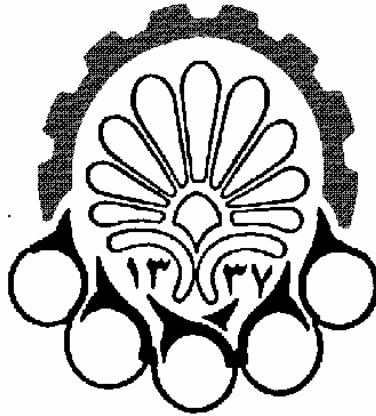


بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ



دانشگاه صنعتی امیرکبیر

دانشکده مهندسی صنایع

پایان نامه برای دریافت درجه کارشناسی ارشد مهندسی صنایع گرایش سیستم های  
اقتصادی اجتماعی

موضوع:

استخراج دانش و بکارگیری تکنیک های داده کاوی از اطلاعات هویتی  
و بدهی کارگاهها و کارفرمایان سازمان تامین اجتماعی

**Extract Knowledge & using Data Mining Techniques from  
Information of Social Security Organization**

تهیه کننده: ونوس شکورنیاز

استاد راهنما: دکتر جمال شهبابی

مهر ۸۶

تاریخ:



دانشگاه صنعتی امیرکبیر  
(پلی تکنیک تهران)

فرم اطلاعات پایان نامه  
کارشناسی - ارشد و دکترا  
کتابخانه مرکزی

شماره مدرک .....

مشخصات دانشجو		نام خانوادگی : شکورنیاز	نام : ونوس	شماره دانشجویی : ۸۴۱۲۵۰۵۲
عنوان		دانشکده: صنایع	رشته تحصیلی: سیستم های اقتصادی - اجتماعی	گروه:
Title :				
استاد راهنما		نام خانوادگی : شهرابی	درجه و رتبه	نام خانوادگی :
		نام: جمال	استادیار	نام:
استاد مشاور		نام خانوادگی : رمضانی خورشید دوست	درجه و رتبه	نام خانوادگی :
		نام: رضا	استادیار	نام:
دانشنامه		کارشناسی <input type="radio"/> ارشد <input checked="" type="radio"/> دکترا <input type="radio"/>	سال تحصیلی: ۱۳۸۶	
نوع پروژه		کاربردی <input checked="" type="radio"/> بنیادی <input type="radio"/> توسعه ای <input type="radio"/> نظری <input checked="" type="radio"/>		
مشخصات ظاهری		تعداد صفحات: ۲۲۰	تصویر: <input checked="" type="radio"/>	جدول: <input checked="" type="radio"/>
		نمودار: <input checked="" type="radio"/>	نقشه: <input type="radio"/>	واژه نامه: <input type="radio"/>
زبان متن		فارسی <input checked="" type="radio"/> انگلیسی <input type="radio"/>	چکیده	فارسی <input checked="" type="radio"/> انگلیسی <input type="radio"/>
یادداشت		لوح فشرده <input checked="" type="radio"/> دیسکت فلاپی <input type="radio"/>		
توصیفگر				
کلید واژه فارسی		داده کاوی ، داده کاوی سازمانی ، کشف دانش از پایگاه داده ، کلاس بندی ، درخت تصمیم گیری		
کلید واژه لاتین		Data Mining, ODM: Organizational Data Mining, Knowledge discovery in database, Classification, Decision Tree		

تمام موفقیت‌هایم را به پدر ، مادر  
و همسرم تقدیم می‌کنم

تقدیم به آرینا و ارشیای عزیزم ،  
آینده ای شگفت انگیز و روشن را  
برایتان آرزومندم

از جناب آقای دکتر جمال شهبازی که جهت انجام این پایان نامه با راهنمایی های ارزشمندشان ، مرا در زمینه های مختلف یاری نمودند تشکر و قدردانی می نمایم .

## چکیده

ODM ( Organization Data Mining ) به عنوان ابزار استخراج دانش اتکا پذیری ازداده ها تعریف شده است و فن آوری است که فرایند تصمیم گیری رابوسیله ی دگرگون ساختن داده ها به سوی دانش ارزشمند درجهت کسب یک مزیت رقابتی سوق می دهد و بعنوان شیوه بکاربردن ابزارهای داده کاوی تعریف شده است .

با توجه به اینکه سازمان ها ، داده های تجاری بسیاری رادر تصرف خوددارند بافلج ساختن اطلاعات یک چالش کلیدی درتصمیم گیری تشکیلات سازمانی ایجادمی نمایند. به این منظور پروسه ی کشف دانش از پایگاه داده سازمانی مطرح شده است که یک فرایند علمی برای شناسایی الگوهای معتبر، نوین، بالقوه مفید و قابل فهم از داده ها می باشد. مهم ترین بخش این فرایند ، داده کاوی (Data Mining) یا کاوش دانش می باشد که با استفاده از الگوریتم های مشخص، الگوهای را از پایگاه داده استخراج می نماید .

هدف از این پایان نامه معرفی تمام مفاهیم مرتبط با واژه داده کاوی ، پروژه و فرایند داده کاوی می باشد و رویکرد اصلی این پایان نامه بهره گرفتن از داده کاوی به عنوان یک ابزار کاربردی و تولیدی در دنیای واقعی (سازمان تامین اجتماعی) در نظر گرفته شده است ؛ با تمرکز بر تکنیک کلاس بندی با استفاده از درخت تصمیم گیری به اجرای واقعی ODM طی فرایند بهبود یافته و خاص در سازمان تامین اجتماعی پرداخته شده است . طرح منطقی ساخت بانک اطلاعاتی داده کاوی ارائه و آماده سازی داده ها و پیش پردازش ها را تا مرحله ای که بتوان الگوریتم ها را بر روی آن اعمال نمود ، انجام شده و ۵ مدل داده کاوی پیاده سازی و مورد تحلیل قرار گرفته است . به منظور ایجاد پروژه و راه حل داده کاوی از محیط *Business Intelligence Development Studio* و موتور بانک اطلاعاتی *Microsoft SQL Server2005* بکار گرفته شده است . به منظور مشاهده ی نتایج داده کاوی استخراج شده از فرایند و طرح پیشنهادی ، مدلها بصورت یک پروژه کاربردی قابل ارائه در *Report Server project* پیاده سازی شده اند .

## کلید واژه ها

داده کاوی ، داده کاوی سازمانی ، کشف دانش از پایگاه داده ، کلاس بندی ، درخت تصمیم گیری

فهرست مطالب

صفحه	عنوان
۱	فصل اول پیش گفتار (توضیحات مقدماتی و پیشینه تاریخی)
۲	۱ + داده کاوی و جایگاه آن
۵	۱ ۴ داده کاوی سازمانی (ODM)
۶	۱-۳ فرایند داده کاوی / کشف دانش از پایگاه داده (KDD)
۸	۱-۴ کلاس بندی / درخت تصمیم گیری
۱۱	۱-۵ مدل سازی و برنامه نویسی ساختارهای داده کاوی
۱۳	1-۶ تعریف مساله و ساختار پایان نامه
۱۴	۱-۷ دستاوردهای پایان نامه
۱۴	۱-۸ انتخاب ابزارها ، پایگاه داده ونرم افزارهای مورد استفاده
۱۸	فصل دوم معرفی داده کاوی ، معماری سیستمهای داده کاوی
۱۹	۲-۱ تعریف داده کاوی
۲۰	۲-۲ داده کاوی چه چیزی محسوب نمی شود؟
۲۰	۲-۳ ضرورت داده کاوی
۲۳	۲-۴ انواع داده کاوی
۲۳	۲-۴-۱ داده کاوی مستقیم یا هدایت شده
۲۳	۲-۴-۲ داده کاوی غیر مستقیم
۲۳	۲-۵ سیر تکاملی پایگاه داده ها و تجمیع داده ها
۲۶	۲-۶ معماری سیستم مبتنی بر داده کاوی
۲۷	۲-۷ طیف کسب و کارهای موجود
۲۸	۲ ۸ نقش انباره داده ها
۳۰	۲-۹ داده کاوی چه وظایفی را انجام می دهد؟
۳۱	۲-۹-۱ کلاس بندی و پیشگویی
۳۱	۲-۹-۲ تخمین
۳۲	۲-۹-۳ پیش بینی
۳۳	۲-۹-۴ قواعد انجمنی یا گروه بندی وابستگی ها

۳۴	۲-۹-۵ خوشه بندی
۳۵	۲-۹-۶ نمایه سازی توصیفی
۳۶	۲-۱۰ دسته بندی الگوریتم ها / زمینه های کاری در داده کاوی
۳۶	۲-۱۱ خلاصه
۳۸	فصل سوم داده کاوی سازمانی ODM
۳۸	۳-۱ مقدمه
۳۹	۳-۲ اصول اولیه تکنولوژی ODM
۳۹	۳-۳ ضرورت و جایگاه ODM در سازمان ها
۴۰	۳-۴ نقش ODM در معرفت سازمانی
۴۲	۳-۵ MTs و ارتباط با ODM
۴۲	۳-۶ نقش ODM در بهبود CRM
۴۵	۳-۷ داده کاوی سازمانی در کشورمان و ضرورت آن و چالش ها و نیازمندیها
۴۶	۳-۸ خلاصه و نتیجه گیری
۴۸	فصل چهارم فرایند استخراج دانش از پایگاه داده
۴۸	۴-۱ مقدمه
۴۸	۴-۲ تعریف فرایند استخراج دانش
۵۱	۴ ۴ ۴ تعریف مشکل و مساله و ترجمه مساله ی تجاری به مساله ی داده کاوی
۵۲	۴ ۴ ۴ ساخت بانک اطلاعاتی داده کاوی / آماده سازی / پیش پردازش داده ها
۶۰	۴ ۴ ۴ بررسی و شناخت ماهیت داده ها
۶۲	۴-۲-۴ ساخت مدل
۶۳	۴ ۴ ۵ بررسی و انتخاب مدل ها ارزیابی مدل و بررسی نتایج
۶۵	۴-۲-۶ استقرار و به روز رسانی مدل ها پیاده سازی مدلها
۶۵	۴-۳ خلاصه و مروری بر بررسی های انجام شده و تعیین مسیر بعدی
۶۸	فصل پنجم مفاهیم مدل سازی و برنامه نویسی داده کاوی
۶۸	۵-۱ مقدمه
۶۹	۵-۲ مدل داده کاوی
۶۹	۵-۳ OLE DB
۷۰	۵-۴ مزایای استفاده از SQL2005 Data Mining



۷۱	Case ۵-۵
۷۱	Case key ۶-۵
۷۱	Nested Key ۷-۵
۷۱	Nested Table و Case Table ۸-۵
۷۲	Shape Operator ۹-۵
۷۴	Data Mining ۱۰-۵ انواع ستونها در
۷۴	DMX ۱۱-۵
۷۴	۱۲-۵ ساخت مدل داده کاوی
۷۴	۱-۱۲-۵ انواع مدل ها
۷۴	۲-۱۲-۵ مجموعه داده ها در مدل
۷۴	۳-۱۲-۵ چه نوع داده‌هایی میتوانند مورد کاوش قرار گیرند؟
۷۵	۱۴ ۵ سه مرحله پایه برای ساخت مدل
۷۵	۱-۱۳-۵ ساخت مدل ، یا ایجاد ساختار داده کاوی
۷۶	۴ ۱۴ ۵ دو روش برای ساخت مدل های کاوش در زبان پرس و جوی DMX
۷۶	۳-۱۳-۵ ساخت مدل کاوش (CREATE MINING STRUCTURE)
۷۷	۴-۱۳-۵ اصلاح مدل کاوش (Alter Structure)
۷۷	۵-۱۳-۵ دلایل مختلفی وجود دارد که ممکن است نیاز باشد تا بیش از یک مدل کاوش را در یک ساختار کاوش اضافه نمود .
۷۸	۶ ۱۴ ۵ آموزش مدل داده کاوی
۷۸	۷-۱۳-۵ پیش بینی مدل
۷۹	۸-۱۳-۵ درک پیش بینی
۷۹	۱۴-۵ خلاصه و رویکرد فصل بعد
۸۲	فصل ششم کلاس بندی با استفاده از درخت تصمیم گیری
۸۲	۱-۶ مقدمه
۸۲	۲-۶ علت انتخاب روش کلاس بندی با استفاده از درخت تصمیم گیری در این پایان نامه
۸۳	۳-۶ علت انتخاب درخت ماکروسافت به عنوان ابزار کار
۸۴	۴-۶ تعریف کلاس بندی
۸۶	۵-۶ دسته کننده
۸۶	۱-۵-۶ ارزیابی شرط دسته کننده بر اساس
۸۶	۶ ۶ مراحل کلاس بندی

۸۶	۱-۶-۶ مرحله یادگیری
۸۷	۲-۶-۶ مرحله پیشگویی
۸۷	۷-۶ کاربرد ها و مثال هایی از کلاس بندی
۸۸	۸-۶ انواع روش های کلاس بندی
۸۹	۹-۶ روش درخت تصمیم گیری
۹۰	۱۰-۶ روش های اندازه گیری خلوص برای ارزیابی تقسیمات مناسب در متغیرهای توابع هدف
۹۰	۱۱-۶ الگوریتم ID3
۹۲	۱۲-۶ تولید درخت تصمیم گیری
۹۳	۱۳-۶ الگوریتم استنتاج درخت
۹۴	۱۴-۶ الگوریتم C4.5
۹۹	۱۵-۶ مفاهیم پایه در رشد درخت ماکروسافت
۱۰۱	۱۶-۶ استخراج قاعده ها
۱۰۲	۱۷-۶ خلاصه و رویکرد فصل بعد
۱۰۵	فصل هفتم معرفی سازمان تامین اجتماعی
۱۰۵	۱-۷ مقدمه
۱۰۵	۲-۷ معرفی خدمات سازمان
۱۰۶	۳-۷ برخی از واژه های اصلی مرتبط
۱۰۸	۴-۷ چگونگی ارتباط بین سازمان تامین اجتماعی و کارفرمایان
۱۰۹	۵-۷ حقوق کارفرمایان
۱۰۹	۶-۷ تسهیلات قانونی برای کارفرمایان
۱۱۰	۷-۷ نام نویسی کارگاه و تشکیل بانک اطلاعاتی هویتی کارگاهها و کارفرمایان در شعب
۱۱۱	۷-۷ مشکلات و ریسک های مرتبط با نام نویسی کارگاه
۱۱۳	فصل هشتم ODM در سازمان تامین اجتماعی
۱۱۳	۱-۸ مقدمه
۱۱۴	۲-۸ طرح جامع تحلیل و طراحی شده به منظور اجرای فن آوری داده کاوی
۱۱۴	۳-۸ فرایند داده کاوی به کار گرفته شده
۱۱۴	۱-۳-۸ تعریف مشکل و مساله، ترجمه مساله ی تجاری به مساله ی داده کاوی

۱۱۶	۲-۳-۸ ساخت بانک اطلاعاتی داده کاوی ، آماده سازی / پیش پردازش داده ها
۱۳۱	۳-۳-۸ بررسی داده ها
۱۳۴	۴-۳-۸ ساخت مدل ها
۱۷۰	۵-۳-۸ بررسی و انتخاب مدل ها، ارزیابی مدل و بررسی نتایج
۱۷۳	۶-۳-۸ اجرای الگوریتم کاوش و به روز رسانی مدل ها ، پیاده سازی مدلها
۱۷۴	۴-۸ نتیجه گیری و بحث
۱۷۴	۱-۴-۸ دستاوردهای عمده
۱۷۴	۲-۴-۸ نتایج
۱۷۹	۵-۸ مروری بر مراحل و کارهای انجام شده در پایان نامه
۱۸۰	۶-۸ پیشنهادات تحقیق آینده
۱۸۲	فهرست منابع
۱۸۵	پیوست ۱
۲۰۳	پیوست ۲
۲۰۸	چکیده لاتین

فهرست شکل ها

صفحه	عنوان
۷	شکل 1-1 فرایند چرخه ای داده کاوی
۸	شکل ۲-۱ فرایند KDD
۱۳	شکل ۳-۱ سه مرحله از داده کاوی
۲۵	شکل ۱-۲ سیر تکاملی پایگاه داده ها و تجمیع داده ها
۲۷	شکل ۲-۲ معماری سیستم های مبتنی بر داده کاوی
۲۹	شکل ۳-۲ بانک مرکزی داده کاوی استخراج شده از انباره داده
۳۰	شکل ۴-۲ بانک مرکزی داده کاوی استخراج شده از پایگاه های اطلاعاتی عملیاتی
۴۰	شکل ۱-۳ دلایل عمده دشوار شدن فهم و شناخت مشتری
۴۳	شکل ۲-۳ ارتباط چرخه ی زندگی مشتری با داده کاوی
۴۹	شکل ۱-۴ نمونه ای از فرایند چرخه ای داده کاوی ۲
۵۱	شکل ۲-۴ تعریف مساله
۶۶	شکل ۳-۴ فرایند داده کاوی سفارشی ارائه شده در پایان نامه
۷۰	شکل ۱-۵ معماری OLEDB برای داده کاوی
۷۲	شکل ۲-۵ ارائه مفاهیم مورد استفاده در مدل سازی داده کاوی
۸۶	شکل ۱-۶ مرحله یادگیری با استفاده از مجموعه ی آموزشی
۸۷	شکل ۲-۶ استفاده مدل دسته کننده در پیش بینی تعیین کلاس داد ها نامشخص و فاقد کلاس
۸۹	شکل ۳-۶ نمونه ای از یک درخت تصمیم گیری ساده با دو مشخصه X,Y
۹۲	شکل ۴-۶ کلاس بندی نمونه های جدید بر اساس درخت تصمیم گیری
۹۸	شکل ۵-۶ مرحله اول شکست مجموعه داده ها
۹۹	شکل ۶-۶ درخت تصمیم گیری نهایی
۱۰۱	۷-۶ الگوریتم درخت تصمیم گیری یک مشخصه را برای شکستن سطح ریشه برمی دارد .
۱۰۲	شکل ۸-۶ استخراج قواعد از درخت تصمیم گیری
۱۱۷	شکل ۱-۸ نحوه ارسال اطلاعات در سطح استان
۱۱۸	شکل ۲-۸ نحوه ارسال اطلاعات به ادارت کل
۱۲۰	شکل ۳-۸ نحوه Convert داده ها

۱۲۱	شکل ۸-۴ ورود داده ها به SQL Server2005
۱۲۷	شکل ۸-۵ طرح منطقی داده ها
۱۲۸	شکل ۸-۶ نمایی از به روز آوری روابط جداول
۱۲۹	شکل ۸-۷ Data source view ساخته شده برای استفاده در ساخت مدل های داده کاوی
۱۳۰	شکل ۸-۸ مراحل اجرای فن آوری داده کاوی
۱۳۱	شکل ۸-۹ انتخاب متغیرهای پیشنهادی
۱۳۲	شکل ۸-۱۰ نمایی از ستون های مجازی تعریف شده
۱۳۳	شکل ۸-۱۱ بررسی ستون مجازی وضعیت بدهی
۱۳۵	شکل ۸-۱۲ نمایی از ساختار داده کاوی به همراه نوع ستون ها
۱۳۶	شکل ۸-۱۳ نمایی از ساختار داده کاوی تعریف شده به همراه نوع ستون ها و نوع الگوریتم
۱۳۶	شکل ۸-۱۴ نمایی از پردازش مدل
۱۳۷	شکل ۸-۱۵ مدل گرافیکی درخت تصمیم گیری شماره ۱
۱۳۸	شکل ۸-۱۶ نمای گرافیکی شبکه وابستگی مربوط به درخت تصمیم گیری ساخته شده
۱۳۸	شکل ۸-۱۷ تنظیم درجه شبکه وابستگی در پایین ترین سطح به منظور انتخاب قدرتمندترین مشخصه پیش بینی کننده
۱۳۹	شکل ۸-۱۸ تنظیم درجه وابستگی
۱۳۹	شکل ۸-۱۹ شبکه وابستگی تنظیم شده در بالاترین سطح و نمایش قدرت وابستگی ها
۱۴۰	شکل ۸-۲۰ گره ریشه درخت شامل همه Case ها
۱۴۱	شکل ۸-۲۱ نمایی از اولین قاعده استخراج شده از گره ریشه تا برگ درخت
۱۴۱	شکل ۸-۲۲ نمایی از مشخصات کارگاههایی که متعلق به کلاس قاعده ۱ می باشند
۱۴۲	شکل ۸-۲۳ نمایی از دومین قاعده استخراج شده از گره ریشه تا برگ درخت
۱۴۲	شکل ۸-۲۴ نمایی از مشخصات کارگاههایی که متعلق به کلاس قاعده ۲ می باشند
۱۴۳	شکل ۸-۲۵ نمایی از سومین قاعده استخراج شده از گره ریشه تا برگ درخت
۱۴۳	شکل ۸-۲۶ نمایی از مشخصات کارگاههایی که متعلق به کلاس قاعده ۳ می باشند
۱۴۴	شکل ۸-۲۷ نمایی از مشخصات کارگاههایی که متعلق به کلاس قاعده ۴ می باشند
۱۴۵	شکل ۸-۲۸ نمایی از مشخصات کارگاه هایی که متعلق به کلاس قاعده ۵ می باشند
۱۴۵	شکل ۸-۲۹ نمایی از مهمترین گره ها که مهمترین قواعد مدل ۱ از آنها استخراج

	می شوند
۱۴۶	شکل ۸-۳۰ نمایی از مشخصات کارگاه هایی که متعلق به کلاس مهمترین قاعده مدل ۱ می باشند
۱۴۷	شکل ۸-۳۱ نمایی از مشخصات کارگاه هایی که متعلق به کلاس مهمترین قاعده مدل ۱ می باشند
۱۴۸	شکل ۸-۳۲ نمایی از ساختار مدل داده کاوی به همراه الگوریتم
۱۴۸	شکل ۸-۳۳ مدل درخت تصمیم گیری تولید شده شماره ۲
۱۴۹	شکل ۸-۳۴ تنظیم وضعیت بدهی به مقدار ۵۰-۱۰۰ میلیون تومان و بررسی مدل
۱۴۹	شکل ۸-۳۵ مشخصات کارگاه های متعلق به گره مشخص شده
۱۵۰	شکل ۸-۳۶ تنظیم وضعیت بدهی به مقدار ۵۰۰ میلیون به بالا
۱۵۰	شکل ۸-۳۷ مشخصات کارگاه های متعلق به کلاس بدهی ۵۰۰ میلیون به بالا
۱۵۱	شکل ۸-۳۸ مشخصات کارگاه های متعلق به این کلاس
۱۵۱	شکل ۸-۳۹ نمایی از ساختار مدل داده کاوی به همراه الگوریتم در درخت تصمیم گیری شماره ۳
۱۵۲	شکل ۸-۴۰ نمایی از مدل درخت تصمیم تولید شده بر اساس تمام case ها - مدل شماره ۳
۱۵۲	شکل ۸-۴۱ نمایی از مدل درخت بر اساس تنظیم وضعیت بدهی به ۵۰-۱۰۰ میلیون تومان
۱۵۳	شکل ۸-۴۲ مشخصات کارگاه های متعلق به کلاس بدهی ۵۰-۱۰۰ میلیون تومان
۱۵۳	شکل ۸-۴۳ نمایی از مدل درخت بر اساس تنظیم وضعیت بدهی به ۵۰-۱۰۰ میلیون تومان
۱۵۴	شکل ۸-۴۴ مشخصات کارگاه های متعلق به کلاس بدهی ۵۰-۱۰۰ میلیون تومان
۱۵۴	شکل ۸-۴۵ مشخصات کارگاه های متعلق به کلاس بدهی بالای ۵۰۰ میلیون بر اساس مدل درخت شماره ۳
۱۵۵	شکل ۸-۴۶ شبکه وابستگی مدل شماره ۳
۱۵۵	شکل ۸-۴۷ ساختار مدل داده کاوی شماره ۴
156	شکل ۸-۴۸ مدل درخت تصمیم گیری شماره ۴

157	شکل ۸-۴۹ مشخصات تمام کارگاه ها بر اساس مدل شماره ۴
157	شکل ۸-۵۰ شبکه وابستگی مدل شماره ۴
158	شکل ۸-۵۱ شبکه وابستگی مدل شماره ۴ ، مشخصه وضعیت کارگاه دارای بیشترین قدرت پیش بینی می باشد
158	شکل ۸-۵۲ مشخصات کارگاه های متعلق به مهمترین قاعده استخراج شده از مدل شماره ۴
159	شکل ۸-۵۳ مشخصات کارگاه های مهمترین قاعده بر اساس بدهی ۱-۵ میلیون تومان
160	شکل ۸-۵۴ نمایی از case مورد استفاده در ساخت مدل خوشه بندی
161	شکل ۸-۵۵ نمایی از ساختار داده کاوی مدل شماره ۵: خوشه بندی
161	شکل ۸-۵۶ مدل خوشه بندی کارگاه های دارای بدهی بالا
۱۶۲	شکل ۸-۵۷ نمایش خوشه هایی که دارای بیشترین شباهت در مدل ۵ هستند
۱۶۲	شکل ۸-۵۸ تنظیم مشخصه وضعیت بدهی به ۵۰۰ میلیون به بالا
۱۶۳	شکل ۸-۵۹ پروفایل کلاسترها بر اساس وضعیت و مقادیر مشخصه ها
۱۶۳	شکل ۸-۶۰ وضعیت پردازش مدل خوشه بندی تمام رکوردهای استان تهران بر اساس مقادیر گسسته بدهی
۱۶۴	شکل ۸-۶۱ مشخصات خوشه بدهی بسیار بالا (بالای ۵۰۰ میلیون تومان) در مدل خوشه بندی
۱۶۴	شکل ۸-۶۲ تنظیم مشخصه وضعیت بدهی به ۱۰۰ - ۵۰۰ میلیون تومان در مدل خوشه بندی
۱۶۵	شکل ۸-۶۳ پروفایل تنظیم مشخصه وضعیت بدهی به ۵۰۰ - ۱۰۰ میلیون تومان
۱۶۵	شکل ۸-۶۴ تفاوت دو خوشه ۶۲ و ۶۳ تنها بر اساس میزان بدهی آنها می باشد
۱۶۶	شکل ۸-۶۵ تنظیم مشخصه بدهی به مقدار ۵۰-۱۰۰ میلیون تومان در مدل خوشه بندی
۱۶۶	شکل ۸-۶۶ پروفایل مربوط به دو خوشه ۷ و ۹
۱۶۷	شکل ۸-۶۷ پروفایل مربوط به خوشه ۷
۱۶۷	شکل ۸-۶۸ پروفایل مربوط به خوشه ۹

۱۶۷	شکل ۸-۶۹ نمایی از تفاوت بین دو خوشه ۹ و ۷
۱۶۸	شکل ۸-۷۰ پروفایل مربوط به خوشه های ۳ و ۵ و ۸
۱۶۸	شکل ۸-۷۱ مشخصات خوشه ها بر اساس تنظیم وضعیت بدهی به ۵۰ - ۱۰۰ میلیون تومان
۱۶۹	شکل ۸-۷۲ مشخصات خوشه شماره ۳
۱۷۰	شکل ۸-۷۳ نگاشت ستون های داده های تستی به ساختار داده کاوی مدل شماره ۱
۱۷۱	شکل ۸-۷۴ نمودار نمودار صعود حاصل از تنظیم مقدار بدهی به ۱-۵ میلیون تومان از مدل شماره ۱
۱۷۲	شکل ۸-۷۵ نمودار نمودار صعود حاصل از تنظیم مقدار بدهی به ۱۰-۵۰ میلیون تومان از مدل شماره ۱
۱۷۳	شکل ۸-۷۶ گزارش پیش بینی حاصل از مدل شماره ۱



# فصل اول

## مقدمه

## فصل اول

### پیش گفتار

#### توضیحات مقدماتی و پیشینه تاریخی

در این پایان نامه 5 محورا اصلی در نظر گرفته شده است که در مورد عنوان هر حوزه شرح مختصری به همراه پیشینه تاریخی و تحقیقات محققین آن در دامنه ی منابع مورد استفاده در ذیل آورده شده است .

#### 1-1 داده کاوی و جایگاه آن

داده کاوی از اعضای کلیدی تجارت هوشمند می باشد که به صورت مکرر به عنوان پروسه استخراج اطلاعات معتبر، نوین و قابل اطمینان از پایگاه داده های حجیم توصیف شده است [1]. به عبارت دیگر داده کاوی روش های استنتاج الگوهایی است که در داده وجود دارد. این الگوها می توانند با یکدیگر یکپارچه شده و به عنوان مدل کاوش<sup>1</sup> تعریف شوند. مدل های کاوش در دوطیف وسیع از علوم از جمله سناریوهای تجاری، مدیریت ارتباط با مشتریان، کنترل فرآیندهای صنعتی، کنترل کیفیت ، مدیریت و کشف جرم و تقلبات ، حقوق ، بانکداری ، علوم اجتماعی، متن کاوی، وب کاوی، پزشکی و ورزش می توانند به کار گرفته شوند [2].

در طول دهه گذشته، حجم زیادی از داده هادر پایگاه داده ها انباشته و ذخیره شده اند. بسیاری از این داده ها از نرم افزارهای تجاری مانند برنامه های حسابداری، مدیریت منابع نیروی انسانی، مدیریت ارتباطات مشتری مداری و وب لاگ ها فراهم شده اند و نتیجه ی این انباشتگی در سازمان ها این است که در داده غنی ولی در کسب دانش بسیار ضعیف گشته اند [1].

SQL Server 2005 موتوربانک اطلاعاتی قدرتمندی است که در بیشتر شرکت ها و سازمان ها انتخابی شایسته و قابل اطمینان برای ذخیره ی داده ها می باشد و سهم عمده ای را در منابع کتابخانه ای این پایان نامه به خود اختصاص داده است. در این راستا تنگ و مکلمن<sup>2</sup> از پیشگامان و طراحان اصلی فن آوری داده کاوی و تحلیل داده ها می باشند [1]. تنگ از مدیران ارشد تیم برنامه نویسی داده کاوی

<sup>1</sup> Mining Model

<sup>2</sup> ZhaoHui Tang, Jamie Maclennan

شرکت ماکروسافت می باشد. تحقیقات عمده ی وی در زمینه ی طراحی ویژگیهای داده کاوی در SQL Server 2005 و SQL 2000 می باشد. مکلنن از توسعه دهندگان موتور داده کاوی در SQL Server 2005 می باشد و طبق گفته خود تحقیقاتش را در زمینه فن آوری داده کاوی از سال ۱۹۹۹ آغاز نموده است [2]. نتایج تحقیقات آنها در کتاب *Data Mining In SQL 2005* آمده است.

گروه مشاوره ای گارتن<sup>۱</sup> داده کاوی را "پروژه استخراج همبستگی های نوین معنادار، الگوها و روش هایی بوسیله واری و کاوش در میان حجم عظیمی از داده های ذخیره شده در انبار داده ها" می داند [3]. این تعریف برگرفته از کتاب کشف دانش از داده، از آثار دانیل لاروس می باشد.<sup>۲</sup> وی در کتاب خود اهم موارد به شرح ذیل برای خوانندگان مطرح نموده است:

- مدلها و تکنیک هایی برای کشف تکه های پنهان در اطلاعات از جمله درخت تصمیم گیری<sup>۳</sup>، شبکه های عصبی<sup>۴</sup>، خوشه بندی سلسله مراتبی<sup>۵</sup>، قاعده های وابستگی<sup>۶</sup> و تکنیک های ارزیابی مدل

- بینشی در خصوص اینکه الگوریتم های داده کاوی چگونه کار می کند  
- تجربه مورد نیاز از اجرای واقعی داده کاوی روی مجموعه داده های حجیم با توصیف کاملی از فرآیند KDD

- تحلیل اکتشافی داده و رویکرد های ریاضی به منظور تخمین و پیش بینی [3].  
داده کاوی عبارتی است که رنج گسترده ای از تکنیک هایی را پوشش می دهد که در صنایع متنوع مورد استفاده قرار می گیرد. افرادی که به تجارت و کسب و کار می پردازند، نیازمند بازاریابی هدفمند و فراگیری دانش راجع به مشتریان خود می باشند. در طول زمان آنها باید بیشتر و بیشتر مشتریان خود را بشناسند تا بتوانند با استفاده از این دانش خدمات بهتری را ارائه نمایند و در نتیجه به یک تجارت هوشمند رسیده و سود بیشتری را کسب نمایند [4] و [5].

در این راستا دیوید و ناتالی [5] تحقیقاتی را خصوص کاوش در مدل های سیستم های پشتیبان تصمیم گیری انجام داده و در مقاله خود طبقه بندی از وظایف تحلیلی مدل های استنتاجی را ارائه نمودند. آنها فن آوری های مختلف کاوش دانش مبتنی بر هوش مصنوعی که برای کاوش دانش از

---

<sup>1</sup> Gartner Group

<sup>2</sup> Daniel T. Larose

<sup>3</sup> Decision Tree

<sup>4</sup> Neural Network

<sup>5</sup> Hierarchical Clustering

<sup>6</sup> Association Rules

مدل های تحلیل پذیر بالقوه مناسبند و نیز تکنیک های قابل کاربرد کاوش دانش را معرفی نموده و مشخص کردند که کدام یک از تکنیک ها برای وظایف تحلیلی خاص، مرتبط می باشند. با توجه به تنوع زیاد مشتریان، بازارها و محیط های کسب و کار، دسترسی به اطلاعات مناسب برای تصمیم گیری صحیح، ضروری می باشد از اینرو استفاده از راهکار های مناسب برای طبقه بندی و تولید اطلاعات از میان انبوهی از داده ها، به ویژه در حوزه بازاریابی الزام آوراست. علم داده کاوی پاسخی به این نیاز به منظور کشف دانش پنهان داده ها و تامین اطلاعات مورد نیاز مدیران می باشد. بهبود سرویس مشتری مداری و کاهش هزینه ها، مثالی از اهمیت تجمیع داده در سازمان ها و شرکت ها به منظور تحلیل می باشد. با اتوماتیک و خودکار نمودن این فعالیت ها و انجام تحلیل داده ها، کوشش ها موثر شده و تمرکز بر نیروی انسانی کمتر می گردد [4].

پاراگراف بالا مقدمه ای بر ضرورت و اهمیت داده کاوی می باشد که برگرفته از کتاب " *Data Mining Techniques For Marketing, Sales, and Customer Relationship Management* " می باشد و نویسندگان آن بری و لینف می باشند. بری<sup>۱</sup> و لینف<sup>۲</sup> [4] در عرصه داده کاوی شناخته شده هستند، آنها همکاری خود را از سال ۱۹۸۰ در سمینار داده کاوی آغاز نموده و بایکدیگر کتاب های بسیاری را در خصوص داده کاوی به زبانهای مختلف نوشته و ترجمه نموده اند و به عنوان مشاوران تخصصی داده کاوی در شرکت های بزرگ در آمریکای جنوبی، اروپا و آسیا در زمینه CRM، ثبت وقایع ورودی های وب، بهبود بازاریابی در روابط مشتری مداری فعالیت داشتند. آنها در حدود یک دهه است که تجربه خودشان را در خصوص کاربرد تکنیک های داده کاوی در تجارت و بازاریابی و CRM به کار برده اند و کتاب تکنیک های داده کاوی در بازاریابی و فروش و CRM از آثار آنهاست.

همچنین در خصوص ضرورت داده کاوی هن<sup>۳</sup> و کمبر<sup>۴</sup> در کتاب خود بیان داشتند که دلیل عمده ای که داده کاوی توجه در صنعت اطلاعات را در سال های اخیر جلب نموده است، این است که فن آوری اطلاعات با حجم بالای داده در سطح گسترده ای سر و کار دارد و نیاز حتمی تبدیل داده به اطلاعات مفید و دانش وجود دارد. اطلاعات و دانش به دست آمده در محدوده مدیریت تجارت، کنترل تولید و تحلیل بازار و طراحی مهندسی و علوم اکتشافی کاربرد دارد و ارتباط تنگاتنگی با

---

<sup>1</sup> Michael J.A. Berry

<sup>2</sup> Gordon S.Linoff

<sup>3</sup> Jiawei Han

<sup>4</sup> Hicheline Kamber