

## تخمین بهای تمام شده کالا با تکنیکهای داده کاوی Data mining for cost estimation

محمدرضا زارع میرکآبادی<sup>۱\*</sup>، فضل الله ادیب‌نیا<sup>۲</sup>

دانشکده مهندسی کامپیوتر

یزد، دانشگاه یزد

mzare@yazduni.ac.ir

### چکیده

در یک کارخانه تولیدی، داده‌های متنوعی در سطح پردازش تراکنش‌ها و به صورت روزمره جمع‌آوری و به سیستم وارد می‌شود. این اطلاعات جهت زمانبندی تولید، مدیریت خط تولید و موارد دیگر مورد استفاده قرار می‌گیرد. آنچه در این میان می‌تواند مفید واقع شود، استفاده از این داده‌های عملیاتی و استخراج دانش از آن می‌باشد. روشهای مختلفی جهت استخراج اطلاعات مورد نیاز در زمینه‌های برنامه‌ریزی تولید، مدیریت ریسک، کنترل فرآیند تولید و کاهش دورریز می‌تواند طراحی و پیاده‌سازی گردد. یکی از موارد مهم در این زمینه تخمین بهای تمام شده محصول تولیدی در کارخانجاتی است که براساس سفارش مشتری، محصول تولید می‌کنند. به کمک این سیستم می‌توان براساس پیش‌بینی مراحل و حجم عملیات تولید که بر اساس اطلاعات موجود در سیستم به دست می‌آید، بهای تمام شده را تخمین زد و بر اساس اصول دیگر حاکم بر قیمت‌گذاری، قیمت محصول نهایی را به مشتری اعلام نمود. این سیستم در بستر یک سیستم تحت web می‌تواند به صورت online قیمت را به مشتری نشان دهد.

آنچه در این مقاله به آن پرداخته می‌شود، استفاده از تکنیکهای داده‌کاوی جهت استخراج دانش و استفاده از اطلاعات موثر در تخمین بهای تمام شده و نهایتاً برآورد قیمت محصول می‌باشد. هدف این است که از اطلاعات موجود در یک سیستم پردازش تراکنش‌های تولید استفاده نمود و قواعدی را از آن استخراج کرد که به کمک آنها و قوانین داده شده دیگر، بهای محصول برای یک سفارش جدید، پیش‌بینی گردد. نهایتاً پیاده‌سازی یک نمونه از چنین سیستمی ارائه گردیده و نشان داده می‌شود چگونه می‌توان از تکنیک‌های داده‌کاوی جهت این امر بهره برد.

واژه‌های کلیدی: تخمین قیمت - داده‌کاوی - استخراج دانش - سیستم‌های پشتیبان تصمیم‌گیری

Keywords : Price estimation - Data mining - Knowledge discovery - Decision support systems

### ۱- مقدمه

کارخانه‌های تولیدی را می‌توان از لحاظ ماهیت تولید و تقاضا به گروه‌های مختلفی تقسیم کرد. یک دسته از آنها، کارخانجاتی هستند که کالا را بر اساس تقاضای قطعی مشتری تولید می‌نمایند [1]. به بیان دیگر کالا بر اساس سفارش مشتری تولید می‌شود. در این کارخانجات سفارش مشتری به بخش سفارشات عرضه گردیده و مراحل تولید بر اساس خصوصیات و مشخصه‌های آن تعیین می‌گردد. در خط تولید مراحل مختلفی وجود دارد که با توجه به نوع سفارش بعضی از آنها و با ترتیبی خاص روی ماده اولیه که آنهم ممکن است وابسته

1- مربی، عضو هیأت علمی دانشکده مهندسی کامپیوتر- دانشگاه یزد

2- استادیار، عضو هیأت علمی دانشکده مهندسی کامپیوتر- دانشگاه یزد

به پارامترهایی در سفارش باشد، صورت می‌گیرد. به بیان دیگر خط تولید شامل  $n$  مرحله تولید است که  $k$  تا از این مراحل ( $k \leq n$ ) با ترتیبی وابسته به سفارش، فرآیند تولید آن سفارش را مشخص می‌نمایند.

به عنوان مثال در یک کارخانه تولید بردهای مدار چاپی، سفارش مشتری که در آن ابعاد برد، چندلایه بودن آن، نوع برد، ترکیبی بودن یا نبودن آن و یا نوع چاپ روی برد، مشخص می‌باشد، دریافت و به سیستم وارد می‌شود. با توجه به هر یک از این خصوصیات است که مشخص می‌گردد چه مرحله یا مرحله‌ای از تولید باید روی فیبر اولیه صورت گیرد تا محصول نهایی طبق سفارش مشتری آماده گردد.

از طرفی یکی از موارد مهم در همه کارخانجات محاسبه قیمت تمام‌شده محصول می‌باشد که بوسیله تکنیک‌های حسابداری صنعتی صورت می‌گیرد. در این تکنیک‌ها با طبقه‌بندی هزینه‌ها به تفکیک عوامل اصلی قیمت تمام‌شده، مستقیم و غیرمستقیم، متغیر و ثابت و با توجه به وظایف سازمانی و به کمک روش‌های ترکیبی قیمت تمام شده محصول محاسبه می‌شود [2]. آنچه این موضوع را پیچیده‌تر می‌سازد، قیمت تمام‌شده محصول در کارخانجات تولیدی مبتنی بر سفارش است. در این کارخانجات باید بر اساس سفارش مشتری و قبل از تولید محصول یک پیش‌بینی یا لااقل تخمینی از قیمت تمام‌شده محصول داشت تا بتوان آنرا به مشتری اعلام نمود. اینجاست که هنر داده‌کاوی (Data Mining) مشخص می‌شود. هدف در این مقاله این است که با توجه به داده‌ها و یافته‌های قبلی موجود در سیستم و با داشتن مشخصات یک سفارش جدید به کمک تکنیک‌های داده‌کاوی قیمت تمام شده محصول را پیش‌بینی نمود. این مقاله به سیستم اطلاعات تولید از این زاویه می‌پردازد.

## ۲- تشخیص اتوماتیک فرآیند تولید

اطلاعات مربوط به یک سفارش در روند عملیاتی تولید استخراج و به سیستم پردازش تراکنش (On-Line Transaction Processing) وارد می‌شود. این اطلاعات به تنهایی مورد استفاده قرار گرفته و یکسری گزارش‌های کاری در زمینه‌هایی همچون حجم سفارش مشتریان، تنوع سفارشات، عملکرد اپراتورها و میزان دورریز را ارائه می‌نماید.

برای تعیین فرآیند تولید، از قواعدی که بین خصوصیات سفارش و مراحل تولید وجود دارد، استفاده می‌شود. مدیر تولید بر اساس این قواعد تصمیم می‌گیرد که مرحله‌ای در فرآیند تولید قرار گیرد یا نه؟ مثلاً اگر ابعاد برد از حدی بزرگتر بود، حتماً مرحله خاصی در فرآیند تولید باید صورت گیرد، یا اگر نوع برد، نوع اول بود، مرحله دیگری نیز حتماً باید انجام شود.

آنچه که یک سیستم داده‌کاوی (Data Mining) در این مورد می‌تواند انجام دهد، این است که از روی انبوه اطلاعات وارد شده به سیستم در دوره‌های گذشته و پیدا کردن وابستگی‌های بین خصوصیات سفارش و مراحل تولید آنها، قواعد بین مشخصات سفارش و مراحل تولید را به صورت پویا استخراج نماید. همچنین با ورود یکسری قواعد ساده که ممکن است صریحاً از روی مشخصات قابل استخراج نباشد، اطلاعات اولیه به سیستم داده می‌شود. این سیستم می‌تواند به تنهایی جهت پیش‌بینی مراحل فرآیند تولید یک درخواست جدید مورد استفاده قرار گیرد یا به کمک یک سیستم خبره، این قواعد پردازش گردیده و فرآیند تولید درخواست جدید استخراج گردد.

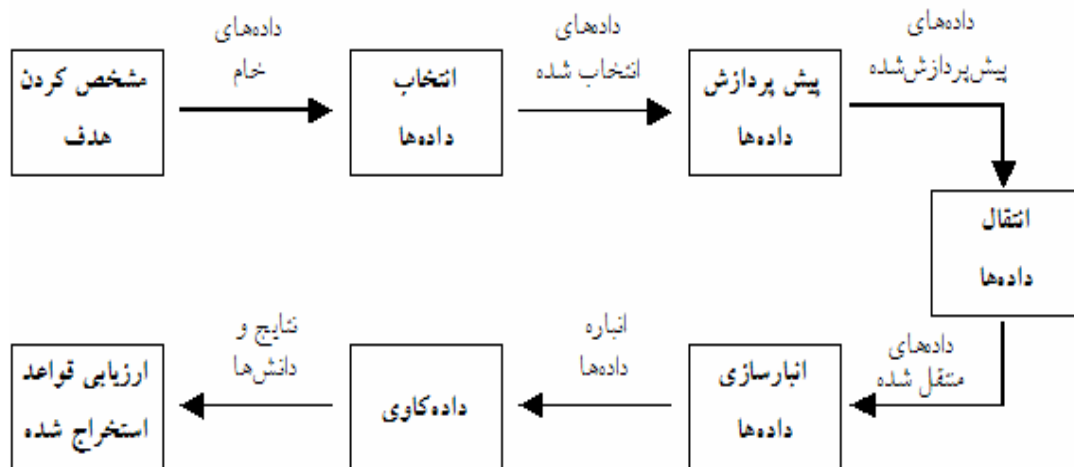
## ۳- داده‌کاوی

داده‌کاوی یکی از مراحل در فرآیند استخراج دانش از پایگاه داده‌ها (KDD : Knowledge Discovery from Database) است که روز به روز استفاده از آن بیشتر می‌شود [3]. با توجه به اینکه این مرحله مهمترین مرحله فرآیند می‌باشد، گاهی به جای یکدیگر نیز استفاده می‌شوند [4]. این مفهوم کمک می‌کند تا از انبوه داده‌های ذخیره شده، اطلاعات مهم و کاربردی را استخراج کرده و به عینیت درآورد. مفهوم داده‌کاوی عبارت است از فرآیند خودکار کشف قوانین، الگوها و مدل‌های سودمند از پایگاه داده‌های مختلف و مخازن داده [4]. داده‌کاوی به شرکت‌ها و مؤسسات کمک می‌کند تا در محیط عملیاتی خود بهتر تصمیم‌گیری نمایند و به همین دلیل یکی از مهمترین ابزارها در پیاده‌سازی سیستم‌های پشتیبان تصمیم‌گیری می‌باشد [5].

داده‌کاوی شامل چندین عملکرد اصلی است که عبارتند از خلاصه‌سازی (Summarization)، وابسته‌سازی (Association)، کلاسه‌بندی یا گروه‌بندی (Classification)، پیش‌گویی (Prediction) و خوشه‌بندی یا دسته‌بندی (Clustering) [6]. هر یک از ابزارهای داده‌کاوی ممکن است بعضی یا همه این روشها را با الگوریتم‌های مختلفی پیاده‌سازی نمایند.

امروزه برنامه‌های کاربردی مختلفی به کمک داده‌کاوی در زمینه‌های مختلفی همچون سیستم‌های فروش، بانکداری، سیستم‌های مالی، سیستم‌های تولید، سیستم‌های مراقبت بهداشتی و خدمات پشتیبانی توسعه یافته‌اند [5] [7]. همچنین در زمینه‌های دیگری همچون

بررسی سری‌های زمانی، سیستم‌های خاص، سیستم‌های ارتباطی، web و چندرسانه‌ای نیز مورد استفاده قرار گرفته است [7]. به طور کلی فرآیند داده‌کاوی و تکنیک‌ها و توابع مورد استفاده، به شدت به نوع مسأله و طبیعت داده‌های محیط عملیاتی مربوط می‌شود [5]. داده‌کاوی معمولاً در فرآیند کشف دانش از پایگاه داده‌ها بعد از مرحله ایجاد انبار داده‌ای قرار دارد. شکل ۱ فرآیند کلی کشف دانش و ارتباط داده‌کاوی با مراحل دیگر را نشان می‌دهد [5].



شکل ۱: فرآیند استخراج دانش و نقش داده‌کاوی در آن

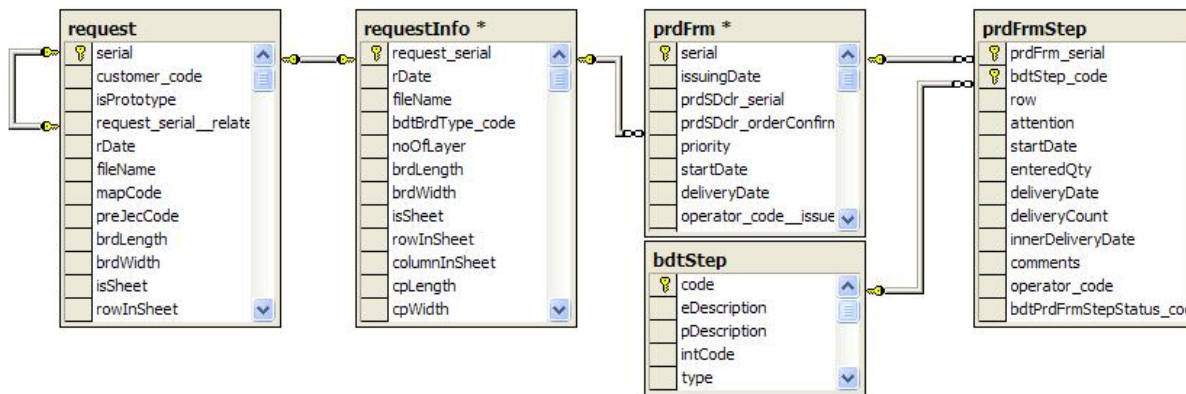
آنچه در شکل آمده است مشخص می‌کند که فرآیند داده‌کاوی جهت استخراج دانش‌های نهفته در حجم زیاد داده‌ها است. این دانش‌ها می‌تواند با بررسی داده‌ها و کشف وابستگی‌های بین آنها و الگوهای جالبی که بین آنها وجود دارد، استخراج گردد. بر این اساس، فرآیند استخراج دانش شامل بخش‌هایی به شرح زیر می‌باشد [5]:

### ۳-۱- مشخص کردن هدف

منظور این است که داده‌کاوی می‌تواند به هدف استخراج دانش‌های مختلفی صورت گیرد. بسته به نوع این اهداف می‌توان از توابع و تکنیک‌های مختلف و ابزارهای متفاوت داده‌کاوی استفاده نمود. آنچه هدف اصلی در این مقاله می‌باشد این است که از اطلاعات موجود در سیستم اطلاعاتی تولید یک کارخانه تولیدی، پروسه تولید یک سفارش جدید به طور اتوماتیک استخراج گردد.

### ۳-۲- انتخاب داده‌ها

معمولاً با توجه به هدف، همه داده‌های یک پایگاه داده عملیاتی در استخراج دانش مفید واقع نخواهد شد. بنابراین به جای درگیری با این انبوه داده‌ها، زیر مجموعه‌ای از آنها که در این راستا مفید خواهد بود، انتخاب می‌شود. در پایگاه داده‌های سیستم تولید جداول اطلاعاتی فراوانی وجود دارد. ما از میان این جداول، جدول درخواست، جدول اطلاعات فنی، جدول اطلاعات فرم تولید، جدول مراحل فرم تولید و جدول اطلاعات پایه مراحل تولید را انتخاب می‌نماییم. این جداول و ارتباط بین آنها در شکل ۲ نشان داده شده است.



شکل ۲: جداول اصلی مورد استفاده جهت پیش‌بینی فرآیند تولید و ارتباطات بین آنها

### ۳-۳- پیش پردازش داده‌ها

هدف کلی این مرحله حذف داده‌های دارای نویز، مشکوک و غیرکامل است. از بین داده‌های موجود در جداول انتخاب شده، ممکن است اطلاعاتی موجود باشد که در استخراج دانش کاربردی ندارد یا حتی ممکن است سیستم را از همگرایی دور کرده و یا استنتاج را گمراه کند. این اطلاعات در سیستم استخراج دانش می‌تواند حذف شود.

### ۳-۴- انتقال داده‌ها

داده‌های مورد نیاز جهت داده‌کاوی در جداول مختلفی قرار دارند. بعضی اوقات استخراج داده‌های محاسباتی و ایجاد دیدهای سطح بالاتر می‌تواند مرحله استخراج دانش را ساده‌تر نماید. این عملیات از عملیات ساده تبدیل واحد گرفته تا ترکیب ستونها و عملیات مختلف روی رابطه‌ها می‌باشد. نمونه‌ای از این انتقال داده‌ها، در نمونه پیاده‌سازی شده، ترکیب جداول درخواست و اطلاعات فنی و قرار دادن اطلاعات مهم و مؤثر تولید در یک دید واحد می‌باشد.

### ۳-۵- انبارسازی داده‌ها

Data warehouse عبارت است از مجموعه‌ای موضوع‌گرا، مجتمع، از زمانهای گوناگون و غیرفرار از داده‌ها که جهت پشتیبانی از تصمیم‌گیری مورد استفاده قرار می‌گیرد [6]. فرآیند برنامه‌ریزی، ساخت، استفاده، مدیریت، نگهداری و در دسترس قراردادن این داده‌ها را Data warehousing گویند [5].

انبار داده‌ای به کمک مختصرسازی، انجام عملیات جمعی و ایجاد ساختارهای داده‌ای جامع‌تر از رابطه‌ها نظیر cube، داده‌ها را در قالب مدل‌های چند بعدی مدل می‌نماید. عملیات مختلفی چون نفوذ به جزئیات (Drill down)، حرکت از جزئیات برای استنتاج کلیات (Roll up)، حرکت به سمت منابع اصلی داده‌ها (Drill across)، Slice&Dice و Pivoting در این مدل‌ها قابل انجام است [8]. تفاوت اصلی این مدل در مقایسه با سیستم‌های پردازش تراکنش‌ها را مفهوم OLAP (On-Line Analytical Processing) بیان می‌نماید [9]. بدین معنا که به جای برخورد با تراکنش‌ها و رویدادهای روزمره محیط عملیاتی، به مفاهیم کلی و تحلیلی سیستم می‌پردازد.

همانطور که در شکل ۱ نشان داده شده، داده‌کاوی معمولاً بعد از مرحله Data warehouse صورت می‌گیرد ولی می‌توان مستقیماً نیز آنرا روی داده‌های پایگاه داده‌ها انجام داد [6]. در سیستم نمونه ارائه شده در این مقاله، عمل داده‌کاوی مستقیماً از روی داده‌های پایگاه داده رابطه‌ای انجام می‌شود و عملاً مرحله Data warehousing مورد استفاده قرار نگرفته است.

### ۳-۶- داده‌کاوی

داده‌کاوی مهمترین مرحله در فرآیند استخراج دانش می‌باشد تا حدی که همچنان که اشاره شد این دو را به جای یکدیگر نیز استفاده می‌نمایند [4]. تکنیک‌ها و توابع مختلفی در این فاز قابل استفاده است که بعضی از آنها در نمونه پیاده‌سازی شده بیان گردیده است.

### ۳-۷- ارزیابی قواعد استخراج شده

آخرین کار در فرآیند داده‌کاوی، مقایسه نتایج حاصل از مدل‌های مختلف استفاده شده می‌باشد. برای این ارزیابی، دانش‌های واقعی در محیط عملیاتی که از ارتباط داده‌های موجود در سیستم وجود دارد را با خروجی دانش استخراج شده از همین داده‌ها که به صورت تستی به سیستم داده می‌شود، مقایسه می‌کند. به بیان دیگر یکسری از داده‌های اولیه به عنوان داده‌های آموزشی و با هدف آموزش به سیستم داده می‌شود و دسته دیگری از این داده‌ها به عنوان داده‌های آزمایشی و به هدف ارزیابی سیستم مورد استفاده قرار می‌گیرد.

#### ۴- وابسته‌سازی

یکی از تکنیک‌های اصلی در داده‌کاوی، کشف وابستگی‌های موجود در دل داده‌ها به شکل قوانین (Rules) می‌باشد که از آن با نام وابسته‌سازی یاد می‌شود. در واقع منظور اصلی، کشف وابستگی‌های موجود در دل داده‌ها به شکل قوانین می‌باشد و نه ایجاد وابستگی، چرا که وابستگی‌ها وجود دارد [6]. استخراج قوانین وابستگی به معنی یافتن الگوهای تکرار شونده، وابستگی‌ها، ارتباط‌های دوجانبه یا ساختارهای علی در میان مجموعه‌هایی از اشیاء یا ایتماها در پایگاه داده‌های تراکنشی، رابطه‌ای یا سایر مخازن داده می‌باشد [8].  
فرم کلی قوانین به صورت زیر است [8]:

Body  $\rightarrow$  Head [ Support , Confidence ]

به عنوان مثال قانون زیر نشان دهنده آن است که اگر درخواستی ترکیبی باشد (IsCompound مقدار True داشته باشد)، آنگاه مرحله برش CNC در مراحل تولید آن وجود دارد.

Request ( X , “ IsCompound” )  $\rightarrow$  Step( X , “CNC Cutting” ) [ 2% , 80% ]

برای یک قانون وابستگی باید یک جامعه، یعنی جمعیت شامل یک مجموعه از نمونه‌ها را معین کرد. در نمونه مورد بحث، جامعه می‌تواند شامل تمام اطلاعات سفارش‌های قبلی وارد شده به سیستم باشد. مسأله‌ای که در مورد یک قانون بسیار مهم است، پشتیبانی و اطمینان یک قانون است. در مثال بالا پشتیبانی و اطمینان قانون ذکر شده به ترتیب ۲ و ۸۰ درصد می‌باشد. پشتیبانی (Support)، عبارت است از اندازه کسری از جامعه که هر دوی مقدم و تالی قانون را ارضاء می‌کند [8]. به عنوان مثال در قانون ذکر شده در سیستم اطلاعات تولید، تنها ۲ درصد درخواستها است که هم مرکب هستند و هم مرحله برش CNC در تولید آنها وجود دارد. بسته به نوع کاربرد، حداقل درجه‌ای از پشتیبانی که معقول باشد، پذیرفته می‌شود.

اطمینان (Confidence)، مشخص می‌کند که از کل مواردی که مقدم درست است، در چه کسری از این موارد تالی نیز درست می‌باشد [8]. یعنی یک احتمال شرطی است که یک تراکنش شامل مقدم، شامل تالی نیز باشد. به عنوان نمونه در مثال ذکر شده مشخص گردیده که اطمینان قانون ۸۰ درصد است. این بدین معناست که ۸۰ درصد از درخواستهای ترکیبی دارای مرحله برش CNC می‌باشند. یک قانون با اطمینان پایین قانون معنی‌داری نیست، لذا همواره حداقلی از اطمینان مورد نظر تعریف می‌شود. بر اساس تعریف مشخص است که اطمینان  $A \rightarrow B$  ممکن است با اطمینان  $B \rightarrow A$  بسیار متفاوت باشد، اگرچه پشتیبانی هر دو قانون یکسان است. برای انتخاب فرآیند تولید به کمک این تکنیک، قواعد ارتباطی بین مشخصات تولید و مراحل تولید استخراج می‌شوند. این قواعد شکل کلی زیر را دارا است: «اگر خصوصیت Pi مقدار Vj را داشته باشد آنگاه مرحله Sk در فرآیند تولید وجود دارد». همچنین بعضی از خصوصیات سفارش که مقدار پیوسته دارد، در مرحله پیش‌پردازش باید آنها را محدود بندی کرد و به هر محدوده‌ای یک مقدار را وابسته نمود. در این موارد جهت استخراج قوانین از تابع تبدیل مربوطه استفاده نموده و به جای مقادیر پیوسته از این مقادیر استفاده می‌شود. مجموعه این قواعد در کنار قواعد احتمالی دیگر که ممکن است از روی داده‌های پایگاه قابل استخراج نباشد و مجزا به سیستم داده می‌شود، منبعی برای یک سیستم خبره است که به کمک آن می‌توان دانش‌های جدید را استخراج نمود.

#### ۵- رده‌بندی

همانطور که قبلاً بیان شد رده‌بندی یکی دیگر از تکنیک‌های اصلی در داده‌کاوی می‌باشد. این تکنیک شامل پیش‌بینی عناوین قطعی کلاس‌ها و رده‌بندی داده‌های جدید در آنها می‌باشد. به بیان دیگر رده‌بندی عبارت است از پیش‌بینی عناوین قطعی کلاسها و رده‌بندی داده‌ها یا ساخت یک مدل بر اساس مجموعه آموزشی و استفاده از آن در رده‌بندی داده‌های جدید [8].

کار اصلی در رده‌بندی عبارت است از تعیین تعلق ایتماها به یکی از چندین کلاس و تعیین نمونه‌ها برای آموزش ( Training instances که مجموعه آموزشی نیز نامیده می‌شود) به همراه کلاس‌هایی که نمونه‌ها به آنها تعلق دارد [8]. به این ترتیب اگر ایتما جدیدی در اختیار قرار گیرد، با توجه به مقادیر صفات آن می‌توان کلاس مربوطه را پیش‌بینی کرد.  
عملیات رده‌بندی در ۲ مرحله کلی زیر صورت می‌گیرد [6]:

الف- ساختن مدل، که عبارت است از توصیف یک مجموعه رده‌های از پیش تعیین شده. هر تعداد نمونه به یک کلاس از پیش تعریف شده تعلق دارد که این مساله توسط صفات رده‌بندی معین می‌شود. برای ساختن مدل، از مجموعه‌ای از نمونه‌ها که هم مقادیر صفات رده‌بندی برای آنها معین است و هم کلاس آنها بر اساس اطلاعات گذشته مشخص می‌باشد، استفاده می‌شود. این مجموعه، مجموعه آموزشی (Training Set) نامیده می‌شود. این مدل‌ها می‌تواند به صورت قوانین رده‌بندی، درختهای تصمیم‌گیری یا فرمول‌های

ریاضی نشان داده شود.

ب- استفاده از مدل، یعنی دادن نمونه ورودی جدید و پیش‌بینی رده آن توسط سیستم. برای این منظور ابتدا با یک مجموعه آزمایشی که متفاوت از مجموعه آموزشی است و کلاس مربوط به آنها مشخص است، دقت مدل تخمین زده می‌شود. این دقت با مقایسه نتیجه حاصل از مدل با جواب صحیح به دست می‌آید. نسبت دقت عبارت است از درصدی از نمونه‌های آزمایشی که به درستی توسط مدل رده‌بندی شده‌اند، یعنی در کلاس مربوط به خود قرار گرفته‌اند.

در مساله رده‌بندی در حقیقت به یک مدل آموزش داده می‌شود. این فراگیری از نوع «نظارت شده» است به این معنی که داده آموزشی شامل مشاهدات و اندازه‌گیری‌ها با عناوینی همراه هست که نشان دهنده کلاس نمونه آموزشی است. سپس داده جدید بر این اساس رده‌بندی می‌شود. این مفهوم وجه تمایز تکنیک رده‌بندی با تکنیک دیگری در داده‌کاوی به نام خوشه‌بندی است که در آن فراگیری «بدون نظارت» است و عناوین کلاسهای داده آموزشی نیز ناشناخته است [6]. در تکنیک خوشه‌بندی هدف یافتن وجود کلاسها در داده‌ها می‌باشد و کلاس‌های مختلف از قبل مشخص نیست.

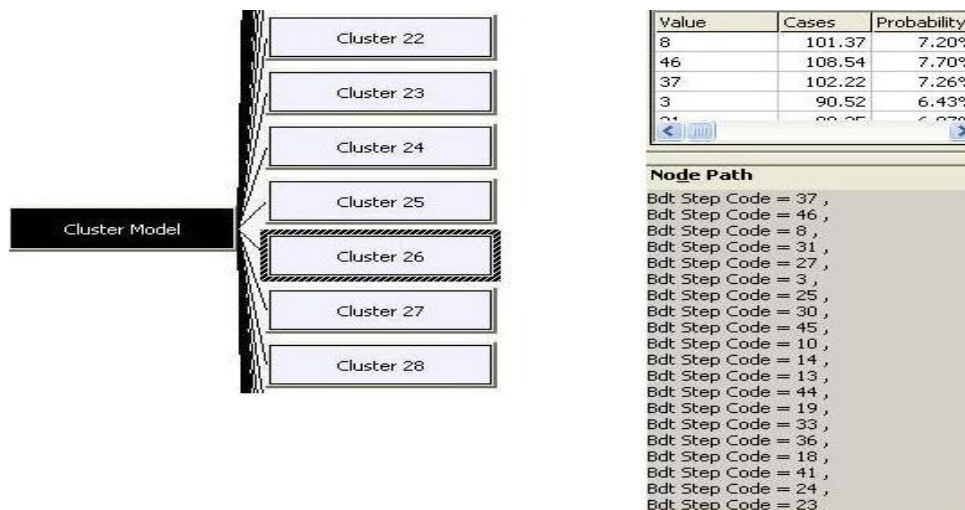
از این تکنیک نیز می‌توان جهت انتخاب اتوماتیک فرآیند تولید برای یک سفارش جدید استفاده نمود. برای این منظور فرآیندهای تولید قبلی بر اساس پارامترهای مؤثر در آن گروه‌بندی می‌شود. به بیان دیگر کلاس‌های مختلفی شامل مقادیر مختلف پارامترها در نظر گرفته می‌شود. هر کلاس از یک سو مشخص کننده مشخصات اصلی یک سفارش است و از سوی دیگر پروسه تولید آن سفارش نیز مشخص می‌باشد. نهایتاً یک ورودی جدید یعنی یک سفارش جدید در یکی از کلاس‌ها قرار می‌گیرد و بدین ترتیب مراحل تولید آن مشخص می‌گردد.

#### ۶- نمونه پیاده‌سازی شده

برای اکتشاف دانش و داده‌کاوی ابزارهای نسبتاً زیاد و قدرتمندی وجود دارد. برخی از این ابزارها بیش از یک تکنیک داده‌کاوی را پشتیبانی می‌کنند، به عنوان مثال Darwin from thinking machine corp. از شبکه‌های عصبی، درخت بازگشتی (CART)، الگوریتم K-means و استدلال Case-based پشتیبانی می‌کند، برخی نیز تنها محدود به یک روش می‌شوند [5] [10].

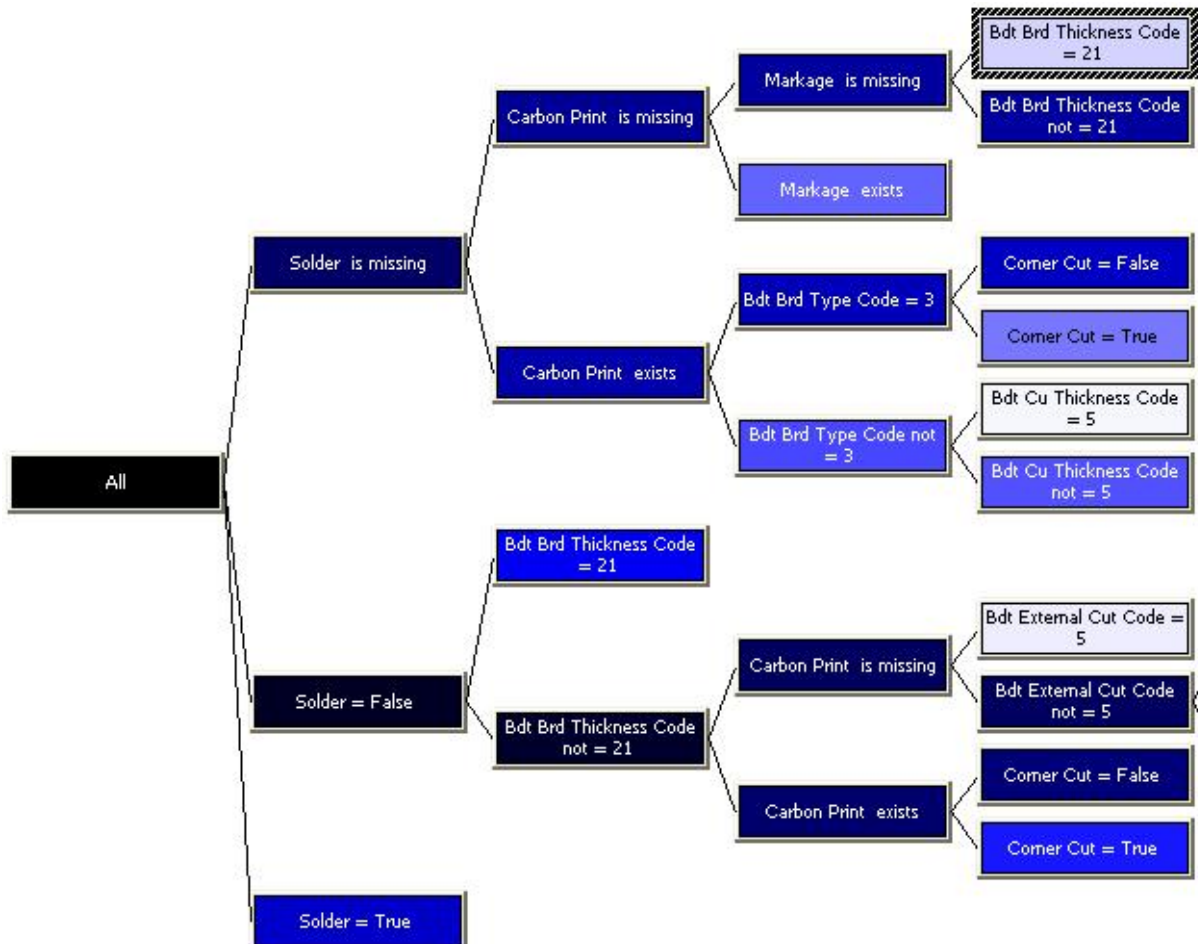
یکی دیگر از ابزارهای داده‌کاوی DB Miner است که بوسیله گروه تحقیق DB Miner از آزمایشگاه سیستم‌های پایگاه‌داده‌های هوشمند در دانشگاه Simon Fraser کانادا طراحی و توسعه یافته است. علی‌رغم قدرت بسیار زیاد این ابزار باید اعتراف کرد که برای کشف دانش از داده‌های غیر ساخت‌یافته همچون داده‌های متنی مناسب نیست [9].

ابزار دیگری که جهت داده‌کاوی کاربرد دارد، Analysis Services می‌باشد که جزو الحاقات SQL Server 2000 است. این سیستم امکان انجام عملیات Data mining و OALP را به کاربر می‌دهد. Analysis Services جهت داده‌کاوی دو تکنیک Microsoft Clustering و Microsoft Decision Tree را استفاده می‌نماید [12]. برای پیاده‌سازی سیستم نمونه پیش‌بینی فرآیند تولید از این ابزار استفاده شده است.



شکل ۳: بخشی از پروسه‌های استخراج شده و مراحل و درصد پوشش در یکی از آنها

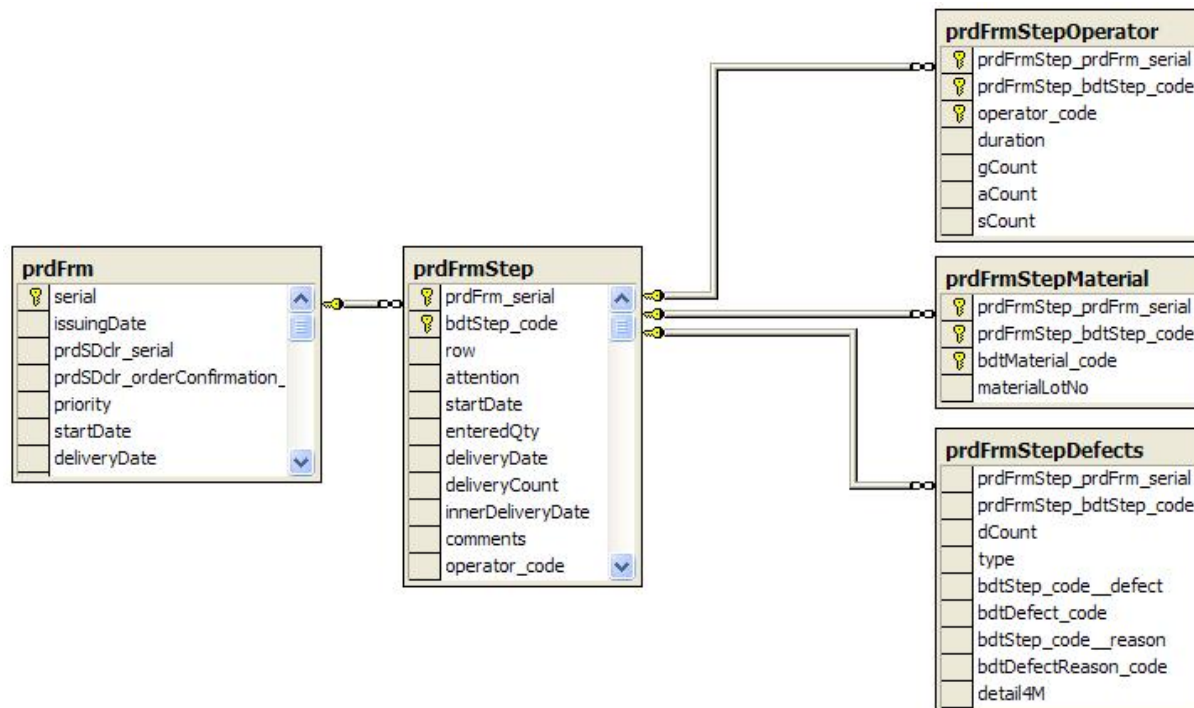
در این نمونه از روش Clustering جهت دسته‌بندی فرآیندهای مختلف تولید استفاده شده است. برای این منظور گروه‌های مختلف شامل مراحل تولید که در تولید یک سفارش همراه هستند استخراج گردیده است. بخشی از کلاسترهای کشف شده و فراوانی و درصد پوشش آنها در شکل ۳ نشان داده شده است. این دسته‌ها عملاً همان فرآیندهای ممکن برای تولید یک سفارش جدید می‌باشد. همچنین برای استخراج فرآیند تولید یک سفارش جدید از روی مشخصات آن، از روش Decision Tree استفاده گردیده است. برای این منظور کلیه پارامترهای سفارش که در اضافه شدن مرحله‌ای به فرآیند تولید مؤثر می‌باشند و طی فاز پیش‌پردازش داده‌ها و انتقال داده‌ها در یک دید مجتمع گردیده، مورد استفاده قرار گرفته است. به کمک این اطلاعات و ایجاد درخت تصمیم بر اساس آن، می‌توان یک ورودی جدید (مشخصات یک سفارش جدید) را به گره برگ مربوطه هدایت نمود. مشخصات این گره مشخص کننده فرآیند و نتیجتاً مراحل تولید آن سفارش می‌باشد. نمونه کوچکی از این درخت تصمیم در شکل ۴ نشان داده شده است.



شکل ۴: بخشی از درخت تصمیم جهت پیش‌بینی فرآیند تولید سفارش بر اساس مشخصه‌های مهم آن

## ۷- فرآیند تولید و قیمت تمام شده محصول

فرآیند تولید از زیرسیستم‌های مختلفی تشکیل می‌گردد که قسمت اصلی و ساده شده آن شامل اطلاعات فرم تولید (جدول prdFrm)، مراحل تولید یک سفارش (جدول prdFrmStep)، اپراتورهای انجام دهنده هر یک از مراحل (prdFrmStepOperator)، مواد مصرفی در هر یک از مراحل (prdFrmStepMaterial) و دلایل خرابی و ضایعات هر یک از مراحل (prdFrmStepDefect) و ارتباط بین آنها در شکل ۵ نشان داده شده است.



شکل ۵: بخشی از جداول اطلاعاتی فرآیند تولید موثر در قیمت

بعضی از مؤلفه‌های مهم و مؤثر از میان مؤلفه‌های نشان داده شده عبارتند از:

- bdtStep\_code: کد مرحله تولید
- enteredQty: تعداد واحد ورودی به مرحله
- deliveryCount: تعداد واحد خروجی از مرحله (که ممکن است متفاوت از واحد ورودی باشد).
- operator\_code: کد اپراتور انجام دهنده مرحله (n نفر به ازاء هر مرحله)
- duration: مدت زمان کار هر اپراتور روی هر مرحله
- sCount, aCount, gCount: تعداد تولید خوب، قابل قبول و ضایعات
- bdtMaterial\_code: کد ماده مصرفی در مرحله (n ماده به ازاء هر مرحله)

بر این اساس عوامل اصلی تعیین قیمت تمام‌شده به صورت زیر قابل استخراج است:

- هزینه مواد: با داشتن مواد مصرفی هریک از مراحل تولید و الحاق اطلاعات قیمت مواد به آن
  - هزینه کار: به کمک اطلاعات اپراتورهای مراحل تولید، میزان تولید هر یک و وابسته‌نمودن اطلاعات مزد به آن
  - هزینه سربار: در اینجا فقط سربار ناشی از ضایعات و خرابی‌های تولید وجود دارد و دیگر سربارها باید مجزا در نظر گرفته شود.
- با استخراج فرآیند تولید، مراحلی که برای تولید سفارش جدید باید صورت گیرد به دست آمده است. از طرفی پارامترهای مؤثر در قیمت تمام شده در هر مرحله نیز مشخص است. با ترکیب این اطلاعات و محاسبه تابع تعیین قیمت بر اساس آنها می‌توان قیمت تمام شده کالا را تخمین زد.

## ۸- نتیجه‌گیری

در این مقاله ضمن بررسی یکسری مفاهیم و تکنیک‌های کاربردی در سیستم‌های استخراج دانش از پایگاه داده‌ها به بررسی داده‌کاوی به صورت خاص جهت استخراج پروسه تولید از روی مشخصات سفارش و محاسبه قیمت تمام شده محصول بر اساس آن پرداخته شده است. بر این اساس، این سیستم می‌تواند در کنار یک سیستم اطلاعات تولید که اطلاعات روزمره را ذخیره و پردازش می‌نماید، استفاده شده و از روی مشخصه‌های مؤثر در تولید، فرآیند تولید سفارش را پیشنهاد دهد و قیمت کالا را به مشتری اعلام نماید. قراردادن



این سیستم در بستر web می‌تواند قیمت را به صورت online به مشتری اعلان نماید. جهت پیاده‌سازی یک نمونه از چنین سیستمی از بین ابزارهای مختلف Analysis Services، محصول مایکروسافت که همراه با SQL Server 2000 عرضه گردیده، استفاده شده است. در این نمونه از تکنیک Clustering جهت دسته‌بندی مراحل و استخراج فرآیندهای ممکن تولید و از تکنیک Decision Tree برای استخراج فرآیند تولید از روی مشخصات سفارش جدید استفاده گردیده است.

## منابع و مراجع

- [1] R.E. Markland,... , "Operation Management", South-Western College Publishing, 1998.
- [2] پرویز بختیاری، "حسابداری و مدیریت مالی برای مدیران"، سازمان مدیریت صنعتی، ۱۳۷۹.
- [3] M.S.Chen, J.Han, P.S.Yu, "Data Mining : An overview from a database perspective", IEEE Transactions on Knowledge and Data Engineering 8, 1996.
- [4] U.M.Fayyad, P.Stolorz, "Data mining and KDD", Jet Propulsion Laboratory, California Institute of Technology, USA, 1997.
- [5] S.C.Hui, G.Jha, "Application data mining for customer service support", Nan Yang technological University of Singapore, November 1999.
- [6] J.Han, M.Kamber, "Data mining : concepts and techniques", Morgan Kaufmann publishers, August 2000.
- [7] K.D.Nuggets , site for data mining and knowledge discovery. <URL : <http://www.kdnuggets.com/>>
- [8] R.J.Roiger, M.W.Geatz, "Data Mining" International Edition, Addison Wesley, 2003.
- [9] J.Han, "Towards On-line Analytical Mining in Large Databases", Intelligent Database Systems Research Laboratory, School of computer science, Database group.
- [10] I.Bose, R.K.Mahapatra , "Business data mining " , Information&Management 39(p211-225), 2001.
- [11] Simon Fraser University , School of computer science , Database group. <URL : <http://db.cs.sfu.ca/>>
- [12] Analysis manager Concepts & Tutorial . Microsoft SQL server 2000 Analysis services .
- [13] G.Fosca, P.Dino, "Knowledge discovery & data mining tools, methods and experiences", Pisa KDD lab CNUCE-CNR & Univ, Pisa, 2000.
- [14] N.S.Koutsoukis, G.Mitra, C.Lucas , "Adapting on-line analytical processing for decision modeling : the interaction of information and decision technologies" , Decision Support Systems 26, 1999.
- [15] M.J.Shaw, C.Subramaniam, G.W.Tan, M.E.Welge, "Knowledge management and data mining for marketing", Decision Support Systems 31, 2001.
- [16] A.Silberschatz, H.F.Korth, S.Sudarshan, "Database system concepts", 4th edition, McGraw-hill, July 2001.
- [17] S.M.Weiss, S.J.Buckley, S.Kapoor, "Knowledge-Based Data Mining", SIGKDD, 03 Washington, DC, USA.