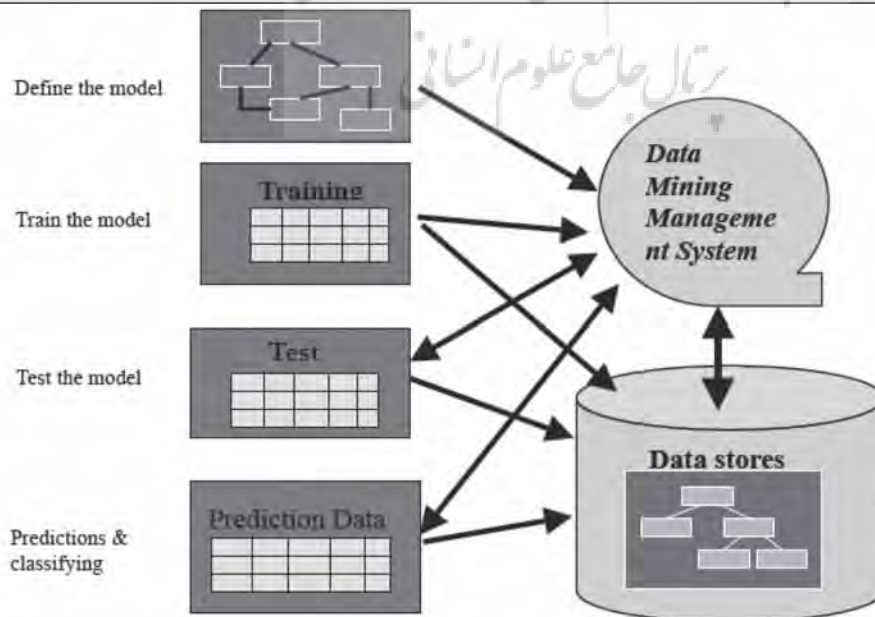


داده کاوی و کاربرد آن در بانکداری

زهرا جلائیان، اداره حراست بانک

مقدمه:

عموماً داده کاوی را که گاهی کشف دانش یا کشف داده، می نامند، فرایند تحلیل داده از منظرهای متفاوت و خلاصه کردن آن به اطلاعاتی است که می تواند برای افزایش درآمد، کاهش هزینه یا هر دو، مورد استفاده قرار گیرد. داده کاوی یکی از ابزارهای تحلیلی برای تحلیل داده هاست. داده کاوی به کاربران این امکان را می دهد تا داده ها را از ابعاد و زوایای متفاوتی تحلیل، طبقه بندی و ارتباطات مشخص شده را خلاصه کند. از لحاظ فنی، داده کاوی عبارت است از فرایند پیدا کردن ارتباطات یا الگوها در میان تعداد زیادی فیلد داده هایی که با یکدیگر در ارتباط هستند. داده کاوی الگوریتمی را به کار می گیرد که آمیخته ای از آمار، منطق فازی، الگوریتم های ژنتیک، ریاضی و هوش مصنوعی است. تعداد بسیار زیادی الگوریتم وجود دارد که به دنبال ارتباطات بین پایگاه های داده می گردد که می توانند به نوعی استنتاج شود یا به صورت جایگزینی برای طراحی پیش بینی، طبقه بندی یا سایر وظایف مورد استفاده قرار گیرد، اما انتخاب بهترین الگوریتم یک فرایند حسی و بصری نیست. الگوریتم شامل تعدادی از گروه روش هایی است که چهار گروه از آنها که بیشتر مورد استفاده قرار می گیرند شامل شبکه های عصبی، درخت تصمیم گیری، منطق فازی و بیسن Bages noive می باشد.



شکل شماره ۱- اجزاء داده کاوی

چالش کل فرایند از این واقعیت منشا می گیرد که:
 ۱- ارتباطات برقرار شده بدون استفاده از روش های داده کاوی به راحتی تشخیص داده نخواهند شد.
 ۲- تراکنش های الکترونیکی در طول چند سال گذشته به سرعت افزایش پیدا کرده است.
 بانکداری الکترونیک امروزه یک مجموعه کامل از کالاها و خدمات تسهیل کننده نه تنها برای مشتریان خرد بلکه برای مشتریان سازمانی فراهم نموده است. بنابراین توانایی کشف قوانین بین خدمات الکترونیکی متفاوت بسیار برای بانک حایز اهمیت است.
ساخت چنین قوانینی مزایای زیر را به دنبال دارد:

- توضیح و برقراری ارتباطات بین انواع متفاوت تراکنش های الکترونیکی
- سهولت استفاده از خدمات الکترونیکی برای عموم به دلیل استفاده گروه های مشتریان خاص از روش های پرداخت
- بهبود خدمات بانکی از آنچه که هم اکنون ارایه می شود به سمت آنچه که بیشتر مورد استفاده قرار می گیرد.

داده کاو آماری موسسات مالی را قادر به موارد زیر می کند:

- تشخیص الگوی تقلب و جعل.
- تعیین دلایل ریسک، ایجاد مدل های ریسک خودکار و پیچیده.
- تقسیم و پیش بینی رفتار گروه های همگن مشتریان.
- آشکارسازی ارتباطات پنهان بین شاخص های متفاوت.
- ایجاد مدل های برای تعیین قیمت سلف، سهام و ...

- بهینه کردن عملکرد سبد سهام

معرفی مدل های داده کاوی در محیط بانکداری به دلیل ماهیت و حساسیت داده های بانکی، می تواند کمک شایانی به بانک ها برای پیش بینی و پشتیبانی از تصمیم نمایند. در ارتباط با تصمیم گیری بسیار مهم است که یک بانک، دانش تشخیص قابلیت سوددهی مشتری و گروه بندی آنان بر اساس این عوامل را داشته باشد و همچنین قوانین وابستگی بین کالا و خدماتی که مشتریان سفارش می دهند، به منظور پشتیبانی از تصمیم به صورت کافی تر، تشخیص و مورد بررسی قرار دهند. ولی تعیین قوانین وابستگی داده های مربوط به بانک چالش برانگیز است چرا که: حجم داده های بانک بسیار زیاد است. بنابراین داده ها باید توسط داده کاو به طور مناسبی آماده شوند.

* **هدف**، باید از ابتدا و به طور واضح مشخص شود. در (بسیاری از موارد نداشتن یک هدف مشخص باعث

نتایج نادرست یا اصلاً عدم حصول نتیجه شده است).
 * **دانش** درست و مفید داده ها نه تنها برای داده کاو بلکه برای تحلیلگر نهایی (مدیر) نیز به عنوان پیش نیاز مطرح است. در غیر این صورت نتایج تولید شده توسط داده کاو اشتباه خواهد بود و نتایج ایجاد شده توسط مدیر نیز غیر قابل اعتماد خواهد شد.
 * **تمام قوانین** موجود حایز اهمیت نبوده و تحلیلگر باید قوانین قوی برای تصمیم گیری را تشخیص دهد.

نمونه ای از کاربرد داده کاوی در بانکداری: به منظور آشنایی بیشتر با الگوریتم های داده کاوی در این مقاله مدل RFM و الگوریتم آن را توضیح می دهیم.

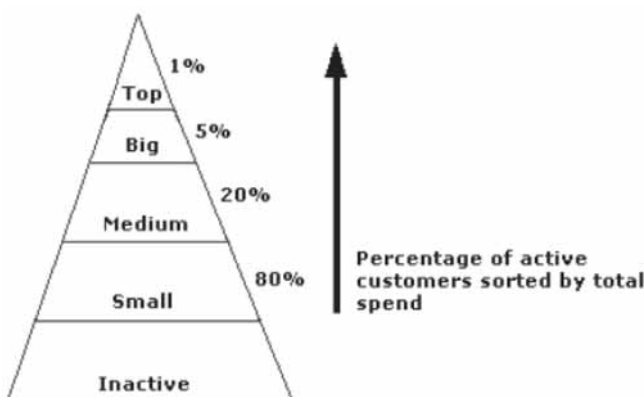
تحلیل RFM^(۱): توضیح در پاورقی

* تحلیل RFM یک روش سه بعدی طبقه بندی مشتریان برای تعیین مشتریان عالی (۲۰٪) است. این تحلیل بر اساس قانون پارتو (۸۰/۲۰) است که ۲۰ درصد مشتریان، ۸۰٪ درآمد را عاید سازمان می کنند. این مفهوم عامل بسیار مهمی در پشتیبانی از تصمیم است.

* در این مدل به منظور گروه بندی مشتریان و انجام تحلیل، یک مدل تقسیم بندی مشتری به نام مدل هرمی استفاده می شود. مدل هرمی مشتریان را بر اساس درآمدی که ایجاد می کنند به طبقه بندی کرده و سپس این گروه ها در تحلیل های متفاوتی مورد استفاده قرار می گیرند.

* مدل هرمی به طور وسیعی در شرکت ها، سازمان ها، و بانک ها مفید بوده است. به طور خلاصه برخی موضوعاتی که می تواند با استفاده از این مدل اصلاح شود شامل موارد زیر است:

- ✓ پشتیبانی از تصمیم و تصمیم گیری
- ✓ پیش بینی درآمد آینده
- ✓ قابلیت سودرسانی مشتری
- ✓ پیشگویی درباره تغییر موقعیت مشتری در هرم
- ✓ درک دلایل تغییرات
- ✓ نگهداری مشتریان بسیار مهم
- ✓ شبیه سازی مشتریان منفعل



شکل شماره ۲- هرم ارزش مشتری

◆ بانکداری

الکترونیک امروزه یک مجموعه کامل از کالاها و خدمات تسهیل کننده نه تنها برای مشتریان خرد بلکه برای مشتریان سازمانی فراهم نموده است

آخرین تراکنش خود را در روز ۲۰۰۲/۱۷/۲۹ انجام داده است، با $R_{new} = 332$ مشخص می‌شود و شخصی که در تاریخ ۲۰۰۲/۴/۴ آخرین تراکنش را انجام داده خواهد داشت $R_{new} = 93$

* **تکرار (F)**: تعداد تراکنش‌های مالی کاربر در دوره مذکور
 * پولی (M) جمع مقادیر تراکنش‌های مالی کاربر در دوره مذکور
 * نمره RFM با استفاده از فرمول زیر محاسبه می‌شود:

$$RFM=R+F+M$$

به منظور داشتن متغیرهای نرمال، جمع مقادیر را بر ۱۰۰۰ تقسیم می‌کنیم
 یک نمونه دیتا که روش‌های داده کاوی در آن استفاده شده است در جدول زیر می‌باشد همانطور که در هیستوگرام می‌بینید تجمع RFM در مقدار کمتر از ۱ می‌باشد. این امر طبیعی است چرا که ۸۰ درصد مشتریان نمره کمی در RFM می‌آورند.

همانطور که در هیستوگرام می‌بینید تجمع RFM در مقدار کمتر از ۱ می‌باشد. این امر طبیعی است چرا که ۸۰ درصد مشتریان نمره کمی در RFM می‌آورند.

User	R _{new}	F	M (*1,000 €)	RFM Factor
User522	330	20	€208,56	558,56
User523	216	6	€169,32	391,32
User524	304	8	€128,66	440,66
User525	92	1	€272,45	365,45

Table 1

* کاربرد الگوریتم K_mean در ۴ شاخه نتیجه می‌شود. بعد از هر شاخه می‌توان تعدادی زیرشاخه و همچنین ارزش میانگین هر متغیر را مشاهده کرد.

* نتیجه خوشه‌بندی بالا در شکل زیر آورده شده است. مدل صفحه بعد مشابه این توزیع در هرم است.
پاورقی:

۱- Recevcy Freqency Monetary

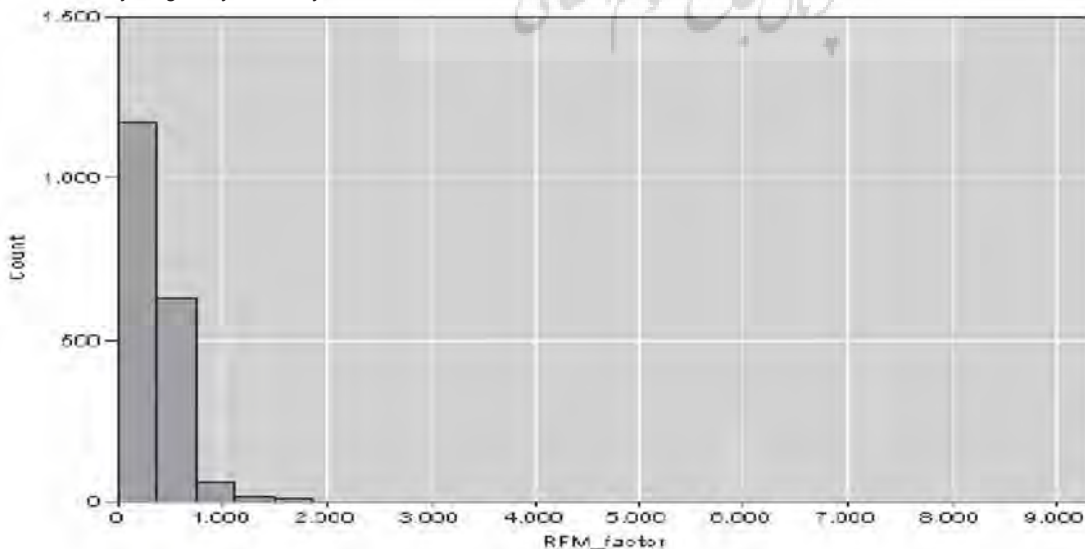


Figure 2



تحلیل RFM اشاره می‌کند که مشتری با نمره RFM بالا باید تراکنش‌ها و نتایج و سوددهی بالایی را برای بانک داشته باشد.

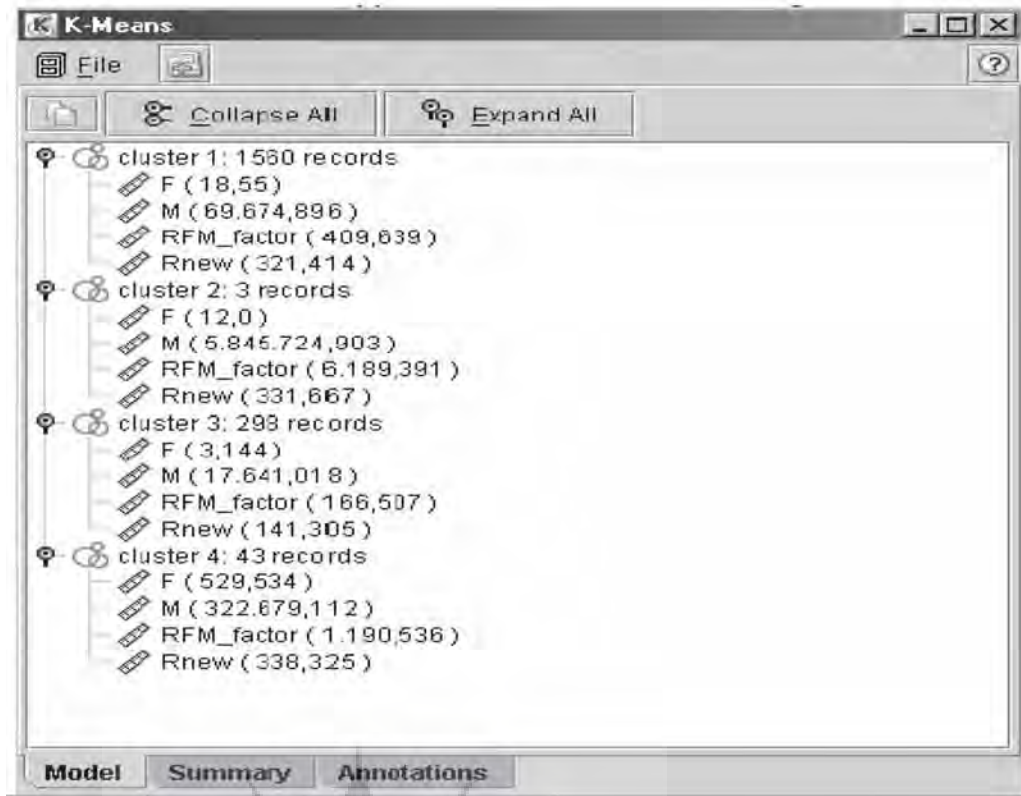
* این تحلیل امروزه مبتنی بر روش‌های داده کاوی مانند خوشه بندی است این روش‌ها به اثر بخشی تعیین و کشف نتایج تحلیل RFM کمک می‌کند.

الگوریتم تحلیل RFM:

نمونه داده‌های استفاده شده مربوط به دوره اول ژانویه تا دوازدهم دسامبر سال ۲۰۰۲ می‌باشد.

* واژه کاربر بانکداری الکترونیک فعال کاربری را توضیح می‌دهد که حداقل به یک تراکنش مالی در طول این دوره کمک کرده باشد. به منظور نشان RFM برای تعیین مشتری سود ده تمام مقادیر تراکنش‌های مالی در طول این دوره مورد ملاحظه قرار گرفته است.

مقادیر زیر برای این دوره زمانی مشخص محاسبه شد:
 * **تاخر (R)**: تاریخ آخرین تراکنش کاربر. از آنجایی که مقدار R بر نمره RFM تاثیر گذار است باید به صورت عدد باشد. بنابراین یک R جدید که به عنوان تعداد روزهای بین اولین تاریخ (۲۰۰۲/۱/۱) تا تاریخ آخرین تراکنش کاربر محاسبه می‌شود. بنابراین کاربری که



Cluster 1	(81,93%)	⇒	Small	80%
Cluster 3	(15,65%)	⇒	Medium	15%
Cluster 4	(2,26%)	⇒	Big	4%
Cluster 2	(0,16%)	⇒	Top	1%

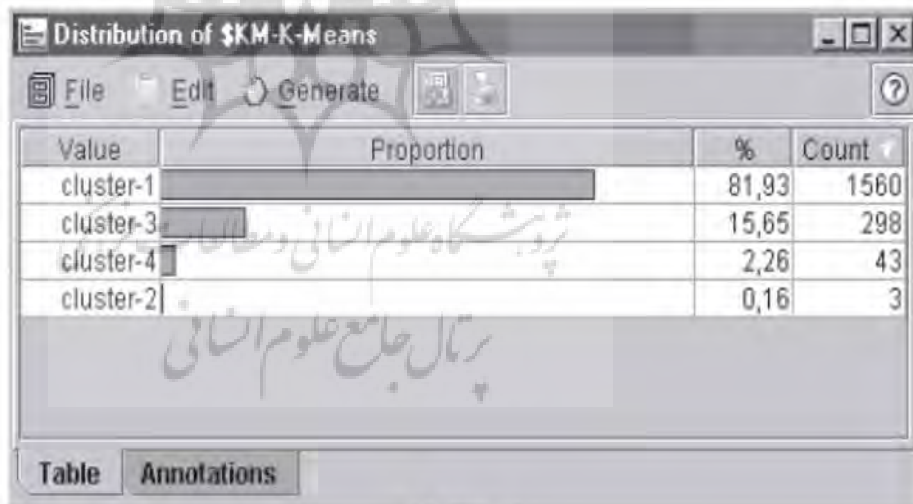


Figure 4

منابع و مراجع:

1-Geotechnical data mining models(GDM) using Microsoft sql server 2005 by N.Nawari P.H.D , PE

2-Intrusion Detection Using Data Mining By Anshu Veda , Prajakata Kalekar , Anirudha Bodhankar

3-Data Mining for Decision Support in e-banking area

Vasilis aggelis University of Patras Department of Computer Engineering and Informatics Rio, Patras, Greece

Vasilis.Aggelis@egnatiabank.gr Dimitris Christodoulakis University of Patras Department of Computer Engineering and Informatics Rio, Patras, Greece

dxri@cti.gr

4-Data Mining in Banking & Finance with Data Miner

5-Data Miner Tools and Techniques worldwide info@ststsoft.com

www.statsoft.com