

#### مقدمه

بازار ابزار استخراج داده ها از دو راه ابتدایی خود در حال ظهور می باشد . بسیاری از ابزارهایی که در اینجا توضیح داده می شوند ، در مرحله اول انتشار می باشند.

موقعیت در بازار CRM که عموماً بخشی از سیستم تجارت الکترونیکی در نظر گرفته می شود ، پیچیده تر می باشد و بنابراین با سرعت وب یا شبکه در حال حرکت می باشد. بازار CRM ، حتی بیشتر از بازار ابزار استخراج داده ها با چندین فروشنده که بر تعریف خود بازار و موقعیت خود در این بازار متمرکزند ، توصیف می گردد.

این اشتباه، با ماهیت بسیار دینامیک خود بازار که یک فعالیت قابل رویت تحکیم مشتری، شرکتهای ادغامی و تملیک ها را تحمل می کند، بیشتر می گردد. علی رغم کل این چالشها، باز رو به تکامل می رود و فروشندگان، پیشرفت مهمی در علمی بودن ابزار، قابلیت استفاده و قابلیت اداره کسب می کنند.

اولین بخش این فصل ، به کاربردهای بسته بندی شده استخراج داده ها می پردازد. این کاربردها ، بر اساس چندین تکنیک استخراج داده ها ادغام شده در ابزارهای بهتر می باشد . همراه با بهترین عملکرد ها ، اسلوب شناسی های خوب تعریف شده و فرآیندها، راه خود را در محیط های تولید شرکتها که در آن استخراج داده ها بخشی از یک فرآیند موسسه ای شده می شود می یابند که شامل رشد و یادگیری سازمانی می شود .

## بازار استخراج داده ها

بیانید بازار استخراج داده ها را از نقطه نظر منحنی اقتباسی تکنولوژی در نظر بگیریم ایمنی به اقتباس کنندگان اولیه ، از تکنولوژی لبه یادگیری برای دستیابی به مزیت رقابتی استفاده می کنند ؛ هنگامیکه تکنولوژی تکامل می یابد ، شرکت های بیشتری آن را اقتباس می کنند ، و در یک حالت تجارت زمانی و عادی درج می نمایند . همچنین مناطق عملی بودن ابزارهای استخراج داده ها بزرگتر و بزرگتر می شوند. به عنوان مثال ، تکنولوژی وایت اوک ( یک شرکت استخراج داده ها در مریلند) از جانب کمیسیون فدرال الکترون ، مجوز فروش سیستم گچین ماینر Capain Miner را کسب کرده است که بی نظمی در دخالت های سیاسی فدرال را کشف می کند . نورتل ، یک بسته کشف کلاهبرداری را توسعه داده است به نام سوپر اسلوت فراود ادوایسور ، که از تکنولوژی های شبکه عصبی استفاده می کند .

صنعت ابزار استخراج داده ها ، برخلاف تکنولوژی های استخراج داده ها ، در مرحله عدم تکامل قرار دارد و می کوشد تا بازار را تعیین نماید . و وجودش را تأیید کند . به همین دلیل است که در می یابیم بازار ابزارهای استخراج داده ها تحت تاثیر موارد زیر قرار دارد:

□ ادغام پیوسته و مداوم ابزارها با اتکاء به تکنولوژی های مکمل و به عنوان

مثال OLAP

□ ظهور کاربردهای بسته بندی شده عمودی و یا اجزاء استخراج داده ها

برای توسعه کاربرد .

□ استراتژی های بسیار اقتباس شده شرکت بین فروشندگان ابزار استخراج

داده ها و فروشندگان تهیه کنندگان راه حل جامع و ادغام کننده های

سیستم ها : فروشندگان مقیاس مؤسسه ، همانند NCR IBM ، اوراکل ،

میکروسافت ) به عنوان مثال ، اوراکل چندین شریک متعدد استخراج داده ها

به عنوان بخشی از او را کل ویرهاوس اینتیشیتیو از جمله آنگاسن دیتا

مایند ، دیتا پیکیج اینفورمیشن دیسکاویری ، SPSS ، SRA اینترنشنال و  
تینکینگ ماشینز را انتخاب کرده است .

### اصول طبقه بندی ابزارهای استخراج داده ها

ما می توانیم کل بازار ابزارهای استخراج داده ها را به سه گروه اصلی  
تقسیم نمائیم ؛ ابزارهای دارای هدف کلی ، ابزارهای ادغام شده استخراج داده ها  
DSSOLAP ، و ابزارهای به سرعت در حال رشد و برای کاربرد.  
ابزارهای هدف کلی بخش بزرگتر و کامل تر بازار را اشغال می کند . آنها بنا  
به تعریف و برای کاربرد نیستند و حوزه آنها از نظر ماهیت افقی است . این ابزار  
شامل موارد زیر می شود .

• SAS اینترپراز ماینر

• IBM اینتلیجنت ماینر

• یونیکا PRW

• SPSS کلمنتین

• SGI ماین ست

• اوراکل داروین

• آنگاس نالج سیکر

بخش ابزار مرکب یا ادغام شده استخراج داده ها بر شرط تجاری بسیار  
واقعی و اجباری داشتن ابزار چند منظوره تقویت تصمیم تاکید می کند که گزارش  
مدیریت ، پردازش تحلیلی روی خط ، و قابیت های استخراج داده ها در یک قالب  
کاری عادی را فراهم می کند . نمونه های این ابزار های مرکب شامل کاکنوس  
سیناریو و بیزینس آبجکت می شود.

بخش ابزارهای ویژه کاربرد ، به سرعت در حال حرکت است ، و  
فروشنندگان در این فضا ، می کوشند تا خود را با ارائه راه حل های تجاری به جای  
جستجوی تکنولوژی برای یک راه حل ، از ساب رین متمایز نمایند . حوزه این ابزار ،

بنا به تعریف از نظر ماهیت عمودی است . در بین این ابزارها ، موارد زیر قرار دارند:

- KD1 ( متمرکز بر خرده فروشی است )
- حق انتخابها و انتخابها ( بر صنعت بیمه متمرکز است )
- HNC ( بر کشف کلاهبرداری متمرکز است )
- یونیکا مدل ۱ ( بر بازاریابی متمرکز است )

#### ارزیابی ابزار : صفات و اسلوب شناسی ها

کل این عوامل ارائه یک توصیف بهینه از ابزارهای استخراج طولانی تر موجود را مشکل ساخته است . بنابراین . بطور کلی ابزارهای استخراج داده ها را می توان با استفاده از صفات زیر ، توصیف نمود :

□ تکامل محصول و ثبات و استحکام شرکت . به دلیل عدم تکامل کلی بازار تجاری برای ابزارهای استخراج داده ها ، این مقوله محصولاتی را توضیح می دهد که برای جنبش از چند سال وجود داشته اند.

□ سکوها و معماری. سکوهای متنوع از نظر تجاری موجود را تقویت کنید ، مدل چند تأبیری مشتری / خادم را تقویت کنید . هدف قابلیت سنجش ، دسترسی به سرعت نزدیک به خطی و سنجش زمان اجرا ، به عنوان میزان داده ها ( اندازه بانک اطلاعاتی ) ، تعداد متغیرها ، و تعداد کاربران متصل به هم می باشد که در حال رشد هستند .

□ تداخلهای داده ها ، قابلیت ابزار برای دسترسی به بانکهای اطلاعاتی ارتباطی ، فایل های یکنواخت و سایر فورمتها.

□ قابلیت های استخراج داده ها از جمله تکنیکها ، الگوریتم ها و کاربردهای تحلیلی ، تکنیک های استخراج داده ( ANN ، CART ، ایفاء قانون و غیره ) با تداخل عادی کاربر که ابزار می تواند آن را تقویت نماید، توانایی ایجاد و



مقایسه چندین مدل ؛ و توانایی برای تقویت تعدادی از انواع مختلف تجزیه و تحلیل از جمله طبقه بندی ، پیشگویی و کشف اتحاد.

□ ایجاد داده ها قابلیت تغییر شکل و طبقه بندی متغیرهای پیوسته ، ایجاد متغیرهای جدید ، استفاده از تاریخها و زمان ، استفاده از ارزشهای از دست رفته و غیره.

□ مدل (خصوصیت تفسیر ، ارزیابی ، صف بندی) . این مقوله ، ابزار را از نظر قابلیت انجام کارهای زیر ، ارزیابی می کند:

- شناسایی مدل بطور اتوماتیک یا دستی ، توسط کاربر
- توضیح نتایج و تعریف مقایسه های اعتمادی یعنی احتمالات طبقه بندی ، حدود اعتماد و غیره

• ارزیابی نتایج مدل با درجه و تناسب

• گزارش دخالت و سهم هر متغیر در مدل ، گزارش درجه افزایش و غیره

• صف بندی مدل برای امتیاز بانکهای اطلاعاتی منبع

• استخراج مقرارت از مدل

• دسته بندی قوانین استخراج شده در یک شکل استاندارد ( به عنوان مثال کد

SQL، اظهارات روشی )

□ عملکرد

• اعتبار پیشگویی اعتبار بر اساس نرخ خطای نمونه پیشگویی می باشد ؛

اعتبار مدل را می توان با درجه افزایش اندازه گیری نمود.

• کارایی پردازش . اجزای بهینه زمان اجرای الگوریتم ها

□ تداخلهای کاربر . این مقوله در می یابد آیا ابزار یک کاربر مبتدی و یا

کارشناس را تقویت می کند یا نه و موارد زیر را تقویت می نماید:

• الگوهای سنجش تجاری . قابلیت ایجاد الگوهای از قبل بسته بندی شده ،

برای تقویت بسته های تجاری ویژه ( به عنوان مثلا بازاریابی هدف ، امتیازدهی

اعتبار، کشف کلاهبرداری ).

• تداخل متادیتا . قابلیت تقویت وظیفه طراحی معنایی ؛ دستیابی به متادیتا  
برای کدگذاری جداول مقادیر مقوله ای ، دستیابی به قوانین استخراج ، تغییر شکل  
و مهاجرت به تعریف دیدگاه تجاری خروجی مدل  
□ قابلیت اداره

• کنترل پیچیدگی . این یکی از روشهای اصلاح تصمیم مدل است ، بنابراین  
مدل را بیشتر قابل اداره می نماید . به عنوان مثال ، استفاده از پارامترهای بسیار  
آزاد منجر به تناسب بیش از حد می گردد ( قبلاً بحث شد ) ؛ این پیچیدگی را  
می توان با زوال وزن ، کنترل نمود . الگوریتم های درخت تصمیم CART از فاز  
شاخه زنی یا ضریب برای کاهش پیچیدگی مدل درخت استفاده می نمایند .  
• سختی . اعتبار درخت و موجود بودن به قابلیت یا بهبود سریع از یک نقطه  
ضعف . استفاده از اجزاء کلیدی به منظور ادغام در قالب کاری مدیریت سیستم ها  
( به عنوان مثال BMCpatrol ، Tivolitme10 و غیره ) .

• مرسوم سازی . قابلیت ابزار برای ایجاد یک معماری باز، انعطاف پذیر و  
قابل گسترش با تداخلهای برنامه ریزی کاربرد منتشر شده و نقاط خروجی کاربر  
که امکان این کاربردهای تقویت تصمیم، ویژه کاربر و همینطور ادغام با سایر  
کاربردها را فراهم می آورد.

هر گاه که ممکن باشد، ما ابزارهایی که از این مقوله ها استفاده می کند را  
مورد بحث قرار خواهیم داد. با این وجود ، به دلیل این که هدف این فصل ، ایجاد  
یک بررسی کلی از ابزارهای برجسته و استخراج داده ها می باشد و صفات یا  
خصوصیات عملکرد که می توانیم تنها با انجام یک ارزیابی جامع از محصول با  
استفاده از داده ها بدست آوریم ، در این بحث در نظر گرفته نخواهد شد . دیگر این  
که فروشنده یک معیار منتشر شده یا اطلاعات بررسی کاربر در مورد عملکرد  
ابزار را فراهم نماید.

## ارزیابی ابزار

ابزارهایی که بطور مفصل در این فصل بحث شد ، شامل کلمنتین (SPSS) ، کاکنوس 4t hought ، کاکنوس سناریو ، دینا ماینر دیتا ماینر ( ، داروین ( اوراکل ) ایستگاه کاری استخراج بانک اطلاعاتی (HNC) سریهای تصمیم ( نئویستا ) ، اینترپرایزماینر SAS . اینتلیجنت ماینر (IBM) ، نالوج سیکر و نالوج استادیو (AGNOSS) ، و مدل 1 و PRW ( یونیکا ) . ابزارهای دیگر بطور مختصر مورد بحث قرار می گیرند . سطح جزئیات موجود برای برخی از این ابزارها به دلیل فقدان تکامل محصول محدود می باشد . این ابزارها به ترتیب حروف الفبا مورد بحث قرار می گیرند.

### کلمنتین (SPSS)

بررسی ، کلمنتین ، یک ابزار کاربرد استخراج داده ها از SPSS می باشد . این ابزار یک تداخل کاربر گرافیکی را با چندین نوع تکنیکهای تجزیه و تحلیل از جمله شبکه های عصبی به قوانین اتحاد و تکنیکهای القاء قانون ، ترکیب می کند . این قابلیت های فنی ، در یک محیط برنامه ریزی دیداری که استفاده از آن ساده است ارائه می شوند .

استعاره گرافیکی که کلمنتین استفاده می کند، استعاره گرفتن ، انداختن و اتصال گره های عمل روی صفحه است. گره هایی برای دسترسی داده ها ، ساخت داده ها ، تجسم داده ها ، یادگیری ماشین ، تجزیه و تحلیل مدل وجود دارد . فرایند توسعه مدلی مرکب از انتخاب گره های صحیح از یک پالت ، قرار داده آنها روی صفحه و اتصال گره ها می باشد.

کلمنتین دارای یک سری غنی از قابلیت های دسترسی اطلاعات از جمله فایل های یکنواخت و بانکهای اطلاعاتی ارتباطی ( از طریق ODBC ) می باشد . بیشتر ، کلمنتین قابلیت دائمی ساختن نتایج طراحی را از طریق نوشتن آنها در ODBC – Compliant DBMS دارد.

ساخت داده های ورودی شامل هماهنگی در نماها و توانایی اشتقاق حوزه های جدید ، می شود . قابلیت های تجسم داده های کلمین شامل نمودارهای پراکندگی ، طرحهای خطی و تجزیه و تحلیل وب می باشد.

کلمنتین روی سیستم های اینتل پنتیو اجرا می شود که سریهای ویندوز ، NT ، HP 9000 که HPUX10 و بالا سیلیکون گرافیکس که IRIX ، Sun SPARC که سولاریس 2.x و دیجیتال APLHA که دیجیتال UNIX 3.X یا VMS 6.X را اجرا می کند .

دسترسی به اطلاعات ساخت و پیش پردازش . کلمنتین فایلهای متن تحریر شده ، فایلهای ارزش مجزا شده با کاما و فایلهای ثبت ثابت (ASCII) را وارد می کند . سایر منابع اطلاعاتی از طریق یک تداخل ODBC حمایت می شوند . سیستم های بانک اطلاعاتی ارتباطی اصلی ، از جمله اوراکل سیباس ، اینفورمیکس و CA-IN gres از طریق ODBC در دسترس می باشد.

قابلیت های ساخت داده های کلمنتین شامل موارد زیر می شود:

- ادغام ثبت ها از طریق توالی ثبت
- تراز داده ها با افزایش نسبت ثبت ها با خصوصیات ویژه
- تراکم تعیین شده کاربر
- تصفیه ثبتهای بیگانه و نامربوط
- اشتقاق جدید حوزه با استفاده از فرمولهای تعیین شده کاربر و اپراتورهای

منطقی

- قابلیت های نمونه گیری داده ها ، از جمله اولین و آخرین N ثبت ، 1 در N نمونه گیری ، و نمونه گیری اتفاقی تکنیک های استخراج داده ها ، الگوریتم ها و کاربردها

- کلمنتین - الگوریتم های القایی قانون - شبکه های عصبی و از جمله شبکه های کوهونن ، او قوانین مربوطه را تقویت می کند :



• درک الگوریتم های القاء قانون کلمنتین ساده می باشد : الگوریتم ها هنگامیکه آموزش داده شوند یک درخت تصمیم را ایجاد می کنند که قانون را نشان می دهد . یک فرآیند دائمی که دنبال می شود ، قرار دارد و متغیرهای مهم در سطوح بالای درخت و سپس آموزش شبکه عصبی با این متغیرها می شود .  
• شبکه های عصبی کلمنتین در توپولوژیهای متنوع و روشهای آموزش ، ارائه می شوند . شبکه معیوب لایه پنهان برای ارزیابی عملکرد  
• کلمنتین شبکه های عصبی سولمونس را برای حل مسائل دسته بندی ارائه می کند .

• قوانین اتحاد همانگونه که از نامش پیداست . یک نتیجه ویژه را به یک سری از صفات ، مربوط می نماید . اتحادها را می توان بین صفات یافت که برای کاربردهایی همانند تجزیه و تحلیل سبد بازار ، بسیار مفید است .  
• کلمنتین برش عمودی مشتری ، تجزیه و تحلیل سریهای زمانی ، تجزیه و تحلیل سه بازار و کشف کلاهبرداری را تقویت می کند .  
کار با ابزار مدل در محیط برنامه ریزی دیداری با استفاده از استعاره گره های متصل ، مشخص می گردد . جعبه های دیالوگ با گره های طراحی مدل وجود دارد که به کنترل الگوریتم ها و روشهای آموزشی کمک می کند .  
کلمنتین به کاربران امکان می دهد تا در یابند کدام ورودیها در مدل دارای اهمیت در پیشگویی کننده هستند ، گر چه تغییر شبکه های عصبی ذاتاً مشکل است . الگوریتم های درخت تصمیم یک بررسی قانون دارای تاثیر متقابل ، که استفاده از آن ساده است را تقویت می کند .

کلمنتین حدودی از وظایف را برای ارزیابی مدل ، ارائه می دهد . اینها شامل یک گره تجزیه و تحلیل می شود که تعداد تشخیص های صبح را برای ورودی مدل گزارش می کند ؛ مقادیر اطمینان متغیر از ۰ تا ۱ ، گره ماتریس که کاربر می تواند جدول سازی در زمینه های انتخابی را در آن انجام دهد .

کلمتین می‌تواند درختان تصمیم، شبکه‌های عصبی، و شبکه‌های کوملونس را به عنوان کد C، صادر نماید. بعد از آنکه مدل در کلمتین ساخته شد، آن را می‌توان به عنوان کد C که باید در محیط خارجی صف بندی شود، صادر نمود. به عنوان مثال، اگر مدل شبکه عصبی بایستی صادر شود. کلمتین سه فایل را صادر خواهد نمود: یک سر فایل، یک فایل وظیفه، و یک فایل تعریف شبکه. قوانینی که الگوریتم القاء قوانین ایجاد می‌کند، قابل صدور می‌باشد. محیط برنامه ریزی دیداری کلمتین، برای یک مبتدی، مناسب است. توالی طراحی به طور مشخص ارائه می‌گردد و حق انتخابهای متعدد، انعطاف پذیری را افزایش می‌دهند. یک کاربر کارشناسی که مایل به ساخت الگوریتم‌های کلمتین می‌باشد، آن آزادی را ندارد. با این وجود، حذف انتخابهایی در شبکه‌های عصبی برای تنظیم نرخ یادگیری و کنترل خستگی مشتری در الگوریتم‌های درخت تصمیم وجود دارد. کلمتین الگوهای تجاری ارائه نمی‌کند. تداخل مقاومتها در میزان نوشتن این بحث وجود نداشت. نتیجه گیری. کلمتین یک محصول قوی است. در معیارهای منتشر شده مشتری، آن برحسب قابلیت سنجش اعتبار پیشگویی کننده و زمان پردازش خوب کار می‌کند. بطور کلی، کلمتین، بسیار خوب با اجراهای تجزیه و تحلیل مقیاس اندک و بزرگ، همانگ بود.

#### ۴ تفکر و سناریو (cognos)

بررسی کوگنوس، به عنوان یکی از رهبران در بازار OLAP، با معرفی دو ابزار استخراج داده‌ها در موجودی دارایی خود از نظر مقام، ارتقاء یافته است: این ابزارها یا بهره برداری از شبکه‌های عصبی و تکنیکهای CHAID قابلیت های استخراج داده‌ها را ارائه می‌کند. بطور کلی این ابزارها در سکوها کاکنوس تقویت می‌شوند. و بطور مشخص در محیط ویندوز مورد توجه قرار می‌گیرند. هر

دو ابزار داده ها را از صفحه گسترده بانکهای اطلاعاتی و فایل های متن Ascii  
تحصیل می کند .

4 thought برای رسیدگی به مسائل تجاری متنوع از جمله بهینه سازی قیمت،  
پیش بینی تقاضا و پیشگویی و سنجش عملکرد ایجاد می گردد . 4 thought از  
پیشگویی چند لایه ای ( MPL ) تکنولوژی شبکه عصبی استفاده می کند که بسیار  
برای تجزیه مسائلی که به اشکال غیر خطی ، داده های اغتشاش و مجموعه های  
کوچک داده ها می پردازد، هماهنگ است . دو تجزیه و تحلیل اصلی که 4 thought  
ارائه داد تجزیه و تحلیل سری های زمانی و برش عمودی مشتری هستند . تجزیه و  
تحلیل سری های زمانی ، به دنبال روندهایی بر اساس رفتار متناوب هستند . در  
صورتی که ، ایجاد برش عمودی مشتری به داده های دموگرافیک ، به عنوان مثال  
برای پیشگویی این که آیا مشتری یک محصول ویژه را خریداری می کند یا نه  
می پردازد.

سناریو برای طبقه بندی و مسائل اتحاد، طراحی می شوند ؛ آن می تواند  
روابطی را بین متغیرها در مجموعه داده ها بیابد . سناریو از کشف تاثیر متقابل  
اتوماتیک مربع خی دو (CHAID) استفاده می کند ، و می تواند اشتقاقهای چند جهته  
را برای متغیرهای پیشگویی کننده فراهم نماید . سناریو ، استراتژی های تجزیه و  
تحلیل متفاوت ، نمونه گیری اتوماتیک و تصمیم و ایجاد درخت را ارائه می دهد . با  
این وجود ، سناریو برای رسیدگی به متغیرهای هدف مقوله ای طراحی نمی شود.  
سناریو در کشف و تجسم داده ها قوی است.

ثبات سناریو و 4 thought ، تداخل کاربر جامع و درک کننده آن است، اساساً  
در یک کاربر مبتدی هدف قرار می گیرد . آن از ترسیم نمودار برای تجزیه و تحلیل  
تک متغیره با استعاره درخت تصمیم استفاده می کند ، و نمایش دیداری خوب را  
برای تغییرات اطلاعاتی بین بخشها برای تجزیه و تحلیل روند و همبستگی عامل  
فراهم می سازد . سناریو ، دارای الگویی برای تجزیه و تحلیل اطلاعات سود از  
داده های رقابتی و فصلی می باشد.

کاکنوس ذکر کرد که آن 4 thought و سناریو را با پاور پلی ( OLAP ) و کاربردهای (گزارشی ) Impromptu ترکیب خواهد نمود.

محصولات کاکنوس در سیستم های عملیاتی ویندوز میکروسافت اجرا می شوند (ویندوز 95 یا با راحتی ، و یا NT ).

دسترسی به داده ها ، ساخت و پیش پردازش . 4 thought می تواند داده ها را به شکل صفحه گسترده های اکسل به فایلهای لوتوس 123 ، فایلهای SPSS ، فایلهای متن صدور شده ، فایلهای ارزش با کاما جدا شده فایلهای دارای پهنای ثابت ، وارد نماید . بعلاوه ، دسترسی به بانکهای اطلاعاتی ارتباطی از طریق محرک odbc از کتابخانه و بانک اطلاعاتی Q&E تقویت می شود.

سناریو می تواند داده ها را از فایلهای یکنواخت ، صفحه گسترده اکسل ، و جداول dBase ، وارد نماید همچنین ، سناریو می تواند از کاکنوس ایمبرومتو برای وارد ساختن داده ها از بورلاند اینترپریس ، سنئورا SQLBase ، اینفورمیکس ، میکروسافت SQL سرور ، اوراکل ، سیبیس SQL سرور ، و مسیرهای بانک اطلاعاتی اصلی از جمله مسیر MDI DB2 ، مسیر omni SQL ، مسیر اوراکل ترانس پرنس ، مسیر سیبیس نت ، استفاده می کند .

قابلیتهای ساخت اطلاعات شامل موارد زیر می شود.

- طبقه بندی
- جایگزینی ارزش از دست رفته ( با یک ارزش مقدم یا استفاده از تفسیر)
- تبدیل ارزش رشته متن به ارزش عددی
- تبدیل خودکار سیاه به صفر
- تصفیه ثبت های بیگانه یا دور دست
- قابلیت های محدود نمونه گیری داده ها و 4 thought می تواند نمونه گیر اتفاقی را تقویت نماید ، اما نمونه گیری لایه لایه وجود ندارد .



### تکنیک های استخراج داده ها ، الگوریتم ها و کاربردها

thought 4 از تکنولوژی شبکه عصبی پیشگویی چند لایه استفاده می کند که طراحی ورودی به خروجی را بر اساس محاسبات گره های متصل بهم ایجاد می نماید . خروجی هر گره یک تابع غیر خطی مجموع وزن دار و ورودیهای از گرمای دو لایه های قبلی می باشد.

سناریو از الگوریتم درخت تصمیم بر اساس CHALD استفاده می کند . آن قوانینی را ایجاد می کند که می تواند برای مجموعه داده های طبقه بندی شده برای پیشگویی این که کدام ثبت ها پیامد مطلوب خواهند داشت که بکار برد . الگوریتم درخت تصمیم سناریو . به اندازه کافی انعطاف پذیر است که حق انتخاب تقسیم هر متغیر یا یک اشتقاق بر اساس اهمیت آماری به کاربر می دهد .

سناریو ، طبقه بندی وسائل اتحاد را تقویت می کند، و می توان در ترسیم برش عمودی و تقسیم بندی بازار ، از آن استفاده نمود . thought 4 برای تجزیه و تحلیل سریهای زمانی طراحی مدل پیشگویی کننده و پیش بینی مسائل بهتر مناسب است . هر دو ابزار ، امکان تجزیه گرافیکی مواد خام از طریق نقشه های خطی میله ای و نمودار پراکندگی را فراهم می آورد. سناریو منظره گراف و درخت را که بر اطلاعات کلی مشابه تکیه دارد ، پیشنهاد می دهد. سناریو می تواند از متغیرهای دو تایی مداوم ، و پیشگویی کننده مقوله ای استفاده نماید ؛ اما آن تنها از متغیر عددی به عنوان اهداف استفاده می نماید.

### سناریو سه استراتژی تجزیه و تحلیل را ارائه می دهد.

- حالت طبقه بندی . حداقل به ۱۰۰۰ ثبت نیاز دارد ، که نیمی از آن برای تست حفظ می شود و به تولید نتایج با اطمینان بالا ، توجه می کند .
- حالت تست یا آزمایش ، نیازمند ثبت های کمتری می باشد ، و از یک سوم این ثبت ها برای تست استفاده می کند . و بطور کلی معیارهای کمتری بی ارزش را برای دسته بندی ارائه می دهد .

• حالت بررسی . به حداقل شرایط ثبت نیاز ندارد و برای بررسی داده های

اولیه طراحی می شود .

کار با ابزار در سناریو ، ، مدلی در یک محیط برنامه ریزی دیداری با استفاده از یک افسونگر مشخص می گردد. سناریو بطور خودکار انواع متغیر را انتخاب می کند ، و به کاربر امکان می دهد تا داده ها را نمونه گیری کند و داده ها را به نمونه های آموزشی و تست، تقسیم نماید . همانگونه که قبلاً ذکر شد ، سناریو سه استراتژی تجزیه و تحلیل را ارائه می دهند که شناسایی ، تست و بررسی سناریو می تواند یک درخت را بطور خودکار ایجاد کند، یا به کاربر امکان می دهد تا هر بخش و قسمت را به یک درخت تبدیل کند . کاربر می تواند بطور مداوم به سادگی محصور شوند و سناریو فهرستی از متغیرهای از نظر آماری مهم درجه بندی شده را ایجاد می کند ، و می تواند بطور خودکار به متغیرهای دارای بالاترین رتبه تقسیم شود.

4 thought به کاربر امکان می دهد تا مدل شبکه عصبی را با اولین تصمیم گیری در این مورد که آیا آن برای تجزیه و تحلیل سریهای زمانی می باشد یا نه ، شناسایی نماید . کاربر در مورد ماهیت سریهای زمانی تصمیم می گیرد (روزها، هفته ها و ماهها و سالها و یا زمان مشخص شده توسط کاربر). برای سریهای زمانی و تجزیه و تحلیل طراحی مدل پیشگویی کننده، کاربر ، مدل را از طریق جعبه شناسایی و یا اگر مدلی ایجاد می نماید . کاربر دارای انعطاف پذیری انتخاب یک شبکه تک لایه دو لایه ، می باشد. 4 thought دو حالت از تست مدل را برای تعیین زمان توقف ارائه می دهد :

تست ساده : ابزار داده ها را به گروههای آموزشی و تست تقسیم می کند ، و از یک گروه تست در سراسر فاز آموزشی برای تعیین نقطه انقضاء و خاتمه ، استفاده می کند .

تست کامل : که ابزار چندین مدل را ایجاد می کند ، و سعی می کند تا یک

نقطه توقف بهینه را بیابد.

4 thought ، یک تفسیر گرافیکی و یک نمودار پراکنندگی مدل در برابر داده های واقعی و همینطور فهرست دینامیکی از متغیرهای بحرانی را برای کاربر فراهم می کند .

در سناریو ، خواه یک نمودار یا منظره درخت استفاده شود ، کاربر می تواند خلاصه ، آمار همانند میانگین متوسط انحراف معیار و غیره را مشاهده نماید 4 thought یک گزارش آماری کامل ، یک گزارش مسطح مقطع و گزارش سناریو را ارائه می دهد . گزارش آخر ، به کاربر امکان می دهد تا مقادیر کل متغیرهای ورودی را برای برگشت پاسخ خروجی انتظاری مدل شناسایی نماید .

مدلهایی که توسط سناریو ساخته شده قابل صدور نیستند ، اما شرحهایی تعقیبی طراحی می شوند که دارای قوانین قابلیت های صدور می باشد . 4 thought می تواند در ارتباط با اکسل لوتوس 123 و SPSS کار نمایند . و می توانند یک مدل آموخته شده را به صورت یک تابع ریاضی در هر یک از این محیط ها ، صف بندی نمایند .

نتیجه گیری . 4 thought و سناریو هر دو قابلیت های حمایت تصمیم کوکناس را توسعه می دهند . و عملی بودن استخراج داده ها را ایجاد می نماید . در معیارهای منتشر شده مشتری سناریو ، نتایج خوب و تداخل دوستانه کاربر را نشان می دهد . 4thought دسترسی به درجه بندی های معقول بر اساس عملکرد و اعتبار را گزارش می کند .

### داروین ( اوراکل )

بررسی . داروین که اغلب ، یکی از ابزارهای ابتدایی استخراج داده ها در نظر گرفته می شود ، شهرتش را تأیید می کند . اخیراً ، اوراکل ، داروین را از شرکت Thinking Machine برای افزایش پیشنهاد و تولید خود بدست آورد و خصوصاً در فضای CRM که استخراج داده ها همانگونه که قبلاً بحث شد ، نقش بحرانی ایفا می کند . بحث زیر بر اساس خصوصیات داروین می باشد ، درست همانگونه که

Thinking Machine توسعه داده و بازاریابی کرد. و ارواکل، تصمیم می‌گیرد تا هر مشخصه، اجزاء و معماری ابزار را تغییر دهد.

ابزار استخراج داده های داروین به عنوان یک محصول پیچیده حاوی ابزار استخراج داده ها طراحی شد: شبکه های عصبی، درخت k نزدیکترین همسایه. ابزار شبکه عصبی داروین (Darwin Net) یک سری جامع از توابع را برای ساخت مدل فراهم می‌کند. آن می‌تواند از پیشگویی کننده های مقوله ای و مداوم و متغیرهای هدف استفاده کند. و می‌توان برای مسائل طبقه بندی، پیشگویی، و پیش بین مورد استفاده قرار داد.

ابزار درخت تصمیم (درخت داروین) از الگوریتمهای CART استفاده می‌کند. می‌توان برای حل مسئله طبقه بندی با متغیرهای مداوم و مقوله ای مورد استفاده قرار داد. ابزار k نزدیکترین همسایه (darwin Match) را می‌توان برای حل مسئله طبقه بندی متغیرهای وابسته مقوله ای و متغیرهای پیشگویی کننده مقوله ای و پیوسته مورد استفاده قرار داد.

اگر چه هر ابزار جزء دارایی نقصهای است. اما Darwin شامل یک سری کامل از توابع برای ارزیابی مدل می‌شود. آن خلاصه آمار، ماتریس اغتشاش و جداول درجه افزایش را برای کل انواع مدلها، ایجاد می‌کند.

داروین تداخلهای نسبتاً قوی کاربر را برای افراد مبتدی همینطور کارشناسها ارائه می‌دهد، اگر چه تداخلها بهتر با یک کاربر کارشناس هماهنگ هستند.

داروین که از یک شرکت معروف به عنوان یکی از اولین سازندگان کامپیوتری موازی حاصل شده است، مزیت زیادی در پردازش کارایی و قابلیت انعطاف پذیری دارد. الگوریتم های آن برای اثبات موازی بهینه می‌شوند و به اندازه کافی بهینه هستند که در معماریهای موازی و سری اجرا می‌شوند. این قابلیت بطور مشخصی از توجه ارواکل فرار نکرد و به ارواکل کمک می‌کند تا به



یک فروشنده اولیه بانک اطلاعاتی و محصولات کاربر تبدیل شوند که می تواند یک مؤسسه جهانی بزرگ را درجه بندی کند.

داروین به عنوان یک سیستم ارباب رجوع / سرور یا خادم طراحی می شود و سکوی خادم از پردازشگرهای متحد تا SMPS تا پردازشگرهای موازی از جمله سیستم های وینتل که در ویندوز NT اجرا می شوند و همینطور محصولات مبتنی بر UNIX از میکور سیستم های SUN ، HP ، IBM ، و NER ، Compag/Digital متغیر است.

دسترسی به داده ها ، ساخت و پیش پردازش. داروین می تواند داده های محدود با طول ثابت را از فایل های یکنواخت (ASCII) و بانک های اطلاعاتی ارتباطی از طریق ODBC وارد نماید . بطور داخلی ، داده ها در یک قالب و فورمت اختصاص ذخیره می شوند که می توان بطور مؤثر در یک سیستم چندین پردازشگر ، طرح بندی نمود .

قابلیت های ساخت داده های داروین شامل موارد زیر می شود.

- قابلیت ادغام برای مجموعه داده های مجزا
- حذف متغیرها از مجموعه اطلاعاتی
- تعیین نوع متغیر ( به عنوان مثال مقوله ای مرتب شده )
- تبدیل به یک مجموعه داده های سری به مجموعه داده های موازی
- نمونه گیری داده ها و جداسازی
- تکنیک های استخراج داده ها ، الگوریتم ها و کاربردها. الگوریتم های اصلی که در زمان چاپ تقویت و تأیید می شوند ، و شبکه های عصبی ، درخت تصمیم نزدیکترین همسایه می باشد.
- الگوریتم های آموزشی شبکه عصبی شامل انتشار رو به عقب ، تندترین فرود یا نزول ، نیوتن اصلاح شده و غیره می باشد . توابع انتقال شامل خطی ، سیگموند ، هیپوتانژانت می شوند.

درخت تصمیم از الگوریتم های CART استفاده می کند که می تواند به طور خودکار درخت را با انتخاب تعداد درختان فرعی که در تجزیه و تحلیل در نظر گرفته می شود . هرس نماید.

الگوریتم K نزدیکترین همسایه بر اساس تکنیک استدلال مبتنی بر حافظه ( MBR) می باشد آن یک ارزش متغیر وابسته را بر اساس واکنشهای K نزدیکترین ثبت هماهنگ کننده در مجموعه آموزشی پیشگویی می کند که در آنجا نزدیکی همسایه با به حداقل رساندن فاصله وزن دار Euclidean بین متغیرها ، تعیین می شود.

شبکه عصبی داورین را می توان برای ساخت مدل های پیشگویی کننده و پیش بینی کننده مورد استفاده قرار دارد و متغیرهای مقوله ای و مداوم را بکار برد . درخت تصمیم و ابزارهای k نزدیکترین همسایه را برای مسائل طبقه بندی مورد استفاده قرار داد.

کار با ابزار داورین ، یک سری غنی از حق انتخابها را برای شناسایی مدل فراهم می نماید . به عنوان مثال برای شبکه عصبی ، داورین امکان شناسایی معماری شبکه عصبی ، توپولوژی ، تابع های انتقال الگوریتم آموزشی و تابع هزینه حالت یادگیری و حداکثر تعداد تکرارهای آموزشی را فراهم می آورد . این حق انتخابها کاربر کارشناس را هدف قرار می دهد.

با این وجود تفسیر شبکه های عصبی ذاتاً مشکل است. داورین ، خلاصه ای از معماری مدل ، توپولوژی الگوریتم ها ، و تابع ها را ارائه می دهد جزء درخت تصمیم مدل را به صورت یک مجموعه ساده از قوانین اگر - پس توضیح می دهد که کاربر می تواند آن را آزمایش نماید.

داورین یک مجموعه جامع از تابعها را که برای ارزیابی مدل از جمله گزارشات در مورد آمار خطا ، فهرستی از خطاهای طبقه بندی نادرست ، گزارشات مقایسه ای از پیامد پیشگویی شده و واقعی ، ماتریس اغتشاش و جدول درجه افزایش فراهم می نماید.

نتیجه گیری . مزیت داروین در تقویت الگوریتم ها و با طرحهایی برای افزودن الگوریتم های ژنتیکی و منطق نامعلوم می باشد . آن را می توان در چندین سکو در شکل ارباب رجوع / خادم اجرا نمود، که خادم ممکن است یک پردازشگر یا چندین پردازشگر متقارن، یا پردازشگر موازی باشد . در حالت خادم چند پردازشگر ، داروین می تواند از مشخصه های قابلیت سنجش سخت افزار سود ببرد . در معیارهای منتشر شده یا مشتری ، داروین عملکرد و قابلیت مقایسه قوی را نشان داد . بطور کلی داروین برای اجراها در قیاس متوسط و بزرگ ، مناسب می باشد . به عنوان مثال اخیراً داروین برای کاربردهای رابطه مشتری وفاداری مشتری توسط GTE و Credit suisse انتخاب گردید.

#### ایستگاه کاری استخراج بانک اطلاعاتی (HNC)

بررسی HNC یکی از موفق ترین شرکتهای استخراج داده ها می باشد . ایستگاه کاری استخراج بانک اطلاعاتی (DMW) یک ابزار شبکه عصبی است که بطور گسترده برای کاربردهای تجزیه و تحلیل کلاهبرداری کارت اعتباری قابل قبول می باشد . DMW مرکب از کاربردهای نرم افزار مبتنی بر ویندوز و یک مورد پردازش مرسوم می باشد . سایر محصولات HNC شامل کاربردهای فاکون و پرافیت ماکس برای خدمات مالی و راه حل کشف کلاهبرداری سیستم کنترل سوء استفاده ارتباطات از راه دور پیشرفته (ATACS) می باشد که HNC قصد دارد تا در صنعت ارتباطات از راه دور ، صف بندی نمایند .

شبکه عصبی DMW ، الگوریتم شبکه عصبی انتشار رو به عقب را تقویت می کند، و می تواند در حالات اتوماتیک و دستی کار کند. مدل آن را می توان با استفاده از آمار جامع و عملی بودن برای محاسبه همبستگی های بین متغیرهای پیشگویی کننده و حساسیت آنها نسبت به یک متغیر وابسته ، تفسیر نمود.

DMW ، تعدادی از حق انتخابهای مشخص کاربر را ارائه می نماید که امکان انعطاف پذیری مهم در اصلاح الگوریتم ، متغیر داده ها و تابعهای ساخت داده ها

را فراهم می‌سازد. DMW بطور مؤثر متغیرهای پیوسته و مقوله ای را مورد استفاده قرار می‌دهد که می‌توان برای مسائل پیشگویی، طبقه بندی و پیش بینی مورد استفاده قرار داد.

DMW، تداخلهایی را برای کاربران مبتدی و با تجربه از جمله حق انتخابهای هماهنگ کننده پیشرفته و امکانات مورد استفاده قرار داد.

DMW، برای ایجاد یک مدل معتبر و مؤثر پیشگویی کننده شهرتی را کسب کرده است، عملکرد پردازش آن، قابلیت سنجش آن و برای تأیید شرایط اصلی پردازش کارت اعتباری کافی می‌باشد.

دسترسی به داده ها، ساخت و پیش پردازش. محصول DMW، بطور مسقیم فایل های ASCII با طول ثابت را تقویت می‌کند. محصول حمل می‌شود و با DDMS/COPY ترکیب می‌شود (شرکت نرم افزار تصویر).

این ابزار می‌تواند DBMS اصلی فورمتهای آماری و صفحه گسترده را در یک فورمت ASCII با طول ثابت مناسب برای DMW را تبدیل نماید.

قابلیتهای ساخت داده های DMW شامل موارد زیر می‌باشد:

- حذف متغیرها از یک مجموعه اطلاعاتی
- نوع متغیر تعریف شده توسط کاربر
- تعداد مقوله های تعیین شده توسط کاربر
- تابع عادی سازی داده های تعیین شده توسط کاربر
- تغییر شکل اتوماتیک متغیرهای مقوله ای به متغیرهای دوتایی
- جایگزینی داده های از دست رفته
- نمونه گیری اطلاعاتی انتخاب شده توسط کاربر
- ضریب تغییر شکل داده های DMW تساویها را حفظ می‌کند که پارامترهای شکل، متغیرها و نوع اطلاعات تابع های عادی سازی نرمال، تعداد مقادیر بی نظیر و ارزشهای اختصاص یافته به داده های از دست رفته را توضیح می‌دهد.



تکنیکهای استخراج داده ها ، الگوریتم ، و کاربردها DMW یک الگوریتم شبکه عصبی انتشار رو به عقب را بکار می برد . یک کاربر می تواند تابع انتقال (آستانه ، منطقی ، خطی ، گواسیان عکس تانژانت ، یا تانژانت هیپربولیک) ، حالت یادگیری و چندین پارامتر دیگر را مشخص کند .

DMW ، مدل های پیشگویی کننده ای را برای مسائل طبقه بندی بر پیشگویی و مسائل پیش بینی ایجاد می کند . کاربرد تحلیلی اصلی آن ، کشف کلاهبرداری ( طبقه بندی ) برای صنعت کارت اعتباری می باشد .

کار با ابزار . DMW یک انعطاف پذیری مهم در شناسایی مدل را ارائه می دهد . مدل را می توان بطور دستی یا خودکار با تعریف معماری اصلی و پارامترهای توپولوژیکی ، مشخص نمود . در حالت اتوماتیک ، DMW می تواند با انجام یک تجزیه و تحلیل دسته بندی در حوزه متغیر ، متغیرهای پیشگویی کند تا مناسب را انتخاب نماید .

DMW ، دو تابع را برای تفسیر مدل ارائه می دهد : تجزیه و تحلیل حسایت و تابع کمک که تک تک سکویی ها را توضیح می دهد . به عنوان مثال این تابع را می توان برای توضیح این که چرا کاربر وام رد شد و عوامل رد تا چه حد قوی بودند مورد استفاده قرار دارد .

برای هر مدل ، DMW یک فایل لوگ و فایل مسابقه را ایجاد می کند که حاوی اطلاعاتی در مورد این که چه تعداد ارزیابی انجام شد ، آمار ارزیابی و جداول پیشگویی های صحیح در برابر پیشگویی های نادرست می باشد . نتایج ارزیابی را می توان در یک صفحه گسترده اکسل برای تجزیه و تحلیل معاملات کارت اعتباری مورد استفاده قرار داد . اگر چه DMW نمی تواند مستقیماً بانک های اطلاعاتی را امتیازدهی نماید ، آن یک API موسوم به Deploy Net را برای صف بندی مدل های ساخته شده با DMW ، ارائه می نماید .

نتیجه گیری . DMW یک محصول کامل و قوی است و در پذیرش بازار ، بسیار موفقیت آمیز بوده است . کاربرد طبقه بندی / کشف کلاهبرداری آن در

زمان واقعی برای تجزیه و تحلیل معاملات کارت اعتباری مورد استفاده قرار می‌گیرد. این یک شاهد قوی برای قابلیت سنجش و عملکرد محصول می‌باشد.

### سریهای تصمیم (Neovisa)

بررسی . سریهای تصمیم راه حل نئوویستا ، یک ابزار استخراج داده های جامع می‌باشد . شرکت از یک شرکت Maspar فروشنده سخت افزار موازی ، ظهور نمود . این نیز مشابه داورین منجر به درک عمیق و تکمیل محکم راه حل استخراج داده های قابل سنجش دارای عملکرد بالا گردید.

سریهای تصمیم یک محیط کشف دانش و استخراج داده ها است که یک سری مرکب از الگوریتم ها را برای تجزیه و تحلیل توصیفی و پیشگویی کننده فراهم می‌سازد . الگوریتم ها بطور مؤثر با کنترل‌های متنوع برای مرسوم سازی اضافی توسط کاربر اجرا می‌شوند . قابلیت های تحلیلی شامل دسته بندی ، قوانین مربوطه ، شبکه های عصبی ، و درختان تصمیم می‌باشند . سریهای تصمیم بطور یکپارچه این الگوریتم ها را با دستیابی به داده ها و موتور تغییر شکل یا تبدیل داده ها ترکیب می‌کند . به شرط سابقه شرکت ، ابزار برای عملیات موازی روی سیستم SMP شرکت ادعا می‌کند به قابلیت سنجش تقریباً خطی دست می‌یابد ، هماهنگ می‌گردد. مدرک قابلیت سنجش ابزار را می‌توان در این حقیقت یافت که در صف بندی آن در صنعت خرده فروشی برای مدیریت موجودی (به عنوان مثال Wal Mort ) ، سریهای تصمیم در یک اساس هفتگی برای تجزیه و تحلیل داده های نقاط فروش در سطح جزئی (SKU) مورد استفاده قرار می‌گیرد . این داده هایی است که حدود ۷۰ گیگا بایت (GB) را نشان می‌دهد.

از نقطه نظر قابلیت استفاده سریهای تصمیم عموماً ، به عنوان ابزاری برای کاربر از نظر فنی آگاه فراهم می‌شوند . نئوویستا ، تداخل‌های GUI را توسعه می‌دهد که استفاده از آنها ساده است و خدمات مشاوره کارشناس را ارائه می‌دهد

و متخصصین ، مهندسین کشف دانش یا KDE نامیده می‌شوند . ) آنها اغلب در محل به عنوان بخشی از نمونه اولیه یا پروژه راهنما کار می‌کنند.

سریه‌های تصمیم در سکوهاى UNIX متنوعی از جمله سان سولاس ، HP-UX و دیجیتال UNIX اجرا می‌شوند .

معماری نرم افزار ابزار مرکب از چندین جزء می‌باشد ، و طرح آن مبتنی بر هدف است. موتورهای استخراج داده ها - شبکه تصمیم ، درخت تصمیم ، دسته تصمیم ، و AR تصمیم در راس دستیابی به اطلاعات و لایه انتقال داده ها ساخته می‌شوند که به عنوان موتور دیگر که دستیابی به تصمیم نامیده می‌شود به اجرا می‌گردند موتورهای استخراج داده ها، خواص دستیابی به تصمیم را دارد ، و بنابراین به سادگی به هم مرتبط می‌گردند.

دستیابی به اطلاعات ، ساخت و پیش پردازش . سریه‌های تصمیم ، هر گونه منبع اطلاعاتی را در فورمت ASCII وارد می‌کند و قابلیت های DBMS توسعه می‌یابد ، و در موتور دستیابی به تصمیم ، ادغام می‌گردد . سریه‌های تصمیم ، متادیتا را در ساختارهای فایل اختصاصی نگاه می‌دارد . این فایلها را می‌توان با یک موتور خواند و سپس به بعدی منتقل نمود . اطلاعات را می‌توان اضافه کرد ، اصلاح نمود و برای استفاده آتی ، ذخیره ساخت قابلیت‌های ساخت داده‌های سریه‌های تصمیم شامل موارد زیر می‌گردد:

- تنظیم برای حذف داده های دور افتاده
- تبدیلهای فورمت
- تبدیل ثبت ها به ثبت با ستونهای بی نظیر
- رمز دار نمودن . تبدیل داده های عددی و کاراکتر به طرح ثبت های ویژه
- قابلیت های نمونه گیری داده ها . از جمله اولین یا آخرین N ثبت ، ۱ در N نمونه گیری، حدود و دامنه ارزشی و نمونه گیری اتفاقی .
- تکنیک های استخراج داده‌ها و الگوریتم ها و کاربردها . سریه‌های تصمیم ، شبکه های عصبی ، درختان تصمیم ، دسته بندی و قوانین اتحاد را تقویت می‌کند :

□ الگوریتم شبکه های عصبی آن در موتور شبکه تصمیم اجرا می شود .  
آن از یک شبکه سه لایه پیش خور با یک گره ورودی تنها برای هر حوزه ورودی تغییر شکل یافته ، استفاده می کند . الگوریتم از تکنیک اختصاصی آموزش شبکه استفاده می کند از این نظر که آن کل ورودیهای شبکه را به رشته های حوزه های دو تایی تبدیل می نماید ، بی نظیر می باشد . این به تسریع فاز آموزش کمک می نماید . شبکه تصمیم تجزیه و تحلیل عامل را تقویت می کند که مشابه با تجزیه و تحلیل حساسیت ، می توان در سطح تاثیری که هر ورودی بر پیشگویی بازده دارد ، گزارش نمود.

□ موتور درخت تصمیم. این موتور از الگوریتم C4.5 با افزایش های مشخص مربوطه به آموزش درخت و تولید قوانین استفاده می کند . مورد آخر ، می تواند قوانینی را ایجاد کند که بطور متقابل انحصاری نیستند ، که تکنیک خوبی برای ارائه یک درخت پیچیده به عنوان یک سری جامع از قوانین تجاری می باشد .

□ دسته تصمیم . این الگوریتم یک تکنیک یادگیری نظارت شده مشابه با دسته بندی آماری را تقویت می کند. آن می تواند از متغیرهای مقوله ای و عددی استفاده نماید، و به کاربران امکان می دهد تا یکی از ماتریسها ( زاویه و Euclidean ) را برای تعیین فاصله بین دسته و مشاهدات تعیین نماید.

□ AR تصمیم. موتور قوانین اتحاد که بر اساس الگوریتم نظارت نشده القاء قانونی می باشد ، که قوانین را به شکل ( A و به معنای B است ) و ( A و B به معنای C هستند ) ایجاد می کند . هر قانون ایجاد شده دارای یک سطح اطمینان و سطح حمایت می باشند.

حدود کاربرهای تحلیلی که با سریهای تصمیم تأیید می شود ، طراحی طبقه بندی ایجاد قوانین پیش بینی و کاربردهای دسته بندی را برای تجزیه و تحلیل توصیفی در بر می گیرد از جمله سیر بازار و تجزیه و تحلیل الگوی متوالی .



این تجزیه و تحلیل ها را می توان در مجموعه اطلاعات بزرگ انجام داد و بازده یک الگوریتم را می توان به عنوان ورودی الگوریتم دیگر مورد استفاده قرار داد (از طریق موتور دستیابی به تصمیم) بعلاوه نئویستا یک کاربرد تجاری را ارائه می دهد (RDS-Profile) که بر مدیریت موجودی برای صنعت خرده فروشی متمرکز است.

کار با ابزار. سریهای تصمیم، یک حدود از حق انتخاب های فزاینده را برای خصوصیت مدل فراهم می کند. آنها شامل حق انتخابهای معماری مختلف برای شبکه عصبی، محدودیتهای ریشه و برگها در درخت تصمیم و غیره می شود. مدلی را می توان از طریق GUI یا با استفاده از زبان چایی دسترسی به تصمیم (DASL) مشخص نمود.

برای درختان تصمیم، مدل را می توان به صورت یک سری از قوانین نسبتاً ساده توضیح داد تفسیر شبکه های عصبی اساساً مشکل است و قابلیت های سریهای تصمیم از این نظر مستثنی نیستند.

شبکه تصمیم. آماری را برای ارزیابی مدل فراهم می کند (خطای میانگین مربع باقی مانده، میانگین، ماتریس اغتشاش و غیره). درخت تصمیم، ماتریس اغتشاش را همراه با فاصله اطمینان برای اعتبار انتظاری تأیید می کند.

مدلهایی که با شبکه تصمیم درخت تصمیم، و دسته تصمیم توسعه یافتند را می توان در برابر دیگر مجموعه اطلاعات با استفاده از مدلها، مادامی که داده ها توسط دسترسی به تصمیم خوانده و رمزدار می شود، اجرا نمود.

نتیجه گیری. سریهای تصمیم یک محصول قوی با معماری متفکرانه و الگوریتم های استخراج داده ها می باشد. ابزار قابل سنجش است و از معماریهای سخت افزار موازی، سود می برد. در معیار منتشر شده مشتری، آن در طبقات مشخص مسائل (به عنوان مثال بانکداری و کاربردهای خرده فروشی) بر حسب قابلیت سنجش اعتبار پیشگویی کننده و زمان پردازش، خوب کار می کند. بطور

کلی، سریهای تصمیم برای اجراهای تجزیه و تحلیل در مقایس بزرگ، مناسب هستند.

### اینترپرایز ماینر (SAS)

بررسی SAS اینترپرایز ماینر یک بازیگر قوی در بازار ابزار استخراج داده ها می باشد. آن به یک قدرت مشخص و تاثیر ضریب های آماری SAS اتکا می کند و آن ابزار را با تعداد الگوریتم های استخراج داده ها، زیاد می کند. SAS از اسلوب شناسی نمونه، بررسی، اصلاح، مدل، اسلوب شناسی دستیابی برای ارائه ابزار استخراج داده ها که می تواند حدود گسترده ای از مدلها از جمله اتحاد دسته بندی درختان تصمیم شبکه های عصبی و رگرسیون آماری را تأیید کند، استفاده می نماید.

SAS اینترپرایز ماینر، برای استفاده کاربران مبتدی و کارشناس طراحی شده است. تداخل GUI آن با جریان داده ها تحریک می گردد و درک و استفاده آن ساده می باشد. آن به تحلیل گر امکان می دهد تا با ساخت یک نمودار جریان اطلاعاتی دیداری که مانده های داده ها را با استفاده از پیوند ها به گره های پردازش متصل می کند، مدلی را بسازد. بعلاوه، تداخل به کاربر امکان می دهد تا پردازش کد را مستقیماً در جریان داده ها درج نماید.

به دلیل این که چندین مدل تأیید می گردند، اینترپرایز ماینر به کاربر امکان می دهد تا مدل های مختلف را مقایسه و ارزیابی نماید. و با استفاده از گره ارزیابی، بهترین تناسب را انتخاب نماید. بعلاوه اینترپرایز ماینر یک گره امتیاز دهی را فراهم می کند که یک مدل امتیاز دهی را ایجاد می کند که می تواند از طریق هر کاربرد SAS آن را بدست آورد.

SAS اینترپرایز ماینر می تواند به شکل ارباب رجوع / خادم یا تنها، اجرا شود. بعلاوه، در حالت ارباب رجوع / خادم، اینترپرایز ماینر به خادم امکان می دهد تا تنها به شکل خادم داده ها، خادم محاسبه، یا ترکیبی از آن دو باشد. ماینر،

برای اجرا در سکوهاى مورد تائيد SAS طراحی می‌گردد . معماری یک شکل ارباب رجوع « چاق » ( که مستلزم مجوز کامل SAS در ارباب رجوع است ) و همینطور ارباب رجوع لاغر (بروسر) ، را تائيد می‌کند .

دسترسی به داده ها ، ساخت و پیش پردازش . تداخل داده های مستقیم ، از طریق مجموعه اطلاعاتی SAS انجام می‌شود . با این وجود داده ها را می‌توان از طریق روشهای اطلاعاتی SAS استاندارد ، بدست آورد به عنوان مثال SAS / ACCESS برای دسترسی به اطلاعات فورمت کامپیوتر خانگی و RDBMS سیستم های اوراکل ، اینفورمیکس، سیبیس، و DB2 RDBMS از طریق SAS/ACCESS تائيد می‌شوند .

قابلیتهای ساخت داده ها شامل کل مشخصه های موجود از طریق موتور اصلی SAS می‌شود . بعلاوه سیکل تکنیکهای نمونه گیری و جداسازی اطلاعات از طریق گره های مناسب اینترپرایزماینر تقویت می‌شود .  
تکنیکهای استخراج داده ها ، الگوریتم ها و کاربردها ، SAS اینترپرایز ، اتحاد، دسته بندی، درختان تصمیم، شبکه عصبی و تکنیکهای رگرسیون آماری کلاسیک را تقویت می‌کند .

□ اتحادها . این الگوریتم امکان کشف قوانین اتحاد ( به عنوان مثال ، تجربه سبد بازار )

□ دسته بندی : تکنیک یادگیری نظارت نشده استفاده شده برای کشف دانش اولیه و تجسم داده ها .

□ درختان تصمیم . تکنیکهای متعدد درخت تصمیم ، تقویت می‌شوند : CHAID و اینترپوی ریواکشن ( متغیرهای دو تایی و مقوله ای ) ، و F-test و واریانس ریواکشن ( برای متغیرهای هدف وقفه ) .

□ شبکه های عصبی . چندین شبکه عصبی از جمله مولتی لایر پرسپتور ( MLP ) و رادیال پیسیس فانکشن (RBF) تقویت می‌شوند . تابع های انتقال و خطا و همینطور روشهای آموزشی ارائه می‌شوند . این قابلیتها برای

ایجاد عملکرد پیشگویی کننده بهتر از انتشار رو به عقب استاندارد ، طراحی می شوند .

□ رگراسیون . چندین تکنیک رگراسیون قبلاً در SAS استاندارد اجرا گردید ، با اینترپرایز ماینر تقویت می گردد.  
اینترپرایز ماینر ، تجزیه و تحلیل سه بازار ، طبقه بندی ، طراحی پیشگویی کننده ، برش عمودی مشتری و حدود تجزیه و تحلیلهای آماری برای سریهای زمانی اقتصاد سنجی ، تحقیق عملیات و سایر موارد ، را تقویت می کند .  
کار با ابزار . اینترپرایز ماینر ، تعدادی از حق انتخابهایی برای ساخت مدلهای پیشگویی کننده را تقویت می کند .

مشخصات خود ، ترکیبی از اعمال اقتباس و حذف موجود از طریق محیط برنامه ریزی دیداری می باشد. یک مجموعه فزاینده از نقصها ، آن را برای کاربران مبتدی نیز مناسب می سازد.  
اینترپرایز ماینر ، فایل های لوگ و کد سنج SAS را برای تفسیر شبکه عصبی ارائه می دهد.

اینترپرایز ماینر دو روش را برای ارزیابی مدل ها تقویت می کند : از طریق مدل منیجر یا گره های ارزیابی مدل منیجر ، یک ابزار خوب برای تجزیه و تحلیل نتایج خروجی و بازده از یک مدل ویژه می باشد . گره ارزیابی برای ارزیابی تعمیم و سختی یک مدل ، مفید می باشد . هر دو ، نقشه درجه بندی نقشه های سود، ROI و سایر شاخصها را تأیید می کند.

مدلها را می توان مستقیماً با استفاده از گره امتیاز صف بندی نمود که مدل را در ساختار های SAS برای کاربرد متوالی در داده ها صف بندی نمود.  
کار برای کارشناس از گره های کد SAS سود می برند که امکان درج کد پیچیده اختیاری را در جریان داده ها ، امکانپذیر می سازد . تا جایی که به متادیتا مربوط می شود . اینترپرایز ماینر از همان متادیتا همانند سایر قسمتهای سیستم SAS ، استفاده می کند .



نتیجه گیری . SAS به مهارت قابل ملاحظه خود در نرم افزار تجزیه و تحلیل آماری برای گسترش یک سیستم با وظیفه کامل ، معتبر ، راحت و قابل اداره ، متکی بود . حدود گسترده ای از حق انتخابهای طراحی مدل و الگوریتم ها ، تداخل خوب طراحی شده کاربر ، قابلیت اتکا به امتیازات اطلاعاتی موجود و سهم بزرگ بازار در تجزیه و تحلیل آماری ( که به شرکت امکان می دهد تا به جای یک ابزار جدید ، جزء افزایش SAS را تحصیل نماید ) سبب می شود تا SAS یک موقعیت بر جسته را در بازار استخراج داده ها بدست آورد .

### اینتلیجنت ماینر (IBM)

بررسی . شرکت IBM ، راه حلهای استخراج داده ها را برای چندی سال ، با استفاده از منابع قابل ملاحظه کتابخانه های تحقیقاتی آن از ایالات متحده ( به عنوان مثال واتسون ریسرچ ، آلماندن ریسرچ ) و سراسر جهان ( به عنوان مثال ECAM در فرانسه ، بابلینگن در آلمان ) توسعه می داد . نتیجه این تلاشها ، یک دست راه حلهای سخت افزاری هوشمند است که تحقیقات کاربردی و اساسی در زمینه های هوش مصنوعی ، یادگیری ماشین ، تجزیه و تحلیل زبان شناسی و کشف دانش را احاطه می کند .

محصولی که اینجا مورد بحث قرار می گیرد . با نام اینتلیجنت ماینر برای داده ها بازار یابی می شود ( محصول خط سر آن ، اینتلیجنت ماینر برای متن در فصل ۲۰ مورد بحث قرار می گیرد .

اینتلیجنت ماینر IBM ، برای مقام رهبری در بازار ابزار استخراج داده ها رقابت می کند ، زیرا آن موارد زیر را ارائه می دهد :

- یکی از فزاینده ترین مجموعه های تکنیکهای استخراج داده ها و الگوریتم ها .
- قابلیت سنجش بالا بر حسب اندازه داده ها و عملکرد مؤسسه ای ؛ در حقیقت ، محصول برای اجرا در سیستم سخت افزار موازی IBM SP بهینه می گردد ، محصول در بسیاری از سکوها IBM و غیر IBM اجرا می شود .

• مجموعه غنی APTs که می توان برای توسعه کاربردهای مرسوم استخراج داده ها مورد استفاده قرار دارد. کل موتور استخراج داده ها و تابع های ساخت داده ها را می توان از طریق کتابخانه های C++ مشترک بدست آورد.

اینتلجنت ماینر الگوریتم ها را برای طبقه بندی / پیشگویی ، ایجاد قوانین اتحاد ، دسته بندی کشف الگویی متوالی و تجزیه و تحلیل سریهای زمانی تقویت می کند . اینتلجنت ماینر ، با بکارگیری تکنیکهای هوشمند تجسم داده ها و تداخل سخت کاربر مبتنی بر جاوا ، قابلیت استفاده خود را افزایش می دهد ( اساساً کاربر با تجربه را هدف قرار می دهد ) . اینتلجنت ماینر ، سیستم مدیریت بانک اطلاعاتی ارتباطی DB2 را تقویت می کند ، و تعدادی از تابعهای ساخت داده های هوشمند را ترکیب می نماید .

اینتلجنت ماینر یک سیستم ارباب رجوع / خادم است که ارباب رجوع تداخل کاربر و تابع های تجسم داده ها را کنترل می کند ، در صورتی که موتورهای ساخت داده ها و استخراج داده ها در یک خادم قرار دارد . جزء خادم در سکوی زیر تقویت می گردد :

IBM RS/6000 ، IBM SP ، که ALX را اجرا می کند ، IBM S/390 ; I MVS را اجرا می کند ، و IBM AS/400 . جزء ارباب رجوع ویندوز ۹۵ و ویندوز IBM OS/2 NT و IBM RS/6000 را اجرا می نماید که AIX را اجرا می کند .

دسترسی به داده ها ، ساخت و پیش پردازش . اینتلجنت ماینر . فایل های یکنواخت را تقویت می کند ، و بطور مستقیم به DB2 دسترسی دارد . مورد آخر ، به کاربر امکان می دهد تا مدلهای پیشگویی کننده را بطور مستقیم از جداول ارتباطی کشف نماید . خادم DB2 برای ارائه کل قابلیت های ساخت داده ها و تغییر شکل داده ها برای محصول استفاده می شود ، و می تواند به صورت دروازه ای برای سایر منابع داده های ارتباطی عمل نماید . اگر فایل های یکنواخت استفاده شوند ، عملی بودن ساخت داده های DB2 ، فراهم نیست و امکانات اضافی یا تلاشهای توسعه ای برای پیش پردازش و انتقال داده ها ، ضروری است .

با این وجود ، کل نتایج مدل ، در فایل‌های یکنواخت ذخیره می‌شوند .  
اینتلجنت ماینر از ساختار فایل موسوم به Mining Base استفاده می‌کند که بطور  
مؤثر یک فایل متادیتا است که فورمتهای کل فایل‌هایی که محصول استفاده می‌کند  
را تعریف می‌نماید.

قابلیتهای ساخت داده های اینتلجنت ماینر که بر اساس قابلیت‌های DB2  
هستند ، شامل موارد زیر می‌شوند:

- تراکم و محاسبه ای که کاربر تعریف می‌کند .
- تبدیل موقعیت بالا به موقعیت پائین
- تصفیه ثبت های بیگانه و ثبت‌های ارزش از دست رفته
- تصفیه حوزه ها
- تصفیه ثبت های با استفاده از مجموعه ارزش
- دسته بندی ثبت ها و اتصال منابع داده ها
- تبدیل چندین حوزه های مقوله به مجموعه حوزه های دوتایی
- رمز دار کردن مقادیر از دست رفته
- قابلیت نمونه گیری داده ها ، ایجاد نمونه رندم یا اتفاقی
- اجرای SQL
- تکنیک‌های استخراج داده ها ، الگوریتم‌ها و کاربرها . اینتلجنت ماینر ،  
شبکه‌های عصبی ، از جمله نقشه مشخصه کوهوتن ، الگوهای سری‌های زمانی ،  
درختان تصمیم ، دسته بندی ، مقرارت اتحاد ، الگوهای متوالی ، تابع اساس رادیال  
را تقویت می‌کند . اکثر الگوریتم ها که توسط امکانات تحقیق IBM توسعه می‌یابند  
تکنیک‌های اختصاصی IBM هستند . و تنها در اینتلجنت ماینر موجود می‌باشند .
- شبکه های عصبی : شامل دو الگوریتم می‌شود : یک شبکه انتشار رو به  
عقب استفاده شده برای طبقه بندی و یک نقشه مشخصه کوهوتن ، که یک  
تکنیک یادگیری نظارت نشده استفاده شده برای جداسازی و تقسیم ثبت‌ها به  
دسته های مشابه می‌باشد.

- درختان تصمیم . این یک تغییر از الگوریتم CART است که برای ایجاد مدل طبقه بندی استفاده می شود و می تواند از داده های مداوم و مقوله ای استفاده نماید.
- الگوهای سریهای زمان . این الگوریتم توسط IBM Almaden Research توسعه یافت ؛ این برای یافتن الگوها در سریهای زمانی هنگام تصفیه اغتشاش و سر و صدا ، مورد استفاده قرار می گیرد.
- دسته بندی . اینتلیجنت ماینر الگوریتم دسته بندی دموگرافیک را ارائه می دهد که ثبت ها را بر اساس امتیاز تشابه آنها ، ثبت می کند ( مشابه الگوریتم KNN که قبلاً توضیح داده شد. )
- قوانین الگوهای متوالی / اعتماد . این الگوریتم توسط IBM Almaden Research توسعه . مشخصه های بی نظیر آن قابلیت یافتن قوانین اتحاد برای مقوله های موجود در فایل معامله و قابلیت یافتن قوانین از جمله قوانین مرکب و سلسله مراتبی می باشد . در یک فایل معامله ای زمان دار ، این الگوریتم می تواند الگوهای متوالی را برای تجزیه و تحلیل رفتار خرید مشتری و ترکیب سبد بازار ردیابی نماید.
- تابع اساس رادیا ، این الگوریتم ارزش متغیر وابسته را بر اساس رابطه آن با مقادیر حوزه های دیگر در ثبت ، برآورد می کند، این تکنیک برای پیشگویی مداوم ارزش مورد استفاده قرار می گیرد .
- اینتلیجنت ماینر از الگوریتمهای متنوع برای تقویت تعداد و تجزیه و تحلیل ها از جمله کشف اتحادها و الگوهای متوالی و تجزیه و تحلیل سهم بازارها و سریهای زمانی ( تجزیه و تحلیل بازار سهام ) ، طبقه بندی و طراحی برش عمودی مشتری ، دسته بندی و ارزشهای پیشگویی کننده استفاده می کند .
- کار با ابزار . اینتلیجنت ماینر ، تعدادی از پارامترهای سطح بازار را ارائه می دهد که کاربر می تواند برای هر تکنیک استخراج داده ها ، شناسایی نماید .
- پارامترهای نمونه شامل حداقل حمایت و عوامل اطمینان / اتحاد و الگوهای متوالی



نرخهای اعتبار مطلوب ( شبکه های عصبی ) و حداکثر و تعداد مطلوب دسته ها ( دسته بندی ) می شود .

الگوریتم های طبقه بندی ، حالات آموزشی، تست و کاربرد را تقویت می کند و یک ماتریس آشفتگی را به عنوان یک ابزار ارزیابی ارائه می نماید . روشهای ارزیابی دیگر نیز توسعه می یابند.

حالت کاربرد را می توان برای صف بندی مدلها در داده های بیرونی و نه نمونه ( مورد مورد استفاده قرار داد . اینتلیجنت ماینر یک کتابخانه برنامه نویسی گسترده را ارائه می دهد که به کاربر برای ساخت و کاربرد مدلها کمک می کند . غنی بودن الگوریتم ها و اختیارات مرسوم سازی متنوع تداخل کاربر اینتلیجنت ماینر را برای کاربر کارشناس مناسب تر از کاربر مبتدی می نماید.

نتیجه گیری. بطور کلی ، اینتلیجنت ماینر ( برای داده ها ) یکی از قابل سنجش ترین و قوی ترین ابزار ها در بازار می باشد . معیارهای بررسی منتشر شده کاربر نشان می دهد که عملکرد کلی ابزار ، خوب بوده و برخی از الگوریتم ها در کاربردهای مختلف، بهتر سایرین عمل می کنند . IBM منابع مهم را برای ایجاد ابزار به عنوان یک راه حل اولیه برای استخراج داده ها در مقیاس مؤسسه ایجاد کرده است .

### جستجوگر دانش و استادیوی دانش ( آنگئوس )

بررسی . جستجوگر دانش /KS از نرم افزار آنگئوس ، یک ابزار استخراج داده های درخت تصمیم می باشد . آن از یک الگوریتم درخت تصمیم بر اساس CART و CHAID برای یافتن روابط بین متغیر پیشگویی کننده و متغیر وابسته در بانک اطلاعاتی استفاده می کند . همینطور ، KS را می توان برای مسائل طبقه بندی با متغیر وابسته و مقوله ای مورد استفاده قرار داد .

تعیین موقعیت ابتدایی ابزار ، ایجاد قابلیت های استخراج داده ها می باشد .  
تداخل کاربر آن ، یک نمایش گرافیکی مدل درخت تصمیم را ارائه می دهد . و کاربر می تواند هر شعبه را انتخاب کند و دسته بندی متغیر پیشگویی کننده را مشخص نماید . یک حالت اتوماتیک که در آن کل شعبه ها را می توان تولید نمود نیز موجود می باشد . KS تعدادی از قابلیت های تنظیم کننده را از جمله قابلیت اصلاح الگوریتم یا تجدید رشد درخت را به کاربر ماهر ارائه می دهد . KS شامل یک API را در موتور استنباط آماری خود وارد می کند که کاربردهای داخلی نوشته شده در C را برای تولید مدلها و وارد ساختن نتایج امکانپذیر سازد .

KS ، در سکوها ی Wintel که ویندوز Ms را اجرا می کند ، یا چندین سکوی UNIX از جمله Hp UX مولاریس و ALX تقویت می شود .

دسترسی به داده ها ، ساخت و پیش پردازش . KS می تواند داده ها را از تعدادی فورمت فایل ، از جمله فایل های ASCII و Splus وارد کند . آن می تواند داده ها را از بانک اطلاعاتی ارتباطی با استفاده از ODBS وارد نماید .

KS یک تابع ساخت داده ها موسوم به Edit view را ارائه می دهد . استفاده از این تابع به کاربران KS امکان می دهد موارد زیر را انجام دهند .

• تغییر نام متغیر

• حذف متغیر

• تعیین مجدد یک فاصله دسته بندی

• شناسایی نوع متغیر

- وارد ساختن ثبت ها از یک منبع ODBC مبتنی بر یک بررسی SQL
- شناسایی ارزشهای از دست رفته
- تقسیم داده ها به نمونه های آزمایش و آموزشی بر اساس توزیع اتفاقی
- تابع دیگر Map Data ، به کاربران KS امکان می دهد تا ارزشهای موجود یک متغیر را در ارزشهای جدید طراحی نمایند.
- تکنیکهای استخراج داده ها ، الگوریتم ها ، و کاربردها ، KS از الگوریتم درخت تصمیم بر اساس CART و CHAID استفاده می کند . این الگوریتم را می توان در دسته یا حالات جامع مورد استفاده قرار داد اگر چه حالت دسته ، برای متغیر پیوسته و CHAID برای متغیرهای مقوله ای ، مشابه CART می باشد ، حالت جامع دسته بندیهای بیشتری را در نظر می گیرد و از نظر آماری مهمترین متغیرها را شناسایی می کند.
- KS می تواند فهرستی از قوانین را تولید نماید که نقاط داده ها را بر اساس تفاوت در توزیع متغیرهای وابسته در دسته های شناسایی می کند . KS می تواند داده های دوتایی پیوسته و دارای چند ارزش را طبقه بندی می نماید ؛ و تقسیمهای چند جهته را برای متغیر پیشگویی کننده ایجاد نماید .
- کار با ابزار KS به کار بر امکان می دهد تا هر بخش را در درخت شناسایی نماید ، و یا از ابزار بخواهد تا بطور خودکار درخت را بسازد . در مورد آخر کاربر فرآیند را در گره ریشه آغاز می کند . KS از نظر آماری مهمترین بخش را از هر گره ، برای ساخت درخت شناسایی می کند. تابع Edit view که قبلاً ذکر شد، به کاربر امکان می دهد تا حق انتخابهای اثر گذار بر مدل را شناسایی نماید .
- KS مدل درخت را تفسیر می کند ، و قوانین حاصل را در حالت گرافیکی با آمار جامع برای هر گروه نشان می دهد . اینها شامل ، احتمال طبقه بندی ، ارزش مربع خی دو، درجات آزادی ، و توزیع متغیرهای پیشگویی کننده و وابسته می شود.

قوانین مدل را می توان به صورت یک سری از اظهارات اگر - پس، کد پرولوگ یا دست خط SQL صادر نمود.

KS یک درخت تصمیم در حالت کلی را به صورت یک سری از گره های متصل بهم نشان می دهد . این تداخل به کاربر امکان می دهد تا کل داده ها را تحت تجزیه و تحلیل بررسی نماید . برای مدلهایی که آنقدر بزرگ هستند که روی صفحه قابل نمایش نیستند ، KS یک تابع موسوم به Tree map را ارائه می کند که می تواند برای نشان دادن کل درخت در یک پنجره کوچک زوم نماید.

نتیجه گیری . جستجوگر دانش، یک محصول کاملی است که یک اساس مشتری در اندازه متناسب را در زمینه بازاریابی هدف ، منتشر نمود . در معیارهای منتشر شده مشتری ، آن عملکرد و مقیاسهای اعتبار معقولی را ایجاد نمود.

برای تداوم حرکت آنی محصول آنگوس ، جستجوگر دانش را تا یک قالب کاری تجزیه و تحلیل بزرگتر موسوم به استادیوی دانش را در ماه می ۱۹۹۸ توسعه داد . هدف استادیوی دانش ادغام اجزاء استخراج داده ها از فروشندگان متعدد در محیط شرکتی می باشد . با ایجاد درختان تصمیم شبکه های عصبی تداخل های شبکه ، و قابلیت حمل و نقل java، آنگوس قصد دارد تا استادیوی دانش را به عنوان یک جز کلیدی برای توسعه ذخیره سازی داده ها قرار دهد . همچنین استادیوی دانش که در ابتدا برای اتکا به حرکت آنی ویندوز تعیین شد ، شامل SDK می شود . SDK برای کمک به قرار دادن محصول در کاربردهای عمودی با استفاده از تکنولوژی Aetiux تولید گردید. آنگوس تعدادی از شراکتهای بسیار قابل رویت با فروشندگان متعدد از جمله کوکنوس ؛ Mel/SHL؛ AT & T و تاندم را توسعه داده است .



### مدل ۱ و کار شناخت الگو (unica)

بررسی . در جدیدترین بازار Data Quest ، unica برای ۹ درصد سهم بازار پشت IBM و کشف اطلاعات ، در نظر گرفته می شود . دلیل چنین نمایش قوی ، این حقیقت است که unica / مدل ۱ (در اصل از گروه ۱) را همراه کار شناخت الگوی خود (PRW) را تحصیل کرده است .

PRW یک ابزار کلی استخراج داده ها می باشد و در حالیکه مدل فارسی ۱، یک کاربرد عمودی را برای unica نشان می دهد . و به عنوان خط رشد محصول برای شرکت ، در نظر گرفته می شود . مدل ۱ یک ابزار استخراج اطلاعات بسیار اتوماتیک می باشد که یکسری از قابلیت های تحلیلی بازاریابی هدف را تقویت می نماید .

ابزار مدل، شامل یک ضریب ( طرح واکنش ) Reaponse Modeler ، ضریب ( تقسیم کننده مشتری ) Customer Segmenter ، ضریب ( فروشنده متقابل ) Cross – seller و ضریب ( ارزیاب مشتری ) Custo,er valuatof ، می شود.

• طرح واکنش یا پاسخ ، مشتریانی را شناسایی می کند که احتمالاً نسبت به صادرات تبلیغاتی پاسخ می دهند.

• تقسیم کننده مشتری ، مشتریان را به گروه هایی با دموگرافیک های مشابه و رفتار خرید مشابه تقسیم می کند .

• ضریب فروشنده متقابل ، مشتریان را با برای خرید محصولات هماهنگ می نماید.

• ضریب ارزیابی کننده مشتری ، مشتریان بالقوه دارای ارزش بالا را شناسایی می کند.

اگر چه ، PRW و مدل ۱ هر دو تعدادی تکنیک استخراج اطلاعاتی هوشمند را از جمله K نزدیکترین همسایه K میانگین ، نزدیکترین دسته ، تابع اساس رادیال (RBF) ، الگوریتم های گواسیان را برای شناخت الگو ارائه می دهد از جمله درختان تصمیم CHAID ، شبکه های عصبی ، الگوریتم های ژنتیکی ، رگرسیون و

الگوریتم های (KFM)، اما ابزارها برای تماشایان نیمه فنی، طراحی شده اند. از نقطه نظر تداخل کاربر PRW و مدل را از جادوگرهایی استفاده می کنند که کاربر را در سریهای مراحل متوالی ناشی از مدل نهایی، هدایت می نماید. ابزارهای unica بطور خودکار در الگوریتمهای متعدد و محلهای پارامتر، برای انتخاب بهترین مدل جستجو می کند. هنگامیکه مدلهای ساخته و ارزیابی می شوند، یک نقشه دیداری درجه بندی بطور خودکار ساخته می شود بطوریکه کاربر می تواند بسادگی طبقه مدلهایی که بهترین اعتبار پیشگویی کننده و عملکرد را ایجاد می کند را شناسایی نماید.

هر دو ابزار، قابلیت های گزارشی غنی را فراهم می کنند که به کاربر امکان می دهد تا مدلهای را در سطوح مختلف جزئیات بررسی نماید. گزارشات، مدلهای را درجه بندی می کنند، و خلاصه آمار مفیدی را در مدلهای داده های ورودی فراهم می نمایند. استفاده فزاینده از جادوگر، الگوهای بازاریابی، و گزارشات، مدل ۱ را برای کار بر مبتدی بسیار مناسب ساخته است. با این وجود، یک کاربر کارشناس، قابلیت های کافی برای ساخت بسیاری از پارامترها برای الگوریتم های ابزار دارد.

بطور کلی، هر دو ابزار، برخی از اتوماتیک ترین ابزارهای استخراج داده ها را ارائه می دهد که برای افزایش سودمندی کاربرهای بازاریابی بطور گسترده تعریف شده، تعیین می شوند.

Unica، برای افزایش عملی بودن بازار و افزایش سهم بازار خود یک سازمان مشاوره استخراج داده ها را ایجاد کرده است که برای بخش مهمی از درآمد آن اهمیت دارد.

PRW و مدل ۱ در سکویهای وینتل سازگار با IBM تقویت می شود. دسترسی به داده ها، ساخت و پیش پردازش یک جادوگر ورود را می توان برای تعیین فایل های ورودی در فرمت های ASCII، MSEXCEL یا فایل های دو تایی مورد استفاده قرار دارد. ابزار DBMS برای انتقال فایلها از سایر فرمتها مورد استفاده

قرار می گیرد . فایل های بانک اطلاعاتی را می توان از طریق ODBC بدست آورد .

قابلیت ساخت داده ها شامل موارد زیر می شود.

- منظره گرافیکی داده های ورودی خام
- پاکسازی داده ها که به کاربر امکان می دهد تا داده های نامعتبر یا ارزشهای ورودی از دست رفته را تعویض نماید.
- محدود کننده های فایل تعریف شده توسط کاربر و کاراکترهای پایان ثبت.
- رسیدگی به داده های از دست رفته
- ادغام چندین فایل ورودی در منبع ورودی تکی
- اشتقاق حوزه جدید با استفاده از فرمولهایی که کار بر تعریف نمود و

اپراتورهای منطقی

PRW ، چندین حق انتخاب جدا سازی داده ها را تقویت می کند، و نمونه

گیری ساده تصادفی را ارائه می دهد .

تکنیک استخراج داده ها ، الگوریتم ها ، و کاربردها. Unica ، الگوریتم های

زیر را در ضریب کار شناخت الگوی آن و ضریب مدل بکار می برد:

- تابع های اساسی رادیال (RBF) برای یادگیری نظارت شده
- گواسیان و گواسیان میکسچر که تابع شدت یا ترکم را محاسبه می کند.
- رگرسیون خطی مرحله گسترده ، سریع ، افزایشی ، و منطقی ارزیابی شده.
- شبکه های عصبی ، از جمله مدل انتشار رو به عقب Multi - Layer Perceptron/ ، یا MLP ، یک سری از این مدل ( حالت افزایشی ) ، شبکه عصبی بررسی ژنتیکی ( که با یک الگوریتم گرافیکی افزایش یافت ) و سایر انواع شبکه .

• درختان تصمیم CHAID

• K نزدیکترین همسایه (KNN) ، نزدیکترین دسته . (NCLUS) و K میانگین

(Kmeans) ، یک خانواده از طبقه بندی و تکنیکهای دسته بندی

تجزیه و تحلیل های PRW ، مسائل طبقه بندی ژنتیک را تأیید می کند در

صورتیکه مدل ۱، بطور گسترده تعریف کاربردهای بازاریابی را هدف قرار می دهد،

و شامل پیشگویی سودمندی واکنش مشتری / مبارزه (که از طرح واکنش استفاده می‌کند) و تجزیه و تحلیل فروش متقابل (که از مدل فروشندۀ متقابل استفاده می‌کنند) می‌شود.

کار با ابزار. مدلها با استفاده از جاودگر طراحی، مشخص می‌شود که کاربر از طریق یک سری از صفحات تعریف هدایت می‌شود که متغیرهای واکنش، ورودی کامل، معماری مدل و حالت (سریع را افزایش و غیره) را مشخص می‌کند. یک مجموعه فزاینده از حق انتخابها و نقص‌ها برای کل الگوریتم ارائه می‌شد. PRW و مدل ۱ یک حالت اتوماتیک را تأیید می‌کند که در آن متغیرها و یا پارامترهای بطور خودکار بر اساس بهینه سازی تابع هزینه انتخاب شده کاربر و به عنوان مثال به حداقل رساندن خطا برای ایجاد بهترین مدل، انتخاب می‌شود. بعد از آنکه مدل ایجاد شد، کاربر می‌تواند تجزیه و تحلیل حساسیت را برای مشاهده مهمترین متغیرها، بکار برد؛ برای یک کاربرد تجزیه و تحلیل فروش متقابل یک گزارش دسته محصول را می‌توان ایجاد نمود. همینطور، نقشه‌های درجه بندی و گزارشات خلاصه مدل، برای تجزیه و تحلیل ارزیابی مدل موجود می‌باشد.

سرانجام مدل ۱ یک الگوی بازاریابی هدفمند موسوم به بهینه کننده مبارزه بازار ایجاد می‌کند. آن به تجزیه و تحلیل مبارزات بالقوه کمک می‌کند، و حتی سود بالقوه خالص را با طبقه بندی مشتری نشان می‌دهد.

مدلهای PRW، را می‌توان به صورت توابع صفحه گسترده یا بطور بیرونی با ایجاد یا کد فورترن صف بندی نمود.

نتیجه گیری: مدل ۱ و کار شناخت الگو (PRW) یونیکا محصولات قوی را نشان می‌دهد که بطور مشخص برای کاربردهای بازاریابی مناسب هستند. هر دو ابزار یک حالت عادی از استفاده و اعتبار خوب پیشگویی کننده را فراهم می‌سازند.



### سایر ابزارهای استخراج داده ها

این بخش به طور بسیار مختصر به چند ابزار استخراج داده های دیگر توجه می کند . به دلیل ماهیت دینامیک تر بازار ابزارهای استخراج داده ها ، فروشندگان این ابزارها در افزایشات اصلی محصولاتشان و همینطور در شراکتهای جدید و متحدین کار می کنند.

IDIS ( کشف اطلاعات ) . سیستم کشف اطلاعات (IDIS) . از الگوریتم اختصاصی برای القای قانون استفاده می نماید . ابزار طراح مدل پیشگویی کننده ، قوانینی را که IDIS بر ای پیشگویی ایجاد نمود بکار می برد . IDIS ، و طراح مدل پیشگویی کننده می توانند به بانک های اطلاعاتی ارتباطی از جمله محیط های بانک اطلاعاتی موازی که در SMP و MPP اجرا می شود ، دست یابند . کشف اطلاعات و همچنین MAPLIDIS را ارائه می دهد که از تکنولوژی MAPLNFO برای شناسایی و نمایش الگوها به صورت نقشه ها استفاده می کند .

کار کشف دانش (NCR) . کار کشف دانش NCR ، یک ابزار یکپارچه است که دسترسی به داده ها برای سیستم های بزرگ بانک اطلاعاتی را با الگوریتم های متعدد کشف دانش ترکیب می کند . NCR قصد دارد تا از ابزارهای استخراج اطلاعات شخص ثالث به عنوان اجزاء در کار کشف دانش استفاده نماید . به عنوان مثال کلمنیتن ( که قبلاً در این بخش مورد بحث قرار گرفت ) برای درج در اجزاء برنامه ریزی می شود . بطور کلی ، الگوریتم استخراج داده ها که با این ابزار ارائه گردید . شامل شبکه های عصبی ، درجه قانون ، و درخت تصیم می شود قابلیت هدایت ODBC ، یک مکانیزم برای دسترسی به بانک اطلاعاتی ترادیتا می باشد.

ماین ست ( سیلیکون گرافیک ) . سیلیکون گرافیک یک رهبر در عرصه تجسم داده ها ( برای تاثیرات متقابل ) می باشند . ابزار استخراج داده های ماین ست آن ، تعدادی از تکنیکهای استخراج داده ها را با تجسم داده های سه بعدی ابتکاری دارای تاثیر متقابل ، و بسیار ادراکی ترکیب می کند . تجسم داده های ماین ست شامل قابلیت های تغییر شکل داده های جامع . الگوریتم طبقه بندی کننده درخت

تصمیم می‌شود که به موتور تجسم داده‌ها مربوط است. تجسم، نیز برای انجام تجزیه و تحلیل حاصل از فرضیه و بررسی، استفاده می‌شود موتور تجسم قانون بطور مؤثر استحکام و تأیید هر قانون را نشان می‌دهد.

پیشگویی شبکه‌های عصبی (افزار عصبی). پیشگویی عصبی، یک ابزار طراحی شبکه عصبی جامع است که می‌توان برای تعدادی از تجارت‌های مختلف و مسائل علمی مورد استفاده قرار داد. استعاره اولیه آن صفحه گسترده اکسل است. مشخصه متمایز کننده اصلی پیشگویی، قابلیت آن برای خودکار سازی وظایف تحلیلی موجود در ساخت یک شبکه عصبی است. محصول، کامل بسیار انعطاف پذیر می‌باشد، و بنابراین برای کاربر کا رشناس مناسب است.

هماهنگ نمودن (تورنت). تورنت (تکنولوژیهای موازی کاربردی قبلی)، ابزار هماهنگ سازی خود را به صورت یک قالب کاری جامع برای مدیریت داده‌ها و استخراج داده‌ها توسعه داده است که می‌توان صف بندی نمود، و از معماریهای کامپیوتری موازی سود می‌برد (MPP, SMP) هماهنگی سازی برای قابلیت سنجش، عملکرد بالا طراحی می‌شود و تداخل مبتنی بر هدف و بر اساس C++ و یک محیط سپر مانند سطح بالا را برای مدیریت داده‌ها و استخراج داده‌ها ارائه می‌دهد. ابزار استخراج داده‌ها، یک تکنیک متنوع از جمله شبکه‌های عصبی و درختان تصمیم را فراهم می‌سازد. بعلاوه، تورنت بدنبال هماهنگ سازی SAS اینترپرایزماینر می‌باشد.

### ابزارهای CRM

عقیده رایج مدیریت رابطه مشتری (CRM) این است که آن فعالیت اصلی تجارت الکترونیکی می‌باشد منطقه تکنولوژیکی مربوط به CRM، خصوصی سازی است. خصوصی سازی بر ایجاد نمایش وب سایت برای اشخاص یا طبقاتی از مشتریان بر اساس اطلاعات برش عمودی، دموگرافیک‌ها، و یا

معاملات قبلی می باشد ، هدف ، بازاریابی و فروش نفر به نفر و افزایش تجربه کار  
بر است ، بطوریکه مشتری به وب سایت تاجر و بازرگان باز می گردد.

### ابزارهای خصوصی سازی

در زیر ، فهرست نمایشی از فروشندگان خصوصی سازی و محصولات  
آنها آمده است . این فهرست بر اساس حروف الفبا سازماندهی می شود و ترتیب  
فروشندگان به معنای سهم بازار مربوط به آنها یا ترجیحات نویسندگان نمی باشد.  
گروه تکنولوژی هنر. ابزار دینامو گروه تکنولوژی هنر (ATG) از تکنیکهای  
هوش مصنوعی برای خصوصی سازی رابطه روی خط بین بازرگانان و مشتری  
می باشد . هسته کاربرد سرور خصوصی سازی دینامو مبتنی بر جاوا می باشد .  
کاربردهای دینامو ، ابزاری را برای مرسوم سازی فروشگاه تقویت معامله و  
ترفیعات هدفمند فراهم می کند .

ایستگاه تبلیغاتی ATG ، برشهای عمودی مشتری را اداره می کند ، رفتار  
مشتری را ردیابی می نماید . و تبلیغات هدفمند را برای سایتهای بزرگ و شبکه  
های تبلیغاتی ارائه می دهد . آن می تواند رفتار مشتری را در زمان واقعی گزارش  
و ردیابی نماید . گزارشات را می توان به صفحه گسترده ها برای تجزیه بیشتر ،  
منتقل نمود .

برایت ویر Advice Agent . برایت ویر برای کمک و معاونت مشتری یک  
نمونه از نرم افزار خصوصی سازی مبتنی بر مورد است که برای رویدادهای  
تجارت الکترونیکی دارای تاثیر متقابل بکار می رود . Advice Agent به مشتری ها  
با درک نیازهای آنها ، کمک می کند و به آنها اطلاعات و راه حلهای مناسب را  
نشان می دهد . آن از یک دیالوگ وب اتوماتیک برای تهیه مشاوره شخصی و  
محو هدفمند استفاده می کند .

براد ویژن این یکی از فروشندگان برجسته در کاربردهای تجارت الکترونیکی مبتنی بر قانون می باشد. مرکز دستور نفر به نفر آن در مؤسسه نفر به نفر براد ویژن و سه شرح عمودی ترکیب می شود:

مالی ، دانش ، تجارت . آن به فروشندگان امکان می دهد تا قوانین تجارت را توسعه دهند که اطلاعات کاربر حاصل در طی معاملات قبلی را ارزیابی می کند و از آن برای توجه به محصولات و خدمات در طی تاثیرات متقابل بعدی استفاده می کند .

فایر فلای. شبکه که میکروسافت در آوریل ۱۹۹۸ بدست آورد . تکنولوژی خصوصی سازی را بر اساس تکنیکهای تصفیة جمعی ، فراهم می کند ، بعلاوه ، فایر فلای ، به کاربران امکان می دهد تا یک لوگ را مشاهده نمایند که نشان می دهند داده های آنها چگونه استفاده شده است ، به آنها امکان می دهد تا اطلاعات برش عمودی را تغییر دهند و در خواست نماید که داده ها در آینده در سایتهایی که از تکنولوژی خصوصی سازی فایر فلای استفاده کردند ، مورد استفاده قرار نگیرد.

نت پرسپشن. موتور توصیه زمان واقعی نت پرسپشن، ترجیحات مشتری را از رفتار مشاهده شده استنباط می کند ، و مقوله های خرید را بیان می کند . به عنوان مثال ، استنباطهایی را بین طول زمانی که فرد صرف خواندن مقاله خبری می کند ، و سطح علاقه فرد به مقاله را ارائه می دهد . اعمال دیگر کاربر که سطح علاقه را به یک مقوله نشان می دهند ، شرط بندی ، چاپ ، پس انداز و تصمیم گیری ضریب هستند . موتور توصیه زمان واقعی ، در یک سرور یا خادم شبکه قرار دارد و با سایر کاربردهایی که داده های مشتری را از طریق مکالمات API ایجاد می کنند با مورد استفاده قرار می دهند . در تاثیر متقابل می باشد. علاوه بر توصیه های مکمل، محصول همچنین حمایت جدول فروش متقابل را فراهم می کند .



ویگنیت . ویگنت ابزارهای مدیریت رابطه مشتری مبتنی بر اینترنت را فراهم می کند . محصولات مؤسسه آن شامل story می باشد که قابلیت ارائه حجم مناسب در هر مرحله چرخه عمر مشتری را ارائه می نماید : از بازدید کننده اولیه تا مشتری بلند مدت . این قابیت از طریق جزئی که خصوصی سازی چرخه عمر نامیده می شود ، ارائه می گردد . نرم افزار می تواند از داده های رفتار برای استنباط علائق بازدید کنندگان و جدید سازی حجم استفاده نماید . این به بازرگانها امکان می دهد تا حجم را بر اساس انتخابهای بازدید کننده با علائق مشابه ، توصیه نمایند. جزء خدمات حمایت تصمیم دیگنت یک تجزیه و تحلیل سطح تجاری رفتار مشتری را برای کمک به تعیین سودمندی حجم برای جذب و حفظ بازدید کنندگان فراهم می کند .

دیگر کاربرد خصوصی سازی CRM ، یک خادم چند کاناله دیگنت است که می تواند کل نقاط تماس مشتری از جمله پست الکترونیکی ، تلفنهای سلولی ، مدیران حسابداری ، مراکز درخواست و وب سایتها را همانگ نماید، و ارتباطات و خدمات خصوصی را برای هر مشتری فراهم آورد . خادم چند کاناله ویگنت ، به مشتریان امکان می دهد تا مشخص نماید چه نوع اطلاعاتی و همینطور واسطه ارتباطی ارجح خود همانند تلفن پیجر ، ایمیل یا فکس را مایلند دریافت نمایند . بعلاوه خادم چندگانه ویگنت می تواند کل ترجیحات ارباب رجوع را با اطلاعات برشی عمودی مشتری واقع در انبار داده های مؤسسه و یا بازار داده های مشتری ، ترکیب نمایند.

به عنوان مثال در کاربرد خدمات مالی، خادم چند کاناله می تواند بطور خودکار مشتریان معاملات حسابداری مهم را همانند سیم های وارد شونده رسیدهای Lock box ، نمایشهای خارج و پرداخت و گزارشات معامله را بیان کند . رویدادهای مدیریت سرمایه گذاری همانند بدهی های اوراق بهادار ، تائیدهای تجاری ، بایگانی های پیشنهاد عمومی و تغییرات قیمت سهام را نیز می توان درجه

نمود . بعلاوه ، مشتریان می تواند دریافت اتوماتیک اطلاعات بازار همانند یادداشتهای همگانی توصیه های خرید و فروش و برگهای نرخ را انتخاب نمایند .  
خادم سندیکا سازی ویگنت (VSS) به ایجاد گروههای آشنای متمایز و ساخت بسته های هدفمند برای توزیع در چندین وب سایت آشنا کمک می کند .  
عامل سندیکا سازی ویگنت یک کاربرد مبتنی بر جاوا است که به گروهها امکان می دهد تا این بسته ها را بطور خودکار و دور دست ، پیاده نمایند . حجم بسته را پردازش کنند ، و از راه دور آنها را به مرور زمان اداره نمایند . خادم سندیکاسازی ویگنت ، بر اساس پروتکل UE در حال ظهور می باشد .  
ویگنت با بیش از ۸۰ شرکت تکنولوژیکی از جمله آدوب ، میکروسافت ، نشنال سمی کانواکتور CNET ، نت پرسپشن و سان میکروسیستم ، کار می کند آن با تهیه کنندگان حجم اصلی ، از جمله خدمات اینترنت نیوز ، پریویو تراول ، تریبون مدیا سرویس و زیف - داویس ، کار می کند . آن پروتکل ICE بر اساس XML را توسعه می دهد که حجم روابط سندیکاسازی را بین خادمهای شبکه توضیح می دهد و مبادلات کنترل شده اتوماتیک و مدیریت دارایی های روی خط بین شرکاء تجاری یک یادداشت برای پروتکل ICE ، ICE که علامت اختصاری تبادل اطلاعات و محتوی می باشد ، یک پرتکل مبتنی بر XML می باشد که هدفش سندیکا سازی حجم با ایجاد استاندارد برای ساخت شبکه های تجارت اینترنتی مانند شبکه های انتشار سندیکا شده ، لوپراستورهای شبکه و کانالهای فروشنده روی خط ، می باشد در این خصوص سندیکاسازی ، به صورت تبادل کنترل شده و مدیریت دارایی های تجارت به تجارت تعریف می شود .

ICE تا فراسوی تنها یک قالب اطلاعاتی XML پیش می رود . ICE ، یک پروتکل سندیکا سازی خادم به خادم و پردازش مدل را تعریف می کند . پروتکل ICE ، با استفاده از XML برای فورمت بندی داده ها و پروتکل های ارتباطی متنوع ساخته می شود . به طور مشخص ICE از XML برای توصیف اشتراکهای سندیکا سازی ، بسته بندی محتوی ، برنامه ریزی اطلاعات ، قوانین تجاری ، و کل جنبه

های دیگر تبادل دارایی فورمت بندی شده استفاده می کند . پروتکل ICE خادم به خادم و پردازش مدل از تکنولوژیهای HTTP ، FTP و سایر تکنولوژیهای استاندارد برای ایجاد یک روش معتبر و کارآمد برای اتوماتیک سازی تبادل حجم، اطلاعات مدیریت ، و داده های لوگ بین سندیکاتورها و مشترکین ، استفاده می کند.

### ابزارهای بازاریابی / مدیریت مبارزه

راه حل های مدرن مدیریت مبارزه و بازاریابی مبارزه و بازاریابی بانک های اطلاعاتی بر سه وظیفه تمرکز می نمایند: برنامه ریزی مبارزه، طراحی تحلیلی و گزارش. اگر چه اکثر ابزارهای موجود امروز کل این سه وظیفه را تقویت می نماید، فروشندگان ابزار خود را با ارائه عملی بودن ویژه ابزار ، خدمات و حمایت متمایز می نمایند.

در توجه به عملی بودن ابزار، ما ارزیابی ابزار را با معیارهای زیر را ترجیح می دهیم :

- منابع داده ها . قابلیت ابزار برای ادغام یا انبار داده های مشتری یا بازار اطلاعاتی دو مخزن متا دیتای مربوطه.

- طراحی مدل قابلیت ابزار برای ادغام با یک موتور استخراج داده ها و به عنوان مثال ، SAS EM ، IBM IM به منظور اتوماتیک سازی فرآیند ایجاد و اصلاح مدلها برای مبارزات جدید.

- قابلیت سنجش . یک ابزار بایستی بتواند یک تعداد بیشتر از مبارزات بازاریابی را تقویت و اداره نماید.

- حمایت رویدادهای جرقه ای . ابزار بایستی بتواند ، رویدادهای جرقه ای را از سیستم های عملیاتی انبار داده های عملیاتی و حتی از تجزیه و تحلیل داده های ذخیره شده در انبار داده ها، بپذیرد ؛ رویداد ها توسط شرکت تعیین می شوند و شامل رویداد خرید اصلی ، تعیین محل مجدد و زندگی ( ازدواج ، تولد کودک ) و غیره می شود .

• گزارش و تجزیه و تحلیل. یک ابزار بایستی برای تجزیه و تحلیل اتوماتیک نتایج مبارزه ، ارزش عمر مشتری و گزارش فعالیت ، فراهم شود .  
در میان ابزارهای بازیابی و مدیریت مبارزه ، ما مختصراً به سه ابزار زیر و فروشندگان آنها توجه می‌کنیم :

VALEX از کاربردهای تبدیلی ، پرایم و نتج از پرایم ریسپونس و وان بای وان از پاراژرن .

□ کاربردهای تبادل (VALEX) . دفتر مرکزی آن در وستون می‌باشد ، و بسیاری از تحلیل گران شرکت را به عنوان یکی از رهبران نرم افزار مدیریت مبارزه در نظر می‌گیرند. ابزار VALEX در UNIX و ویندوز NT وجود دارد ، و بسیار قابل سنجش می‌باشد ( آن را می‌تواند جریان‌ات پیوسته بازیابی را در بانک اطلاعاتی بسیار بزرگ اجرا نماید ) . VALEX از تداخل GUI ویندوز خوب طراحی شده استفاده می‌کند که در آن تمایل یا آیکونهایی که فعالیت‌های مختلف را نشان می‌دهند، توسط کاربر بر ای ایجاد جریان‌ات مبارزه ، بهم متصل می‌شوند. معماری VALEX ، یک خادم / ارباب رجوع سه تایی می‌باشد که ارباب رجوع تداخل کاربر را اجرا می‌کند ، خادم تایر میانی نرم افزار مدیریت مبارزه را اجرا می‌کند ، و خادم داده ها می‌تواند هر DBMS استاندارد ارتباطی باشد . VALEX به دقت با SAS برای حمایت طراحی مدل ادغام می‌شود ، و مبارزات مبتنی بر رویداد جرقه ای را تقویت می‌نماید VALEX می‌تواند ردیابی مبارزه را گزارش نماید و با ابزارهای متعدد OLAP برای قابلیت‌های گزارشی اضافی و تجزیه و تحلیل ، ادغام گردد .

□ پرایم ریسپونس ( پرایم و نتج ) ، تاکید زیادی بر مشخص های طراحی شده برای تقویت بازیابی هوشمند مبتنی بر رویداد جرقه ای می‌نماید . محصول بر اساس معماری دوتایی ارباب رجوع / خادم می‌باشد که حجم زیادی از کاربرد در همان سکوی بانک اطلاعاتی که انبار داده های مشتری قرار دارد ، اجرا می‌شود ( ویندوز NT و UNIX ) تقویت می‌شود . محصول اداره امتیازات



را تقویت می کند ، می تواند کارتهای امتیاز را مجدداً ایجاد نماید و می توان آن را با SASEM برای مدیریت مدل ترکیب نمود . رویدادهای جرقه ای شامل درختان رویداد می شوند که دارای شاخه های بسیار هستند بطوریکه ارتباط بعدی با مشتری با سابقه ارتباطات قبلی و واکنشهای مشتری تعیین می شود .

□ پاراژرن ( یک به یک ) ، یک شرکت فرعی مبتنی بر وسرچیانای آپاک ( یک بازارباب از راه دور بزرگ ) به خاطر استراتژی برای افزایش ابزارهای مدیریت مبارزه با استخراج داده ها و کمک های تغییر شکل و طرحهای مدلی ، معروف می باشد . طراح مدل می تواند از ضریب کشف کننده داده ها یکی یکی برای تعریف مجموعه مشتق شده از متغیرها برای SASEM ، داروین یا ضریب مدیر مبارزه تک به تک استفاده نماید . محصول به صورت یک کاربرد ارباب رجوع / سرور ( خادم ) دو تایی معماری می شود و نرم افزار خادم با انبارداده های مشتری مبتنی بر NT یا UNIX ، معماری می شود و بنابراین با استفاده از دسترسی مستقیم به DBMS به جای ارتباطات ( ICBC ) . وان بای وان ، تسهیلات اختصاصی را برای ساخت داده ها بعد از استخراج داده ها از فایل های SAS به جای تکیه بر SAS برای این کاربرد ها ، فراهم می نماید . محصول ، مبارزات مبتنی بر رویداد جرقه ای را تقویت می کند و با اوراکل اکسپرس یا Hyperion Essbase برای گزارشات پیچیده ، ادغام می شود .

#### ابزارهای خدمات مشتری و اتوماتیک سازی فروش

تجارت الکترونیکی ، بازاری را برای بسته های نرم افزار ایجاد کرده است که به سازمانها در اجرای وظیفه مدیریت رابطه مشتری کمک می کند این بسته ها برای استفاده نمایندگان خدمات مشتری و سایر کارمندانی که با مشتریان در تاثیر مقابل هستند ، واکنش نشان می دهند . تاثیرات متقابل به اشکال وب و کانالهای پست الکترونیکی محدود نمی گردد . تکنولوژیهای خدمات مشتری مبتنی بر وب

توسعه یافته اند تا شامل تلفن اینترنتی و گزینه هایی همانند چت های متنی و درخواست های مکالمه شوند.

استفاده از تلفن IP در مراکز مکالمه اینترنتی به مشتریان امکان می دهد تا مستقیماً با نماینده های مرکز مکالمه صحبت نمایند در حالیکه از پروسه های خود برای دستیابی به وب سایت سازمان استفاده می کنند . تلفن IP بطور مداوم در مراکز مکالمه از قبل موجود ریسک می شوند . همچنین ، برای یک مشتری شخص ، استفاده از تلفن IP ، راحت ، ارزان است و نیازمند دانش سخت افزار و نرم افزار فزاینده نمی باشد (یک کامپیوتر شخصی مولتی مدیا و یک سری تلفن IP ، و نرم افزار تلفن اینترنتی که از اکثر محیط های عملیاتی موجود می باشد . ) در بین راه حل های معروف که تاثیرات متقابل مشتری یا سازمانها را هدف قرار می دهند . سیبل و وانتیو قرار دارد.

سیبل . سیبل ۹۹ ، یک ابزار کاربرد مبتنی بر شبکه ، فروشها ، بازار یابی و خدمات مشتری را تقویت می کند. راه حل مدیریت منبع مؤسسه ، شامل کاربردهای زیر می شود.

- مؤسسه فروش سیبل ، نمایندگان فروشی را که نیاز به دسترسی به داده ها از محل های دور دارند با فراهم آوردن امکان تبادل بین کامپیوترهای دستی و خادم شرکتی ، به نمایندگان فروش کمک می کند .

- مؤسسه سرویس سیبل ، ابزار را برای تقویت مدیریت درخواست خدمات ، ترسیم برش عمودی محیط خدمت و حل مسئله با استفاده از تجربه و تحلیل حالت ، خصوصیات محصول بانک اطلاعاتی FAQ و راه حل خدمات مشتری ، مشخص می نماید.

- خدمات حوزه سیبل ( سی بل ) عملیات خدمات حوزه همانند برنامه های نگهداری ، درخواستهای تعمیر و تعویض قطعات و برگشتها را تقویت می کند و دستورات مفصلی را در خصوص مواد لازم برای یک سفارش خدمات مشخص فراهم می نماید.

• مرکز درخواست سی بل ، فروشها و وظایف خدمات مشتری را با ایجاد فروش به درخواست با کمک و معاونت خرید برای مشتریان ، ادغام می نماید .  
نمایندگان خدمات مشتری می توانند به سوابق و گزارشات مشتری دست یابند و سپس فروش محصولات و خدمات را بر اساس خریدهای تاریخی مشتری توجه نمایند . کاربرد ادغام تلفن کامپیوتری (CTI) ، خدمات فهرست سازی تعداد مکالمات (DNIS) ، شماره یا شناسنامه خودکار (ANI) و سیستمهای واکنش صوتی مؤثر بر هم (IVR) را تقویت می کند .

• مؤسسه بازاریابی سی بل شامل بازار اطلاعاتی پیش ساخته و خادم پردازش تحلیلی روی خط (OLAP) برای تجزیه و تحلیل های داده ها می شود .  
آن همچنین شامل تقسیم بندی بازار ، مبارزه چاپ مکالمه ، و ابزارهای مدیریت جریان کار می شود .

• SIEBEL Inter Active از تکنولوژی فایل برای جمع آوری اطلاعات شخصی در مورد مشتریان بالقوه ، صنایع ، و غیره استفاده می کند . اطلاعات مرسوم به کامپیوتر شخصی کاربر ارائه می شود .

• شکل دهنده محصول سی بل بازار یابی ، فروش و عملیات خدمات مشتری را تقویت می کند . آن اشکال محصول را اداره می کند تا این اطمینان را ایجاد کند که قیمت گذاری محصول و خدمات معتبر است و ایجاد اشکال محصول ترفیعی را که می توان به فروشهای مناسب و کانالهای خدمات مشتری توزیع نمود . تسهیل می نماید .

ونتیو ، vantage 8 یک ابزار کاربردهای CRM مبتنی بر شبکه است که موجودی ، خدمات مشتری ، خرید ، فروش ، خدمات حوزه ، و بازاریابی را تقویت می کند ، شرح رایج ، Vantage8 مرکب از پنج کاربرد اصلی می باشد:

• مؤسسه ونتیو ، برای استفاده کارمندی که بطور مستقیم با مشتریان در تاثیر متقابل هستند طراحی می شود . آن را می توان با سیستمهای مشروع و حمایتی دفتری ، ادغام نمود .

- فروشهای و تئو حاوی نرم افزار پیشنهاد و سهمیه و همینطور پیش بینی و نرم افزار گزارش مؤسسه ، کاربرد دارای قابلیت های موبایل می باشد.
- خدمات حوزه و تئو عملیاتی مانند ردیابی و لیست درخواستهای خدمات ، اختصاص سفارشات خدمت ردیابی زمان و مواد صرف شده برای یک سفارش مشخص و اداره موجودی را تقویت می کند .
- حمایت و تئو ، یک کاربرد CRM است که عملیات خدمات مشتری را از طریق تلفن فاکس ، پست الکترونیکی و وب ، تقویت می کند .
- Vantive Help Desk به سایتهای بازرگان کمک می کند تا یک سیستم کمک فنی مرسوم را ایجاد نماید . آن می تواند نسبت به گرایشهای مسئله که در وب یا اینترنت شرکتی ارائه می شود ، واکنش نشان دهد.
- علاوه بر سیبل ، و تئو و بسیاری دیگر ، تعمیرکنندگان ارتباط از راه دور به فعالان در خدمات مشتری مبتنی بر شبکه تبدیل شده اند . به عنوان مثال ، خدمات متصل به کمک MCI ، تکه ای را در سایتهای وب قرار می دهد که یک مکالمه تلفن IP را از طریق دروازه اینترنت راه اندازی خواهد نمود که تا تلفن عامل مرکز مکالمه دنبال می شود . پاسخهای مؤثر بر همه AT&T ، سرویس Sprint's Give و Me Acall و راه حل IPT در یک سوی (یک بسته کاربر مبتنی بر اینترنت که امکان ارتباطات تلفن - تلفن ، فکس به فکس ، PC به تلفن و PC به PC روی شبکه های IP را فراهم می آورد). نمونه ای دیگر ارائه خدمات مشابه هستند.



## مقدمه

این فصل بطور مختصر به حوزه گسترده تر ظرفیت های کشف اطلاعات توجه می کند ، و بر نیاز به درک نه تنها انواع داده های سنتی (اعداد و خصوصیات) بلکه درک انواع اطلاعات پیچیده ، همانند متن، تاکید می کند . در حقیقت حجم اطلاعات موجود در تعداد زیادی از اسناد شرکتی (به عنوان مثال فایل های پردازشگر واژه، فایل های نمایش) به نظر می رسد از حجم اطلاعات موجود در انبار های الفبایی سنتی با عامل سه به پنج ، متجاوز می باشد . علاوه بر این ، کل اسناد وب و این اعداد ندرتاً بزرگ هستند. بنابراین قابلیت تجزیه و تحلیل و درک آن اقیانوس پهناور داده های پیچیده به سرعت به یک حکم تجاری تبدیل می شود . زمینه تکنولوژی اطلاعات که بر تجزیه و تحلیل اطلاعات متنی متمرکز است . به استخراج متن ( برخلاف استخراج داده ها که به معنای داده های عددی است ) معروف می باشد و در این فصل مورد بحث قرار می گیرد.

بعلاوه ، این افزایش در حجم اطلاعات نه تنها به تعداد کامل بیت و بایت هایی که باید با الگوریتم استخراج پردازش شوند ، مربوط می گردد، بلکه به تعداد متغیر های مهم نیز مرتبط می گردد. حقایقی که بر پیشگویی کننده اثر می گذارد .

هنگامیکه تعداد متغیر ها رشد می کند ، توانایی ، برای تجزیه و تحلیل کل آن در یک زمان در یک حالت رادیکال ، زیاد شود . بنابراین دومین موضوع مربوط به این بخش ، بحث در مورد روش های جدید برای استفاده از قدرت مغز انسان برای کمک به درک فضای چند بعدی چند متغیره می باشد که در غیر آن صورت رسیدگی به آن مشکل است . در بین روش هایی که این فصل بطور مختصر به آن می پردازد ، چیزی است که به محاسبه معنایی معروف گردیده است.

### هوش تجاری و استخراج اطلاعات

در عملیات روزمره شرکتها کل آنها مقادیر زیادی از اطلاعات را جمع آوری می کنند : داده های مربوط به سفارشات ، موجودی حسابهای قابل پرداخت ، معاملات نقطه فروش ، و مشتریان ، بعلاوه ، تجارت اغلب داده هایی همانند فهرستهای دموگرافیک و پستی را از منابع خارج بدست می آورد . توانایی تحکیم و تجزیه و تحلیل این داده ها برای تصمیمات تجاری بهتر اغلب منجر به مزایای رقابتی می گردد و یادگیری برای آشکار سازی و اتکاء به آن مزایا ، کانون هوش تجاری استراتژیک می باشد.

در محیط تجاری جاری و در حال ظهور رقابتی و بسیار دینامیک ، تنها رقابتی ترین شرکتها به موفقیت بازار حفظ شده ، دست خواهند یافت . به منظور سرمایه گذاری بر روی فرصتهای تجاری ، این سازمانها ، با قابلیت اتکاء به اطلاعات در مورد بازار خود ، مشتریان خود و عملیات خود ، خود را متمایز خواهند نمود . بخشی اصلی این استراتژی برای موفقیت بلند مدت قابل حفظ ، مخزن اطلاعات فعال، یک انبار اطلاعاتی پیشرفته است که در آن اطلاعات از قابلیت های متعدد یا بخشهای تجاری درک می شود و بهم می پیوندد.

کوتاهترین مسیر از داده های پیچیده به کشف دانش ، استخراج اطلاعات است . توجه کنید که ، آن را استخراج اطلاعات می نامیم نه فقط ، استخراج داده ها . تا اشکال بسیار متنوعی که اطلاعات لازم برای هوش تجاری می تواند کسب نماید را منعکس نمائیم . استخراج داده ها به معنای استفاده از ابزار قوی و هوشمند برای انجام کارهای زیر می باشد:

- آشکار سازی اتحادها ، الگوها و روندها
- کشف انحرافات
- دسته بندی و طبقه بندی اطلاعات
- توسعه مدلهای پیشگویی

همانگونه که در فصول قبلی بحث شد ، استخراج اطلاعات به شکل استخراج داده ها مزیت رقابتی مهمی را برای بسیاری از سازمانها ، در صناعی مانند ، مالی ، بهداشتی ، بیمه ، خرده فروشی و ارتباطات از راه دور ، ایجاد کرده است . تکنیکهای جدید استخراج اطلاعات می توانند دانشی را آشکار نمایند که دستیابی به آن از طریق روشهای دیگر مشکل می باشد و به موجب آن مزایای رقابتی واقعی را ارائه می دهد.

از جنبه فنی ، کلیدهای واقعی استخراج اطلاعات موفق ، الگوریتم های آن هستند: فرآیندهای ریاضی پیچیده که داده ها را مقایسه و مرتبط می نمایند . الگوریتم ها به یک کاربرد استخراج داده ها امکان می دهد تا بهترین مشتریان برای تجارت - چه کسانی هستند ، و چه چیزی را مایلند خریداری نمایند . آنها همچنین می توانند تعیین کنند که در چه زمانی از روز و در چه ترکیبی و یا چگونه ، یک سازمان می تواند موجودی ، قیمت گذاری و تجارت را به منظور حفظ این مشتریان بهینه نماید و سبب شود تا بیشتر و با حاشیه سود افزایش یافته ، خرید نمایند.

تعدادی از فصول قبلی در مورد الگوریتم ها و تکنیکهای استفاده شده در استخراج داده ها ، بطور مفصل ، بحث کردند. با این وجود - یک حجم بزرگی از اطلاعات به یک شکل غیر عادی ذخیره می گردد : اسناد ، تصاویر و فایل های ویدئویی . بخش بعدی به فرآیند استخراج اطلاعات متنی می پردازد که می توان در تعداد زیادی از اسناد، یادداشتها نامه ها ، قراردادهای ، علائم انحصاری ، رونوشت های سخنرانیها ، پیامهای ایمیل و منابع مشابه یافت .

### مدیریت دانش و استخراج متن

استخراج متن ، یک زیرمجموعه از تکنولوژی استخراج اطلاعات است که به ترتیب ، جزئی از مقوله کلی تر مدیریت دانش (K.M) می باشد . دانش ، در این حالت ، به مهارتهای جمعی ، تجارب ، دانش فنی و عقل یک سازمان اشاره دارد . دانش چیزی بیش از داده ها یا اطلاعات ساده می باشد ، آن شامل زمینه ، گزینه ها

و حقایقی است که به فرایند تصمیم گیری کمک می نماید . در دنیای تجاری دانش نه تنها با داده های ساختار بندی شده یافت شده در بانکهای اطلاعاتی سنتی ارائه می شود ، بلکه در منابع ساختار بندی نشده متنوع همانند اسناد واژه ، یادداشتها و نامه ها ، پیامهای ایمیل ، اخبار ، صفحات وب و غیره نیز ارائه می گردد.

اینفورمیشن ویک ، مدیریت دانش را به صورت فرآیند تسخیر مهارت جمعی شرکت در جایی که قرار دارد - در بانک اطلاعاتی ، مقاله یا در افکار مردم - و توزیع آن در جایی که می تواند به ایجاد بیشترین پرداخت کمک نماید ، توضیح می دهد . ابزار مدیریت موقتی دانش ، با اهداف متنی موجود کار می کند ، و کار جمعی را به منظور تسهیم متن های نوشته نشده در افکار اشخاص تشویق می کند . ابزار و محصولات در بازار مدیریت دانش شامل موتورهای جستجو ، سیستم های مدیریت سند ، و محصولات گروهی می گردد ، اما به اینها محدود نمی شود .

استخراج متن به عنوان یک تکنولوژی کلیدی که مدیریت دانش را توانا می سازد . از این نظر که روابط اطلاعات را آشکار می سازد ، استخراج داده هاست . اگر چه آن شبیه استخراج داده هاست اما از نظر آن نیز متفاوت می باشد . در حقیقت ، استخراج داده ها و یک کاربرد ابزارهای یادگیری ماشینی و آماری برای مجموعه ای از داده ها به منظور آشکار سازی روابط و همبستگی های قبلاً مشخص نشده ، می باشد . تا به امروز ، استخراج داده ها آگاهی های ارزشمندی را برای سازمانهایی که به تفسیر رفتار مشتری و ساخت مدل های پیشگویی کننده توجه می کند ، فراهم نموده است . با این وجود ، استخراج داده ها اساساً با داده های ساختار بندی شده کار می کند که اغلب از نظر ماهیت عددی هستند که در یک بانک اطلاعاتی پاکسازی شده ، استاتیک و معقول ، که معمولاً در قلب انبار داده ها یا مرکز داده ها قرار دارد ذخیره می شود (برای جزئیات بیشتر فصل ۲ را ببینید) . برخلاف استخراج داده ها ، استخراج متن با اطلاعات ذخیره شده در یک مجموعه ساختار بندی نشده اسناد متن کار می کند . بطور مشخص ، استخراج متن روی خط ، فرایند جستجو در داده های ساختار بندی نشده در اینترنت و اقتباس



معنا از آن اشاره دارد. استخراج متن تا فراسوی کاربرد مدلهای آماری در فایل‌های داده‌ها پیش می‌رود: در حقیقت، استخراج متن، روابط در مجموعه متن را آشکار می‌کند و به خلاقیت کارگر دانش برای بررسی این روابط و کشف دانش جدید، اتکا دارد. بسیاری از الگوریتم‌های استخراج داده‌ها به کشف دانش جدید، از طریق تکمیل عقاید و منطقی که در مغز کارگر دانش قرار دارد، کمک می‌نماید.

استخراج متن، بطور مشخص به دلیل میزان دانش اشتباه که در اسناد متن قرار دارد که درون سازمان یا خارج آن ذخیره می‌شوند، به تجارت امروز مربوط می‌باشد، ظهور شبکه و انتشار روی خط بطور مؤثر میزان اطلاعات ذخیره شده به شکل متن را افزایش داده است. سازمان‌هایی که به تحقیق آتی و اطلاعات موجود از منابع داخلی و خارجی تکیه می‌کنند، باید بر چالش‌های مهم در کار با این حجم زیاد متن، غلبه نمایند. کل مجموعه اطلاعات متنی موجود برای یک سازمان مشخص آنقدر زیاد است که به سادگی خوانده و تجزیه نمی‌شود. بعلاوه، آن به طور ثابت و بطور مشخص اطلاعات مبتنی بر وب را تغییر می‌دهد و اگر جایی باقی بماند، مستلزم بررسی و تجزیه و تحلیل آتی می‌باشد. استخراج داده‌ها با تعریف ابزار و تکنیک‌های طراحی شده برای تجزیه و تحلیل و درک این نوع اطلاعات دینامیک، این مسائل را مورد تاکید قرار می‌دهد.

### مزایای استخراج متن

مزایای ممکن از استفاده از یک راه حل استخراج متن شامل موارد زیر

می‌شود:

- ارزش افزایش یافته و اطلاعات شرکتی با صف بندی استخراج متن، سازمان‌ها می‌توانند ارزش سرمایه‌گذاری‌های شرکتی در سیستم‌های اطلاعاتی موجود را افزایش دهند. استخراج متن به شرکت‌ها امکان می‌دهد تا بطور مؤثر تر دانش شرکتی پنهان در مجموعه‌های بزرگ متن ساختار

بندی نشده را جمع آوری نمایند و مورد استفاده قرار دهند . و بنابراین نیاز

به ترکیب مجدد یا تعویض مخازن داده های قدیمی تر حذف می گردد.

□ هزینه های کمتر ادغام در برابر سایر تکنولوژی های پردازش متن .

بسیاری از تکنولوژیهای پردازش متنی موجود ( به عنوان مثال مدیریت

سند ) به میزان زیادی از ادغام سیستم و مشاوره کارشناسی قبل از صف

بندی ، نیاز دارند . بسیاری از محصولات جستجو سنتی مستلزم این است

که شرکتها منابع مهمی را صرف ساخت دستی درختان موضع یا شاخصها

نمایند . محصولات پیشرفته استخراج متن که در این فصل بررسی شد، راه

حلهای جریان کاری جمعی که نیاز به ادغام بسیار کمی دارند . ترکیب

می گردند . آنها ، بطور خودکار متن را پردازش می کند و بنابراین راه

اندازی و تشکیل پر هزینه را حذف می نمایند .

□ سودمندی افزایش یافته کارگران دانش ، راه حل های استخراج متن ،

یافتن اطلاعات در مخازن متن شرکتی بزرگ را برای کارگران دانش ساده

تر می سازد ؛ این راه حلها به کاربر امکان می دهد تا اطلاعات مورد نیاز را

با به پیروی از مفاهیم معنایی کلیدی که برایشان مهم است ، منتقل نمایند .

□ سودمندی اثبات شده . مشابه ادعای استخراج داده ها ، استخراج متن

می تواند تصمیم گیری بهتر و سریع تر را تسهیل نماید.

### تکنولوژیهای استخراج متن

بطور کلی ، دو تکنولوژی کلیدی وجود دارد که استخراج متن روی خط را

امکان پذیری می سازد. یکی ، قابلیت جستجوی اینترنتی است و دیگر اسلوب

شناسی تجزیه تحلیل متن می باشد . برخی محصولات هر دو را ترکیب می کنند .

### جستجوی اینترنتی

جستجوی اینترنتی تنها برای چندین سال وجود داشته است یا / نفی بر شبکه های گسترده جهانی ، در چند سال گذشته موتورهای جستجوی متعددی برای کمک به کاربران کمک می کنند تا محتوی که تنها در یک شب در اینترنت ظهور می کند را بیابند . آلتا ویتا ، واکسایت و یا هر سه مورد از قدیمی ترین آنها هستند . موتورهای جستجو و خدمات دیراکتوری با فهرست سازی محتوی در یک وب سایت ویژه کار می کند ، و به کاربران امکان می دهد تا شاخصها را جستجو نمایند . اگر چه اولین نسل این ابزارها مفید هستند ، اما اشباه بودند ، زیرا آنها محتوی را که بازآوری می کردند ، به درستی فهرست بندی نمی کردند . پیشرفتهای در استخراج متن بکار رفته در جستجوی اینترنتی منجر به استخراج متن روی خط گردید که نسل جدید ابزارهای جستجوی اینترنتی را نشان داد . با این محصولات ، کاربران می توانند از اطلاعات مناسب با پردازش میزان کمتری از پیوندها ، صفحه ها و شاخصها ، سود ببرند .

### تجزیه و تحلیل متن

تجزیه و تحلیل متن ، به عنوان یک حوزه ، مدت طولانی تری نسبت به جستجوی اینترنت ، وجود داشته است . در حقیقت ، دانشمندان می کوشیده اند تا کامپیوتر زبانهای طبیعی را برای چندین دهه ، درک نمایند ! تجزیه و تحلیل متن ، یک بخش مکمل این تلاشهاست اگر چه بسیاری از محققان برجسته به مسئله اقتباس و نمایش معنا از اسناد زبان طبیعی ، حمله کرده اند اما روشهای فنی و استانداردهای متنوعی برای استخراج داده ها امروزه وجود دارد که منجر به غیبت یک برنده غیر قابل انکار ، می گردد.

تجزیه و تحلیل خودکار اطلاعات متن را می توان برای چندین هدف کلی

توضیح داد:

• ایجاد یک بررسی از محتویات مجموعه بزرگ اسناد ؛ به عنوان مثال ، یافتن دسته های مشخصی از اسناد در مجموعه بازخور مشتری می توانست نشان دهد که محصولات شرکت یا خدمات نیاز به اصلاح دارند.

• شناسایی ساختارهای پنهان بین گروه های اشیاء ؛ این به سازمان کمک می کند تا یک سایت اینترنت را سازماندهی نماید ، بطوریکه اسناد مربوطه با هایلینکها ، بهم متصل می شوند.

• افزایش کارایی و سودمندی فرآیند جستجو ، برای یافتن اطلاعات مناسب مشابه ؛ به عنوان مثال ، جستجوی مقالات از سرویس خبری و کشف کل اسناد بی نظیری که حاوی نشانه ها در مورد روندهای ممکن یا تکنولوژیایی است که تاکنون در مقالات دیگر ذکر شده اند.

#### جمع آوری اسناد نسخه برداری شده ذکر نشده اند.

بطور خلاصه استخراج متن را می توان در هر کجا که مقدار زیادی از متن وجود دارد که باید تجزیه و تحلیل شود ، مورد استفاده قرار داد . اگر چه پردازش اتوماتیک امکان تجزیه و تحلیل عمیق خواندن روان را فراهم نمی آورد. اما آن را می توان برای استخراج نقاط کلیدی ، طبقه بندی اسناد و ایجاد خلاصه ها مورد استفاده قرار داد . اینجا برخی از نمونه های زندگی واقعی وجود دارند که ساختار متن با استفاده از تجزیه و تحلیل خودکار متن را می توان بطور موفق صف بندی نمود.

مدیریت پست الکترونیکی ( ایمیل ) . استفاده معروف از تجزیه و تحلیل متن ، برای ردیابی پیام است که در آن کامپیوتر پیام را می خواند تا مشخص کند آن به چه کسی مربوط می شود . کاربرد دیگر استخراج متن ، تجزیه و تحلیل آماری ماهیت پیام است .

مدیریت سند. استخراج متن به شرکتها کمک کرده است تا به هزاران سندی که در انبار دارند ، رسیدگی نمایند. با استخراج اسناد مختلف برای کسب معنا ،



هنگامیکه آنها در انبار و مخزن سند قرار می گیرند ، و یک شرکت می تواند یک شاخص کامل ایجاد کند که تعیین محل اسناد مربوطه را در هر زمان امکانپذیر می سازد.

کمک اتوماتیک ، برخی از شرکتها از استخراج داده ها برای واکنش نسبت به درخواستهای مشتری استفاده می کند . نامه ها و ایمیل های مشتری با کاربرد استخراج متن ، پردازش می شود . اگر کاربر بتواند با درجه معقولی از اطمینان مشخص کند که مشتری چه می خواهد ، بطور خودکار اطلاعات مناسب را برایش ارسال می نماید.

تحقیق بازار ، یک محقق بازار از استخراج متن روی خط برای جمع آوری آمار مربوط به وقوع واژه های مشخص ، عبارات مفاهیم با موضوعات در WWW ( شبکه گسترده جهانی ) استفاده می نماید . این اطلاعات برای برآورد و دموگرافیکهای بازار و منحنی های تقاضا مفید می شود.

جمع آوری هوش تجاری . این پیشرفته ترین استفاده از استخراج متن می باشد . معمولاً بسیاری از شرکتها اطلاعات در مورد بازارها ، رقبا و محیطهای تجاری را با استفاده از کراولرهای هوشمند شبکه جمع آوری می کنند که برخی از پیشرفته ترین اشکال روی خط استخراج داده ها را نشان می دهد . شرکت هایی همانند اینتونالیتک محصولاتی را ارائه می دهند که اینترنت را برای اخبار مربوط به موضوعات از قبل مشخص شده تماشا می کند و گزارش خلاصه ای را ارائه می نمایند.

### شبکه های معنایی و سایر تکنیکها

یک عنصر کلیدی ساخت یک سیستم پیشرفته برای تجزیه و تحلیل اطلاعات متنی، خلاصه سازی و جستجو، توسعه یک شبکه معنایی برای متن تحقیق شده می باشد . یک شبکه معنایی یک سری از مهمترین مفاهیم - واژه ها و ترکیبات واژه بدست آمده از متن های تجزیه و تحلیل شده همراه با روابط معنایی بین این

مفاهیم - واژه ها و ترکیبات واژه - بدست آمده در متن های تجزیه و تحلیل شده ، همراه با روابط معنایی بین این مفاهیم در متن می باشد . یک شبکه معنایی یک خلاصه دقیق و بسیار معتبری از متن تجزیه و تحلیل شده را فراهم می سازد .

مشابه با شبکه های عصبی مصنوعی ، هر عنصر از شبکه معنایی - یک مفهوم - با وزن آن و یک سری روابط با عناصر دیگر شبکه - یک گره زمینه ، طبقه بندی می گردد . به هر رابطه بین عناصر شبکه یک وزن اختصاص می یابد .

طیف کلی وظایف تجزیه و تحلیل متن را می توان به محض ساخت یک مجموعه از شبکه های معنایی برای متن های تحقیق شده ، اجرا نمود . در حقیقت ، یک شبکه معنایی به دقت ساخته شده ( کشف شده ) کاربردهایی را که می توانند اجرای اتوماتیک وظایف بسیاری را تقویت نمایند ، امکانپذیر می سازد . این وظایف ، شامل خلاصه سازی متن ، ایجاد پایه دانش شخصی که حرکت آن ساده است ، دسته بندی اسناد ؛ طبقه بندی پیامهای وارد شونده بر جستجوی معنایی هوشمند برای اطلاعات در یک سری از متن ها یا اینترنت ، مکانیسم حرکت برای یک کتاب الکترونیکی برای کاربر ؛ و بسیاری از موارد دیگر می شود .

برای تاکید ، اولین مرحله تجزیه و تحلیل مؤثر یک متن ، توسعه شبکه های معنایی این متن است ، البته سودمندی سیستم تجزیه و تحلیل متن بطور بحرانی بر سودمندی الگوریتم های استفاده شده برای ساخت یک شبکه معنایی بستگی دارد . در اکثریت روشهای موجود ، شبکه معنایی ، بر اساس قوانین و مفاهیم از قبل تعیین شده توسعه می یابد . با این وجود ، یک الگوریتم قوی تر می تواند یک شبکه معنایی کاملاً اتوماتیک را بر اساس یک متن تحقیق شده ، بدون هر گونه دانش قبلی در مورد موضوع ، ایجاد نماید . در این حالت ، اهمیت نسبی مفهوم متنی است که تنها با ارتباطهایش با سایر مفاهیم موجود در شبکه تعریف می گردد .

این یک ایده در نسبت چند محصول می باشد یکی از آنها ، موسوم به سیستم تحلیل گر متن (سیستم های مگاپیوتر روسیه) ، یک شبکه معنایی خودکار را بر اساس پردازش پیشرفته یک متن یکنواخت ، می سازد . کاربر نیازی ندارد تا

هر قانون را شناسایی نماید، یا اساسهایی را برای توسعه شبکه ایجاد نماید . یک شبکه معنایی ایجاد شده تنها به ساختار ، فهرست لغات و حجم متن تجزیه و تحلیل شده بستگی داد . در حقیقت ، تحلیل گر متن الگوریتم های مشابه الگوریتم های استفاده شده برای تجزیه متن در مغز انسان را اجرا می کند.

تکنیکهای دیگر . البته فروشندگان مختلف ، با یکدیگر در چندین مرز از جمله تکنیکهای استخراج داده ها رقابت می کنند.

بطور مشخص ، کامپیوتر از تعیین محل مطلق ، شناخت الگوها ، بر چسب های ثابت و شناور استفاده می کند.

سمیومپ ، از سمیوتیکها ( علوم معنایی ) محاسبه ای، بررسی رسمی علائمی که با ارتباطات طراحی شده منتقل می گردد ، استفاده می کند . نرم افزار سمیومپ ، کل عبارات مربوطه از مجموعه متن را استخراج می کند . آن یک شبکه معنایی از رویدادهای همزمان را با دسته بندی عبارات مربوطه و افزایش برجسته ترین مشخصه های این دسته بندیها می سازد . بطور مختصر ، سمیو ، به نزدیکی های عبارات در ترکیب در متن توجه می کند . تکنولوژیهای اساس پردازش معنایی ، دسته بندی اطلاعات و نمایش نموداری را ترکیب می کنند.

معماری اینجنت ویر شرکت آتونومی، الگوریتم های دارای عملکرد بالا و هماهنگ کننده الگو را با تجزیه و تحلیل زمینه ای و استخراج برای اتوماتیک ساختن طبقه بندی و اشاره به اطلاعات مربوط به یک موضوع اصلاح کارایی بازآوری اطلاعات و توانایی خصوصی سازی دینامیک محتوی دیجیتالی با هم ترکیب می کند . این محصولات استخراج متن در بخش بعدی این فصل ، مورد بحث قرار می گیرند.

### محصولات استخراج متن

تعدادی از محصولات استخراج متن در بازار امروز وجود دارد ( جدول ۱ -  
۲۰ را ببینید ) . برخی از این محصولات که بطور مفصل مورد بحث قرار می گیرند  
عبارتند از :

جدول ۱ : ۲۰

Company	Headquarters	Product
Aptex Software, Inc.	San Diego, California	SelectResponse
Autonomy	Cambridge, England	Agentware
Data Junction	Austin, Texas	Cambin
Exonlithur Technologies Corp.	Vienna, Virginia	RetrievalWare
Fulcrum Technologies, Inc.	Ottawa, Canada	DOCSFulcrum, SearchServer
IBM Corporation	White Plains, New York	Intelligent Miner for Text
InsightSoft-M	Moscow, Russia	Cross-Reader
Intercon System, Ltd	Jerusalem, Israel	DataSet
Megaputer, Inc.	Moscow, Russia	TextAnalyst
Sernio Corporation	San Mateo, California	SernioMap
Sovereign Hill Software, Inc.	Hindley, Massachusetts	InQuery
Verity, Inc.	Sunnyvale, California	KeyView, Intranet Spider

ایجنت ویر ( آتونومی ) ، آتونومی سه نوع محصول مربوط به استخراج متن  
روی خط را ارائه می کند و اینها یک خانواده « ایجنت ویر » از محصولات را  
تشکیل می دهند.

□ خادم دانش . یک ابزار دقیق و کاملاً خودکار از طبقه بندی ، فهرست  
سازی و ارائه اطلاعات را برای کاربر فراهم می سازد.

□ جدید سازی دانش. سایتهای مشخص اینترنت و اینترنت ، اخبار ، و  
مخازن داخلی سند را کنترل می کند ، و گزارش اختصاصی در مورد  
محتویات آن گزارش می نماید.

□ سازنده دانش. یک جعبه ابزار که به شرکتها امکان می دهد تا قابلیت های  
ایجنت ویر را در سیستم های خود ترکیب نمایند . ایجنت ویر با اسناد در



اینترنت تداخل می‌کند به جای آنکه بعد از جمع آوری آنها ، تداخل نماید ، آن دارای قابلیت تجزیه و تحلیل معنایی نیز می‌باشد.

معماری محصول ایجنت ویر ، الگوریتم های ابتکاری دارای عملکرد بالای و هماهنگ کننده الگو را با تجزیه و تحلیل زمینه ای هوشمند و استخراج مفهوم ترکیب می‌نماید تا طبقه بندی و مرجع سازی اطلاعات را خودکار نماید . کارایی بازآوری اطلاعات را اصلاح کند ، و خصوصی سازی دینامیک محتوی دیجیتالی را امکانپذیر نماید.

قدرت اتونومی در الگوریتم های هماهنگ کننده الگو با عملکرد بالای آن قرار دارد. این الگوریتم ها ، یا اصول نظریه اطلاعاتی شانون ، احتمالات بیسیمان و جدیدترین تحقیقات در مورد شبکه های عصبی و تقویت می‌گردند . طراحی مفهوم احتمالی و انطباقی ایجنت ویر (APCM) می‌تواند یک متن را تجزیه و تحلیل کند و مفاهیم اصلی در سند را شناسایی نماید، زیرا آن درک می‌کند چگونه فرکانس و روابط عبارات با معنا ارتباط دارد . ایجنت ویر ، تکنولوژی هماهنگ کننده الگو پیشرفته ( پردازش غیر خطی، دیجیتالی، انطباقی سیگنال ) را برای استخراج، چکیده دیجیتالی هستند، و در تعیین خصوصیات که به متن معنا می‌دهند، بکار می‌برد . بعد از آنکه تکنولوژی APCM اتونومی در امضاء بی نظیر مفاهیم اصلی را شناسایی کرد، و سایر اسناد ایجاد می‌کند . به دلیل این که آن به لغات کلیدی تکیه ندارد ، می‌تواند با هر زبانی کار کند.

موتور استدلال دینامیک (DRE) ، در مرکز سیستم ایجنت ویر اتونومی، قرار دارد . DRE بر اساس تکنولوژی پیشرفته هماهنگ سازی الگو قرار دارد که از تکنیکهای شبکه عصبی با عملکرد بالا که در ابتدا توسط شرکت مادر اتونومی یعنی شرکت نیورودونامیک توسعه یافت ، بهره برداری می‌کند . DRE و APCM را برای انجام چهار وظیفه اصلی بکار می‌برد: هماهنگ سازی مفهوم ، ایجاد عامل ، آموزش مجدد عامل و جستجوی متن استاندارد DRE یک واژه Boolean یا بررسی

زبان عصبی را می‌پذیرد و به فهرستی از اسناد حاوی عبارات مرتب شده با ارتباط با بررسی و درخواست ، باز می‌گردد.

کامپیو ( شرکت دیتا جانکشن ) ، کامبینو ، یک محصول است که اسناد را بررسی می‌کند، و داده های معنادار یک فایل بانک اطلاعاتی را استخراج می‌نماید . این یک ابزار است که تحت کنترل سیستم مرکزی نیست ، اما به سادگی در ارتباط با یک مجموعه از اسناد جمع آوری شده توسط کراولر وب ، کار می‌کند . کامبینو، از تعیین موقعیت مطلق ، شناخت الگو ، برچسبهای ثابت و شناور و سایر روشها برای علامتگذاری عناصر داده ها در یک فایل متن استفاده می‌کند . کامبینو، تابع های تجزیه و تحلیل معنایی محصولات دیگر را ندارد . آن ، همراه با کراولر وب ، اسناد را جمع آوری می‌کند و سیستم استخراج متن روی خط را در سطح ورودی، ارائه می‌دهد .

اینترنت ماینر برای متن (IBM) . اینترنت ماینر ، برای متن (IMT) یک جعبه ابزار توسعه نرم افزار است کاربردهایی که با IMT توسعه یافت ، می‌توانند اطلاعات را از منابع متنی همانند نامه ها ، صفحات وب ، سرویسهای خبری روی خط ، جمع آوری نمایند . IBM ، قابلیت استخراج الگوها از متن سازماندهی اسناد با موضوع و جستجوی اسنادی که با یک موضوع مشخص هماهنگ هستند را ارائه می‌دهد IMT دارای ابزار تجزیه و تحلیل متن و موتور جستجوی پیشرفته می‌باشد که با عملی بودن استخراج و قابلیتهای تجسم نتایج ، افزایش یافته اند . این ابزارهای وب ، کل عناصر را برای توسعه راه حل‌های استخراج متن آماده می‌نماید . IMT برخی کاربردهای نمونه را فراهم می‌کند که می‌توان با نیازهای کاربران آن را اصلاح نمود . این محصول ، هدف دارای بازار استخراج متن روی خط را با قابلیت‌های جامع جستجوی شبکه و تجزیه و تحلیل معنایی پیشرفته ، نشان می‌دهد.

ابزار شناسای زبان IMT، از سرنخها در محتویات سند برای شناسایی زبانها بر اساس یک سری از اسناد آموزشی در تجزیه و تحلیل معنادار، استفاده می نماید.

استخراج کننده مشخصه IMT، می تواند در دو حالت ممکن کار نماید. در ابتدا، آن سند را به تنهایی تجزیه و تحلیل می نماید. در حالت ارجح، آن فهرست لغت را در سند موجود در دیکشنری قرار می دهد که قبلاً بطور خودکار از مجموعه اسناد مشابه ساخته شد. هنگام استفاده از مجموعه سند، استخراج کننده مشخصه قادر است تا مدرک را از بسیاری از اسناد برای یافتن فهرست لغات بهینه، جمع آوری کند. سپس می تواند یک مقیاس اهمیت آماری را به هر مقوله فهرست لغات اختصاص دهد.

استفاده از تخمین های سریع و سخت، ضریب استخراج نام، رویدادهای نامها را در متن مشخص می کند، و تعیین می کند که نام به چه نوع ماهیتی اشاره دارد: شخص، مهمی سازمان و یا سایرین (همانند یک نشریه، پاداش، جنگ و غیره). ضریب استخراج نام، به یک بانک اطلاعاتی از نامهای از قبل موجود نیاز ندارد. بلکه آن نامها در متن را بر اساس تخمین های دارای انگیزه زبانشناسی که از تیپوگرافی و مقررات زبانی بهره برداری می کنند، کشف می کند. آن در سرعتهای بالا کار می کند، زیرا آن نیازی به انجام عبارت بندی نحوی متن، ندارد. ضریب استخراج عبارت IMT از یک سری از تخمین های ساده برای شناسایی عبارت های فنی چند واژه ای در سند استفاده می کند. آن تخمین ها که بر اساس یک دیکشنری حاوی اطلاعات بخش سخنرانی برای واژه های انگلیسی می باشد، که شامل هماهنگی الگوهای ساده به منظور یافتن عباراتی که دارای ساختارهای عبارت اسمی اختصاصی عبارات فنی هستند، می باشد، این فرآیند بسیار سریع تر از روشهای جایگزین می باشد.

دسته بندی، یک فرایند کاملاً اتوماتیک است که مجموعه اسناد را به گروهها تقسیم بندی می کند. اسناد در هر گروه از جهاتی مشابه یکدیگر هستند. هنگامیکه

محتوی اسناد به عنوان اساس دسته بندی استفاده می شوند ، گروه های مختلف با موضوعات مختلفی مطابقت دارند که در مجموعه مورد بحث قرار می گیرد . بنابراین - دسته بندی یک روش یافتن محتویات مجموعه می باشد . ابزار دسته بندی برای کمک به شناسایی موضع یک دسته - فهرستی از عبارات یا واژه ها را شناسایی می کند که در اسناد گروه ، عادی می باشد . دسته بندی را می توان در خصوص ترکیبات خصوصیات اسناد همانند طول ، زمینه تاریخ و غیره و تمایز داده ابزار های دسته بندی در شرح استخراج داده های ابزار ( اینتلیجنت ماینر برای داده ها ) در آن نوع مسئله عملی می باشد.

طبقه بندی کننده IMT در جعبه ابزار با اقتباس مشخصه های اختصاصی از اسناد طبقه بندی شده ، کار می کند ، سپس این مشخصه ها را با یک سری از مشخصه ها برای هر مقوله مقایسه می کند که از اسناد نمونه در طی فاز آموزش ، استخراج گردید . این روش ، یک شاخص فشرده و پردازش سریع را تامین می کند . موتور جستجوی متن IMT ، تعدادی از مشخصه های پیشرفته را ارائه می کند ، و آن را به یکی از پیشرفته ترین محصولات از نوع خود در بازار امروز ، تبدیل می نماید . آن تعداد از نمونه های مختلف جستجو را در داخل همان موتور جستجو اجرا می کند . قلب موتور تحقیق یک ساختار شاخص است که بررسیهای هیبریدی ، متن آزاد ، Boolean را تقویت می کند . جستجوهای فونتیک در همان ساختار شاخص ممکن ست . یک شاخص دارای هدف ویژه ، جستجوهای نامعلوم و زبانهای مبتنی بر کاراکتر بایت دابل را تقویت می کند : ژاپنی ، چینی و کره ای . نمونه های IMT : فایناسن وایز ، یک جستجوی توسعه یافته توسط نشریات ریسک و اوراق بهادار IBM ، و بازارهای سرمایه ، کاربر را قادر می سازد تا اطلاعات جامع بازار مالی در اینترنت را جستجو نماید . یک کاربرد موسوم به تکنولوژی واچ از الگوریتم های استخراج متن IMT برای تجزیه و تحلیل کاربردهای انحصاری به منظور کشف روندهای اصلی در تحقیق جهان در



صنایع ویژه ، تمرکز سرمایه گذاری تحقیق و محتوی تحقیق در دموگرافیک های مشخص ، استفاده می کند.

تکنولوژی واچ : یک کاربرد IMT است که از تکنیک تجزیه و تحلیل داده های ارتباطی که توسط مرکز اروپایی IBM برای ریاضیات کاربردی در پاریس (ECAM) ، تکمیل شد استفاده می کند . آن علائم انحصاری انتخاب شده با اسناد متنی را تجزیه و تحلیل می کند . و بطور خوار آنها را از طریق حجم به چند گروه طبقه بندی می نماید.

تکنولوژی واچ IBM ، این اطمینان را ایجاد می کند که علائم انحصاری با اسناد در هر گروه تا جایی که امکان دارد ، تشابه هستند ، و گروه های مختلف ، تا جایی که امکان دارد از هم متمایز می شوند. بطور خلاصه تکنولوژی واچ IBM ، یک خلاصه تعداد علائم انحصاری در هر گروه و واژه های کلیدی توصیه کننده هر گروه را فراهم می کند . یک شاخص دیداری از روابط بین گروه ها ، با خطوط رنگی که آنها را بهم پیوند می زند، ارائه می گردند . و رنگها استحکام رابطه را نشان می دهد . برآمدگیها روی گره ها (توپولوژی . اندازه ، توصیف کننده ها) یا برآمدگیها در ارتباطها ، نیز موجود هستند.

سیمومپ ( شرکت سیمو ) . سیمومپ ، یک محصول استخراج متن که حجم های بزرگ متن را اسکن می کند و نقشه های مفهوم قابل جستجوی اطلاعات را می سازد . پس شما می توانید مفاهیم را در یک مجموعه بزرگ از اسناد متن بدون نیاز به اسکن در کل فایل بررسی کنید . سیمو ، مستلزم این است که کل اسناد با استفاده از کراوله وب جمع آوری شود . بعد از آنکه سیمو کل اسناد را تجزیه و تحلیل نمود ، از طریق آنها یکی یکی کار می کند و معنای آنها را تفسیر می نماید . آن تجزیه و تحلیل معنایی خودکار را انجام می دهد، اما آن برای استفاده با تاثیر متقابل کاربر طراحی شده است ، نه برای یک حالت دسته ای .

سمیومپ ، از تکنولوژی اختصاصی استخراج متن (SEML OLEXM) استفاده می کند که سمیو فوندر کلاود وگل استفاده می کند . آن معانی محاسبه ای ، بررسی رسمی علائمی که با ارتباطات طراحی شده، منتقل می گردد را بکار می برد . نرم افزار سیمیومپ ، کل عبارات مربوط را از مجموعه متن استخراج می کند . آن یک شبکه معنایی رویدادهای همزمان را با دسته بندی عبارات مربوطه و افزایش برجسته ترین مشخصه های این دسته بندی ها ، می سازد . الگوهایی که سیمیومپ ایجاد می کند ، یک اساس تصویری از مجموعه متن را نشان می دهد . اصل تصویری ، مفهوم کلیدی است که روابط معنادار ، مناطق اصلی دسته بندیهای پیچیده و موضوعات گذرا که کل شبکه معنایی دانش را بیان می کند ، را جذب و بر آن تمرکز می نماید . تکنولوژیهای اصلی ، پردازش معنایی ، دسته بندی اطلاعات و نمایش گرافیکی را ترکیب می کند .

تحلیلگر متن ( شرکت مگاپیوتر). تحلیل گر متن یک استخراج متن هوشمند و سیستم جستجوی اطلاعاتی پرمعنایی است که یک تکنولوژی شبکه عصبی بی نظیر را برای پردازش ساختاری تنهای نوشته شده در زیان طبیعی اجرا می کند . عملی بودن آن مشابه عملی بودن سمیوست . آن برای علمیات دارای تاثیر متقابل و با مجموعه ها و اسنادی است که قبلاً جمع آوری شده اند . تحلیل گر متن را می توان برای ایجاد پایگاههای دانش جستجوی معنی اطلاعات دو اقتباس اتوماتیک متن مورد استفاده قرار دارد .

این نرم افزار استخراج متن یک شبکه معنایی را در سه مرحله ایجاد می کند . متن از درون یک پنجره تک متغیره در هر زمان به صورت یک سمبول رانده می شود . سمبولها ، ممکن است حروف ، علامت گذاری ، یا فضاها ی خالی باشند . پنجره رامی توان از دو تا ۲۰ سمبول تنظیم نمود . هنگامیکه متن از درون پنجره عبور می کند، عکسهای فوری گرفته می شود که برای ایجاد نمایشی از واژه ها، ریشه های کلمه ، و دسته بندی لغات در متن مورد استفاده قرار می گیرند .

مرحله بعدی، شناسایی چگونگی مواجهه این مفاهیم در قطعات معنایی متن همانند جمله می باشد. بعد از این مرحله، سیستم دارای یک شبکه معنایی مقدماتی است که در آن هر واژه دارای وزن و هر مفهوم موجود دارای یک وزن مطابق با آن بر اساس تجزیه و تحلیل فرکانس، می باشد این روابط بین واژه ها نیز بر اساس چگونگی وقوع آنها با همدیگر، وزن دار هستند.

سومین مرحله در فرآیند شامل استفاده از شبکه عصبی مشابه با شبکه های هاپفیلد می شود که شبکه های عصبی تک بعدی هستند که همه عصب ها به هم متصل هستند.

شبکه معنایی اولیه به عنوان ورودی شبکه عصبی مورد استفاده قرار می گیرد. نتیجه یک شبکه معنایی پاکسازی شده می باشد. از طریق سنجش روابط، شبکه مجدداً عادی می شود و یک شبکه معنایی نمایی را فراهم می سازد. ایجاد یک شبکه معنایی مهم ترین عنصر اساسی است. کاربرد های تحلیل گر متن شامل ساخت پایگاه های دانش، تجزیه و تحلیل محتوی متن های اختیاری و خلاصه سازی متن ها، طبقه بندی متن ها به موضوعات مشخص و انجام جستجوهای معنایی اطلاعات، می شود.

### استفاده از قدرت مغز انسان

در طی ۲۵ سال گذشته سازمانها در تمام بخشهای صنعت حجم های بزرگی از داده ها را جمع آوری نموده اند. آنها انواع متعددی از ابزارهای پردازش داده ها و تکنولوژیها را برای سازماندهی، اجرا و دسترسی به آن، خریداری نموده اند. میزان رو به افزایش داده ها در ارتباط با رقابت افزایش یافته و ساختار بندی مجدد سازمانی، موسسات را مجبور ساخته اند تا استفاده از تکنیکهای استخراج داده ها را آغاز نمایند، و آنها را به منظور جستجو و کشف بخشهای با اهمیت اطلاعات مدفون در بانکهای اطلاعاتی، عمل آوری نمایند.

با این وجود این ذخایر داده ها و دانش مدفون در آن ، ساخت انسانها هستند . می توان استدلال نمود که به دلیل این که ارتباط بین کامپیوتر و کاربر همواره در زمینه معنایی درک انسان است یک طبقه جدید از ابزار - موسوم به ابزار محاسبه معنایی - برای استفاده کامل از این حقیقت لازم است .

یک شرکت روشی را برای استخراج داده ها پیشنهاد کرده است که به همان اندازه آرکئولوژی ، از ابزارهای استخراج داده ها متفاوت است . شرکت ASOC نامیده می شود که علامت اختصاری ( Associative computing ) و آن ابزار محاسبه معنایی را توسعه می دهد . بخش زیر از انتشارات ASOC اقتباس شده است تا بطور معتبر نمونه محاسبه معنایی را توضیح دهد.

در ظهور عصر اطلاعات، محیطی را که شرکتها در آن برای بقا مبارزه می کنند، تغییر می دهد . توسعه ابزارهای استخراج داده ها یک دیدگاه مفید اما باریک برای واکنش نسبت به آن چالش می باشد . اگر چه قیاسی با استخراج یک ستاره عموماً استفاده شده می باشد ، اما موقعیت بسیار مستعدانه با فشار انقلابی مقایسه می شود که منجر به توسعه سنسورها ، سیستم های عصبی و مغز برای ایجاد ارگانیزم هایی با دسترسی سریع و بسیار مفید به اطلاعات غنی در محیط های طبیعی شده است و واکنشهای سریعی را به آنها ارائه نمود که حیوانات را برتر از گیاهان می سازد .

به نظر می رسد که سازمانها بایستی یک دهه انقلابی مشابه را برای مشارکت در محیط اطلاعاتی غنی تر تحمل نمایند . این توسعه تکنولوژی را برای طراحی در ایجاد سیستم های تصور مصنوعی را برای سازمانها ضروری می سازد . اینها ، آنها را قادر خواهند ساخت تا اشیاء ، مقرارات ، فرصتها و خطرات در فضاهای اطلاعاتی با ابعاد بیشتر که برای بقای آنها مهم هستند مشاهده نمایند و احساس کنند.

به عنوان اولین مرحله برای این هدف ، ما حواس طبیعی خود را با حوزه های داده های مصنوعی توسعه خواهیم داد اگر چه این هنوز یک روش مبهم به



نظر می‌رسد ، اما حداقل یک شرکت وجود دارد که این نمونه اولیه تجاری سیستم دید داده ها را بر اساس آن فلسفه تکمیل کرده است . شرکت ، ASOC نامیده می‌شود و محصول آن SPHINXVISION است.

شرکت بیان می‌کند که ابزار آن یک طبقه از الگوریتم های انطباقی را که اینک اکثر فرایندهای اساسی تشکیل نقشه های حسی در مغز است را توضیح می‌دهد اینها همان الگوریتم هایی هستند که قبلاً در بسیاری از کاربردها موفق بوده اند . ASOC این روش را در طی سالهای گذشته توسعه داده است و آن را با تداخل کاربر گرافیکی دارای عملکرد بالاتر ترکیب نموده است که می‌تواند کاربر را مستقیماً در تماس با فضاهای اطلاعاتی از انواع متنوع قرار دهد .

اگر چه اکثر روشهای دیگر جعبه ابزارهایی با یک مجموعه متنوع از روشهای مختلف می‌باشد (متغیر از شبکه های عصبی و الگوریتم های ژنتیکی تا الگوریتم های آماری متعدد)، اما محصول ASOC از استراتژی مختلف پیروی می‌کند . یکی از توانایی های مهیج حواس انسان این است که آنها به ما امکان می‌دهند تا اشیاء ، فرصتها و خطرات را در یک روش مستقیم و تقریباً بدون زحمت، بدون مشورت با کتابچه راهنمایی کاربر شناسایی نمائیم ، و از هر توالی هوشمند ارزیابی پیروی نمائیم ، بدیهی است که به یک شخصیت مشابه از سیستم نیاز داریم که تصور طبیعی ما را از فضاهای اطلاعاتی مصنوعی گسترش می‌دهد ، این امر را توسعه دهندگان شرکت بیان می‌کند . اگر چه بیان این امر بسیار ساده تر از عمل است ، شرح فعلی Sphinx vision قبلاً مسیر طولانی را در جهت این هدف مبهم طی کرده است .

کلید آن این مفهوم است که شرکت آن را محاسبه معنایی می‌نامد . برای کاربر ، محاسبه معنایی بدین معناست که وی می‌تواند تا حد زیادی جدا از سطح الگوریتم های ویژه کار کند و در عوض بر مفاهیم معنایی تکیه نماید که برای حوزه وی دارای اهمیت هستند.

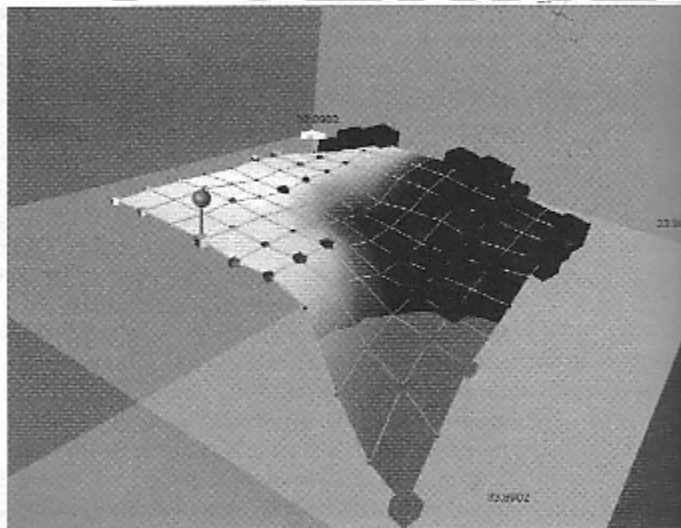
یک نکته در مور مفهوم معنایی بازدید نمایشی دانش انسان را در نظر بگیریم که به وسیله زبان بین اشخاص منتقل می‌گردد. زیان مرکب از یک سری تک واژه هاست که هر یک از آنها به معنای ویژه ای مربوط می‌گردند. ترکیب هر چند واژه در یک جمله (تشکیل یک ساختار گرامری)، درجه بالاتری را از معنا با محاط نمودن واژه ها در زمینه، ایجاد می‌کند. ترکیب این زمینه، تحت تسلط قوانین معنایی مشخص برای ترکیب لغات در جملات معنادار قرار دارد. دانش در یک دنیای معنای تحت سلطه زمینه های معنایی منتقل می‌گردد. مفاهیم معنایی در دنیای زمینه های معنایی قرار دارند.

اگر کاربر از قبل بداند کدام مفاهیم معنایی مناسب هستند، Sphinx vision به کاربر امکان می‌دهد تا مستقیماً وابستگی ها و بی نظمی‌ها در مفاهیم را مورد تحقیق قرار دهد. آن با ارائه دورنماهای اطلاعات با ابعاد کاهش یافته و کلی که می‌تواند به روشهای متعدد برای بهره برداری از توانایی های توسعه یافته مغز برای شناخت الگوی برای ایجاد روابط دقیق در فضای بسیار بعدی قابل ردیابی، این کار را انجام می‌دهد. بعد از آنکه چنین الگوهایی شناسایی شدند، Sphinx vision روشهای متعددی را برای کاربرد روشهای سنتی ایجاد شده برای کمک به تجزیه و تحلیل بیشتر، ارائه می‌نماید.

با این وجود، شرکت تاکید می‌کند که مفهوم محاسبه معنایی بطور مهمی بیشتر از این است. در بسیاری موارد مفاهیم معنایی دقیق برای یک تجزیه و تحلیل از اول مشخص نیست، اما در عوض بایستی به عنوان بخشی از خود تجزیه و تحلیل، یافت شود. ین را نمی‌توان بطور کامل اتوماتیک نمود. آن همواره مستلزم اتکاء به مهارت کاربر می‌باشد. اکثر کارهای مهندسی ASOC در مورد محاسبه معنایی، در جهت ارائه یک پیوند ادراکی بسیار کنترل شده بین دانش حوزه کاربر و بانک اطلاعاتی تحت تحقیق، هدایت می‌شود. یک شرط مهم برای این امر قابلیت ایجاد دور نماهایی از داده هایی است که قبلاً بطور مؤثر در برابر شماره کاربر قرار گرفت.

به عنوان مثال یک کاربرد نمونه را در خدمات مالی در نظر بگیرید . یک کاربردی که برای شناسایی مشتریان خطر ، طراحی گردید . نتیجه مطلوب این کاربرد ، شناسایی مقوله مشتریان خطر ، با استفاده از فرآیند اقتباس گزارشات مناسب از بانک اطلاعاتی مشتری دنیای واقعی می باشد که حاوی ثبت های مشتری با حوزه هایی همانند درآمد ، سن ، در میان سایر موارد ، می باشد . اینجا ، کل جمعیت مشتریان را می توان به صورت مجموعه ای از اشیاء با اندازه و رنگ متفاوت ارائه دارای اندازه ، رنگ ، و سایر صفات بطور اختیاری بر اساس ترکیبات ارزش صفات و ثبت مشتری ، اختصاص می یابد .

این نمایش نموداری یک دورنمای داده ها بطور مشخص و سریع در میان سایر موارد ، در جایی که مشتریان دارای اعتبار خوب قرار دارد ، ترسیم می شود . و آن این کار را در انیمیشن 3-D انعطاف پذیر چند بعدی انجام می دهد که شما را قادر می سازد تا تنها بیش از دو متغیر را مشاهده نمائید ، درست همانگونه که با نمودارهای کلیدی و دایره ای تماشا می کنید ( شکل ۱ را ببینید ) .



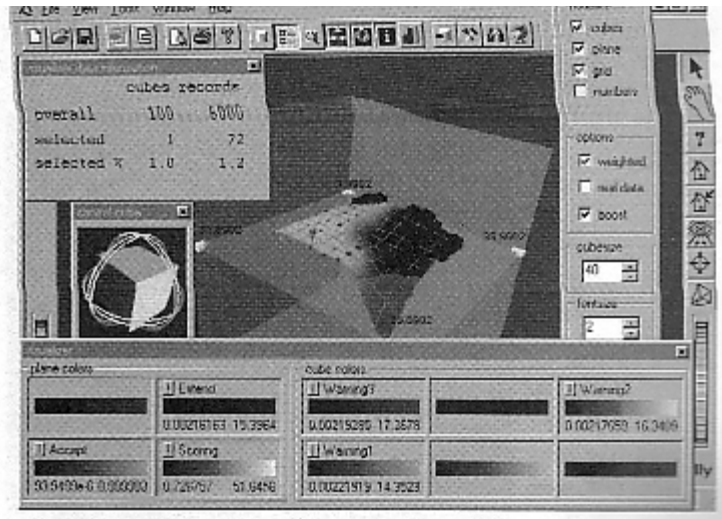
شکل ۱

در محاسبه معنایی، یک کاربر اینک می‌تواند زیر مجموعه اشیا را که ترکیبات رنگ آنها نشان می‌دهد که آنها باید مفهوم مشتریان خطر را ارائه دهند، شناسایی نماید. این مفهوم بطور آزمایش شکل گرفته را می‌توان به روشهای متعدد ارزیابی نمود. به عنوان مثال آن را می‌توان برای پیشگویی های ارزیابی خطر مشتری در یک سیاست پذیرش ویژه، مورد استفاده قرار داد. آن را همچنین می‌توان به عنوان یک مفهوم اضافی تعریف شده توسط کاربر که وابستگی هایش با مفاهیم موجود قابل تحقیق می‌باشد، معرفی کرد. بر اساس چنین ارزیابی هایی آن را می‌توان با تغییر مرزهای آن در دورنمایی اطلاعاتی رایج، بطور مؤثر، پاکسازی نمود. البته، در هر مرحله، امکان استخراج توضیحات عددی قراردادی همانند ارزشهای متوسط. یا طرحهای وابستگی تک یا دو بعدی وجود دارد.

سرانجام، هنگام تکمیل مفهوم ساخت کاربر، آن را می‌توان به صورت یک مدل معنایی بطور مجرد قابل استفاده برای کاربرهای بعدی یا ارزیابی های بعدی حفظ نمود. این امر امکان ایجاد یک کتابخانه از گونه های مختلف از یک مفهوم معنایی را برای کاربر فراهم می‌سازد. به عنوان مثال، یک بانک می‌توانست چندین مفهوم خطر مشتری را به مرور زمان توسعه دهد که هر یک از آنها را می‌توان بطور مشخص برای انواع مختلف خدمات مالی ایجاد نمود. در این حالت، محاسبه معنایی ممکن است بسیار انعطاف پذیر باشد و استاندارد سازی یک بخش مهم از دانش تجربی داخل سازمان را تقویت کند و این دانش را در سایتهای دوردست نیز فراهم نماید.

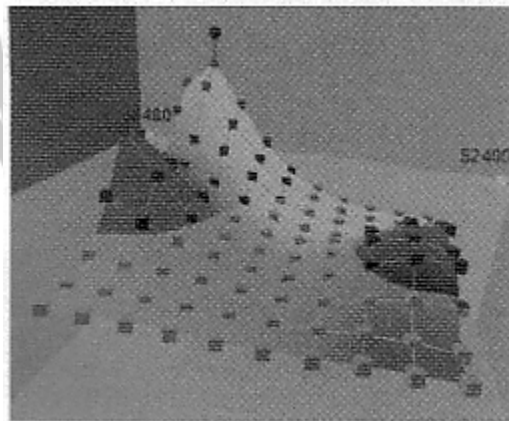
کاربران از طریق یک تداخل کاربر نموداری 3-D، دارای تاثیر مقابل و ابداعی، می‌توانند یک نمایش را برای تنظیم فرضیه های جدید متحرک نمایند، و برای استفاده از تصوراتشان با یا بدون طرح هر گونه سؤالی در مورد تجزیه و تحلیل شما در داده ها، راهنمای شوند. Sphinx vision به کاربران برای حرکت در این دور نمای اطلاعاتی پیچیده، کمک می‌کند (شکل ۲) را ببینید.





شکل ۲

ASOC از یک تکنولوژی بی نظیر محاسبه معنایی علامت انحصاری استفاده می کند که بر اساس اهداف دانش وابسته متفاوت موسوم به عاملهای داده ها (Data Agents)، سناریو و مدل های معنایی می باشد. عاملها و سناریو همواره در یک جفت عامل- سناریو ترکیب می شوند. شکل ۳ جفت های عامل سناریو برای استخراج زمینه ها از یک فضای داده های بسیار بعدی، مورد استفاده قرار می گیرند، آنها معادل کامپیوتری شده از معانی را شکل می دهند. عاملهای داده ها بهم مرتبط بر طبق زمینه شان با همدیگر در سناریو دسته بندی می شوند.



شکل ۳

### نتیجه گیری

اگر چه بازار ابزارهای استخراج داده ها نسبتاً کوچک است اما در همان زمان ، بازار راه حل کاربرد استخراج داده ها بطور نمایش در حال رشد است . حداقل یک بخش از این رشد ، به اقتباس گسترده اینترنت و صعود گریز ناپذیر کل اشکال تجارت الکترونیکی نسبت داده می شود . بنابراین ، پیشگویی آینده مشکل است ، بیائید به برخی از روندهای شایع در بازار توجه کنیم - روندهایی که بر تکنولوژی استخراج داده ها تکیه دارند ، و از آن سود می برند .

### مدیریت دانش

مدیریت دانش نظامی است که یک روش کامل برای ترویج ایجاد حفظ ، دستیابی ، و اتکاء به دارایی های نامحسوس یک مؤسسه برای سود تجاری ، می باشد.

یک انبار داده ها ، ترکیب بطور سلسه مراتبی ساختار بندی شده منظم از خدمات مبتنی بر دانش است که امکان حل یک طیف گسترده از مسائل نگهداری داده ها و تجزیه و تحلیل را فراهم می آورد که از خانه داری تا ایجاد مدلهای پیش بینی هوشمند ، متغیر می باشد . نمونه ها شامل ، طرح فروش آتی محصول بر اساس داده های مشتری فعلی که از تجارت روزانه جمع آوری گردید ، یا یک سیستم اخطار برای کشف خستگی های احتمال مشتریان برای یک شرکت ارتباطات از راه دور می باشد.

از دور نمای کاربر نهایی ، انبار دانش ، به صورت مجموعه ای از خدمات به نظر می رسد .

### مدیریت رابطه مشتری در جهان تجارت الکترونیکی

هدف ذکر شده CRM، قابلیت آن برای استفاده از تاثیرات متقابل مشتری در کانالها و وظایف می باشد. این قابلیت برای ایجاد روابط سودمند و صادقانه مشتری، قاطع است هنگامیکه شرکتها سرمایه گذاریهای کلانی را در بازارهای داده های، واسطه، و کاربردهای جدید می نمایند، انتظار دارند کاربردهای CRM بتواند از استراتژیهای آنها حمایت نماید.

هنگامیکه تجارت الکترونیکی به تکامل خود ادامه می دهد و بر تغییرات رادیکال در تمام جنبه های تجارت اثر می گذارد. قانون نرم افزار کاربردی مبتنی بر تجارت الکترونیکی، از سکوهای تجاری دقیقاً تعریف شده، به سمت دید گسترده تر اداره روابط مشتری پیش می رود. بعلاوه بازار یابی به ظهور خود در یک فعالیت شخصی حلقه بسته خارج از مرز ادامه خواهد داد که این با افزایش وب سایتهای شخصی بازاریابی ایمیل و ردیابی هوشمند تر و تجزیه و تحلیل روند مشخص می گردد.

اکثر محصولات CRM امروزی قابلیت های قوی در ردیابی تاثیرات متقابل ارباب رجوع مبتنی بر تلفن دارند که کارمندان در کاربردها ثبت می نمایند. با این وجود، سایر داده های مشتری (همانند معاملات تاریخی اعتباری) تنها از طریق APIS اختصاصی، قابل دسترسی هستند. در نتیجه، کاربردهای CRM نمی توانند یک منظره از کل رابطه را ایجاد نمایند - منظره های که می توان در تاثیرات متقابل آینده مورد استفاده قرار داد.

بعلاوه بسیاری از کاربردهای CRM، مبتنی بر وظیفه حفظ ثبت را فراهم می سازند. اینها کیفیتهای خوبی هستند که به تنهایی به شرکتها در تشکیل روابط شخصی، متقابل با مشتریان در شبکه کمک نمی نمایند اغلب. سایتهای تجاری موفق، از کاربردهای مجزای شبکه با داده های مشتری مشخص تجارت و قوانین تجاری استفاده می نمایند. هنگامیکه مشتری به یک کانال روی خط منتقل می شود. نویسندگان معتقدند که محصولات CRM جدید و ابزار های CRM فیلور مشخص

افزایش خواهند یافت تا به یک نقطه کانونی مدیریت رابطه چند وظیفه ای و چند کانالی تبدیل گردند.

ابزار آتی CRM چندین سیستم برخورد با مشتری را با موتور ساخت رابطه، از طریق انجام موارد زیر، همزمان می‌سازد:

- ایجاد یک درک کامل زمان واقعی مشتری
- ارائه کنترل بیشتر جنبه های رابطه به مشتریان و عرضه کنندگان به عنوان مثال نوع خدمتی که آنها ترجیح می‌دهند یا اطلاعاتی که مشتاقند با دیگران تسهیم نمایند.

• ساخت محصولات مبتنی بر شبکه برای تغییر اقتصاد حمایت از مشتری از طریق خدمات بیشتر و دسترسی به مقادیر زیادی از داده ای مشتری که در شبکه موجود است.

بطور خلاصه، شرکتها بایستی چگونگی اداره روابط را تغییر دهند، مدلی جدیدی که شرکت فورستر ریسرچ، مدیریت روابط الکترونیک (Erm) می‌نامد، به صورت زیر تعریف می‌شود:

“یک روش متمرکز بر شبکه برای همزمان سازی روابط مشتری در سراسر کانالهای ارتباطی، تابعهای تجاری و بینندگان.”  
برای اجرای این مدل تجارت الکترونیکی CRM، شرکتها بایستی کارهای زیر را انجام دهند.

- ایجاد یک زمینه مشتری دینامیک که می‌تواند هر تاثیر متقابل مشتری را که متفاوت از ایده مشتری است که از داده های موجود در کاربردهای ساخته شد، مورد تاکید قرار دهد. این را می‌توان با جمع آوری سازماندهی داده های تاریخی مشتری / محاسبه متریک های سطح بالا برای هر مشتری و یعنی سودمندی مشتری، رضایت و پتانسیل خستگی مشتری و ایجاد و ارائه زمینه دینامیک به نقاط تماس مشتری بدست آورد.



□ ایجاد واکنشهای یکنواخت و مرسوم ارائه موتور قوانین محکم برای ردیابی، جریان کاری، خصوصی سازی حرکت هوشمند و عملکرد یکنواخت مشتریان

□ ایجاد و حفظ یک دایرکتوری محتوی برای اشاره به شرکت، محصول و رضایت شریک تجاری و ارائه آن به کارمندان، شرکاء تجاری و مشتریان

بطور مشخص، کل فعالیتهای به تکنولوژی موتور تصمیم کامل و محکم احتیاج دارند که بر اساس تکنیکهای استخراج داده ها و کاربردهای بحث شده در سراسر این بحث می باشد. بسیاری از شرکتها، در مسیر حوزه CRM حرکت می کنند. ما احتمالاً بسیاری از فعالان سنتی CRM و همینطور فروشندگان ERP اصلی را مشاهده می کنیم که مشغول تحصیل و یا توسعه متمرکز برای تحکیم موقعیت در این بازار در حال رشد می باشند.

#### تهیه کنندگان خدمات کاربرد

بخش قبلی قصد داشت تا به روندهای بلند مدت و کوتاه مدت برای مدیریت رابطه مشتری توجه نماید. توسعه جالب دیگر در زمینه CRM به این حقیقت مربوط می گردد که CRM یکی از زمینه های کاربرد اصلی برای تهیه کنندگان خدمات کاربرد (ASPS) می باشد.

کارتترگروپ چهار مدل برای خدمات ASP را تعریف می کند:

مدل کانسریتوم ASP: بین کاربر نهایی و فروشنده نرم افزار قرار می گیرد، و خدمات و نرم افزار را فراهم می سازد.

□ مدل شریک ASP: فروشنده نرم افزار و ASP بطور مستقیم به مؤسسه می پردازند.

□ مدل حمایت از افزایش یا تطبیق ASP: ASP یک تماس ابتدا - انتها برای مؤسسه است، و از فروشنده نرم افزار برای حمایت استفاده می کند.

□ مدل حمایت مشترک ASP : ASP و فروشندۀ نرم افزار از طریق یک شریک تجاری ، معمولاً یک ترکیب کنندۀ سیستم ، یا مؤسسۀ کار می کنند .

در اکثر معاملات مربوط به ASP این تهیه کنندگان خدمات و تکنولوژی مسئول نصب ، مرسوم سازی محدود ، نگهداری ، و حمایت نرم افزار CRM می باشند اگر چه این خدمت را می توان برای کاربران ارباب رجوع / خادم اجرا نمود ، بسیاری از معاملات جدید . بر دستیابی شبکه به کاربرد ها توجه کرده اند .

به عنوان مثال ، یکی از خدمات به سرعت در حال رشد ASP CRM برای تهیه کنندگان خدمات اینترنتی می باشد که خدمات موجود در وب سایت CRM را فراهم می سازد. یکی دیگر از خدمات ASP به سرعت در حال رشد ، خدمت موتور تصمیم می باشد . مورد آخر برای ایجاد خدمات کاربرد برای کاربردهایی همانند کشف کلاهبرداری در خدمت کاربرد HNC Falcon که صادر کنندگان اصلی کارت اعتباری استفاده کردند سفارشات فروشی متقابل و تقسیم بندی بازار می باشد،

بطور مشخص ، این خدمات کاربرد ، اجرای مستقیم کاربردهای استخراج داده ها هستند و به عقیده ، مؤلفان این منطقه فرصت رشد دیگری را برای کاربردهای استخراج داده ها نشان می دهند.

محصولات جدید ASP استخراج داده ها و خدمات ، باید انعطاف پذیر ، قابل تطبیق ، و ایمن باشد . تنها چند شرط برای این طبقه از کاربرد بیان شد . بعلاوه ، خدمات استخراج داده ها ASP باید به سؤالاتی در مورد ارائه مجوز و معماری فنی با توجه ویژه به روش تناسب کاربردهای موجود در محیط ASP ، و چگونگی ترکیب کاربرد ASP با کاربردهای مشروع حفظ شده در داخل ، پاسخ دهد .

سازمانهایی که سرمایه مهمی برای سرمایه گذاری در ابتکارات CRM ندارند یا نرخ بالای تغییر در روابط مشتریانش پیش بینی نمی کنند ، احتمالاً بهترین نامزدها برای استفاده از خدمات ASP CRM می باشند .

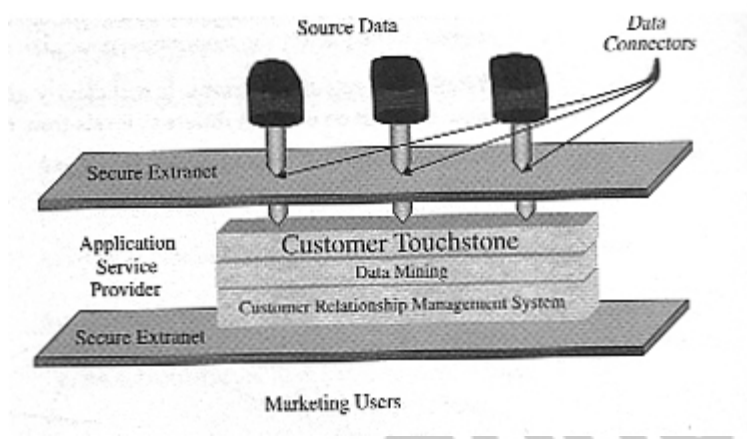
این اجراهای سیستم CRM از طریق تهیه کنندۀ خدمات کاربرد شما را از کنترل سیستم استخراج داده ها منع نمی کند . در مقابل ، سیستم های استخراج

داده ها را می توان با سیستم های CRM که با قراردادن عملی بودن و استخراج داده ها توسط ASP در اینترنت و همینطور با محلی نمودن سیستم استخراج داده ها در اینترنت ایجاد شد با سیستم های CRM ترکیب کرد و از طریق اگزترانت منبع ، یا سیستم CRM ASP ادغام نمود.

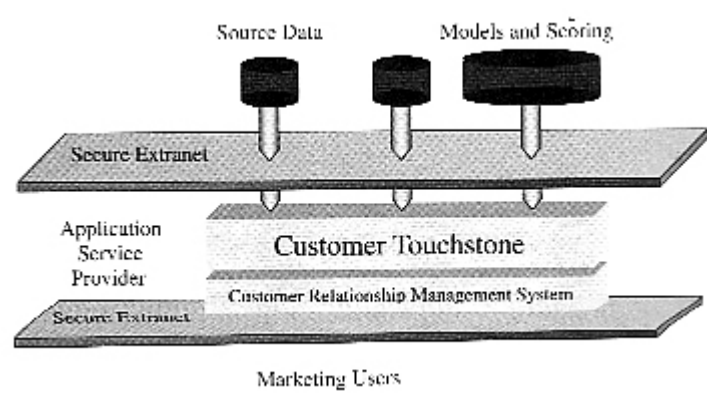
این روش دوم امکان حداکثر آزادی در توسعه مدل و انتخاب ابزار استخراج داده ها را فراهم می آورد در حالیکه به پمپ های اطلاعاتی موجود و ارتباط دهندگان اطلاعاتی که ذخایر داده های محلی شما را به سیستم CRM خارج از منبع متصل می کنند ، اتکاء می نماید . این دو روش را می توان بطور گرافیکی در اشکال ۴ و ۵ مشاهده نمود .

در خاتمه در می یابیم که هزاره میلادی جدید نه تنها نگرانی های مربوط به Y2K را معرفی کرده است بلکه همچنین یک انقلاب اطلاعاتی را همراه با آن ایجاد نمود و مسیر زندگی و تجارت را تغییر داد . این عصر تجارت الکترونیکی ، با سرعت وب در حال حرکت است و اغلب شگفتی های زیادی را با خود به همراه می آورد که اکثر آنها خوشایند نیستند ، کسانی که فرصتها را از دست می دهند، جا خواهند ماند .

کسانی که برای بقا روز خود را اداره می کنند و به قابلیت های ارائه شده توسط استخراج داده ها و تکنولوژی های قوی مشابه متکی هستند ، روندهای ارزشمندی را در داده های خود کشف می کنند ، و روابطی را با مشتریان شرکاء و عرضه کنندگان می سازند و تقویت می کنند ، و خود را در یک موقعیت غوطه آور مزیت رقابتی حفظ شده قرار خواهند گرفت . در حقیقت آینده برای استخراج داده ها و CRM به دقت به پتانسیل فزاینده تجارت الکترونیکی مرتبط است و بنابراین بی نهایت درخشان به نظر می رسد . آن بطور مشخص به شرکتها امکان می دهد تا در سطوح کاملاً متفاوت از آن زمان تا به امروز رقابت کنند و برنده باشند.



شکل ۴



شکل ۵