی از فیلدها به خصوص فیلدهای کمی<sup>۱</sup> می‌باشند. کاوش قوانین انجمنی در چنین پایگاه داده‌هایی کاوش قوانین انجمنی کمی<sup>۲</sup> نامیده می‌شود [Srikant, 1996]. روش‌های متفاوتی جهت انجام فرآیند کاوش در این پایگاه داده‌ها ارائه شده‌اند که در برخی از مفاهیم منطق فازی استفاده می‌شود، روش‌های کاوش فازی، مفهوم مجموعه‌های فازی و الگوریتم‌های کلاسیک کاوش قوانین انجمنی را مجتمع می‌کنند تا قوانین انجمنی فازی مفید استخراج کنند [Hong, 1999]. یکی دیگر از مهم‌ترین روش‌های داده‌کاوی، خوشه‌بندی است. در خوشه‌بندی مجموعه‌ای از داده‌ها گروه‌بندی می‌شوند. خوشه‌بندی داده‌ها بر اساس اصل مفهومی زیر صورت می‌گیرد:

" حداکثر کردن شباهت‌های بین اعضای هر کلاس و حداقل کردن شباهت‌ها بین اعضای مربوط به کلاس‌های مختلف "

خوشه‌بندی معمولاً به عنوان پیش درآمدی برای انواع دیگری از داده‌کاوی یا مدل سازی بکار می‌رود. به عنوان مثال مجموعه‌ای از کالاها را می‌توان در ابتدا به صورت مجموعه‌ای از کلاس‌های مختلف خوشه‌بندی کرده و سپس مجموعه‌ای از قوانین انجمنی را بر اساس این چنین دسته‌بندی نتیجه‌گیری کرد. [Salazar, 2005] در برخی موارد نیز می‌توان از خوشه‌بندی تراکنش‌ها به منظور کاوش قوانین انجمنی نادر<sup>۳</sup> بهره جست [Koh, 2008]. گاهی هم بالعکس، از نتایج حاصل از کاوش قوانین انجمنی بر روی داده‌ها، برای خوشه‌بندی تراکنش‌ها استفاده می‌شود [Wang, 1999].

## ۲-۱ تعریف مسأله و بیان سؤالیهای اصلی تحقیق

الگوریتم‌های فعلی کاوش قوانین انجمنی به تعداد گذرهای بالا از پایگاه داده نیاز دارند، به همین دلیل در پایگاه‌های بزرگ سربرار ناشی از عملیات ورودی/خروجی در این الگوریتم‌ها بسیار بالا است. در الگوریتم پیشنهادی این تحقیق، به منظور حل مشکل فوق، کل پایگاه داده فقط یک بار مورد بررسی قرار می‌گیرد. در این گذر، تراکنش‌های پایگاه داده به خوشه‌هایی تقسیم می‌شوند، به نحوی که تراکنش‌های هر خوشه بیشترین شباهت را به یکدیگر داشته باشند. سپس به جای اعمال الگوریتم‌های کاوش قوانین انجمنی بر روی داده‌های کل پایگاه داده، این الگوریتم‌ها به صورت جداگانه و موازی بر روی داده‌های هر خوشه اجرا خواهند شد.

---

<sup>۱</sup> Quantitative

<sup>۲</sup> Quantitative Association Rules Mining

<sup>۳</sup> Rare Association Rules Mining

هدف بررسی تأثیر الگوریتم مورد استفاده بر روی زمان اجرای الگوریتم‌های کاوش قوانین انجمنی و دقت و صحت قوانین کشف شده است.

شرکت ملی پخش فرآورده های نفتی ایران وظیفه تامین، توزیع و فروش روزانه بیش از ۲۲۰ میلیون لیتر انواع فرآورده های نفتی را (مانند بنزین معمولی، بنزین سوپر، نفت گاز یا گازوئیل، نفت سفید، نفت کوره و سوخت هواپیما) در سراسر خاک ایران عهده دار است. از آنجا که فرآورده‌های نفتی در داخل کشور به صورت بارانه ای توزیع می شوند، در بخش فروش این محصولات حساسیت های خاصی در جهت تعیین الگوی مصرف نواحی مختلف کشور وجود دارد. همچنین نیازسنجی مصرف سوخت در بخش های مختلف از جمله حمل و نقل، صنایع، کشاورزی، خانگی، اداری و نیروهای مسلح از اهداف این شرکت است.

کاوش قوانین انجمنی مقداری به کمک یک روش فازی و بر اساس خوشه بندی، در پایگاه داده فروش روزانه فرآورده های نفتی در ۱۳ سال گذشته به عنوان موضوع این تحقیق انتخاب شده است.

به منظور افزایش سرعت فرآیند کاوش ابتدا نواحی فروش فرآورده ها بر اساس الگوی مصرف فرآورده در کاربردهای گوناگون، خوشه بندی می شوند.

خوشه بندی به کمک پایگاه داده تجزیه میزان فروش ماهیانه فرآورده های نفتی در نواحی، به تفکیک مصرف آنها در بخش های مختلف حمل و نقل، صنایع، کشاورزی، خانگی، اداری و نیروهای مسلح انجام می شود.

تکنیک های کاوش قوانین انجمنی مقداری به روش فازی، بطور جداگانه در داده های فروش روزانه نتایج هر یک از خوشه ها اعمال خواهند شد و سعی بر این خواهد بود که الگوهای استخراج شده از خوشه بندی به همراه قوانین انجمنی هر خوشه به سوالهای زیر پاسخ دهند:

۱. آیا اعمال تکنیکهای کاوش قوانین انجمنی پس از خوشه بندی تراکنش ها، به افزایش سرعت فرآیند کاوش خواهد انجامید؟
۲. آیا به کمک روش پیشنهادی، اعمال تکنیکهای کاوش قوانین انجمنی بصورت برخط<sup>۱</sup>، بدون افت چشمگیر دقت و صحت قوانین استخراج شده، با سرعت بالاتری انجام خواهد شد؟
۳. الگوهای مصرف و قوانین استخراج شده تا چه اندازه در بهبود عملیات توزیع و فروش فرآورده های نفتی در نواحی مختلف مفید واقع خواهند شد؟

---

<sup>۱</sup> Online

### ۱-۳ سابقه و ضرورت انجام تحقیق

کاوش قوانین انجمنی از سال ۱۹۹۳ توسط آگراوال مطرح شد [Agrawal, 1993]. این تکنیک اولین بار با استفاده از آنالیز داده های فروش، برای کشف الگوهای خرید مشتریان بدون در نظر گرفتن میزان خرید آنان، استفاده شد. به همین دلیل به این روش ها، کاوش قوانین انجمنی دودویی نیز گفته می شود. امروزه با گسترش استفاده از سیستم های پایگاه داده، جداول رابطه ای اکثراً دارای داده های مقداری و قیاسی هستند. به همین دلیل الگوریتم های معمول کاوش قوانین انجمنی نمی توانند بر روی این داده ها بکار گرفته شوند. روش های متعددی برای کاوش قوانین انجمنی مقداری پیشنهاد شده اند.

برخی از این روش ها تلاش می کنند تا با نداشت مسئله کاوش قوانین انجمنی مقداری به مسئله کاوش قوانین انجمنی دودویی این مشکل را برطرف نمایند. لازمه این نگاشت، پارتیشن بندی کردن دامنه مقادیر هر خصیصه به تعدادی فاصله خاص و سپس نگاشت هر زوج <فاصله، خصیصه> به یک خصیصه دودویی می باشد. پس از این نگاشت می توان از هر کدام از الگوریتم های کاوش قوانین انجمنی دودویی بر روی این داده های جدید استفاده نمود. برای پارتیشن بندی می توان از متدهای مختلف بخش بندی دامنه مقادیر [Srikant, 1996] و یا از خوشه بندی تراکنش ها استفاده کرد [Tong, 2005].

برخی دیگر از روش های کاوش قوانین انجمنی مقداری از منطق فازی برای کاوش قوانین استفاده می کنند. این الگوریتم ها، مفهوم مجموعه های فازی و الگوریتم های کاوش قوانین انجمنی دودویی را مجتمع می کنند تا قوانین انجمنی فازی مفید را استخراج کنند [Salazar, 2005].

از طرفی به دلیل بالا بودن تعداد گذرها از پایگاه داده در اکثر الگوریتم های کاوش قوانین انجمنی، شاهد افت کار آیی الگوریتم ها هستیم. هزینه محاسبات کاوش قوانین انجمنی به چهار روش زیر می تواند کاهش پیدا کند [Kotsiantis, 2006]:

- ✓ کاهش تعداد گذرها از پایگاه داده
- ✓ نمونه برداری از پایگاه داده
- ✓ اضافه کردن محدودیت های بیشتر بر روی ساختار الگوها
- ✓ موازی نمودن انجام عملیات کاوش

با توجه به پراکندگی داده های فروش روزانه فرآورده در نواحی گوناگون و دردست نبودن الگوی دقیق مصرف فرآورده ها در بخش های مختلف آنها، توصیف دقیق و درست این الگوها جهت بهبود توزیع فرآورده و یا استقرار مناسب انبارهای فرآورده امکان پذیر نیست.

آنچه به طرح ریزی درست می انجامد دقت در استخراج دانش مفید جهت ترسیم وضعیت موجود و وضعیت مطلوب است.

با توجه به آنچه ذکر شد، تجمیع داده های فروش روزانه فرآورده‌ها، تقسیم بندی نواحی مختلف براساس الگوی مصرف و کاوش قوانین انجمنی به روش‌های فازی به منظور اخذ تصمیمات مدیریتی دقیق تر از ضرورت‌های انجام این تحقیق است.

در این تحقیق برای افزایش سرعت کاوش قوانین، از تکنیک‌های خوشه بندی استفاده خواهد شد. بدین ترتیب به جای کاوش کل پایگاه داده، عملیات کاوش مبتنی بر منطق فازی و بر روی خوشه های مختلف، به صورت موازی قابل انجام است.

#### ۴-۱ فرضیه‌ها

- پایگاه داده فروش روزانه فرآورده‌های نفتی شرکت ملی پخش فرآورده های نفتی مفروض است.
- پایگاه داده فروش ماهیانه فرآورده‌ها به تفکیک کاربرد آن‌ها در مصارف گوناگون مفروض است.
- تعداد تراکنش‌های پایگاه داده بالا است.
- داده های موجود در هر دو پایگاه، ماهیت مقداری (کمی) دارند.

#### ۵-۱ اهداف

- کاوش قوانین انجمنی مبتنی بر منطق فازی، به کمک تکنیک‌های خوشه بندی در داده‌های فروش روزانه فرآورده های نفتی در سراسر ایران
- توصیف دقیق وضعیت کنونی و پیش بینی وضعیت آینده جهت بهبود توزیع فرآورده‌های نفتی

#### ۶-۱ کاربردهای تحقیق

- ارائه راهکارهای اصولی جهت کاهش مصرف سوخت، بر اساس الگوهای کشف شده در مصرف
- بهبود توزیع فرآورده های نفتی در استان‌ها و شهرستان‌های مختلف، متناسب با الگوی مصرف هر خوشه
- بررسی متناسب بودن توزیع انبارهای نفت در ایران، با توجه به الگوهای کشف شده
- کمک به اخذ تصمیمات مدیریتی دقیق تر به کمک دانش‌های مفید کشف شده



## ۷-۱ جنبه جدید بودن و نوآوری طرح

کاوش قوانین انجمنی با خوشه بندی داده‌ها بر اساس یک روش فازی برای فرآورده های نفتی در ایران.

## ۸-۱ روش انجام تحقیق

- بررسی و مطالعه مباحث قبلی که در مطالب آکادمیک به ثبت رسیده است
- بررسی وضعیت داده‌ها در شرکت ملی پخش فرآورده های نفتی ایران
- استفاده از منطق فازی جهت کاوش قوانین انجمنی در داده های مذکور

## ۹-۱ ساختار پایان نامه

در فصل دوم و سوم این پایان نامه، ابتدا توضیحات بیشتری در مورد داده‌کاوی، مراحل مختلف آن و دو راهکار مهم داده‌کاوی، یعنی خوشه‌بندی و کاوش قوانین انجمنی و روش‌های مختلف انجام آن‌ها بر روی انواع مختلف داده‌ها آورده شده است. فصل چهارم به معرفی منطق فازی و مفاهیم آن می‌پردازد. در فصل پنجم روش‌های مختلف ارائه شده جهت کاوش قوانین انجمنی کمی و حل مشکلات مربوط به آن‌ها بررسی می‌شوند. در فصل ششم به توضیح الگوریتم کاوش قوانین انجمنی مقداری به کمک یک روش فازی و بر اساس خوشه‌بندی، در پایگاه داده فروش روزانه فرآورده‌های نفتی در ۱۶ سال گذشته پرداخته می‌شود و در فصل هفتم ارزیابی نتایج حاصل از اجرای الگوریتم فوق بر روی این پایگاه داده انجام می‌شود.

## ۱-۲ داده‌کاوی و مراحل آن

به داده‌کاوی عملیات اکتشاف دانش در پایگاه‌های داده هم گفته می‌شود، اگرچه برخی محققین داده‌کاوی را بخشی از عملیات کلی‌تر اکتشاف دانش و برخی آن را هسته عملیات اکتشاف دانش در پایگاه داده می‌دانند [Zhao, 2003].

فرآیند کشف دانش به معنای کشف نیمه اتوماتیک دانش از پایگاه داده و یافتن اطلاعات و الگوهای مفید از داده‌ها است که باید معتبر، از قبل ناشناخته و بالقوه مفید باشند [Abari, 2008]. این فرآیند یک فرآیند تکرارپذیر و نیمه اتوماتیک است. عبارت نیمه اتوماتیک به این دلیل بکار می‌رود که اغلب در برخی شرایط، تعاملات کاربری نیز لازم است. عملیات اکتشاف دانش در پایگاه داده در شکل ۱-۲ نشان داده شده است و شامل اجرای تکراری مراحل زیر است [Piatetsky, 1991]:

- پاک‌سازی داده‌ها<sup>۱</sup>: در این مرحله داده‌ها از وجود خطاها، داده‌های مشکوک و داده‌های غیر مرتبط پاک می‌شوند.
- مجتمع‌سازی داده‌ها<sup>۲</sup>: در این مرحله داده‌هایی که از منابع مختلف گردآوری شده‌اند در یک منبع واحد مجتمع می‌شوند.
- گزینش داده‌ها<sup>۳</sup>: در این مرحله داده‌هایی را که عملیات تحلیل روی آن‌ها انجام می‌شود از پایگاه داده بازیابی می‌شوند.
- تبدیل داده‌ها<sup>۴</sup>: در این مرحله داده‌ها به قالب و فرم مناسب برای انجام عملیات داده‌کاوی تبدیل می‌شوند.
- داده‌کاوی: مرحله‌ای ضروری است که در آن روش‌های هوشمند به منظور استخراج الگوها، به داده‌های تبدیل یافته اعمال می‌شوند.
- ارزیابی الگوها<sup>۵</sup>: در این مرحله الگوهای بدست آمده توسط یک سری معیارهای مقبولیت سنجیده می‌شوند و در نهایت از میان الگوهای بدست آمده آن‌هایی که واقعاً با ارزش هستند تشخیص داده می‌شوند.
- ارائه دانش<sup>۱</sup>: در این مرحله نتایج بدست آمده به شکل معناداری به وسیله عملیات گرافیکی و سایر روش‌های نمایش دانش به کاربر عرضه می‌شوند.

---

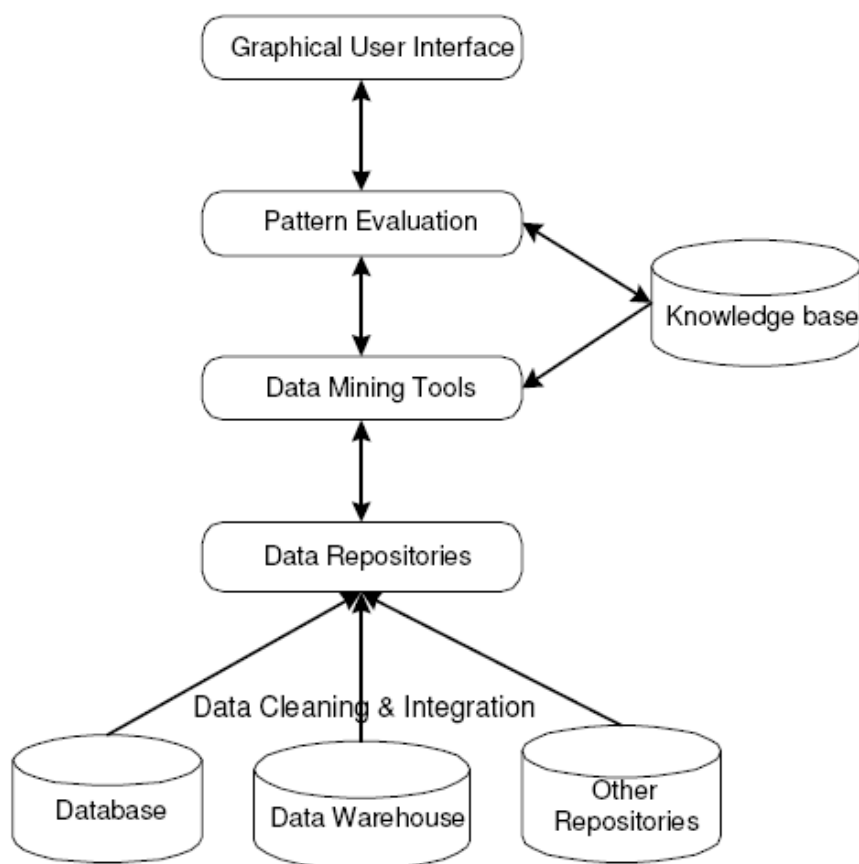
<sup>۱</sup> Data Cleaning

<sup>۲</sup> Data integration

<sup>۳</sup> Data selection

<sup>۴</sup> Data transformation

<sup>۵</sup> Pattern evaluation



شکل ۱-۲ کشف دانش در پایگاه داده [Piatetsky, 1991]

با گسترش روز افزون استفاده از پایگاه‌های داده و انبارهای داده، چهار مرحله نخست شامل خالص‌سازی، سرهم‌سازی، گزینش و تبدیل، با ساختن انباره‌های داده و انجام برخی عملیات گزارش‌گیری مربوط به انباره‌های داده قابل انجام هستند. عملیات داده‌کاوی، ارزیابی الگوها و ارائه دانش گاهی در یک مرحله کلی به نام داده‌کاوی انجام می‌شود.

<sup>۱</sup> Knowledge presentation

## ۲-۲ تکنیک‌های داده‌کاوی

داده‌کاوی موضوعی وابسته به کاربرد است و کاربردهای مختلف نیازمند روش‌ها و تکنیک‌های داده‌کاوی مختلفی هستند. کاوش قوانین انجمنی، دسته‌بندی، خوشه‌بندی، پیش‌بینی و تحلیل سری‌های زمانی از جمله مهم‌ترین کارهایی هستند که می‌توان با داده‌کاوی انجام داد. در ادامه هرکدام از این عملیات به صورت خلاصه توضیح داده می‌شوند، در ابتدا باید گفت که داده‌کاوی در دو نوع ظاهر می‌شود:

- داده‌کاوی هدایت شده<sup>۱</sup>
- داده‌کاوی غیر هدایت شده<sup>۲</sup>

داده‌کاوی هدایت شده می‌خواهد برخی از زمینه‌های هدف خاص را بیان یا دسته‌بندی کند، حال آنکه هدف داده‌کاوی غیر هدایت شده یافتن الگوها یا تشابهات بین گروه‌هایی از اطلاعات، بدون استفاده از زمینه هدف خاص یا مجموعه‌ای از دسته‌های از پیش تعیین شده می‌باشد.

دسته‌بندی، پیش‌بینی و تحلیل سری‌های زمانی مثال‌هایی از داده‌کاوی هدایت شده هستند که در آن‌ها هدف یافتن ساختاری در داده‌ها بدون توجه به یک متغیر هدف خاص است.

کاوش قوانین انجمنی در پایگاه داده‌های رابطه‌ای یا تراکنشی که هدف اصلی این پایان نامه نیز است، جزو اعمال غیر هدایت شده است که اخیراً جذابیت زیادی را در انجمن‌های مربوط به پایگاه‌های داده به وجود آورده است. در این راهکار داده‌کاوی وابستگی‌ها و ارتباطات بین داده‌های موجود در یک پایگاه داده بدست می‌آیند. نتیجه این عملیات داده‌کاوی دسته‌ای از قوانین است که به آن‌ها قوانین انجمنی گفته می‌شود. اگرچه قوانین انجمنی را می‌توان در موارد متعددی بکار برد، ولی ریشه آن‌ها به تجزیه و تحلیل تعاملات نقطه فروش برمی‌گردد.

یکی دیگر از روش‌های مهم داده‌کاوی توانایی انجام عملیات دسته‌بندی در حجم زیاد داده‌هاست که در آن اشیاء موجود در یک پایگاه داده بر اساس مقادیر چند خصوصیت از آن‌ها، به دسته‌های مجزا تقسیم می‌شوند. در دسته‌بندی داده‌ها مجموعه‌ای از داده‌های آزمایشی تحلیل می‌شوند. برای مجموعه داده‌های آزمایشی برچسب کلاس‌ها مشخص است. برای هر کلاس داده‌های آزمایشی مدلی بر اساس خصوصیات داده‌ها ساخته می‌شود. برخی از تکنیک‌هایی که به کمک آن‌ها می‌توان دسته‌بندی انجام داد عبارتند از [Zhao, 2003]:

- درخت‌های تصمیم<sup>۳</sup>
- تکنیک‌های نزدیک‌ترین همسایه

---

<sup>۱</sup> Directed Data Mining

<sup>۲</sup> Indirected Data Mining

<sup>۳</sup> Decision Tree