



مطالب مورد بحث در کلاس

- تعاریف و مفاهیم (جلسه اول و دوم و سوم)
 - مفاهیم
 - متدولوژی اجرای پروژه های داده کاوی
- خوشه بندی (جلسه چهارم و پنجم)
- دسته بندی
 - درخت تصمیم (جلسه ششم)
 - شبکه های عصبی (جلسه هفتم)
 - نزدیکترین همسایگی (جلسه هشتم)
 - دسته بندی بیزی (جلسه نهم)
 - رگرسیون (جلسه دهم)
 - جمع بندی و اعتبار سنجی دسته بندی (جلسه یازدهم)

مطالب مورد بحث در کلاس

- قواعد تلازمی (جلسه دوازدهم و سیزدهم)
- انباره داده ها (جلسه چهاردهم)
- آماده سازی داده ها، پیش پردازش داده ها (جلسه ۱۵ و ۱۶)
- سریهای زمانی (۱۷ و ۱۸)
- متن کاوی (۱۹)
- وب کاوی (۲۰)
- پیوندکاوی و تحلیل شبکه های اجتماعی (۲۱)

سمیه علیزاده هیات علمی دانشکده صنایع
دانشگاه خواجه نصیر طوسی

مطالب مورد بحث در کلاس

- کاربردهای داده کاوی در بازاریابی و CRM (۲۲)

سمیه علیزاده هیات علمی دانشکده صنایع
دانشگاه خواجه نصیر طوسی



مفاهیم داده کاوی

❖ داده کاوی چیست؟ چرا داده کاوی نیاز است؟

❖ چه موقع و کجا داده کاوی نیاز است؟

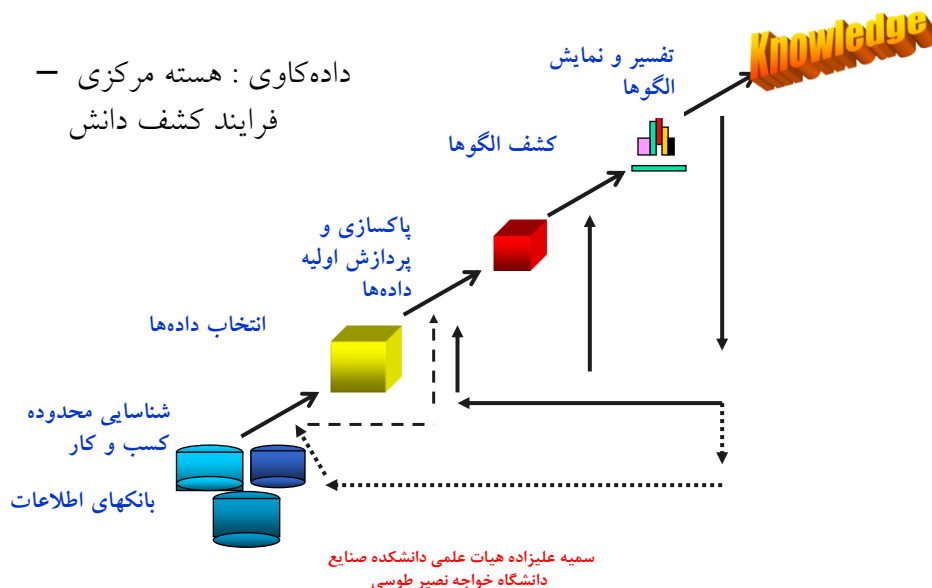
❖ چگونه داده کاوی را به کار گیریم؟ فرایند

داده کاوی

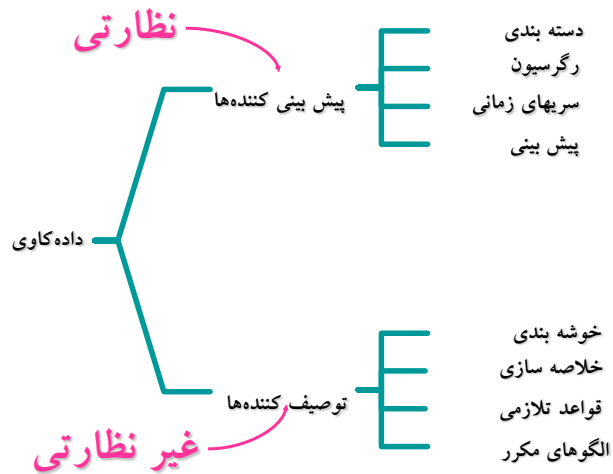
❖ روشهای داده کاوی

سمیه علیزاده هیات علمی دانشکده صنایع
دانشگاه خواجه نصیر طوسی

فرایند داده کاوی



تکنیکهای داده کاوی



سمیه علیزاده هیات علمی دانشکده صنایع
دانشگاه خواجه نصیر طوسی

روشها و ابزارهای داده کاوی



دسته بندی



خوشه بندی



رگرسیون



مدلهای وابستگی و قواعد انجمنی



خلاصه سازی

سمیه علیزاده هیات علمی دانشکده صنایع
دانشگاه خواجه نصیر طوسی

روشها و ابزارهای داده کاوی



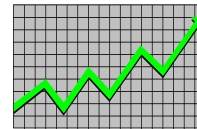
الگوریتمهای
ژنتیک



آمار



متن کاوی



سریهای زمانی

سمیه علیزاده هیات علمی دانشکده صنایع
دانشگاه خواجه نصیر طوسی

عملیات ، تکنیک ، کاربرد

دسته بندی پیش بینی خوشه بندی تخمین تعیین ارتباط بین متغیر ها	عملیات
دسته بندی مشتریان نگهداری مشتریان [1] شناسایی تخلفات تحلیل سایتهای اینترنتی تحلیل داده ها در بازاریابی تحلیل و پیشبینی روندهای بازار بورس	کاربردها
شبکه های عصبی درخت تصمیم گیری الگوریتمهای نزدیکترین همسایه تحلیل خوشه استنتاج قانون الگوریتم ژنتیک	تکنیکها

مثالهایی از کاربردهای داده کاوی در ایران

سمیه علیزاده هیات علمی دانشکده صنایع
دانشگاه خواجه نصیر طوسی

خوشه‌بندی مشتریان و انتخاب بازار هدف با استفاده از داده کاوی

**مشتری شناسی
هدف‌گیری
مشتری مداری**

همکاران: زهرا فردوسی، دکتر محمد مهدی سپهری

تعریف مسئله

✓ شرکت تعداد زیادی مشتری دارد که شناخت خوبی از آنها ندارد.

➤ کدام مشتری سودآوری بیشتری برای شرکت دارد؟

➤ کدام مشتری آمادگی دریافت خدمات جدید ما را دارد؟

➤ شباهت و تفاوت گروههای مشتریان در چیست؟

✓ می توان با استفاده از تکنیکهای داده کاوی و با توجه به سوابق مشتریان آنها را به چند بخش معنی دار تقسیم کرد.

✓ آیا می توان بر اساس خواسته ها و سلیقه مشتریان به آنها سرویس دهی کرد و یا پیشنهاد سرویس جدید داد.

بکارگیری تکنیکهای داده کاوی برای بخش بندی مشتریان

به منظور

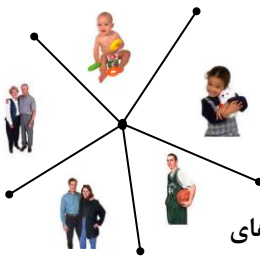
✓ شناسایی مشتریان و تهیه پروفایلی از آنها

✓ کشف رابطه معنی داری بین مشخصات مشتریان و نوع خدمات مورد استفاده آنها

✓ شناسایی سوده ترین مشتریان

سمیه علیزاده هیات علمی دانشکده صنایع
دانشگاه خواجه نصیر طوسی

بخش بندی بازار برای شرکت خدمات اینترنت



• بخش بندی بازار یک طرح بازاریابی است که برای جذب

و برآوردن نیازهای زیربخش های خاص بازار توصیه می شود.

• یک شرکت ممکن است استراتژی های بازاریابی خود را برای

بخش های متعددی از بازار تدوین کند.

• بخش بندی به معنی تقسیم مشتریان فعلی و محتمل به گروه های

کوچک تر

دلایل بخش بندی:

• محدودیت های مختلف: مالی، نیروی انسانی، محدوده فعالیت و ...

• هدف گیری بخش های مطلوب تر برای بعضی محصولات و خدمات

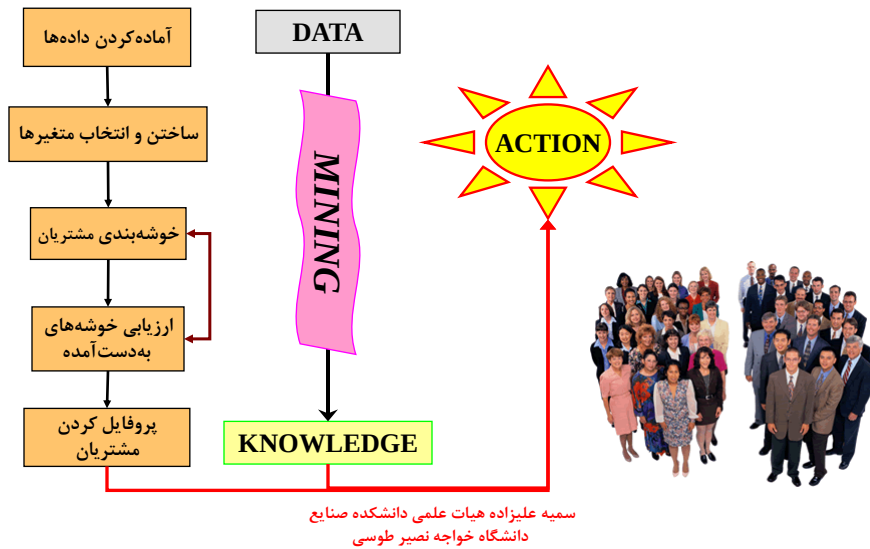
• بخش بندی و رتبه بندی مشتریان و تنظیم روابط بر اساس رتبه آنها

• متمایز سازی محصول، موقعیت یابی محصول و ...



سمیه علیزاده هیات علمی دانشکده صنایع
دانشگاه خواجه نصیر طوسی

متدلوژی انجام تحقیق



کام‌های خوشه‌بندی بازار، هدف‌گیری و موقعیت‌یابی



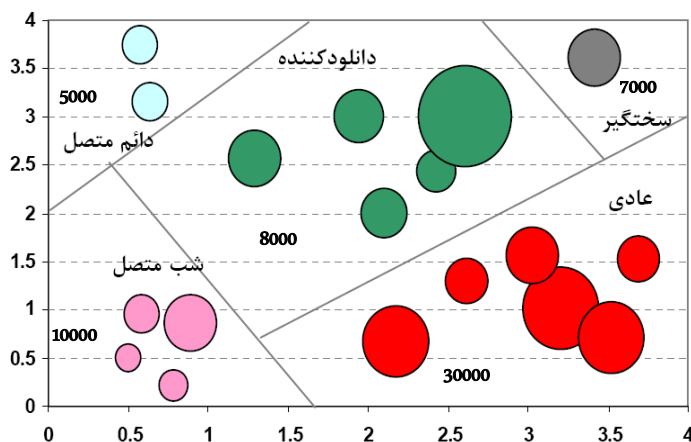
سمیه علیزاده هیات علمی دانشکده صنایع
دانشگاه خواجه نصیر طوسی

داده‌های تحقیق

- شرکت ارائه دهنده خدمات اینترنتی در تهران
- مجموعه داده‌های خام شامل:
 - سن، جنسیت، سطح تحصیلات، منطقه مسکونی، تراکنش‌ها
 - ۶۵ هزار مشتری
 - ۵۰۰ هزار کارت
 - تقریباً ۱۶ میلیون اتصال
 - نوع و قیمت کارت‌ها
 - مقدار دریافت و ارسال
- در مجموع حجم داده‌ها بیش از ۴ گیگابایت

سمیه علیزاده هیات علمی دانشکده صنایع
دانشگاه خواجه نصیر طوسی

نتایج خوشه‌بندی مشتریان با تکنیک SOM



سمیه علیزاده هیات علمی دانشکده صنایع
دانشگاه خواجه نصیر طوسی

شناسایی مشتریان خوش حساب و بد حساب مخابرات

مشتری شناسی و مشتری مداری

همکاران: حمید فروش، دکتر محمد مهدی سپهری

سمیه علیزاده هیات علمی دانشکده صنایع
دانشگاه خواجه نصیر طوسی

تعریف مسئله

✓ بخش قابل توجهی از درآمد شرکت‌های ارائه دهنده خدمات، از جمله شرکت مخابرات، به وسیله مشتریان بد حساب از بین می‌رود (۶۶ میلیارد مخابرات تهران، بیش از ۲۰۰ میلیارد تلفن همراه).

✓ غالباً دلایلی منطقی برای رفتار بد حسابی مشتریان وجود دارد.

✓ می‌توان با استفاده از تکنیک‌های داده‌کاوی و با توجه به سوابق مشتریان بد حسابی آنان را پیش‌بینی کرد.

بکارگیری تکنیک‌های داده‌کاوی برای دسته‌بندی مشتریان به دو دسته

مشتریان خوش حساب و مشتریان بد حساب

به منظور

✓ شناسایی مشتریان و تهیه پروفایلی از آنها

✓ پیش‌بینی رفتار بد حسابی و پیشگیری از زیان‌های ناشی از آن

✓ شناسایی مشتریان خوش حساب و ارائه خدمات بهتر و بیشتر

سمیه علیزاده هیات علمی دانشکده صنایع
دانشگاه خواجه نصیر طوسی

سؤال‌ها و اهداف تحقيق

سؤال‌های اصلی تحقيق

- آیا بین الگوهای رفتاری مشتریان بدحساب و بدحساب اختلاف وجود دارد؟
- آیا می‌توان مشتریان بدحساب را تا سرحد ممکن و با دقتی مقبول از بقیه مشتریان تشخیص داد؟
- آیا امکان ارائه مدلی برای تشخیص مشتریان بدحساب با استفاده از تکنیک‌های داده‌کاوی و تحلیل‌های آماری و نیز فرایند کشف دانائی وجود دارد؟

اهداف تحقيق

- تشخیص بیشترین تعداد مشتری بدحساب
- کمینه‌سازی هشدارهای اشتباه
- اعلان‌دهی در زمانی مناسب به ارائه‌دهنده خدمت
- ارائه چهارچوب سیستم پشتیبانی تصمیم‌گیری برای شناسایی مشتریان بدحساب.

سمیه علیزاده هیات علمی دانشکده صنایع
دانشگاه خواجه نصیر طوسی

داده‌کاوی در مخابرات

پایگاههای داده مخابرات از بزرگترین پایگاههای داده جهان می‌باشند (Gary M. Weiss, 2004)

انواع داده‌های مخابراتی:

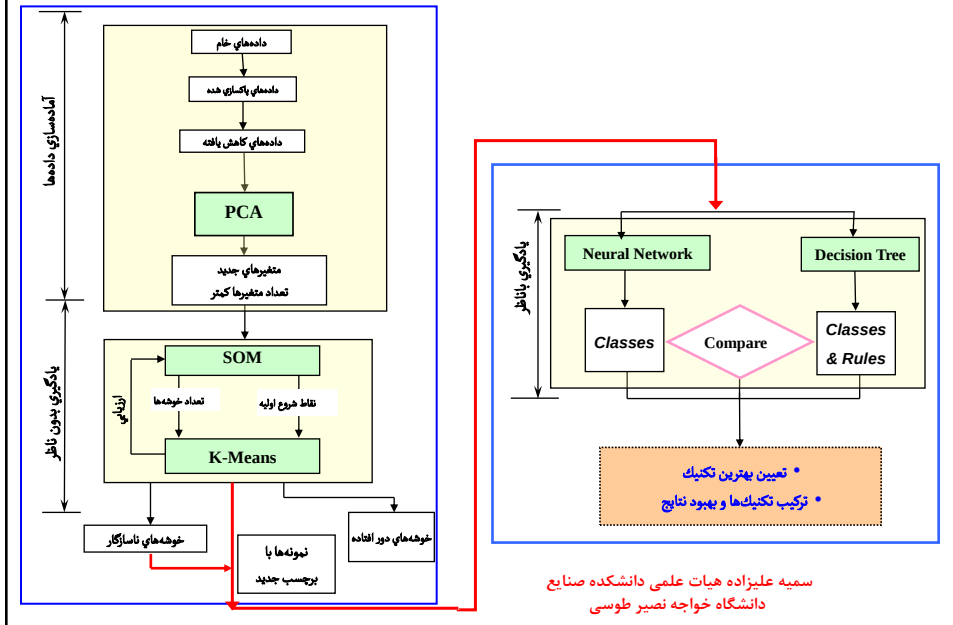
- داده‌های ریز مکالمات
- داده‌های شبکه
- داده‌های مشتری

کاربردهای داده‌کاوی در مخابرات:

- تشخیص فریبکاری
- تشخیص بدحسابی
- بازاریابی / پروفایل مشتری
- ایمن‌سازی شبکه در مقابل عیوب

سمیه علیزاده هیات علمی دانشکده صنایع
دانشگاه خواجه نصیر طوسی

شمای کلی مدل‌سازی تحقیق



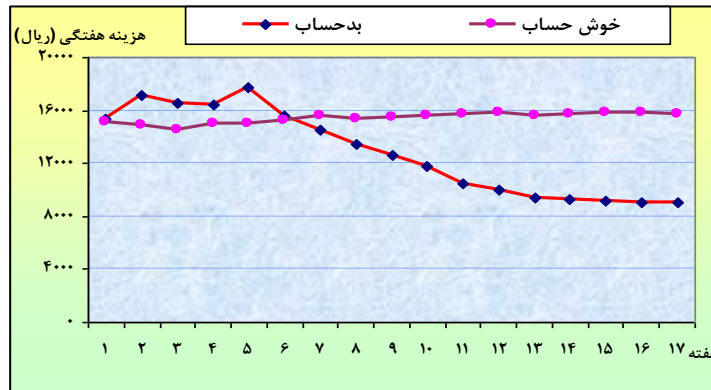
داده‌ها و آماده‌سازی آنها

مجموعه داده اصلی:

- شامل ۳۵۰۰۰ مشتری شرکت مخابرات تهران
- شامل ۱۶ ماه رفتارهای تماس مشتریان و ۸ دوره صورت‌حساب (از ۱ بهمن ۱۳۸۲ تا ۱ خرداد ۱۳۸۴)
- شامل ۸ نوع مختلف مشترک (مسکونی، تجاری و ...)
- شامل ۳ نوع جدول (مشخصات مشتریان، ریز مکالمات و صورت‌حساب)
- شامل مشتریان مناطق ۲، ۵ و ۷ شهر تهران
- اطلاعات به صورت فایل‌های متنی (فایل‌های Notepad)
- حجم اطلاعات در حدود ۲ گیگا بایت
- بیش از ۲۲ میلیون رکورد برای مجموع ۳۵۰۰۰ مشترک

سمیه علیزاده هیات علمی دانشکده صنایع
دانشگاه خواجه نصیر طوسی

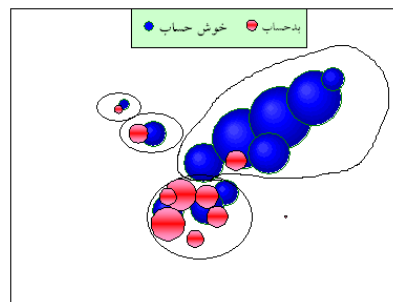
تفاوت رفتاری دو دسته مشتری (هزینه هفتگی)



سمیه علیزاده هیات علمی دانشکده صنایع
دانشگاه خواجه نصیر طوسی

نتایج - خوشه‌بندی داده‌ها

شماره خوشه	1	2	3	4
1	-	486/6	431/7	372/13
2	486/6	-	112/8	425/12
3	242/7	112/8	-	583/5
4	372/13	425/12	583/5	-
مجموع فواصل	1/27	023/27	126/21	38/31



در پایان خوشه‌بندی:

- ✓ خوشه شماره ۴ از مجموعه داده حذف می‌شود.
- ✓ با تعریف ۲ متغیر جدید بر اساس شماره خوشه و فاصله از مرکز خوشه، نتیجه خوشه‌بندی را حفظ می‌کنیم.
- ✓ برچسب نمونه‌ها را به آنها اضافه می‌کنیم.
- ✓ مجموعه داده جدید را که شامل ۴۰۲۳ نمونه و ۲۷ متغیر می‌باشد، بازسازی می‌کنیم.
- ✓ ترتیب و توالی نمونه‌ها را مجدداً بطور تصادفی می‌چینیم و چگالی یکنواخت خوش حساب و بد حساب را در کل توالی حفظ می‌کنیم.

سمیه علیزاده هیات علمی دانشکده صنایع
دانشگاه خواجه نصیر طوسی

خروجی نهایی مدل

وضعیت	مدل نهایی		مجموع
	بد حساب	خوش حساب	
بد حساب	تعداد	۶۳۸	۸۸۲
	درصد	٪۷۲.۳۴	٪۲۷.۶۶
خوش حساب	تعداد	۲۳	۳۱۳۱
	درصد	٪۰.۷۳	٪۹۹.۲۷
دقت مدل		٪۹۳.۳۵	۴۰۱۳

سمیه علیزاده هیات علمی دانشکده صنایع
دانشگاه خواجه نصیر طوسی

بررسی و مقایسه مالی مدل

دسته پیش‌بینی شده برای مشتریان بد حساب	درصد بدهی مشتری از کل بدهی
مشتریان بد حساب	٪۸۹.۵۷
مشتریان خوش حساب	٪۱۰.۴۳

سمیه علیزاده هیات علمی دانشکده صنایع
دانشگاه خواجه نصیر طوسی

با تشکر از حضورتان

سمیه علیزاده هیات علمی دانشکده صنایع
دانشگاه خواجه نصیر طوسی