# مروری بر مدل‌های فرآیند هوش تجاری

حمیدرضا بویری[۱]، سید رئوف خیامی[۲]

[۱] دانشجوی دکتری، آزمایشگاه معماری سازمانی و هوش تجاری، دانشکده مهندسی فناوری اطلاعات، دانشگاه صنعتی شیراز، شیراز
hr.boveiri@sutech.ac.ir

[۲] استادیار، آزمایشگاه معماری سازمانی، دانشکده مهندسی کامپیوتر و فناوری اطلاعات، دانشگاه صنعتی شیراز، شیراز
khayami@sutech.ac.ir

## چکیده

**امروزه کمتر سازمان موفقی در دنیا وجود دارد که از هوش تجاری به عنوان یکی از مهمترین ابزارهای راهبردی در راستای نیل به اهداف و چشم‌انداز خود استفاده نکرده باشد. هوش تجاری یا هوشمندی کسب و کار به معنی استفاده از تمامی منابع، تکنیک‌ها، روش‌ها و ابزارهای موجود بطور هوشمندانه و در راستای نیل به اهداف سازمانی است. مهمترین این منابع، همانا داده‌های سازمانی است که در دنیا از آن به عنوان ابر کالای استراتژیک یاد می‌شود و بدین ترتیب پروژه‌های داده‌کاوی و استخراج دانش در هوشمندی کسب و کار جایگاه ویژه‌ای دارند. یکی از بزرگترین چالش‌ها در اجرای موفق اینگونه پروژها، انتخاب متدولوژی مناسب فرآیند هوش تجاری است بطوریکه متناسب با شرایط و راهبرد سازمان بوده و امکان موفقیت پروژه را به حداکثر برساند. در این مقاله انواع مختلف متدولوژی‌ها و مدل‌های فرآیندی پروژه‌های داده‌کاوی و هوش تجاری معرفی شده و نقاط ضعف، قوت و نوآوری ارائه‌شده توسط هر یک بحث و بررسی می‌شوند. این مقایسه بستر مناسبی است برای محققان، مهندسان و فعالان این حوزه تا به فراخور موقعیت و شرایط اجرای پروژه، مدل فرآیندی مناسب خود را انتخاب، توسعه یا ترکیب کنند.**

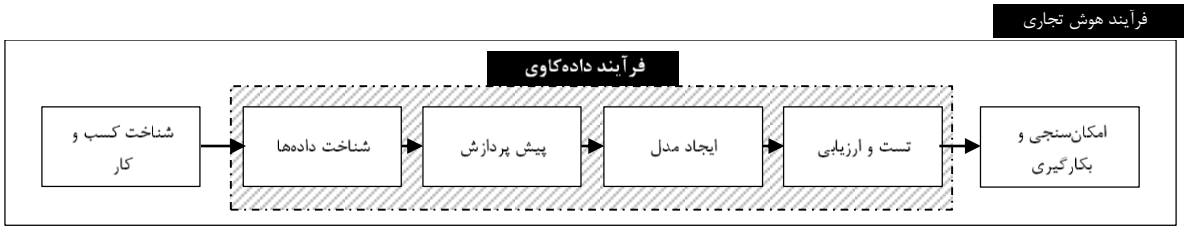
## کلمات کلیدی

**هوش تجاری (هوشمندی کسب و کار)، داده‌کاوی، متدولوژی، مدل فرآیندی هوش تجاری، زنجیره ارزش داده‌ها.**

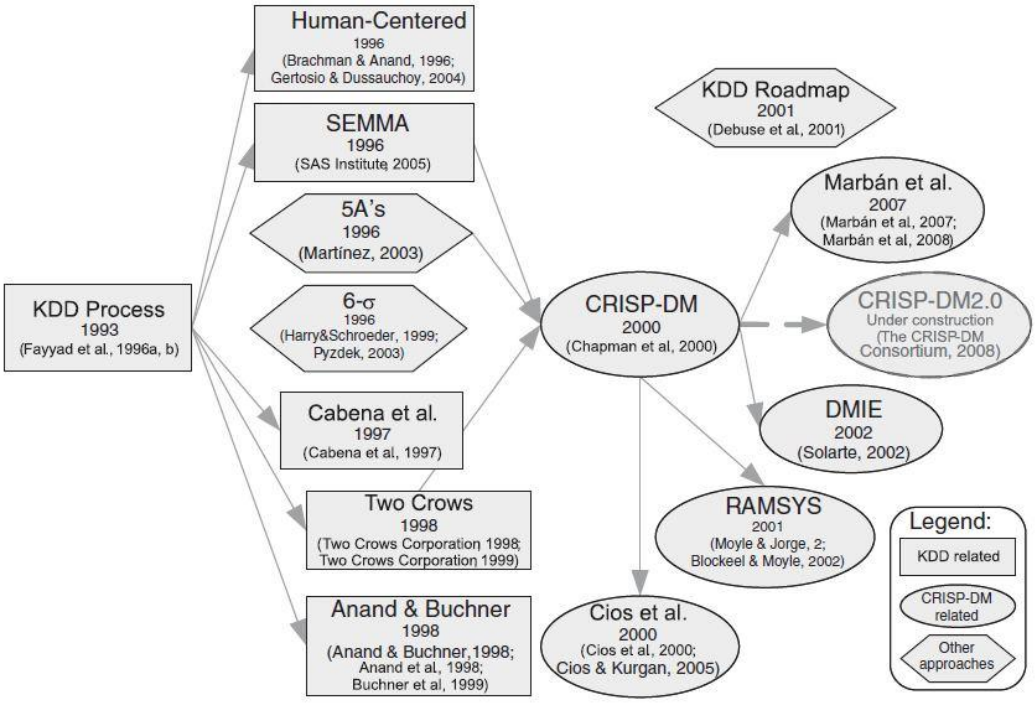
## ۱- مقدمه

امروزه کمتر سازمان موفقی در دنیا وجود دارد که از هوش تجاری به عنوان یکی از مهمترین ابزارهای راهبردی در راستای نیل به اهداف و چشم‌انداز خود استفاده نکرده باشد. هوش تجاری یا هوشمندی کسب و کار به معنی استفاده از تمامی منابع، تکنیک‌ها، روش‌ها و ابزارهای موجود به طور هوشمندانه و در راستای رسیدن به اهداف سازمانی است [۱].

یکی از مهمترین این منابع، داده‌های سازمانی است بطوریکه در بعضی موارد هوش تجاری با استخراج دانش یا داده‌کاوی اشتباه گرفته می‌شود. در حقیقت می‌توان گفت استخراج دانش یا داده‌کاوی خود به عنوان یک تکنیک یا ابزار درون هوش تجاری قابل استفاده است. شکل (۱) این رابطه را به خوبی نشان می‌دهد. در حقیقت می‌توان گفت وجود فاز شناخت کسب و کار نشان می‌دهد که قبل از شروع به انجام هر عملیاتی در هوش تجاری، ابتدا باید شناخت کاملی از سازمان مطبوع و کسب و کار آن بدست آورد و سطح بلوغ فناوری اطلاعات را در آن سنجید. چه بسیار سازمان‌هایی که ممکن است نیازی به سرمایه‌گذاری در این زمینه نداشته باشند و یا امکانات سازمان امکان اجرای پروژه‌های مبتنی بر داده را آنطور که برای سازمان مثمرثمر و قابل استفاده باشد فراهم نمی‌کنند. همچنین فاز امکان‌سنجی و بکارگیری مشخص می‌کند با توجه به شرایط و محدودیت‌های سازمانی، تا چه حد از نتایج بدست-آمده می‌توان استفاده کرد [۲].




**شکل (۱): فرآیندهای هوش تجاری و داده‌کاوی**



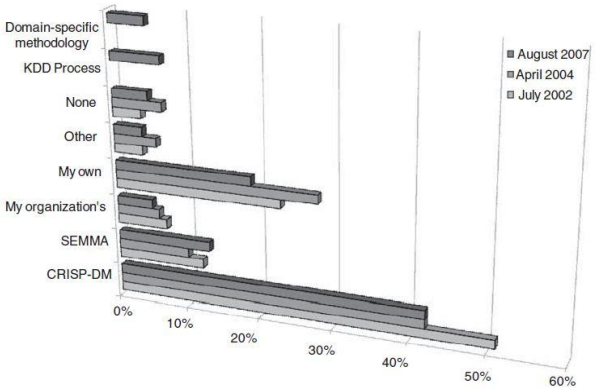

**شکل (۲): تحقیقات انجام‌شده و مدل‌های ارائه‌شده در زمینه فرآیند داده‌کاوی و هوش تجاری [۴]**

هوش تجاری دارای مدل‌های فرآیندی متنوعی است که به آنها متدولوژی هوش تجاری یا زنجیره ارزش داده‌ها گفته می‌شود. مهمترین عوامل ارائه چنین متدولوژی‌هایی عبارتند از کاهش ریسک‌های احتمالی شکست در پروژه‌های هوش تجاری انجام‌شده در سراسر دنیا (متدولوژی‌های صنعتی) و ایجاد نظم و دیسیپلین در تحقیقات آکادمیک داده‌کاوی (متدولوژی‌های آکادمیک) بطوریکه با هر بار اجرای عملیات همان نتایج پیشین گرفته شود. در حقیقت می‌توان گفت هوش تجاری یک فرآیند پیچیده است و برای آنکه با موفقیت انجام شود نیازمند چارچوب‌ها، متدولوژی‌ها، مدل‌ها، استانداردها، الگوها و به‌روش‌هاست تا ریسک‌های احتمالی شکست پروژه‌های اجراشده در این زمینه کاهش یابد. همچنین بسیار اتفاق می‌افتد که نتایج گزارش‌شده از یک تحقیق روی داده‌های مشخص قابل بدست‌آمدن نیست و بعضاً متفاوت می‌باشد و لذا نیاز به یک چارچوب و متدولوژی یکپارچه و واحد برای تمامی تحقیقات آکادمیک نیز امری ضروری بنظر می‌رسد [۳].

شکل (۲) تحقیقات انجام‌شده و مدل‌های ارائه‌شده در زمینه فرآیند داده-کاوی و هوش تجاری را نشان می‌دهد [۴]. در حقیقت، اولین مدل بنام KDD Process در سال ۱۹۹۶ ارایه شد که مقبولیت بسیاری در حوزه علمی و دانشگاهی بدست آورد؛ عملاً بسیاری از مدل‌های بعدی مانند Human Centered، SEMMA، 5A's، 6-σ، Cabena، Two Crows و Anand & Buchner شالوده خود را از این مدل اقتباس کردند. همچنین مدل CRISP-DM قلب مدل‌هایی است که بیشتر برای اجرای پروژه های صنعتی مناسب هستند. این مدل نیز پایه و اساس تعدادی از مدل-های فرآیندی بعدی مانند DIME، RAMSYS، Cios و Marban قرار گرفت.

شکل (۳) میزان استفاده مدل‌های فرآیندی مختلف داده‌کاوی و هوش تجاری در پروژه‌ها و تحقیقات امروزی را مطابق گزارش شرکت گارتنر نشان می‌دهد [۵]. در حقیقت می‌توان گفت اگر چه مدل‌های فرآیندی زیادی در این خصوص در دنیا مطرح شده است که با توجه به هدف ارائه به دو دسته تحقیقاتی و صنعتی دسته‌بندی می‌شوند؛ KDD Process و CRISP-DM به ترتیب پرارجاع‌ترین مدل‌های تحقیقاتی و صنعتی هستند که اکثر مدل‌های بعدی شالوده خود را از آنها اقتباس کرده‌اند. همچنین به لحاظ میزان استفاده در پروژه‌های تحقیقاتی و صنعتی طبق گزارش گارتنر CRISP-DM در رده اول و My own و SEMMA به ترتیب در رتبه‌های بعدی قرار دارند که My own به مجموع مدل‌های بهبودیافته و ترکیبی گفته می‌شود که توسط افراد و موسسات کوچک ارائه شده و بکار گرفته می‌شود و SEMMA یک

متدولوژی نسبتاً ساده صنعتی است که توسط شرکت SAS، غول پروژه‌های هوش تجاری، ارائه شده و استفاده می‌شود.




**شکل (۳): میزان استفاده از مدل‌های فرآیندی مختلف هوش تجاری در تحقیقات و پروژه‌های اجراشده در سراسر دنیا مطابق گزارش گارتنر [۵]**

سازماندهی ادامه مقاله بدین شکل است: در ادامه و در بخش بعدی متدولوژی‌های هوش تجاری و مدل‌های فرآیندی داده‌کاوی معرفی و بررسی می‌شوند. بیش از ۱۴ مدل فرآیندی که در تحقیقات و پروژه‌های مختلف استفاده شده‌اند یافته شده و با جزییات معرفی خواهند شد و در نهایت مقاله با ارائه نتایج و دست‌آوردهای تحقیق در بخش آخر خاتمه می‌پذیرد.

## ۲- مدل‌های فرآیندی هوش تجاری و داده‌کاوی

در این بخش، متدولوژی‌های هوش تجاری و مدل‌های فرآیندی داده‌کاوی معرفی و بررسی می‌شوند.

### ۲-۱- مدل فرآیندی KDD Process Model

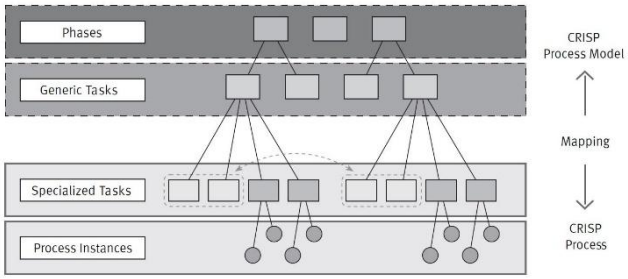
آنچه که بدیهی بنظر می‌رسد اولین بار در اولین کنفرانس و کارگاه کشف دانش از پایگاه داده (IJCAI-89) [۶] در سال ۱۹۸۹ لزوم وجود یک چارچوب فرآیندی واحد و یکپارچه در این خصوص میان محققان، خبرگان و فعالان این حوزه مورد بحث و بررسی قرار گرفت. اما اولین مدل، معروف به KDD Process Model، توسط Fayyad و همکاران در کتاب Advances in Knowledge Discovery and Data Mining [۷] در سال ۱۹۹۶ منتشر شد که محبوبیت و مقبولیت بسیاری نیز پیدا کرد.

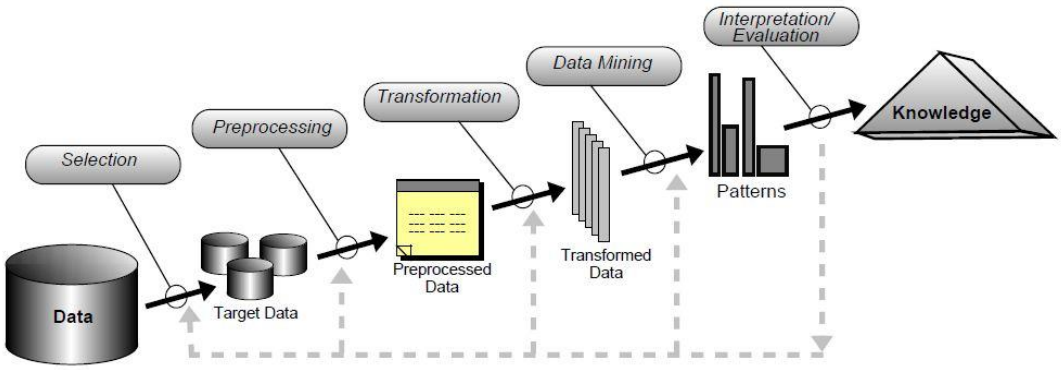
این مدل از ۹ فعالیت به شرح زیر تشکیل شده است که فازها و خروجی‌های مربوطه در شکل (۴) نمایش داده شده‌اند:

1- Developing and Understanding of the Application Domain
2- Creating a Target Data Set
3- Data Cleaning and Preprocessing
4- Data Reduction and Technology Projection
5- Choosing the DM Task
6- Choosing the DM Algorithm
7- DM
8- Interpreting Mined Patterns
9- Consolidating Discovered Knowledge

### ۲-۲- مدل فرآیند CRISP-DM

مدل KDD Process اگرچه یک مرجع مناسب برای استفاده آکادمیک و تحقیقاتی به شمار می‌آمد ولی از همان ابتدا جستجوها برای ایجاد یک مدل که مناسب پروژه‌های صنعتی باشد شروع شد و سنگ‌بنای CRISP-DM در سال ۱۹۹۶ با ائتلاف چهارگانه شرکت‌های SPSS، NCR، Daimler و OHRA گذاشته شد که اولی تولیدکننده محصولات تجاری داده‌کاوی، دومی فعال در زمینه پایگاه داده‌های تجاری و شرکت‌های سومی و چهارمی زمینه مالی اجرای تحقیق و مطالعات موردی را فرآهم کرده بودند. در نهایت در سال ۲۰۰۰ اولین نسخه CRISP-DM برای انجام پروژه‌های صنعتی داده‌کاوی ارائه شد و کم‌کم مقبولیت بسیاری بدست آورد [۹ و ۸].



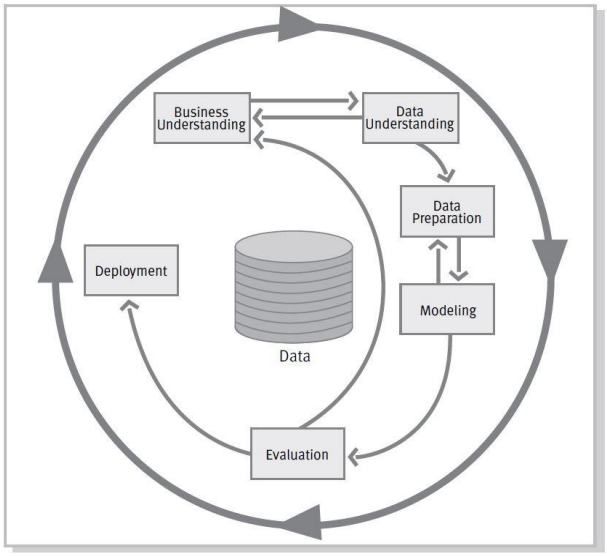

**شکل (۴): مدل فرآیند KDD Process Model [۷]**

**شکل (۵): CRISP-DM در چهار سطح جزییات مختلف [۹]**

مطابق آنچه در شکل (۵) نشان داده شده است این مدل از چهار سطح جزییات مختلف تشکیل شده که در بالاترین سطح فازهای اصلی متدولوژی متشکل از ۶ فاز مختلف قرار دارند و عبارتند از:

1- Business Understanding
2- Data Understanding
3- Data Preparation
4- Modeling
5- Evaluation
6- Deployment

مطابق شکل (۶) این فازها بصورت یک مدل محاوره‌ای تکرارشونده اجرا می‌شوند که اجازه بازبینی درون فاز، بین چند فاز و کل متدولوژی را فراهم می‌آورد.




**شکل (۶): چرخه حیات CRISP-DM [۹]**

در سطح جزییات بعدی، فعالیت‌های عمومی قرار دارند که در شکل (۷) نشان داده شده‌اند؛ مثلاً فاز آماده‌سازی داده خود ممکن است از دو فعالیت عمومی یکپارچه‌سازی داده در یک انبار داده و پیش‌پردازش داده‌ها جهت افزایش کیفیت داده تشکیل شده باشد. البته مدل CRISP-DM فقط تا این سطح را بطور دقیق مشخص می‌کند و الگوها و دستورالعمل‌ها بعلاوه ۲۴ فعالیت عمومی استاندارد که می‌بایست در هر پروژه داده‌کاوی بطور مرسوم انجام شود را ارائه می‌کند. هر چند این مدل دست استفاده‌کننده را باز می‌گذارد تا در صورت لزوم در این فعالیت‌ها یا ترتیب اجرای آنها دخل و تصرف کند و بتواند فعالیت‌هایی را حذف یا اضاف نماید.

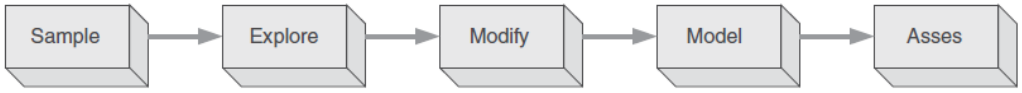
سطح سوم فعالیت‌های اختصاصی مربوط به هر فعالیت عمومی را مشخص می‌کند؛ مثلاً فعالیت عمومی پیش‌پردازش خود ممکن است شامل حذف رکورهای نامرغوب یا پرکردن فیلدهای خالی باشد و در سطح آخر دقیقاً مشخص می‌شود که هر فعالیت اختصاصی چگونه و از طریق چه الگوریتم و با چه پارامترهایی انجام شود؛ مثلاً استفاده از توزیع نرمال با پارامترهایی خاص. بهرحال CRISP-DM اگر چه متدولوژی غالب اجرای پروژه‌های داده‌کاوی است اما یکه‌تاز در بازار نیست و کم‌کم میزان استفاده آن در حال کم‌شدن است یا بهتر بگوییم میزان استفاده دیگر مدل‌ها رو به رشد است که از جمله دلایل این پدیده می‌توان به موارد زیر اشاره کرد:

۱- انحصار طلبی IBM بطوریکه این مدل فقط در ابزار Clementine متعلق به این شرکت پیاده‌سازی شده است.

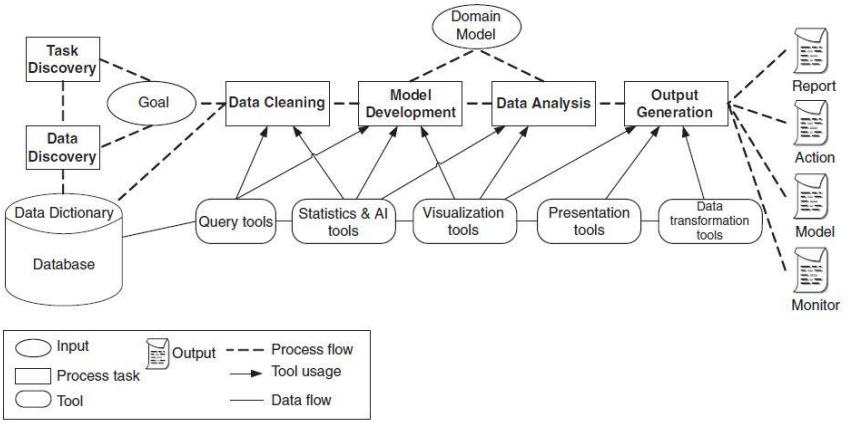
۲- CRISP-DM فقط مشخص می‌کند که چه فعالیت‌هایی انجام شود اما در مورد چگونگی انجام آنها بطور دقیق حرفی زده نشده است (مثلاً فرم‌های مربوطه جهت اجرا یا الگوریتم‌های پیشنهادی برای هر نمونه‌ی خاص وجود ندارد). در حقیقت این مدل فقط تا سطح جزییات دوم یعنی فعالیت‌های عمومی سخن برای گفتن دارد. همچنین این مدل بسیاری از فعالیت‌های مدیریت پروژه مانند مدیریت ریسک یا مدیریت کیفیت را در دل

| Business Understanding | Data Understanding | Data Preparation | Modeling | Evaluation | Deployment |
|---|---|---|---|---|---|
| **Determine Business Objectives**<br>*Background*<br>*Business Objectives*<br>*Business Success Criteria*<br>**Assess Situation**<br>*Inventory of Resources*<br>*Requirements, Assumptions, and Constraints*<br>*Risks and Contingencies*<br>*Terminology*<br>*Costs and Benefits*<br>**Determine Data Mining Goals**<br>*Data Mining Goals*<br>*Data Mining Success Criteria*<br>**Produce Project Plan**<br>*Project Plan*<br>*Initial Assessment of Tools and Techniques* | **Collect Initial Data**<br>*Initial Data Collection Report*<br>**Describe Data**<br>*Data Description Report*<br>**Explore Data**<br>*Data Exploration Report*<br>**Verify Data Quality**<br>*Data Quality Report* | **Select Data**<br>*Rationale for Inclusion/ Exclusion*<br>**Clean Data**<br>*Data Cleaning Report*<br>**Construct Data**<br>*Derived Attributes*<br>*Generated Records*<br>**Integrate Data**<br>*Merged Data*<br>**Format Data**<br>*Reformatted Data*<br>*Dataset*<br>*Dataset Description* | **Select Modeling Techniques**<br>*Modeling Technique*<br>*Modeling Assumptions*<br>**Generate Test Design**<br>*Test Design*<br>**Build Model**<br>*Parameter Settings*<br>*Models*<br>*Model Descriptions*<br>**Assess Model**<br>*Model Assessment*<br>*Revised Parameter Settings* | **Evaluate Results**<br>*Assessment of Data Mining Results w.r.t. Business Success Criteria*<br>*Approved Models*<br>**Review Process**<br>*Review of Process*<br>**Determine Next Steps**<br>*List of Possible Actions*<br>*Decision* | **Plan Deployment**<br>*Deployment Plan*<br>**Plan Monitoring and Maintenance**<br>*Monitoring and Maintenance Plan*<br>**Produce Final Report**<br>*Final Report*<br>*Final Presentation*<br>**Review Project**<br>*Experience Documentation* |

**شکل (۷): فازها و فعالیت‌های عمومی CRISP-DM [۹]**




**شکل (۸): مدل فرآیندی SEMMA [۱۰]**




**شکل (۹): مدل فرآیندی Human Centered [۱۱]**

خود ندارد و این باعث شده است تا بسیاری از فعالان بازار به فراخور حال خود نسخه‌های ترکیبی یا بروزشده از این مدل ارائه و استفاده کنند که به مجموعه آنها My own گفته می‌شود.

۳- غول بزرگ محصولات هوش تجاری یعنی SAS بخش مهمی از سهم این بازار را در اختیار دارد که محصولات آن از یک مدل فرآیندی ساده بنام SEMMA استفاده می‌کنند.

## ۲-۳- مدل فرآیندی SEMMA

مدل فرآیندی sample, explore, modify, model, assess (SEMMA) در سال ۲۰۰۵ توسط شرکت بزرگ SAS معرفی شده و در ابزار تجاری داده‌کاوی محبوب SAS Enterprise Miner یکپارچه‌سازی شد. این متدولوژی و ابزار فوق‌الذکر در اغلب پروژه‌هایی که توسط این شرکت اجرا می‌شوند استفاده می‌گردد [۱۰].

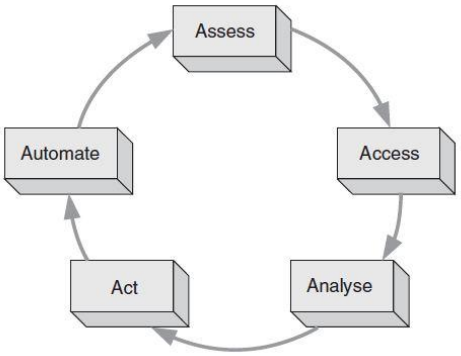
تفاوت این مدل با مدل‌های مرجع KDD و CRISP-DM در آنست که این مدل بیشتر روی جنبه‌های فنی داده‌کاوی تمرکز داشته و عملاً فاز اول یعنی شناخت کسب و کار و همچنین فاز آخر یعنی اعمال و استفاده از نتایج اجرای پروژه را نادیده گرفته است که جهت اجرای موفق یک پروژه داده‌کاوی بسیار مهم قلمداد می‌شوند.

## ۲-۴- مدل فرآیندی Human Centered

این مدل توسط برچمن و آناند در سال ۱۹۹۶ به عنوان یک دیدگاه عملگرایانه از فرآیند داده‌کاوی و با تاکید بر ماهیت تکرارشوندگی این فرآیند پایه‌ریزی شده و در ادامه توسط گرتوزیو و داساوچوی در سال ۲۰۰۶ بسط داده شد [۱۱]. گام‌های اساسی این مدل در شکل (۹) نمایش داده شده است که عبارتند از شناسایی پروژه، شناسایی داده، پاکسازی داده، توسعه مدل، تحلیل داده و تولید خروجی‌های لازم. آشکارا مشخص است که این مدل بر پایه مدل KDD Process Model بنا شده و منطبق بر آن است؛ با این تفاوت که تمرکز این مدل بر انجام کارها از دیدگاه مهندس داده‌کاو است در حالیکه که KDD Process Model بر جریان داده بین فعالیت‌های مختلف داده‌کاوی تاکید دارد.

## ۲-۵- مدل فرآیندی 5A's

متدولوژی 5A's توسط شرکت SPSS در سال‌های ۲۰۰۳ و ۲۰۰۷ و مبتنی بر ۵ مفهوم assess، access، analyze، act و automate به ترتیب ارائه و توسعه یافت که یک نمای کلی و عمومی از فرآیند تحلیل داده و داده‌کاوی ارائه می‌کرد [۱۲]. شکل (۱۰) یک نمای کلی از فازهای این متدولوژی را نمایش می‌دهد. مهمترین نوآوری مدل 5A's در فاز automate نهفته است بطوریکه خودکارسازی فرآیند اجازه می‌دهد افراد غیرخبره در زمینه داده‌کاوی با اجرای مدل موجود بر داده‌های جدید کسب دانش کنند و نیازی به تکرار فرآیند عریض و طویل داده‌کاوی مرتفع گردد. هر چند نکته منفی آنجاست که تکرار مجدد فرآیند ممکن است به بهبود نتایج یا درک بهتر از داده‌ها یا دانش قابل‌اقتباس کمک کند که با اعمال مدل قبلی اینگونه پتانسیل‌ها نادیده گرفته می‌شود. در نهایت این متدولوژی با ادغام شرکت SPSS با IBM به کار خود خاتمه و جای خود را به CRISP-DM داد.




**شکل (۱۰): مدل فرآیندی 5A's [۱۲]**

## ۲-۶- مدل فرآیندی 6-Sigma

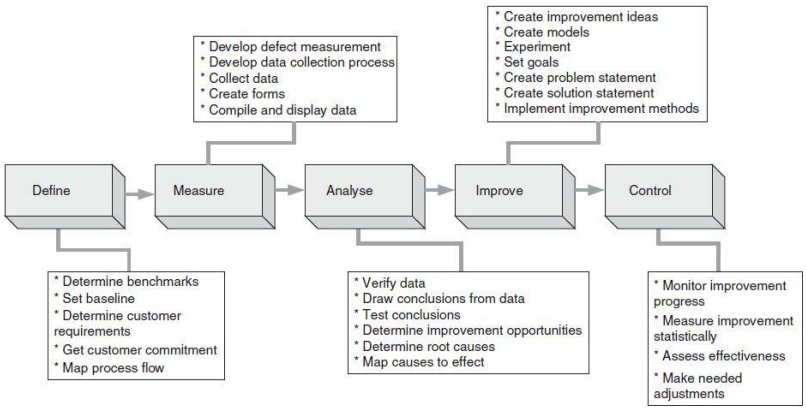
این متدولوژی توسط شرکت موتورلا در سال ۱۹۹۶ به عنوان یک رویکرد برای افزایش کیفیت مدیریت بطور عام ارائه شد [۱۳]؛ بطوریکه کیفیت پروژه و رضایت مشتریان و ذی‌نفعان بطور همزمان دیده شود. مراحل و فعالیت‌های این متدولوژی در شکل۱۱ نشان داده شده است و بر پایه DMAIC یعنی define، measure، analyze، improve و control استوار است. در پی موفقیت شرکت موتورلا، این متدولوژی توسط شرکت‌های مختلفی پذیرفته و اعمال شد و در نهایت با پروژه بزرگ StatSoft در سال ۲۰۰۵ به پروژه‌های داده‌کاوی نیز راه یافت و جزء مدعیان این حوزه قلمداد می‌شود.

## ۲-۷- مدل فرآیندی Cabena

کابینا و همکاران در سال ۱۹۹۷ یک مدل فرآیندی هوش تجاری ارائه کردند که از آن به فرآیند کامل استخراج دانش از پایگاه داده یاد می‌شود [۱۴]؛ یعنی علاوه بر مراحل و فازهای اصلی داده‌کاوی شامل فاز نهایی تحلیل و بازنگری خروجی و دست‌آوردهاست که بعدها در پروژه‌های هوش تجاری جزء ملزومات و فازهای اصلی قرار گرفت. مراحل این متدولوژی در شکل (۱۲) نشان داده شده است و عبارتند از انتخاب داده، پیش‌پردازش، تبدیلات، داده‌کاوی، تحلیل نتایج و استخراج دانش قابل‌استفاده. در حقیقت این متدولوژی در ساختار با KDD Process Model یکسان بوده و در تعداد مراحل تا حدودی متفاوت است.

## ۲-۸- مدل فرآیندی Two Crows

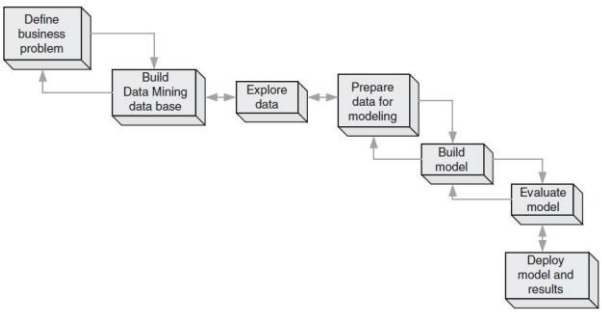
این متدولوژی فرآیندی هوش تجاری در سال ۱۹۹۹ توسط شرکت Two Crows و بر پایه مفاهیم نسخه‌های اولیه CRISP-DM ارائه شد که بطور کامل از فازها و فعالیت‌های هوش تجاری پشتیبانی می‌کند [۱۵]. مراحل و فازهای این متدولوژی در شکل (۱۳) نشان داده شده است و عبارتند از:




**شکل (۱۱): مدل فرآیندی 6-Sigma [۱۳]**

1- Define Business problem
2- Build data mining database
3- Explore data
4- Prepare data for modeling
5- Build model
6- Evaluate model
7- Deploy model and results

اگر چه این متدولوژی خطی به نظر می‌رسد اما چرخه‌ها و تکرارهای زیادی ممکن است بین مراحل بوجود آید.




**شکل (۱۳): مدل فرآیندی Two-Crows [۱۵]**

## ۲-۹- مدل فرآیندی Anand & Buchner

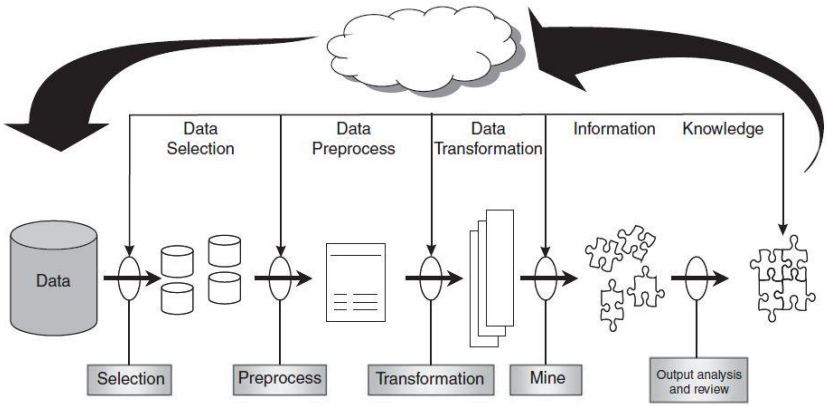
آناند و بوچنر در سال ۱۹۹۹ مدلی را ارائه کردند که ادعا می‌شد تمامی چرخه حیات یک پروژه هوش تجاری از طریق اینترنت (بطور دقیقتر از طریق وب) را شامل می‌شود و برای پروژه‌های وب‌کاوی بومی‌سازی شده است [۱۶]. مراحل و فعالیت‌های این مدل در شکل (۱۴) نشان داده شده است و عبارتند است:

1- Human resource identification
2- Problem specification
3- Data prospecting
4- Domain knowledge elicitation
5- Methodology identification
6- Data preprocessing
7- Pattern Discovery
8- Knowledge postprocessing

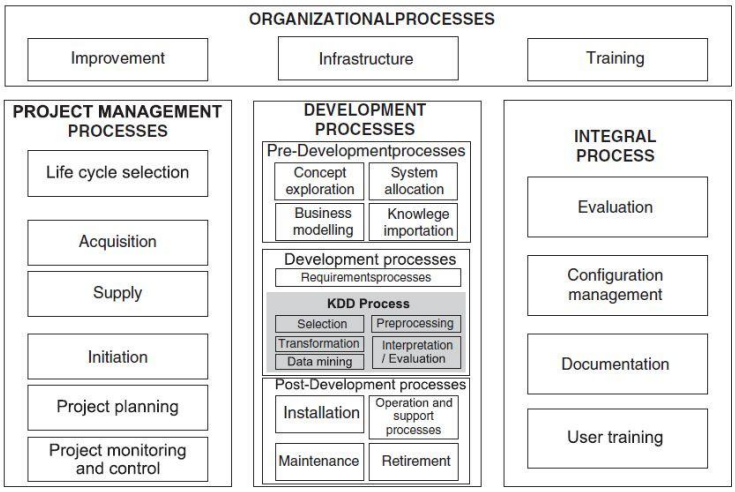
متخصصان عقیده دارند که اگر چه این مدل تمامی تلاش خود را برای ارائه جزییات اجرای فعالیت‌ها در فازها و گام‌های اولیه متدولوژی انجام داده است اما فاقد دستورالعمل‌های لازم جهت اعمال نتایج و خروجی‌های پروژه در سازمان است.

## ۲-۱۰- مدل فرآیندی Marban

این متدولوژی برگرفته از مدل فرآیندی CRISP-DM اذعان می‌دارد که اگر

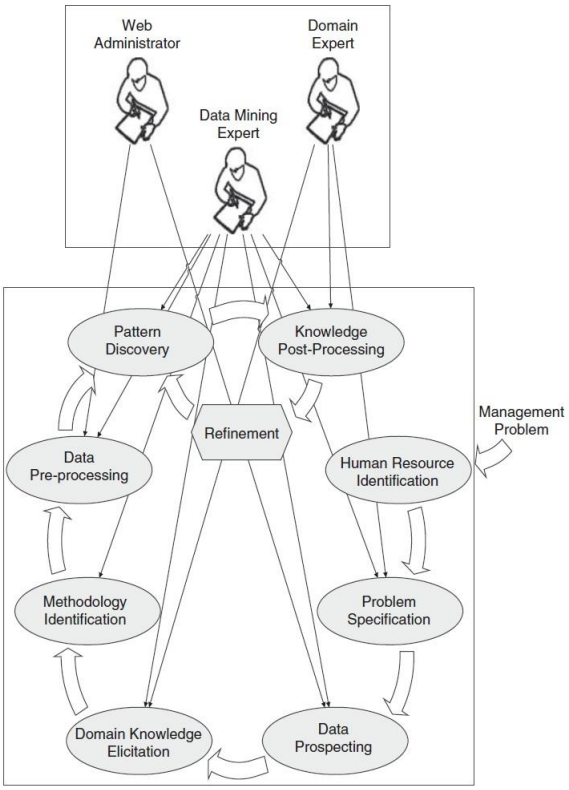



**شکل (۱۲): مدل فرآیندی Cabena [۱۴]**




**شکل (۱۵): مدل فرآیندی Marban [۱۷]**

داده‌کاوی یک فعالیت مهندسی و ساخت‌یافته است پس می‌بایست تمامی فعالیت‌ها با جزییات شامل دستورالعمل‌ها و استانداردهای لازم تعریف شده باشد [۱۷]؛ چیزی که CRISP-DM فاقد آنست. شمای کلی این متدولوژی در شکل (۱۵) نشان داده شده است. در حقیقت این متدولوژی استانداردها، به‌روش‌ها و دستورالعمل‌های مدیریتی و توسعه‌ای لازم از سایر نظام‌های مهندسی، مانند استانداردهای IEEE 1074 و ISO 12207 مدل فرآیندی توسعه نرم افزار، را انتخاب کرده و با فرآیند داده‌کاوی ترکیب می‌کند تا الگویی مشخص و با جزییات حاصل شود.



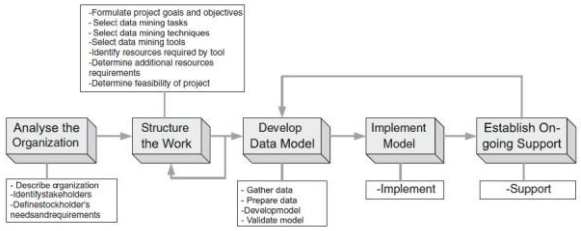

**شکل (۱۴): مدل فرآیندی Anand & Buchner [۱۶]**

## ۲-۱۱- مدل فرآیندی DMIE

این مدل بر پایه تحقیقات سولارته در سال ۲۰۰۲ جهت اجرای پروژه‌های داده‌کاوی در صنعت معرفی شد که بر پایه CRISP-DM استوار بود با این تفاوت که در تعداد و عنوان فازها تغییراتی اعمال شد [۱۸]. مراحل و شمای کلی این متدولوژی در شکل (۱۶) نشان داده شده است و عبارتند از:

1- Analyze the organization
2- Structure the work
3- Develop data model
4- Implement model
5- Establish on-going support

در حقیقت مهمترین نوآوری این متدولوژی در فاز آخر یعنی حمایت مستمر نهفته است جایی که حمایت و نگه‌داری مستمری از فرآیند داده‌کاوی از جمله پشتیبان‌گیری و نگه‌داری داده‌ها، بروزرسانی مدل‌ها و نرم‌افزارها (در صورت لزوم) همواره در حال اجراست.




**شکل (۱۶): مدل فرآیندی DMIE [۱۸]**

## ۲-۱۲- مدل فرآیندی RAMSYS

این متدولوژی توسط مویل و جورگه در سال ۲۰۰۱ متناسب با پروژه‌هایی ارائه شد که در آن ذینفعان مختلف پروژه داده‌کاوی بطور جغرافیایی پراکنده بودند

[۱۹]؛ لذا در این متدولوژی سعی شده است که فرآیند حل مساله با موضوعات اشتراک دانش و سهولت در ارتباطات ترکیب گردد. شمای کلی این متدولوژی در شکل (۱۷) نشان داده شده است. آنگونه که پیداست این متدولوژی نیز یک شکل پالایش شده از CRIDP-DM است تا چارچوبی مهیا گردد که دست‌اندرکاران پروژه بتوانند براحتی از راه دور فعالیت خود را انجام داده و به اشتراک گذارند.

## ۲-۱۳- مدل فرآیندی Cios

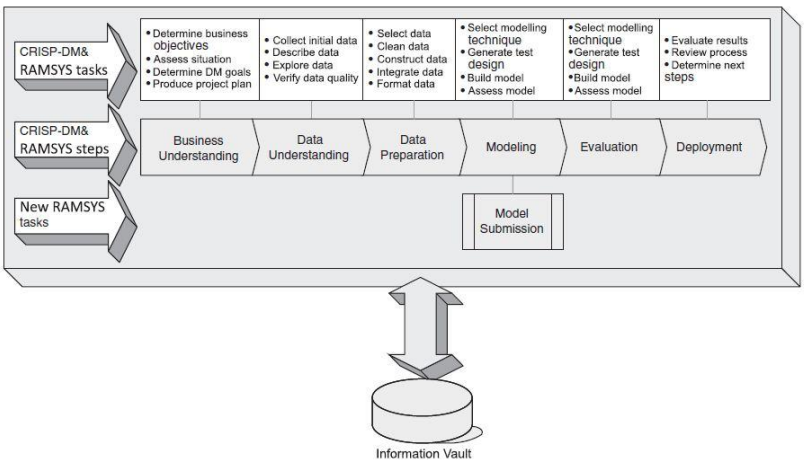
همانگونه که گفته شد CRISP-DM یک متدولوژی جهت اجرای پروژه‌های داده‌کاوی در بخش صنعت است که مقبولیت بسیاری دارد. بر همین اساس مدل فرآیندی Cios در سال ۲۰۰۰ بر پایه مدل CRISP-DM ارائه شد تا متناسب با نیازهای پروژه‌های تحقیقاتی جامعه اکادمیک باشد [۲۰]. شمای کلی این متدولوژی در شکل (۱۸) نشان داده شده است. این متدولوژی یک متدولوژی محاوره‌ای، بازخوردی و تکرارشونده است که در آن از تکنیک‌ها و تکنولوژی‌های جدید متنوعی مانند XML، PMML، SOAP، UDDI و OLE BD-DM استفاده شده است.

## ۲-۱۴- مدل فرآیندی KDD Roadmap

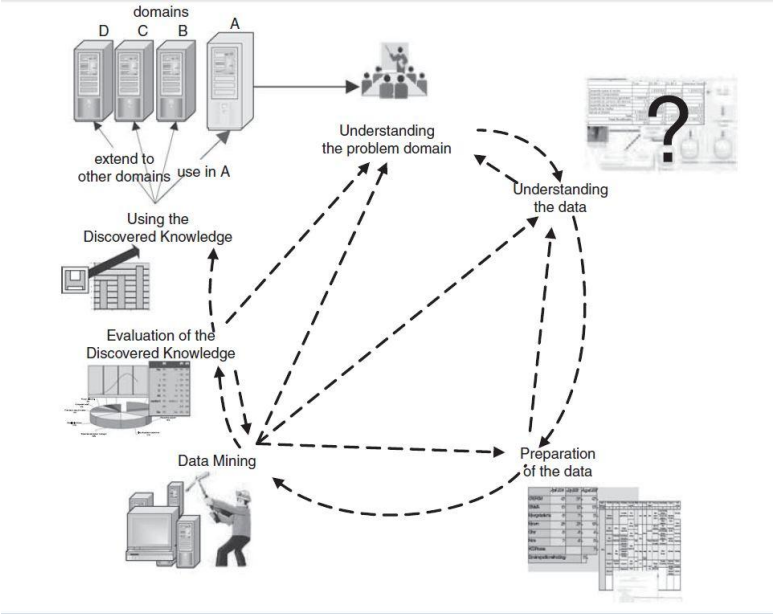
این متدولوژی توسط دبیوس و همکاران در سال ۲۰۰۱ ارائه شد و بعدها در سال ۲۰۰۶ به عنوان یک مدل فرآیندی داده‌کاوی در ابزار Witness Miner جاسازی شد [۲۱]. شکل (۱۹) فازهای اصلی این متدولوژی را نشان می‌دهد که عبارتند از:

1- Problem specification
2- Resourcing
3- Data cleansing
4- Preprocessing
5- Data mining
6- Evaluation
7- Interpretation
8- Exploitation

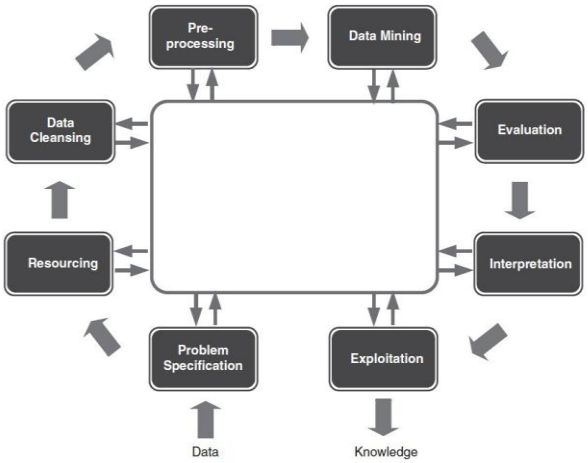
در حقیقت نوآوری این متدولوژی در فاز Resourcing است که تاکید می‌کند ادامه اجرای پروژه با توجه به منابع موجود مدیریت شود.




**شکل (۱۷): مدل فرآیندی RAMSYS [۱۹]**




**شکل (۱۸): مدل فرآیندی Cios [۲۰]**




**شکل (۱۹): مدل فرآیندی KDD Roadmap [۲۱]**

## ۳- نتیجه‌گیری

در این مقاله متدولوژی‌ها و مدل‌های فرآیندی پروژه‌های داده‌کاوی و هوش تجاری معرفی شده و نقاط ضعف و نوآوری ارائه‌شده توسط هر یک بحث و بررسی شد. در حقیقت مدل‌های فرآیند داده‌کاوی زیادی در دنیا مطرح شده است که با توجه به هدف ارائه به دو دسته تحقیقاتی و صنعتی دسته‌بندی می‌شوند. KDD Process و CRISP-DM به ترتیب پرارجاع‌ترین مدل-های تحقیقاتی و صنعتی هستند که اکثر مدل‌های بعدی شالوده خود را از آنها اقتباس کرده‌اند. به لحاظ میزان استفاده در پروژه های تحقیقاتی و صنعتی طبق گزارش گارتنر، CRISP-DM در رده اول و My own و SEMMA به ترتیب در رتبه‌های بعدی قرار دارند که My own به مجموع مدل‌های بهبودیافته و ترکیبی گفته می‌شود که توسط افراد و موسسات کوچک ارائه شده و استفاده می‌شوند. بدین‌ترتیب نتیجه گرفته می‌شود که مدل‌هایی در دنیا مقبولیت و محبوبیت پیدا می‌کنند که یا توسط افراد پیشرو در این زمینه و بر بستر برآیند نظرات محققان و متخصصان این حوزه شکل گرفته باشند (مانند KDD Process) و یا توسط شرکت‌های پیشرو و بر پایه تجربیات اجرای پروژه‌های متعدد در این زمینه اقتباس شده و از حمایت این شرکت‌ها برخوردار باشند (مانند CRISP-DM و SEMMA که به ترتیب توسط شرکت‌های IBM و SAS تحت حمایت قرار دارند).

## مراجع

[1] Chen, Hsinchun, Roger HL Chiang, and Veda C. Storey, "Business intelligence and analytics: from big data to big impact," *MIS Quarterly*, vol. 36, no. 4, pp. 1165-1188, 2012.

[۲] رئوف خیامی، الهام پروین نیا، امیر درجه، *معماری سازمانی در عمل*. جلد ۱، موسسه فرهنگی هنری دیباگران تهران، ۱۳۹۵.

[3] L. A. Kurgan and P. Musilek, "A survey of Knowledge Discovery and Data Mining process models," The Knowledge Engineering Review, vol. 21, no. 1, pp. 1–24, 2006.

[4] Mariscal, Gonzalo, Oscar Marban, and Covadonga Fernandez. "A survey of data mining and knowledge discovery process models and methodologies." *The Knowledge Engineering Review* 25, no. 2 (2010): 137-166.

[5] KdNuggets.Com, Data Mining Methodology, 2007.

[6] Piatetsky-Shapiro, Gregory. "Knowledge discovery in real databases: A report on the IJCAI-89 Workshop." *AI magazine* 11, no. 4 (1990): 68.

[7] Piatetsky-Shapiro, Gregory. *Advances in knowledge discovery and data mining*. Edited by Usama M. Fayyad, Padhraic Smyth, and Ramasamy Uthurusamy. Vol. 21. Menlo Park: AAAI press, 1996.

[8] Wirth, Rüdiger, and Jochen Hipp. "CRISP-DM: Towards a standard process model for data mining." In *Proceedings of the 4th international conference on the*

*practical applications of knowledge discovery and data mining*, pp. 29-39. 2000.

[9] Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C. & Wirth, R. 2000. CRISP-DM 1.0 Step-by-Step Data Mining Guide. Technical report, CRISP-DM.

[10] SAS Institute, SEMMA Data Mining Methodology, 2005.

[11] Brachman, R. J. & Anand, T. 1996. The process of knowledge discovery in databases. Advances in Knowledge Discovery and Data Mining. American Association for Artificial Intelligence, 37–57.

[12] de Piso´ n Ascacibar, F. M. 2003. Optimizacio´n Mediante Te´cnicas de Mineri´a de Datos Del Ciclo de Recocido de Una Li´nea de Galvanizado. PhD thesis, Univeridad de la Rioja.

[13] Harry, M. & Schroeder, R. 1999. Six Sigma, the Breakthrough Management Strategy Revolutionizing the World's Top Corporations. Currency.

[14] Cabena, P., Hadjinian, P., Stadler, R., Verhees, J. & Zanasi, A. 1997. Discovery Data Mining. From Concept to Implementation. Prentice Hall.

[15] Two Crows Corporation 1998. Introduction to Data Mining and Knowledge Discovery, 2nd edition. Two Crows Corporation. ISBN 892095-00-0.

[16] Anand, S. & Buchner, A. 1998. Decision Support Using Data Mining. Financial Times Management, 184.

[17] Marba´ n, O., Mariscal, G., Menasalvas, E. & Segovia, F. J. 2007. An engineering approach to data mining projects. Lecture Notes in Computer Science 4881, 578–588. Springer.

[18] Solarte, J. 2002. A Proposed Data Mining Methodoloy and Its Aplication to Industrial Engineering, Master's thesis, University of Tennessee, Knoxville.

[19] Moyle, S. & Jorge, A. 2001. Ramsys—a methodology for supporting rapid remote collaborative data mining projects, ECML/PKDD 2001 Workshop on Integrating Aspects of Data Mining, Decision Support and Meta-Learning: Internal SolEuNet Session, 20–31.

[20] Cios, K. J. & Kurgan, L. A. 2005. Trends in data mining and knowledge discovery. In Advanced Techniques in Knowledge Discovery and Data Mining, Pal, L. C. Jain, N. (eds), Advanced Information and Knowledge Processing. Springer, 1–26.

[21] Debuse, J. C. W., de la Iglesia, B., Howard, C. & Rayward-Smith, V. 2001. Building the KDD Roadmap: A Methodology for Knowledge Discovery. Industrial Knowledge Management. Springer-Verlag, 179–196.