

## داده کاوی؛ ابزاری در خدمت تدوین استراتژی

شاھین رشتیان<sup>۱</sup> – بابک نصیری<sup>۲</sup>

### چکیده:

موفقیت یا عدم موفقیت یک برنامه استراتژیک، رابطه مستقیم با نحوه تدوین استراتژی آن دارد. استراتژی را میتوان اساس و بنیان مدیریت استراتژیک دانست. این مقاله سعی دارد با تشریح چرخه تدوین استراتژی، ابزاری بنام داده کاوی را برای بررسی بهتر و جامع تر عوامل تاثیرگذار بر سازمان و روابط بین آنها ارائه کنند در چارچوب پیشنهادی، مدلی مبتنی بر الگوریتم ژنتیک – بعنوان یکی از تکنیکهای داده کاوی – پیشنهاد شده و مورد پیاده سازی قرار گرفته است که با کاوش بر روی انبوه داده های محیطی مربوط به یک شرکت فرضی، مجموعه ای از داشتهای جالب را در قالب یکسری الگو و قانون ارائه میکند. استفاده از این الگوها و قوانین میتواند به استراتژیست در شناخت بهتر عوامل تاثیرگذار بر سازمان کمک نموده و در نتیجه، استراتژی تدوین شده را کارآفر و واقع بینانه تر نماید.

**واژه های کلیدی:** مدیریت استراتژیک – برنامه ریزی استراتژیک – تدوین استراتژی – داده کاوی – الگوریتم ژنتیک

### ۱- مقدمه

از زمانهای گذشته که استراتژی را بعنوان "هنر برنامه ریزی و هدایت تحرکات نظامی و عملیات جنگی" تعریف میکرده اند تا به امروز که استراتژی را "یک الگوی حرکتی برای نیل به وضعیت مطلوب سازمان" میدانند، کمتر کسی در اصل لزوم مدیریت استراتژیک برای یک فعالیت گروهی هدفمند شک داشته است.

موفقیت مدیریت استراتژیک بستگی به استراتژی بی دارد که بر آن مبتنی است. امروزه مدیریت استراتژیک بر پایه استراتژیهای بنا میشود که جامع نگ بوده، بر مسائل و وقایع پیامونی احاطه داشته باشد و بطور منظم (در مقاطع زمانی مشخص) و نامنظم (به هنگام بروز تغییرات عمده) پیوسته بازنگری شود.

بحث اصلاح پذیری و بازنگری مذاوم استراتژیها در مفهوم "استراتژیهای غلتان" تبلور می یابد.

لحاظ کردن اهم عوامل تاثیرگذار بر سازمان و روابط متقابل بین آنها را شاید بتوان اصلی ترین شاخص یک استراتژی اثربخش دانست.

محیط به شدت رقابتی دنیا امروز، پیشرفت روزافزون تکنولوژی، هم پوشانی دامنه مسائل اقتصادی، سیاسی و ...، علاج باعث شده عوامل تاثیرگذار بر سازمانها برآیندی از پارامترهای بسیار گسترده باشند که ادراک یک استراتژیست (ولو مبتنی بر ابزارهای تجزیه و تحلیل مرسوم) نمی تواند تضمینی برای لحاظ شدن مهمترین آنها در استراتژی سازمان باشد.

<sup>۱</sup> فوق لیسانس مدیریت اجرایی (MBA) – دانشگاه علوم و فنون مازندران

<sup>۲</sup> فوق لیسانس هوش مصنوعی – دانشگاه آزاد اسلامی، واحد علوم و تحقیقات

چالش بزرگ بر سر راه بررسی عوامل موثر بر سازمان، پیدا کردن الگوهایی از میان انبوه داده‌های محیطی میباشد که در راستای فعالیت سازمان قابل تحقیب بوده و بتوانند بصورت معنی دار بعنوان قوت، ضعف، تهدید و یا فرصت (بالقوه یا بالفعل) برای سازمان معرفی شوند.

چالش دیگر پیش روی استراتژیستها این است که شناسایی و انتخاب عناصر مهم از میان انبوه داده‌های محیطی که در یک محیط پویا، بصورت آنی در حال تغییر و تحول هستند و انتخاب تاثیرگذارترین و مهمترین این عوامل و همچنین چگونگی وزن دهی (تعیین درجه اهمیت) هر یک از عناصر، در آنرا موقع یک مقوله ادراکی و شهودی بوده و حتی با ابزارهای تکنولوژیکی و محاسباتی مرسوم نیز نمیتوان از بررسی همه جانبه عوامل موثر و نیز اعتبار وزن دهنی انجام شده به هریک از عوامل انتخابی مطمئن بود<sup>۷</sup>.

در این مقاله سعی خواهد شد با معرفی ابزاری بنام داده کاوی و ارائه روشها و تکنیکهای مرسوم در این ابزار، راهکاری برای اثربخشی کردن بحث بررسی عوامل تاثیرگذار بر سازمان و در نتیجه، افزایش اثربخشی برنامه ریزی و مدیریت استراتژیک ارائه شود.

ادامه مقاله بصورت زیر سازماندهی شده است: در بخش دوم به معرفی مدیریت استراتژیک و فرآیند مربوطه پرداخته شده، در بخش سوم، داده کاوی و روشها و تکنیکهای آن مورد بررسی قرار گرفته است، بخش چهارم به ارائه چارچوب پیشنهادی اختصاص یافته، در بخش پنجم چارچوب معرفی شده مورد ارزیابی قرار گرفته و در بخش ششم نیز نتیجه گیری و راهکارهای پیشنهادی برای محققین دیگر ارائه گشته است.

## ۲- مدیریت استراتژیک چیست؟

یک سازمان با شناخت موقعیت فعلی خود، تعریف وضعیت مطلوب (آرمانی) و تبیین راستای حرکت برای دستیابی به آن اهداف، در حقیقت یک الگوی حرکتی خواهد داشت که آنرا "استراتژی سازمان" می‌نامیم.

استراتژی، یک تصمیم‌گیری و سیاست گذاری مرحله مند برای رسیدن از وضع موجود به وضع مطلوب میباشد. مدیریت استراتژیک نیز علم و هنر تدوین، اجرا و ارزیابی تصمیمات وظیفه ای چندگانه میباشد که سازمان را در نیل به اهداف آرمانی خود توانا می‌سازد. فرآیند مدیریت استراتژیک بر استوار است و تدوین یک استراتژی نیز بر تجزیه و تحلیل استوار میباشد. با توجه به اینکه محیط هر سازمان در دنیای واقع متاثر از عوامل بیشمار داخلی و خارجی میباشد که هر کدام میتواند برای سازمان یک تهدید و یا یک فرصت بالقوه باشند، مدیریت استراتژیک موفق باید بر پایه استراتژیهایی بنا شود که با تجزیه و تحلیل کردن قوتها و ضعفهای سازمان و همچنین فرصت‌ها و تهدیدهای محیطی، اولاً اهداف سازمان را واقع گرایانه تدوین کند، ثانیا راستای حرکت به سمت وضعیت آرمانی را بصورت بهمنه تدوین نماید.

شاید پربرهای نباشد اگر بگوییم بحث بررسی عوامل داخلی و خارجی تاثیرگذار بر سازمان، یکی از کلیدی‌ترین مراحل تدوین استراتژی و در نتیجه، نقطه حساس مدیریت استراتژیک میباشد.

بررسی این عوامل با ابزارهایی از قبیل ماتریس بررسی رقابت<sup>۸</sup> (برای بررسی نقاط قوت و ضعف رقبای سازمان نسبت به موضع استراتژیک سازمان)، ماتریس ارزیابی عوامل خارجی<sup>۹</sup> (برای بررسی عوامل تاثیرگذار خارجی از قبیل عوامل اقتصادی، سیاسی، اجتماعی، فرهنگی و ...) و ماتریس ارزیابی عوامل داخلی<sup>۱۰</sup> (برای بررسی استراتژیک عوامل داخل سازمانی از قبیل مدیریت، بازاریابی، امور مالی، تولید و ...) قابل انجام میباشد. برای شناخت مواضع و موضوعات حساس و بحافی سازمان نیز ابزارهایی از قبیل تجزیه و تحلیل شکاف<sup>۱۱</sup> (برای اندازه گیری وسعت شکاف موجود بین عملکرد گذشته و حال سازمان با عملکرد موردنظر)، ماتریس تحلیل بدره<sup>۱۲</sup> (برای شناسایی حوزه‌هایی که تاثیر آنها روی نرخ سودآوری یا رشد سازمان بیشتر است) و ... استفاده میشود که تشریح این ابزارها از بحث ما خارج است.

<sup>۷</sup> CPM (Competitive Profile Matrix)

<sup>۸</sup> EFE (External Factor Evaluation)

<sup>۹</sup> IFE (Internal Factor Evaluation)

<sup>۱۰</sup> Gap Analysis

<sup>۱۱</sup> Portfolio Analysis

یک استراتژیست با استفاده از ابزارهای مقایسه گر مانند ماتریس تهدیدها، فرصتها، نقاط ضعف و قدرت<sup>۸</sup>، ماتریس ارزیابی موقعیت استراتژیک<sup>۹</sup>، ماتریس گروه مشاوران بوسون<sup>۱۰</sup> و ... اقدام به تصمیم گیری و تدوین استراتژی می نماید.

استراتژی تدوین شده پس از اولویت بندی فعالیتها عملیاتی شده و به اجرا گذاشته میشود آخرين مرحله از فرآيند مدیریت استراتژیک نيز بررسی میزان اثربخشی تصمیمات اخذ شده می باشد که مرحله ارزیابی استراتژی نامیده میشود.

تمرکز این مقاله بر مرحله تدوین استراتژی میباشد.

### ۳- داده کاوی<sup>۱۱</sup> چیست؟

داده کاوی ابزاری است برای پیدا کردن مجموعه ای از داشن های جالب نظر قوانین ، نظم ها و الگوها از درون داده های موجود. به بیان دیگر داده کاوی را می توان استخراج اطلاعات و الگوهایی از داده های بالقوه مفید و کارا که قبل از شناخته بوده و بصورت ضمنی در بین داده ها موجود بوده است تعریف نمود.

داده کاوی یعنوان یک مفسر انبوی داده ها را مورد بررسی قرار داده و سیل عظیم و بی قاعده اطلاعات را به داشن قابل تفسیر تبدیل میکند. در نتیجه لایه های پنهان اطلاعات پیشگویانه<sup>۱۲</sup> از میان توده داده ها قابل رویت خواهند شد و الگوهای معنادار مشخص خواهند شد.

داده کاوی یکی از مهمترین مراحل پروسه استخراج داشن<sup>۱۳</sup> از مجموع داده های موجود بشمار می آید. برای داده کاوی نامهای مختلفی همچون اکتشاف داشن<sup>۱۴</sup> ، آنالیز الگو و داده<sup>۱۵</sup> ، باستان شناسی داده ها<sup>۱۶</sup> و ... نیز ارائه شده است.

تا امروز کاربردهای بسیار مختلفی برای داده کاوی پیشنهاد شده است [۷] که میتوان مختصرا به این موارد اشاره نمود:

- مدیریت و آنالیز بازار: تشخیص الگوهای خرید مشتریان در زمانهای مختلف، طبقه بندی کردن مشتریان بر اساس یک سری مشخصات ( سطح درآمد ، عادات و... )، خلاصه اطلاعات آماری بر اساس زمان و مکان.
- مدیریت و آنالیز ریسک: پیش بینی و آنالیز جریان مالی ، آنالیز نرخ رشد سرمایه و سری های زمانی و ...
- تقسیم بندی بازار: تعیین و تقسیم دقیق تر بازار هدف .
- یافتن خصوصیات مشترک بین مخاطبین بالفعل محصول یا خدمات سازمان: یافتن خصوصیاتی از مشتریان بالفعل که بتواند یعنوان یک عامل پیش گویانه مشتریان بالقوه را تشخیص دهد.
- تعیین راهکار فروش ( یا ارائه محصور ) : امکان ارائه استراتژیهای خاص و منطقه ای برای عملکرد بهتر در هر قسمت از بازار .

...

### ۱-۳- روش های داده کاوی

به طور کلی روش های داده کاوی [۷] را می توان به صورت زیر تقسیم بندی نمود:

<sup>۸</sup> SWOT

<sup>۹</sup> Space

<sup>۱۰</sup> BCG

<sup>۱۱</sup> Data Mining

<sup>۱۲</sup> Predictive Information

<sup>۱۳</sup> Knowledge Extraction

<sup>۱۴</sup> Knowledge Discovery

<sup>۱۵</sup> Pattern and Data analysis

<sup>۱۶</sup> Data Archeology

## - طبقه بندی<sup>۱۷</sup> :

هدف از طبقه بندی مشخص نمودن ویژگی هایی است که بتوان توسط آن، گروههای مختلف را شناسایی کرده و از یکدیگر تشخیص داد. طبقه بندی در داده کاوی در دو مرحله انجام می گیرد. ابتدا از روی داده های قدیمی، گروههای مختلف تشخیص داده شده (مرحله اول) و سپس تعلق داشتن داده های جدید به گروههای موجود پیش بینی می گردد (مرحله دوم). این روش جزء روشهای یادگیری با ناظر<sup>۱۸</sup> و پیش بینی کننده<sup>۱۹</sup> بشمار می آید.

## - خوشه بندی<sup>۲۰</sup> :

به پروسه دسته بندی داده ها و نسبت دادن آن به گروههایی از داده های مشابه خوشه بندی می گویند. این روش با تفکر تقسیم و حل<sup>۲۱</sup> به دسته بندی داده ها پرداخته و در نتیجه، مجموعه داده اولیه به دسته های کوچکتر تقسیم می گردد. خوشه بندی جزو روشهای یادگیری بدون ناظر و تشریح کننده<sup>۲۲</sup> به حساب می آید.

یک خوشه بندی را مناسب گویند هر گاه اشیاء داده ای درون هر خوشه به یکدیگر بسیار شبیه و با اشیاء خوشه های دیگر بسیار متفاوت باشند مثلاً شباهت و تفاوت بین اشیاء داده ای توسط یکتابع فاصله مشخص میگردد. بسته به نوع داده، توابع فاصله متفاوتی موجود می باشند که از آن جمله میتوان به تابع فاصله Manhattan<sup>۲۳</sup>، تابع فاصله Minkowski<sup>۲۴</sup> و تابع فاصله اقلیدسی اشاره نمود. همچنین یک خوشه بندی را مناسب گویند چنانچه بتواند کلیه یا تعداد زیادی از الگوهای پنهان در مجموعه داده را کشف کند.

## - کاوش قوانین داده های بهم وابسته<sup>۲۵</sup> :

یکی دیگر از روش های داده کاوی، کاوش قوانین برای کشف وابستگیها و همبستگی های موجود در بین داده ها می باشد. این روش به یافتن الگوهای پنهان داده ها و همچنین پیدا کردن ساختارهای سببی آیتم ها و داده های موجود می پردازد. برای جواب دادن به سوالاتی نظیر چه محصولاتی اغلب با هم خریداری می شوند؟ بعد از خرید محصول A معمولاً چه چیز دیگری خریداری می شود؟ و.. نیاز به کاوش در مجموعه داده و کشف همبستگی های موجود بین آنها داریم.

## - خلاصه سازی و کلی نگری داده ها در سطوح مختلف<sup>۲۶</sup> :

بطور کلی اغلب داده های موجود دارای جزئیات فراوانی می باشند. برای مثال در مجموعه داده های مربوط به قسمت فروش، جدول اطلاعات کالا شامل فیلد های اطلاعاتی نظری شماره کالا، نام کالا، سال ساخت، قیمت و.. می باشد. جزئیات زیاد سبب پائین آمدن سطح ادراک می گردد و برای تصمیم گیری بر اساس اطلاعات قبلی نیاز به سطح ادراکی بالاتری می باشد. داده کاوی با انجام خلاصه سازی و کلی نگری در داده ها در سطوح مختلف به کمک شناخته و با دسته بندی داده ها، سطح ادراکی بالاتری را ایجاد می نماید.

<sup>۱۷</sup> Classification

<sup>۱۸</sup> Supervised Learning

<sup>۱۹</sup> تکنیکهای پیش بینی کننده : این تکنیکها با ساخت مدلی برای داده های موجود، پیش بینی ناشناخته ها را بر عهده دارند. مثلاً پیدا کردن یک الگوریتم قابل تکرار و یا خصوصیات مشترک بین مشتریان فعلی خدمات یا محصولات سازمان و سپس ایجاد الگویی برای طبقه بندی مخاطبان بالقوه.

<sup>۲۰</sup> Clustering

<sup>۲۱</sup> Divide and Conquer

<sup>۲۲</sup> تکنیکهای تشریح کننده : این تکنیک ها وظیفه گروه بندی بین داده های موجود و در نتیجه تقسیم بندی داده ها به دسته های متمایز را بر عهده دارند.

<sup>۲۳</sup> Association Rule Mining

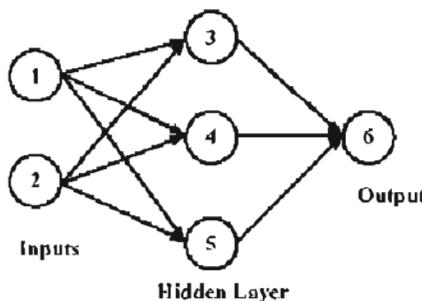
<sup>۲۴</sup> Summarization and Globalization of Data

### ۳-۴- تکنیک های داده کاوی :

از معروفترین تکنیک های داده کاوی [۶] می توان به موارد زیر اشاره نمود.

#### - شبکه های عصبی<sup>۲۵</sup> :

شبکه های عصبی با توجه به کارایی بالای آن در یادگیری داده های آموزشی (داده های اولیه) در سالهای اخیر بسیار مورد توجه دانشمندان و محققین قرار گرفته است. در این راهکار که بر اساس سیستم عصبی بدن انسان تعریف شده است، شبکه عصبی سعی در ایجاد نگاشتهای مناسب بین الگوهای ورودی و خروجی دارد. از مزایای شبکه عصبی می توان به دقت پیش‌بینی بالا و توان مقاومت در برابر خطا های داده های آموزشی اشاره نمود از معایب آن نیز می توان به زمان یادگیری طولانی و مشکل بودن فهم تابع یادگرفته شده توسط شبکه اشاره کرد. در شکل (۱) نمای کلی یک شبکه عصبی نمایش داده شده است.



شکل (۱): نمای کلی یک شبکه عصبی

نحوه فعالیت شبکه عصبی به این صورت است که تاثیر وزن هر یک از سیناپس ها (مسیرهای بین نرونها) بر روی نرون های ورودی، در نرون های لایه بعدی نمود می یابد و این روند تا رسیدن به نرون خروجی ادامه حواهد یافت<sup>[۷]</sup>. نحوه آموزش در این شبکه به این صورت است که ابتدا یک سری وزن های تصادفی به شبکه داده می شود، پس از آن از نرون های ورودی شروع کرده و به خروجی می رسمیم، سپس بر می گردیم و خطای کشف شده در خروجی را در وزن های شبکه بصورت عقب گرد تاثیر می دهیم. این پروسه تا رسیدن به یک شبکه ایده آل که بتواند کلیه داده های آزمایشی را کمتر از یک حد مطلوب از میزان خطای در نظر گرفته شده تخمین بزند ادامه می یابد.

#### - رگرسیون<sup>۲۶</sup> :

رگرسیون یکی از روش های آماری برای طبقه بندی شمار می آید. هدف از رگرسیون پیش بینی مقدار یک متغیر پیوسته بر اساس مقادیر متغیرهای دیگر می باشد. برای مثال پیش بینی میزان فروش یک محصول جدید بر اساس میزان تبلیغات صورت گرفته بر روی آنرا می توان از روش رگرسیون انجام داد.

رگرسیون به دو دسته خطی و غیر خطی تقسیم می شود. از آنجا که بسیاری از مسائل دارای پیچیدگی های بسیار می باشند و متغیرهای مختلفی در آن دخیل هستند، نمی توان فقط با استفاده از رگرسیون خطی به پیش بینی پرداخت و نیاز به رگرسیون های پیچیده تری نظیر رگرسیون درخواست می باشد. رگرسیون های غیر خطی به وفور برای مسائل طبقه بندی مورد استفاده قرار می گیرند.

<sup>۲۵</sup> Neural Network

<sup>۲۶</sup> Regression

## - الگوریتم های زنگنه<sup>۳۴</sup> :

در میان روش‌های یادگیری ماشین الهام گرفته از طبیعت جاندار، الگوریتم زنگنه از تکامل یافته ترین آنها به شمار می‌رود. الگوریتم زنگنه یک روش بهینه سازی عددی است که بر اساس اصول داروین و با الهام از انتخاب طبیعی و زنگنه طبیعی کار می‌کند. این روش یکی از روش‌های موجود در مجموعه الگوریتم‌های تکاملی<sup>۳۵</sup> می‌باشد.

برای حل هر مسئله با استفاده از الگوریتم زنگنه ابتدا لازم است با توجه به پارامترهای مسئله یک کروموزوم تشکیل گردد (یک کروموزوم از یکسری ژن تشکیل می‌شود که در اینجا هر ژن، یکی از پارامترهای مسئله می‌باشد) سپس یک جمعیت از راه حل‌های تصادفی (کروموزوم‌ها) تولید می‌گردد. در مراحل بعد هدف، یافتن بهترین کروموزوم در این مجموعه می‌باشد.

در الگوریتم زنگنه عملگر انتخاب، همبستگی<sup>۳۶</sup> و جهش<sup>۳۷</sup> به عنوان مهمترین و اصلی ترین عملگرها جهت جستجو و استخراج به کار می‌روند.

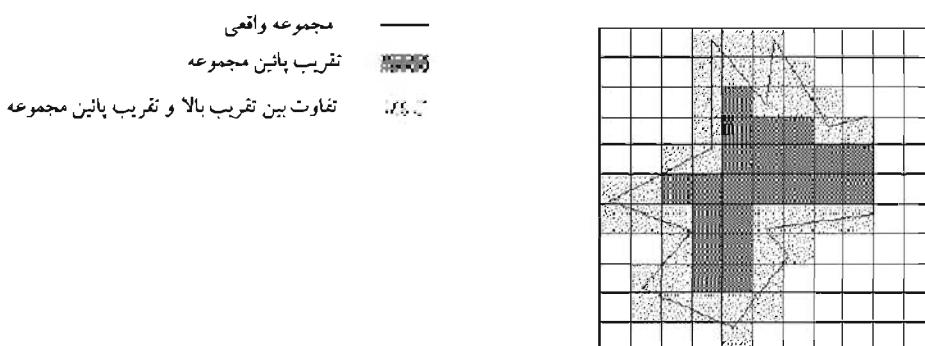
عملگر انتخاب: این عملگر بمنظور انتخاب بهترین کروموزوم‌ها برای انتقال آنها به نسل بعد مورد استفاده قرار می‌گیرد.

عملگر همبستگی: در این عملگر ابتدا دو کروموزوم به صورت تصادفی انتخاب می‌گردد. سپس یک نقطه قطع بصورت تصادفی مشخص می‌گردد و در نهایت قسمتهای هر دو کروموزوم از نقطه قطع با یکدیگر تعویض می‌شوند. از این عملگر به منظور جستجوی بهتر استفاده می‌شود.

عملگر جهش: این عملگر نیز همچون همبستگی، عملگری تصادفی است ابتدا کروموزومی به تصادف انتخاب می‌شود، سپس از کروموزوم انتخاب شده بیتی بصورت تصادفی انتخاب می‌گردد و مقدار بیت انتخاب شده بر عکس می‌شود، یعنی اگر مقدار این بیت (۰) باشد این بیت تبدیل به (۱) و اگر (۱) باشد تبدیل به (۰) می‌شود.

## - تئوری مجموعه های دانه درشت<sup>۳۸</sup> :

مجموعه های دانه درشت عنوان یک ابزار محاسباتی مهم برای مدیریت نمودن عدم قطعیت به حساب می‌آید. از مجموعه های دانه درشت در پروسه اکتشاف دانش و داده کاوی استفاده می‌شود. ایده اصلی در این روش عبارت از کشف ناسازگاریها و وابستگیها بین پارامترهای هر یک از داده‌ها می‌باشد. در شکل (۲) شمای کلی مجموعه های دانه درشت نشان داده شده است.



شکل (۲): شمای کلی مجموعه های دانه درشت.

<sup>۳۶</sup> Genetic Algorithm

<sup>۳۷</sup> Evolutionary Algorithm

<sup>۳۸</sup> Crossover

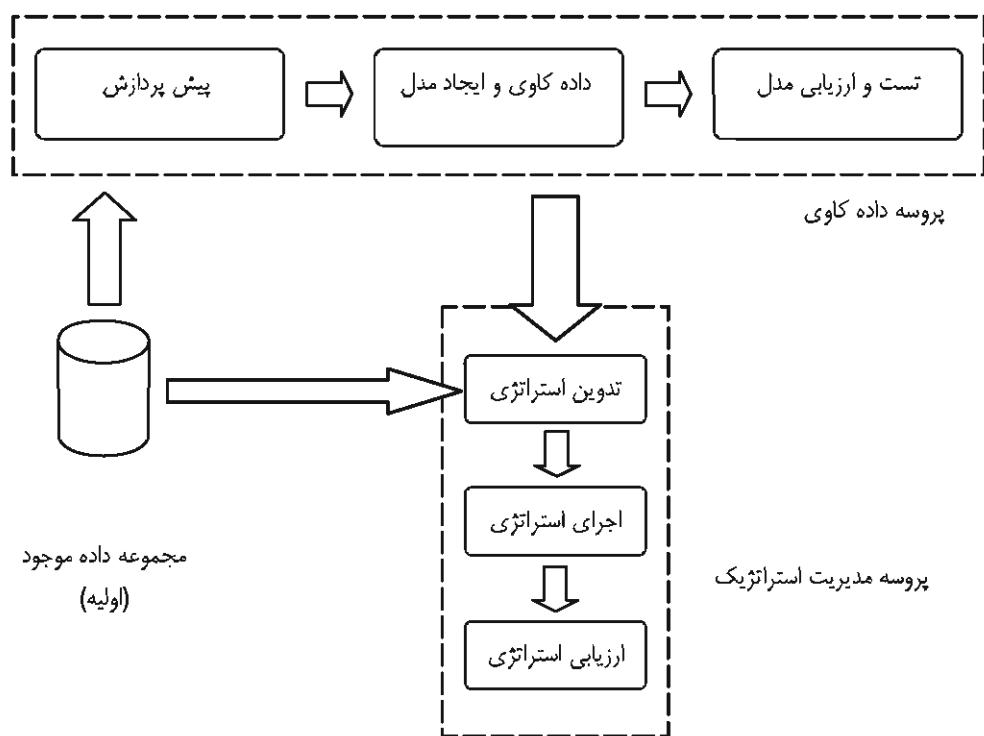
<sup>۳۹</sup> Mutation

<sup>۴۰</sup> Rough Set Theory

همانطور که مشاهده می شود در این روش، یک تقریب بالا و یک تقریب پائین برای مجموعه در نظر گرفته می شود که این تقریب بالا و پائین توسط مجموعه ای از دانه ها ایجاد شده است. مجموعه های دانه درشت یک راه مناسب برای بدست آوردن قوانین آنگاه از مجموعه داده می باشد. با استفاده از این روش، داده ها درون جداول تصمیم طبقه بندی شده و به یکسری دانش تبدیل می شوند.

#### -۴- ارائه چارچوب پیشنهادی

چارچوب پیشنهادی در شکل (۳) نمایش داده شده است.



شکل (۳) : چارچوب پیشنهادی

#### - پیش پردازش:

پس از جمع آوری داده ها از منابع داده ای مختلف ، لازم است یک سری پیش پردازش ها بر روی داده ها انجام گیرد . انجام پیش پردازش بر روی داده ها باعث می شود تا الگوریتم های داده کاوی با مشکلات کمتری هنگام برخورد با داده ها روبرو شوند . از پیش پردازش های مهم که معمولاً بر روی داده ها انجام می گیرد می توان به موارد زیر اشاره نمود :

الف / گستته سازی داده ها : برخی الگوریتم های داده کاوی فقط با مقادیر داده ای گستته کار می کنند لذا لازم است که مقادیر متغیرهای با بازه پیوسته به مقادیر گستته تبدیل گردد. برای این کار از الگوریتم های گستته سازی که یکی از مهمترین آنها ChiMerge می باشد ، استفاده می شود.

ب) سازگار نمودن مقادیر: چنانچه داده ها از چندین منبع داده ای جمع آوری شوند، در بعضی اوقات لازم است که مقادیر یک سری فیلدهای مشترک بصورت سازگار تعریف گردد برای مثال چنانچه فیلد اطلاعاتی جنسیت در دو منبع داده ای مختلف یکی بصورت (۱) و (۰) و دیگری بصورت  $M$  و  $F$  پر شده باشد لازم است پس از جمع آوری داده ها یکی از آنها به دیگری تبدیل شود ((۰) و (۱) به  $M$  و  $F$  یا برعکس).

ج) انتخاب مشخصه های مهم؛ زمانیکه حجم داده های جمع آوری شده بسیار زیاد باشد می توان با اعمال الگوریتم های انتخاب مشخصه<sup>۳۳</sup>، مهمترین و تأثیرگذارترین مشخصه ها را از بین داده های انتخاب نمود و بقیه مشخصه های زاید را از مجموعه داده حذف کرد با این کار حجم داده ها به صورت فوق العاده ای کاهش می یابد و در عین حال از ارزش داده ها چیز زیادی کم نمی شود، انتخاب مشخصه های مهم همچنین می تواند دقیق الگوریتم های داده کاوی را افزایش دهد.

#### - ایجاد مدل داده کاوی :

پس از اعمال یک سری پیش پردازش ها بر روی داده ها، یکی از روش های معرفی شده داده کاوی (با توجه به هدف داده کاوی) انتخاب می گردد سپس از بین الگوریتم های آن روش یک الگوریتم مناسب با توجه به نوع داده های انتخاب می شود و در نهایت الگوریتم بر روی داده های اعمال می گردد.

#### - تست و ارزیابی مدل :

نتایج حاصل از داده کاوی در مرحله قبل زمانی می تواند دانش تلقی گردد که بصورت شهودی با ادراک انسانی مورد قبول قرار گیرد در این مرحله نتایج مورد ارزیابی قرار می گیرد چنانچه نتایج قابل قبول نبود پروسه از یکی از مراحل قبل تکرار می گردد . تکرار پروسه داده کاوی و تغییر دادن پارامترهای مدل برای ارتباط بهتر با داده های می تواند در رسیدن به موقیت و کشف دانش مفید باشد

#### - فرآیند مدیریت استراتژیک:

نتایج حاصل از مراحل قبل، بعنوان عوامل موثر بر سازمان (عوامل داخلی، خارجی و یا محیط رقابت) میتوانند جهت تدوین یک استراتژی موثر در اختیار استراتژیست قرار گیرند .

#### ۵- ارزیابی چارچوب پیشنهادی

در اینجا، از اطلاعات دو شرکت تولید کننده فرضی  $A$  و  $B$ <sup>۳۴</sup> برای ارزیابی مدل پیشنهاد شده استفاده میشود مجموعه داده فرضی از طریق پرسشنامه جمع آوری شده اند که در شکل (۳) قسمتی از پرسشنامه مذکور نشان داده شده است.

<sup>۳۳</sup> Feature Selection Algorithms

<sup>۳۴</sup> مدل ارزیابی شده واقعی، مبتنی بر اطلاعات اخذ شده در مورد دو شرکت تولید کننده مواد غذایی بوده که به دلیل نمونه آماری کوچک، عدم امکان آمارگیری علمی و نیز عدم امکان پردازش پیشرفته داده ها (به واسطه استفاده از برنامه داده کاوی آزمایشگاهی و غیرحرفه ای) ترجیح داده شد اطلاعات اولیه و نتایج اخذ شده به دو شرکت فرضی  $A$  و  $B$  نسبت داده شوند

مشخصات فردی:									
جنسیت :	<input type="checkbox"/> مرد	<input type="checkbox"/> زن							
وضعیت تأهل :	<input type="checkbox"/> مجرد	<input type="checkbox"/> متأهل							
قد :	ووزن:								
تعداد افراد خانوارده:									
درآمد ماهانه: کمتر از <input type="checkbox"/> ۲۰۰ هزار تومان <input type="checkbox"/> بین ۲۰۰ هزار تا <input type="checkbox"/> ۳۰۰ هزار تومان <input type="checkbox"/> بین <input type="checkbox"/> ۳۰۰ هزار تا <input type="checkbox"/> ۵۰۰ هزار تومان <input type="checkbox"/> بینتر از <input type="checkbox"/> ۵۰۰ هزار تومان <input type="checkbox"/>									
تحصیلات: زیر دبیل <input type="checkbox"/> دبیل <input type="checkbox"/> غریق دبیل <input type="checkbox"/> لیسانس <input type="checkbox"/> بالاتر از لیسانس <input type="checkbox"/>									
سوالات:									
۱- آیا تبلیغات شرکت A بیشتر از تبلیغات شرکت B می باشد؟ <input type="checkbox"/>									
۲- تبلیغات شرکت A در کدام رسانه نظر شما را جلب کرده است؟ <input type="checkbox"/> تلویزیون <input type="checkbox"/> رادیو <input type="checkbox"/> اینترنت <input type="checkbox"/> روزنامه									
۳- گفایت محصول کدام شرکت A در مقایسه با محصول شرکت B در چه سطحی قرار دارد؟ <input type="checkbox"/> بهتر از B <input type="checkbox"/> بینتر از A <input type="checkbox"/> همچنیدان									
۴- قیمت محصول کدام شرکت به نسبت کفایت آن مناسب است؟ <input type="checkbox"/> B <input type="checkbox"/> A <input type="checkbox"/> همچنیدان									
۵- گفایت کدام محصول عرف ظاهر از قیمت آن مناسب است؟ <input type="checkbox"/> B <input type="checkbox"/> A <input type="checkbox"/> همچنیدان									
۶- چنانچه به دلایل سرویس دهنده بهتر به مشتریان و بالاترین کیفیت، قیمت محصول بالاتر رود، تا پنده توجه، تناول غذت حوز نسایل به خرید محصول از شرکت A می باشد؟ <input type="checkbox"/> ۰ درصد <input type="checkbox"/> کمتر از ۱۰ درصد <input type="checkbox"/> کمتر از ۲۰ درصد									
۷- پس از جمع آوری داده ها، یک سری پیش پردازش های اولیه برای آماده سازی داده ها به شرح ذیل انجام می گیرد									

شکل (۴): قسمتی از پرسشنامه جمع آوری شده

این پرسشنامه توسط یک جامعه آماری ۰۰۰ انفرادی پاسخ داده شد (تاكید میگردد که در این مقاله، هدف انجام یک آمارگیری علمی و بدست آوردن نتایج معتبر نیست و این داده ها صرفا برای انجام پروسه های مربوطه و دستیابی به شکل قوانین ذروجی جمع آوری شده اند).

پس از جمع آوری داده ها، یک سری پیش پردازش های اولیه برای آماده سازی داده ها به شرح ذیل انجام می گیرد

(الف) گسته سازی داده ها: مقادیر متغیرهای قد و وزن و ... که بصورت پیوسته می باشند، گسته می گردد و به بازه های کوچکتری تقسیم بندی می شود. برای مثال برای گسته سازی متغیر وزن، چهار بازه کمتر از ۴۰ کیلوگرم، بین ۴۰ تا ۷۵ کیلوگرم، از ۷۵ تا ۹۰ کیلوگرم و بالای ۹۰ کیلوگرم در نظر گرفته می شود

(ب) سازگار نمودن داده ها: از آنجا که الگوریتم داده کاوی استفاده شده فقط با اعداد کار می کند لذا لازم است مقادیر حرفی بصورت عددی درآیند. بدین منظور برای متغیر جنسیت به جای مرد و زن از (۱) و (۰) و برای متغیر سن به جای کمتر از ۱۷، بین ۱۷ تا ۲۵، بین ۲۵ تا ۳۵ و بالاتر از ۳۵ برتریب (۰)، (۱)، (۲) و (۳) در نظر گرفته می شود. برای متغیر وزن نیز بجای چهار بازه ذکر شده در بالا، برتریب (۰)، (۱)، (۲) و (۳) استفاده می شود و بهمین ترتیب برای دیگر متغیرها. به این صورت کلیه متغیرها دارای مقادیر عددی می شوند.

پس از انجام پیش پردازش، داده ها برای انجام پروسه داده کاوی آماده می باشند. جدول (۱) نشان دهنده قسمتی از مجموعه داده پس از انجام پیش پردازش می باشد.

جدول (۱): قسمتی از مجموعه داده پس از پیش پردازش های لازم

ردیف	جنسیت	سن	وضعیت تأهل	قد	-	تبلیغات A بیشتر از B	رسانه تبلیغات	کیفیت محصول نسبت به قیمت	B	تبلیغات B نسبت به A	کیفیت A	رسانه تبلیغات	کیفیت B	ردیف
۱	۱	۱	۱	۱	۱	۱	۱	۱	۱	۱	۱	۱	۱	۱
۲	۰	۰	۰	۰	۰	۰	۰	۰	۰	۰	۰	۰	۰	۲
۳	۰	۱	۱	۱	۱	۱	۱	۱	۱	۱	۱	۱	۱	۳
۴	۰	۲	۲	۲	۲	۲	۲	۲	۲	۲	۲	۲	۲	۴
۵	۰	۳	۳	۳	۳	۳	۳	۳	۳	۳	۳	۳	۳	۵
۶	۰	۰	۰	۰	۰	۰	۰	۰	۰	۰	۰	۰	۰	۰
۷	۰	۰	۰	۰	۰	۰	۰	۰	۰	۰	۰	۰	۰	۰

در این مقاله بمنظور داده کاوی از روش "الگوریتم زنتیک" استفاده شده است [۴۹] . با کمک الگوریتم زنتیک میتوان یک مجموعه قوانین را تولید نمود که قوانین تولید شده مثلا میتواند به فرمت زیر باشد:

*if ( $x_1 = 1$ ) and (( $x_1 = 1$ ) or ( $x_1 = 2$ )) then    ( $x_{..} = 1$ )*

متغیر سن:  $x_1$

متغیر تحصیلات:  $x_2$

متغیر اولویت جلب نظر کننده بودن تبلیغات رسانه ای:  $x_{..}$

(تعییر قانون فوق در مجموعه داده فوق الذکر این است که اگر مخاطب در محدوده سنی بین ۲۵ تا ۳۵ سال باشد و تحصیلاتش لیسانس یا بالاتر از لیسانس باشد، آنگاه تبلیغات شرکت **A** در روزنامه بیشتر از بقیه رسانه ها نظرش را جلب میکند)

برای بدست آوردن قوانینی به این شکل، نیاز به ساختار کرومزمومی به صورت شکل (۵) داریم:

زنهای تشکیل دهنده کرومزموم	{...}	{۰,۱}	{۰,۱}	{۰,۱}	{۰,۱}	{۰,۱}	{۰,۱}	{۰,۱}	{۰,۱}	...	{۰,۱}	
متغیرها	$x_1$		$x_2$							...	$x_{..}$	مقدار $x$
قانون												قسمت آنگاه

شکل (۵): ساختار کرومزموم پیشنهادی

(متغیرهای نظیر جنسیت، سن و ... با  $x$  نشان داده شده اند)

اگر قانون فوق الذکر را نتیجه اجرای الگوریتم ژنتیک بر روی مجموعه داده بدانیم، کرومزموم حاصله (که قانون از روی آن تعییر شده است) به صورت شکل (۶) میباشد:

۰	۰	۰	۱	۱	...	۰	۰	۱	۱	...	۱۰	۳
$x_1$		$x_2$			...	$x_8$					$x_{..}$	مقدار $x$
												شماره $x$

شکل (۶): کرومزموم نمونه برای قانون ذکر شده در بالا

با اجرای الگوریتم ژنتیک، بدنبال بدست آوردن کرومزمومهایی از این قبیل و در نتیجه ایجاد قوانینی مانند قانون فوق الذکر هستیم، برای الگوریتم ژنتیک دو پارامتر "درصد صحت" و "درصد پوشش" بصورت زیر تعریف میشود:

- درصد صحت یک قانون را **a** گویند درصورتیکه اگر قسمت "اگر" قانون درست بود، قسمت "آنگاه" قانون در **a**٪ موقع درست باشد.
- درصد پوشش یک قانون را **b** گویند هرگاه قسمت "اگر" قانون در **b**٪ از کل مجموعه داده صحیح باشد

پارامترهای الگوریتم در این مقاله بصورت زیر مقداردهی شده است:

- درصد صحت: حداقل ۹۰٪ - درصد پوشش: حداقل ۱۰٪
- نرخ همبزی: ۰/۷ - نرخ جهش: ۰/۰۳

پس از پیاده سازی **۳** و اعمال الگوریتم ژنتیک بر روی داده های مهیا شده، ابوهی از قوانین کشف شده اند که الزاماً کارا و مفید نیستند تفسیر تعدادی از این قوانین به شرح زیر میباشد:

- افرادی که دارای سطح درآمد کمتر از ۱۰۰ هزار تومان در ماه بودند و تعداد افراد خانواده شان بیشتر از ۳ نفر بودند، افزایش بیشتر از ۱۰٪ قیمت محصول را برای سرویس دهی بهتر غیر قابل قبول می دانند.
- خانمهای متاهل، کیفیت محصول تولید شرکت **B** را بالاتر از محصول شرکت **A** ارزیابی میکنند.

درصد صحت: ۹۲٪ درصد پوشش: ۳۴٪

درصد صحت: ۹۲٪ درصد پوشش: ۲۲٪

<sup>۲۴</sup> نسخه آزمایشگاهی برنامه فوق الذکر با استفاده از نرم افزار **Matlab** (وبرايشن <sup>۷</sup>) پیاده سازی گردیده است

- مجردهای با درآمد بین ۲۰۰ هزار تا ۳۰۰ هزار تومان در ماه، قیمت محصول شرکت A را به نسبت کیفیت آن، مناسبتر از محصول شرکت B میدانند.
- درصد صحت: %۹۷ درصد پوشش: %۲۴
- افراد با سطح تحصیلات فوق دبیلم، دبیلم و پائین تر از دبیلم، با شرکت A بیشتر از طریق تبلیغات تلویزیونی آشنا هستند.
- درصد صحت: %۹۰ درصد پوشش: %۶۳
- افراد ۱۷ تا ۲۵ سال جزء مخاطبین تبلیغات رادیویی شرکت A نیستند.
- درصد صحت: %۹۶ درصد پوشش: %۳۰
- آقایان با قد بین ۱۶۵ تا ۱۷۵ سانتی متر، قیمت محصول شرکت A را مناسبتر از قیمت محصول B میدانند.
- درصد صحت: %۹۲ درصد پوشش: %۲۴
- خانمهای مجردی که کیفیت محصول شرکت A را بهتر از محصول شرکت B میدانند و مخاطب تبلیغات روزنامه یی شرکت A میباشند، وزنشان از ۷۵ کیلوگرم بیشتر نیست.
- درصد صحت: %۹۵ درصد پوشش: %۱۱
- متاھلین با درآمد بیش از ۵۰۰ هزار تومان در ماه، کیفیت هیچکدام از محصولات تولیدی شرکت A و B را مناسب نمی دانند.
- درصد صحت: %۹۲ درصد پوشش: %۲۷
- ...

تعدادی (و نه همه) قوانین فوق الذکر، میتوانند تعبیر به فرست، تهدید، نقطه ضعف و یا نقطه قوت سازمان شوند ادراک انسانی میتواند با انتخاب اهم این قوانین و بررسی آنها با استفاده از ابزارها و ماتریس‌هایی که بیشتر ذکر شده اند و نیز با بررسی مستقیم مجموعه داده موجود، تجزیه و تحلیل عوامل تاثیر گذار بر سازمان را بهتر انجام داده و در نتیجه استراتژی مناسبتری تدوین نماید

## ۹- نتیجه گیری

مشاهده میشود که کاوش اطلاعات در دنیای امروز، بدون ابزارهای هوشمند و کارآمدی از قبیل داده کاوی و ... عملای نمیتواند تهدیدها و فرصتها را برای استراتژیست بخوبی نمایان ساخته و او را به تدوین استراتژی بسی کارآمد رهنمون سازد همانطور که پیشرفت روزافزون علم و دانش و همینطور تجربیات مدیران، آنها را به سمت استفاده از ابزارهای تحلیل اطلاعات (و حتی ابزارهایی برای تصمیم گیری) سوق داده است، لازم است تلاش‌هایی در جهت فیلتر کردن اطلاعات، دسته بندی، خلاصه سازی، کلی نگری و کشف الگوها و قوانین مستقر در اطلاعات اولیه نیز برای درک بهتر محیط انجام گردد.

آنچه در این مقاله ارائه شد، پیش درآمدی بود بر آنچه که میتواند عنوان یک واسطه بین مراحل "جمع آوری اطلاعات" و "پردازش اطلاعات" در تدوین استراتژی مطرح باشد.

پیشنهاد میشود در تحقیقات آینده بجای یک تکنیک داده کاوی، از ترکیب دو یا چند تکنیک برای نیل به نتایج بهتر استفاده گردد. با توجه به سرعت عملکرد پائین الگوریتم تکنیک، استفاده از روش‌های نوینی مانند "الگوریتم کلونی مورچگان"<sup>۲۵</sup> و ... توصیه میگردد همچنین، استفاده از مجموعه داده های واقعی تر و حجمی تر میتواند برای ارزیابی بهتر مدل پیشنهادی مفید باشد.

## ۷- منابع

- [۱] Andrew Pettigrew , Howard Tomas , Richard Whittington , "Handbook of Strategy and Management " , 2002 .
- [۲] William B. Rouse , "Essential Challenges of Strategic Management" , N.Y:John Wiley&sons , 2001 .
- [۳] A.R. Ramos Rodriguez , J. Ruiz Navarro , "Changes in the intellectual structure of strategic management Research : A bibliometric study of the strategic management Journal, ۱۹۸۰-۲۰۰۴" , Strategic Management Journal , Vol ۲۵ , No ۱۰ October 2004 .
- [۴] M. E. M. Pourzarandi, R.Shabahang, C.Lucas , "Designing and Developing an Appropriate Inventory Management Model Through the Use of Data Mining and Knowledge Discovery Technology the application of Neural Networks" , Economics Management , Quarterly Journal of the Islamic Azad University, Vol ۲۱ Summer 2004 .
- [۵] D. E. Goldberg , "Genetic algorithms in search, optimization, and machine learning" :Addison-Wesley, Reading, Mass, 1989.
- [۶] M. Kantardzic , "Data Mining Concepts , Models , Method and Algorithms" , Wiley-InterScience , 2003 .
- [۷] R.J. Roiger , M.W.Geatz , "Data Mining A Tutorial – Based Primer" : Addison Wesley 2003 .
- [۸] Han, Jiawei: Lamber , Micheline : "Data Mining concepts and Techniques" : Morgan Kaufmann publishers 2001 .
- [۹] "Matlab" Software Help, Section: Genetic Algorithm Toolbox.

[۱۰] "روش جدیدی برای داده کاوی مبتنی بر مجموعه های دانه درشت و پردازش تکاملی" ، عبدالرضا سالاری ، مهران محسن زاده ، بلک نصیری؛دهمین کنفرانس انجمن کامپیوتر ایران ، زمستان ۱۳۸۳ .