

به نام خداوند حکیم

دانشگاه فردوسی مشهد

پایان نامه کارشناسی ارشد حسابداری

موضوع

**پیش بینی گزارش حسابرس مستقل : رویکرد داده کاوی**

استاد راهنما

دکتر محمد علی باقرپور ولاشانی

اساتید مشاور

دکتر محمد جواد ساعی

دکتر علی مشکانی

نگارش

مصطفی باقری

زمستان ۱۳۹۰

**Ferdowsi University**

**M.A Thesis in Accounting**

**Title**

**Prediction of Independent Auditor Opinion : Data Mining Approach**

**Supervisor**

**Dr. Mohammad Ali Bagherpour velashani**

**Advisors**

**Dr. Mohammad Javad Saei**

**Dr. Ali Meshkani**

**By**

**Mostafa Bagheri**

**Winter 2011**

## تقدیر و سپاس

خداوند بزرگ را سپاسگزارم که بر این بنده توفیق ارزانی داشت تا انجام این تحقیق را به

پایان برسانم. به رسم ادب و احترام، بر خود واجب می دانم که نسبت به زحمات پدرو

مادم، کلیه اساتید و دوستانم در دوره کارشناسی ارشد شکر قلبی خویش را ابراز دارم.

## خدای رابی ساکرم

که پرومادی فداکار نصییم ساخته تا از ریشه آنها شاخ و برگ کیرم و در سایه وجود پر مهرشان

برای کسب علم و دانش تلاش نمایم...

آموزگارانی که برایم ایستادن و انسان بودن را معنا کردند...

حال این برگه تحفه ای است تقدیم به آنان...

از استاد راهنمای گرامی جناب آقای دکتر محمد علی باقر پور و لسانی، که بارها سہامی ها و نظرات  
ارزنده و حمایتہای ہمیشگی شان، نقش چہمگیری در بہ ثمر رسیدن این تحقیق داشتند، صمیمانہ  
تشکر می نمایم. ہمیشہ و شاکردی این استاد بزرگوار از افتخارات من می باشد و پاسکزار  
محبت های بیدریغ ایشان، مسم.

بہمچنین از اساتید مشاور بزرگوار، جناب آقای دکتر محمد جواد ساعی و دکتر علی مشکانی و اساتید  
داور گرامی، جناب آقای دکتر محمد حسین ودیعی و دکتر مہدی مرادی بہ خاطر مساعدت ہا،  
راہنمایی ها و ہمکفری های ارزشمندشان پاسکزارم.

## فهرست

۱	فصل اول : کلیات تحقیق
۱-۱	۱-۱. مساله اصلی تحقیق
۲	۲-۱. تشریح و بیان موضوع
۳	۳-۱. اهمیت و ضرورت انجام تحقیق
۴	۴-۱. استفاده کنندگان از نتایج تحقیق
۵	۵-۱. فرضیه تحقیق
۵	۶-۱. قلمرو تحقیق (مکانی و زمانی)
۵	۷-۱. روش انجام تحقیق
۵	۱-۷-۱. شرح روش تحقیق
۵	۲-۷-۱. شرح روش و ابزار گردآوری داده ها
۶	۳-۷-۱. جامعه آماری، روش نمونه گیری و حجم نمونه
۶	۴-۷-۱. روش و فرآیند تجزیه و تحلیل داده ها
۹	فصل دوم : مبانی نظری تحقیق
۱۰	مقدمه
۱۰	۱-۲. گزارش حسابرسان مستقل
۱۰	۱-۲-۱. تاریخچه حسابرسی
۱۰	۲-۱-۲. تعریف حسابرسی
۱۱	۳-۱-۲. انواع اظهارنظر حسابرسی
۱۱	۱-۳-۱-۲. اظهارنظر مقبول
۱۲	۲-۳-۱-۲. اظهارنظر مشروط
۱۲	۳-۳-۱-۲. اظهارنظر مردود
۱۲	۴-۳-۱-۲. عدم اظهارنظر
۱۲	۴-۱-۲. اجزای اصلی گزارش حسابرسی
۱۴	۵-۱-۲. شرایطی که باعث صدور گزارش حسابرسی غیرمقبول می شود
۱۴	۶-۱-۲. عوامل موثر بر نوع گزارش حسابرسی

۱۵.....	۱-۱-۲-۶. اهرم مالی
۱۶.....	۲-۱-۲-۶. سودآوری
۱۶.....	۳-۱-۲-۶. ساختار سرمایه
۱۶.....	۴-۱-۲-۶. عملکرد
۱۷.....	۵-۱-۲-۶. نقدینگی
۱۸.....	۶-۱-۲-۶. ریسک ورشکستگی
۱۸.....	۷-۱-۲-۶. حاکمیت شرکتی
۱۹.....	۸-۱-۲-۶. اقلام تهedy اختیاری
۲۰.....	۹-۱-۲-۶. اندازه شرکت مورد رسیدگی
۲۱.....	۱۰-۱-۲-۶. سایر عوامل مرتبط با گزارش حسابرس
۲۱.....	۲-۲. تکنیک های مورد استفاده در تحقیقات قبلی
۲۲.....	۱-۲-۲. درخت تصمیم
۲۲.....	۲-۲-۲. شبکه های عصبی مصنوعی
۲۳.....	۳-۲-۲. رگرسیون لجستیک
۲۴.....	<b>فصل سوم: روش تحقیق</b>
۲۵.....	مقدمه
۲۵.....	۱-۳. مروری اجمالی بر مفاهیم داده کاوی و فرآیند کشف دانش
۲۵.....	۱-۱-۳. تاریخچه داده کاوی
۲۵.....	۲-۱-۳. تعریف داده کاوی
۲۶.....	۳-۱-۳. دلایل استفاده از داده کاوی
۲۶.....	۴-۱-۳. پیش نیاز یک داده کاو موفق
۲۷.....	۵-۱-۳. قابلیت های اساسی داده کاوی
۲۷.....	۱-۵-۱-۳. طبقه بندی
۲۸.....	۲-۵-۱-۳. تشریح مفاهیم
۲۸.....	۳-۵-۱-۳. پیش بینی و برآورد
۲۸.....	۴-۵-۱-۳. تحلیل خوشه ای
۲۹.....	۵-۵-۱-۳. تعیین نقاط انحراف



۲۹	۳-۱-۵-۶. تحلیل وابستگی
۲۹	۳-۱-۶. مراحل فرآیند داده کاوی بر استاندارد CRISP – DM
۳۰	۳-۱-۶-۱. شناخت کسب و کار
۳۱	۳-۱-۶-۲. شناخت داده ها
۳۳	۳-۱-۶-۳. آماده سازی داده ها
۳۶	۳-۱-۶-۴. مدلسازی
۳۸	۳-۱-۶-۵. ارزیابی مدل
۳۸	۳-۱-۶-۶. بکارگیری مدل
۳۹	۳-۱-۷. دسته بندی الگوریتم های داده کاوی مورد استفاده در این تحقیق
۳۹	۳-۱-۷-۱. الگوریتم درخت تصمیم
۴۱	۳-۱-۷-۲. الگوریتم شبکه های عصبی مصنوعی
۴۳	۳-۱-۷-۳. الگوریتم رگرسیون لجستیک
۴۵	<b>فصل چهارم : تجزیه و تحلیل داده های تحقیق</b>
۴۶	مقدمه
۴۶	۴-۱. توصیف داده ها (آمار توصیفی)
۴۶	۴-۱-۱. نحوه توزیع متغیرهای مستقل و وابسته در جامعه آماری
۴۷	۴-۱-۱-۱. نحوه توزیع متغیرهای مستقل
۶۸	۴-۱-۱-۲. نحوه توزیع متغیر وابسته
۷۰	۴-۲. تحلیل داده ها
۷۰	۴-۲-۱. انتخاب متغیرهای تحقیق
۷۴	۴-۲-۲. نتایج حاصل از فرآیند داده کاوی
۷۴	۴-۲-۲-۱. مدل حاصل از تکنیک درخت تصمیم C5.0
۸۱	۴-۲-۲-۲. مدل حاصل از تکنیک شبکه های عصبی مصنوعی
۸۵	۴-۲-۲-۳. مدل حاصل از تکنیک رگرسیون لجستیک
۸۹	۴-۳. بررسی نتایج داده کاوی
۸۹	۴-۳-۱. مقایسه دقت حاصل از بکارگیری تکنیک های داده کاوی
۹۰	۴-۳-۲. مقایسه نتایج داده کاوی با صحت پیش فرض جامعه آماری

۹۰	۴-۴ . تجزیه و تحلیل ریسکهای نوع اول و دوم .....
۹۰	۴-۴-۱ . ریسکهای نوع اول و دوم تکنیک درخت تصمیم C5.0 .....
۹۱	۴-۴-۲ . ریسکهای نوع اول و دوم تکنیک شبکه های عصبی مصنوعی .....
۹۱	۴-۴-۳ . ریسکهای نوع اول و دوم تکنیک رگرسیون لجستیک .....
۹۲	۴-۴-۴ . مقایسه ریسکهای نوع اول و دوم .....
۹۴	<b>فصل پنجم : نتیجه گیری و پیشنهادها</b> .....
۹۵	مقدمه .....
۹۶	۵-۱ . نتایج و یافته های تحقیق .....
۹۶	۵-۱-۱ . دقت مدل های ساخته شده و توانایی آنها جهت کاهش ریسک حسابرسی .....
۹۶	۵-۱-۲ . متغیرهای با اهمیت .....
۹۹	۵-۲ . استفاده کنندگان از نتایج تحقیق .....
۱۰۰	۵-۳ . محدودیتهای تحقیق .....
۱۰۰	۵-۴ . پیشنهادات جهت تحقیقات آتی .....
۱۰۲	<b>منابع</b> .....
۱۰۷	<b>پیوستها</b> .....
۱۰۷	پیوست شماره ۱ : توزیع متغیرهای تحقیق .....
۱۲۳	پیوست شماره ۲ : میانگین متغیرهای تحقیق .....
۱۳۰	پیوست شماره ۳ : ماتریس همبستگی متغیرهای مستقل .....
۱۳۲	پیوست شماره ۴ : نتایج آزمون VIF .....
۱۳۳	پیوست شماره ۵ : نتایج آزمون تحلیل واریانس یکطرفه (ANOVA) .....
۱۳۵	پیوست شماره ۶ : خروجی نرم افزار کلمنتاین .....
۱۴۸	پیوست شماره ۷ : لیست شرکت - سالهای مورد بررسی .....

## چکیده :

هدف این پژوهش پیش بینی گزارش حسابرسان مستقل با استفاده از تکنیک های داده کاوی می باشد. گزارش حسابرسان مستقل در این تحقیق به دو دسته مقبول و غیرمقبول (شامل گزارشات مشروط، مردود و عدم اظهارنظر) طبقه بندی شده است. به منظور پیش بینی گزارش حسابرسان مستقل از سه تکنیک طبقه بندی داده کاوی شامل، درخت تصمیم C5.0، شبکه های عصبی مصنوعی و رگرسیون لجستیک استفاده شده است. جامعه آماری پژوهش شامل تمامی شرکت های پذیرفته شده در بورس اوراق بهادار تهران طی سال های ۱۳۸۲ الی ۱۳۸۸ می باشد. در این تحقیق از ۲۹ متغیر مالی و غیرمالی در قالب ۱۰ طبقه نقدینگی، عملکرد، اهرم مالی، ساختار سرمایه، سودآوری، ریسک ورشکستگی، مدیریت سود، حاکمیت شرکتی، اندازه شرکت و سایر متغیرها (شامل نوع صنعت و عمر شرکت) به منظور آموزش و آزمون مدل استفاده شده است. نتایج تحقیق نشان می دهند که میانگین دقت مدل حاصل از تکنیک درخت تصمیم C5.0 جهت پیش بینی نوع گزارش حسابرسی (میانگین دقت ۸۸.۶۴٪) از دو تکنیک دیگر بیشتر بوده و این تکنیک توانایی بالاتری جهت کاهش ریسکهای حسابرسی دارد. مهمترین متغیرها جهت پیش بینی نوع گزارش حسابرسی به ترتیب شامل نوع گزارش حسابرسی سال قبل (مقبول یا غیر مقبول)، نسبت سودآوری (سود خالص به درآمد)، نسبت بدهیها (بدهیها به داراییها)، نوع صنعت، نوع عملکرد (سوده یا زیانده)، نوع حسابرس (دولتی یا خصوصی) و ریسک ورشکستگی (ورشکسته یا سالم) می باشند.

# فصل اول

## کلیات تحقیق

## ۱-۱. مساله اصلی تحقیق :

در دنیای اقتصادی امروز با افزایش روز افزون تقاضا برای ارائه اطلاعات قابل اتکاء، نیاز به افرادی با کفایت برای اعتباردهی به گزارش ها و اطلاعات ارائه شده امری حیاتی به نظر می رسد. جهت نیل به هدف مذکور، استفاده از خدمات حسابرسان مستقل ضروری خواهد بود (ستایش و جمالیان پور، ۱۳۸۸). حسابرسی جزء لاینفکی از فرآیند گزارشگری مالی است که با اعتبار بخشی به اطلاعات ارائه شده، نقش مهمی در قضاوت و تصمیم گیری آگاهانه استفاده کنندگان دارد. تحقیقات صورت گرفته نشان می دهند که محتوای گزارش های حسابرسی بر قیمت سهام شرکت های بورس اوراق بهادار تهران تاثیر گذار بوده (ازگلی، ۱۳۷۶) و بازارهای سرمایه نسبت به انتشار گزارش حسابرسی مشروط و یا رفع بندهای شرط گزارش حسابرسی واکنش نشان می دهند (علیخانی دهقی، ۱۳۸۵). نتایج تحقیقات صورت گرفته در خارج از ایران هم نشان داده که بندهای شرط گزارش حسابرسی بر روی قیمت سهام و میزان پاداش دریافتی توسط مدیران اثر منفی (DeAngelo, ۱۹۸۱) و شروط حسابرسی مربوط به تداوم فعالیت و ارزشیابی دارایی ها، تاثیر معناداری بر تصمیم گیری اعتبار دهندگان برای اعطای اعتبار دارد (Firth, ۱۹۸۰). بنابراین، نوع و محتوای گزارش های حسابرسی دارای بار اطلاعاتی بوده و پیش بینی نوع گزارش حسابرسی سال بعد می تواند در تصمیم گیری استفاده کنندگان مفید واقع شود.

از سویی با افزایش مبادلات تجاری، اقتصادی و پیشرفت تکنولوژی اطلاعات، داده های مالی به سرعت انباشت شده و این موضوع به نوبه خود یکسری محدودیت هایی را برای استفاده بهینه و کارا از این داده ها بوجود آورده است. از اینرو به منظور استفاده بهینه و موثر از داده های مالی در تصمیم گیری ها، از تکنیک های داده کاوی استفاده می شود. داده کاوی<sup>۱</sup> فرآیندی است که با نگرشی نو به مسئله استخراج اطلاعات از داده های حجیم می پردازد (مشکانی و ناظمی، ۱۳۸۸) و به کمک مجموعه ای از روش های آماری و مدل سازی، می تواند الگوها و روابط پنهان موجود در داده ها را در کمترین زمان ممکن و با دقتی بالا تشخیص دهد. هدف داده کاوی جستجو برای یافتن اطلاعاتی با ارزش تجاری در یک پایگاه داده است و از آن می توان برای پیش بینی رویه ها و رفتارهای آتی در بازارهای مالی استفاده کرد (مهديخانی و کیان راد، ۱۳۸۵). هدف ما این است که با استفاده از اطلاعات شرکت ها، حسابرسان مستقل و با بکارگیری تکنیک های "طبقه بندی داده کاوی" و پیروی از فرآیند کشف دانش<sup>۲</sup>، مدلی را برای "پیش بینی نوع گزارش حسابرس مستقل" ارائه نماییم. مساله اصلی تحقیق این است که آیا با استفاده از فرآیند داده کاوی (تکنیک های طبقه بندی) می توان مدلی با توانایی بالا جهت پیش بینی نوع گزارش حسابرس مستقل ارائه داد؟

<sup>۱</sup> Data Mining

<sup>۲</sup> Knowledge Discovery

## ۲-۱. تشریح و بیان موضوع :

تحقیق حاضر با توسعه و مقایسه مدل های حاصل از تکنیک های طبقه بندی داده کاوی، مدلی را برای پیش بینی گزارش سال بعد حسابرسان مستقل ارائه خواهد داد. در این تحقیق نتایج و مدل های حاصل از بکارگیری سه تکنیک : شبکه های عصبی مصنوعی<sup>۳</sup>، درخت تصمیم گیری C5.0<sup>۴</sup> و رگرسیون لجستیک<sup>۵</sup> با هم مقایسه شده و مدلی با بالاترین توانایی (قدرت) جهت پیش بینی گزارش سال بعد حسابرس انتخاب می شود. مطابق با تحقیق های قبلی (Laitinen & Laitinen، ۱۹۹۸؛ Ronning، ۲۰۰۰؛ Spathis و دیگران، ۲۰۰۳؛ Ireland، ۲۰۰۳؛ Pasiouras و دیگران، ۲۰۰۷؛ Gaganis و دیگران، ۲۰۰۷؛ Kirkos و دیگران، ۲۰۰۷؛ Martens و دیگران، ۲۰۰۸) و استاندارد حسابرسی ایران بخش ۷۰، گزارش های حسابرسی را به دو دسته مقبول<sup>۶</sup> و غیرمقبول<sup>۷</sup> تقسیم بندی می کنیم.

## ۳-۱. اهمیت و ضرورت انجام تحقیق :

شرکت های سهامی نقش عمده ای در اقتصاد هر کشوری بازی کرده و موفقیت این شرکت ها به میزان زیادی به تمایل سرمایه گذران و اعتبار دهندگان به سرمایه گذاری در این شرکت ها بستگی دارد. گزارش های حسابرسان اهمیت و جایگاه خاصی در تصمیم گیری سرمایه گذاران و اعتبار دهندگان دارند. بنابراین اهمیت ارائه اطلاعات با کیفیت، با نگاه به اتفاقات رخ داده اخیر برای شرکت ها (ورشکستگی و انحلال، رسوایی های مالی، ضرر و زیان های متحمل شده توسط اعتبار دهندگان و ...) مشخص می شود و گزارش های حسابرسی با ایجاد سیگنال هایی می تواند از وقوع چنین اتفاقاتی جلوگیری کند (ستایش و جمالیان پور، ۱۳۸۸).

از سویی با رشد فناوری اطلاعات و روش های تولید و جمع آوری داده ها، پایگاه مربوط به داده های تبادلات تجاری، کشاورزی و ... سریع تر از هر روز جمع آوری و انبار می شوند. نتایج یک تحقیق در این خصوص نشان می دهد که :

" ۶۱ درصد از مدیران اعتقاد دارند که سرریزی اطلاعات در محل کارشان وجود دارد، ۸۰ درصد عقیده دارند این وضع بدتر خواهد شد، بیش از ۵۰ درصد از مدیران از داده ها در مراحل تصمیم گیری فعلی به واسطه سرریزی اطلاعات چشم پوشی می کنند، ۸۴ درصد از مدیران این اطلاعات را برای آینده ذخیره می کنند و از آن ها برای تحلیل های جاری استفاده نمی کنند و ۶۰ درصد آنان عقیده دارند که هزینه جمع آوری اطلاعات سنگین تر از ارزش خود اطلاعات است" (علیخانزاده، ۱۳۸۵).

<sup>۳</sup> Artificial Neural Network

<sup>۴</sup> Decision Tree C.5

<sup>۵</sup> Logistic Regression

<sup>۶</sup> Unqualified Opinion Class

<sup>۷</sup> Qualified Opinion Class

به همین دلیل بشر به فکر دستیابی به اطلاعات نهفته در این پایگاه داده های حجیم افتاد، زیرا سیستم های سنتی قادر به این کار نبودند. بدلیل رقابت در عرصه های سیاسی، نظامی، اقتصادی و علمی و اهمیت دستیابی به اطلاعات در کمترین زمان بدون دخالت انسان، علم تجزیه و تحلیل داده ها یا داده کاوی پا به عرصه گذاشت (مشکانی و ناظمی، ۱۳۸۸). در حال حاضر داده کاوی مهمترین فناوری جهت بهره برداری موثر از داده های حجیم است و اهمیت آن رو به فزونی است. به طوریکه تخمین زده شده است که مقدار داده ها در جهان هر ۲۰ ماه به حدود دو برابر می رسد<sup>۸</sup>. هدف داده کاوی جستجو برای یافتن اطلاعات با ارزش تجاری در یک پایگاه داده است و این کار را با استفاده از یک سری تکنیک ها و مدل سازی ها انجام می دهد. از جمله این موارد می توان به تکنیک هایی همچون : طبقه بندی و پیشگویی (رگرسیون خطی چندگانه، رگرسیون لجستیک، تجزیه و تحلیل چندگانه، بیز ساده، شبکه های بیزی، شبکه های عصبی، درخت های تصمیم، K- نزدیک ترین همسایه و ...)، خوشه بندی، تحلیل روابط و وابستگی ها ، پیش بینی و ... اشاره نمود (مهدیخانی و کیان راد، ۱۳۸۵).

با توجه به اهمیت گزارش حسابرس در فرآیند تصمیم گیری استفاده کنندگان از صورت های مالی (به ویژه سرمایه گذاران و اعتبار دهندگان) در این تحقیق ما سعی در ارائه مدلی به منظور پیش بینی گزارش حسابرسی سال بعد با استفاده از تکنیک های طبقه بندی داده کاوی داریم، تا بتوانیم تصمیم های استفاده کنندگان از گزارش های حسابرسی را بهبود دهیم.

#### ۱-۴. استفاده کنندگان از نتایج تحقیق :

عمده گروه هایی که می توانند از نتایج این تحقیق بیشترین استفاده را ببرند شامل : سرمایه گذاران (به منظور استفاده در تصمیم های سرمایه گذاری)، اعتبار دهندگان (به منظور استفاده در تصمیم های مربوط به اعطای تسهیلات و اعتبارات)، حسابرسان مستقل (جهت کاهش ریسک حسابرسی<sup>۹</sup> و همچنین به عنوان ابزار کنترل کیفیت در فرآیند حسابرسی)، مدیران (شناسایی نقاط ضعف و برنامه ریزی جهت پیشگیری و رفع آن)، نهادهای نظارتی و قانون گذاری مثل سازمان امور مالیاتی و سازمان بورس اوراق بهادار و ... ، می باشند. همچنین شرکت های سرمایه گذاری و بیمه با الگو برداری از روش تحقیق و مدل های استفاده شده در این تحقیق می توانند تصمیم های بهتری را اتخاذ کنند.

#### ۱-۵. فرضیه تحقیق :

<sup>۸</sup> An Introduction to Data Mining : <http://www.theartling.com>

<sup>۹</sup> ریسک حسابرسی: احتمال اینکه حسابرس از رسیدگی هایش نتیجه گیری غیر معتبر کند. ریسک حسابرسی به دو گروه تقسیم می شود: اول ریسک آلفا : یعنی آنکه حسابرس ممکن است در مواردی که صورت های مالی تصویری مطلوب منعکس می کنند، گزارش غیر مقبول ارائه دهد؛ دوم ریسک بتا : یعنی آنکه حسابرس ممکن است در مواردی که صورت های مالی تصویری مطلوب منعکس نمی کنند، گزارش مقبول ارائه دهد. ریسک آلفا بر کارایی و ریسک بتا بر اثربخشی حسابرسی تاثیر می گذارد (آقایی، سید نژاد فهیم، ۱۳۸۵)

به علت این که تحقیق پیش رو به دنبال ارائه مدلی برای پیش بینی گزارش حسابرس مستقل است و یکی از حالت های تحقیق (بررسی رابطه، بررسی تاثیر، بررسی وضع موجود) را دارا نمی باشد، لذا فرضیه آماری برای این تحقیق تدوین نمی شود.

### ۱-۶. قلمرو تحقیق (مکانی و زمانی) :

قلمرو مکانی و زمانی این تحقیق به ترتیب شامل، کلیه شرکت های پذیرفته شده در بورس اوراق بهادار تهران طی دوره زمانی ۱۳۸۲ الی ۱۳۸۸ می باشد. در این بین، شرکت های طبقه بندی شده در صنعت واسطه گری مالی مثل شرکت های سرمایه گذاری و بانک ها بدلیل نوع فعالیتشان مستثنی خواهند شد.

### ۱-۷. روش انجام تحقیق :

#### ۱-۷-۱. شرح روش تحقیق :

با توجه به این که در تحقیق حاضر از فرآیند داده کاوی استفاده خواهد شد، لذا مراحل انجام تحقیق و مدل اجرایی آن متناسب با استاندارد CRISP-DM<sup>۱۰</sup> خواهد بود. فرآیند CRISP-DM یک متدولوژی استاندارد داده کاوی می باشد که در اواخر سال ۱۹۹۶ توسط سه شرکت بزرگ دایملر کرایسلر (بنز)<sup>۱۱</sup>، اس پی اس اس<sup>۱۲</sup> و ان سی آر<sup>۱۳</sup> ایجاد گردید. این متدولوژی چرخه عمر یک پروژه داده کاوی را به ۶ مرحله انعطاف پذیر (۱. شناخت کسب و کار، ۲. شناخت داده ها، ۳. آماده سازی داده ها، ۴. مدل سازی، ۵. ارزیابی مدل و ۶. توسعه مدل) تقسیم بندی کرده است<sup>۱۴</sup>. نگرش این تحقیق به این شکل است که از کار با داده ها شروع شده و سعی بر آن است تا مواردی را که قبلاً آگاهی نسبت به آن ها وجود نداشته است، کشف کرده و برای آن ها قوانینی ساخته شود.

#### ۱-۷-۲. شرح روش و ابزار گردآوری داده ها :

داده های مورد نیاز این تحقیق عمدتاً از صورت های مالی شرکت ها، موجود در سایت بورس اوراق بهادار تهران<sup>۱۵</sup> و مراجعه به نرم افزارها و بانک های اطلاعاتی موجود از جمله تدبیر پرداز و ره آورد نوین جمع آوری شده اند.

#### ۱-۷-۳. جامعه آماری، روش نمونه گیری و حجم نمونه :

<sup>۱۰</sup> Cross Industry Standard Process for Data Mining

<sup>۱۱</sup> Daimler Chrysler (then Daimler-Benz)

<sup>۱۲</sup> SPSS

<sup>۱۳</sup> NCR

<sup>۱۵</sup> www.rdis.ir

<sup>۱۴</sup> SPSS Clementine 12.0 Help



با توجه به این که نتایج داده کاوی با داشتن داده های بیشتر، از صحت بالاتری برخوردار می شوند، بنابراین مشاهدات این تحقیق بر اساس معیارهای زیر انتخاب شده اند :

۱. شرکت می باید در گروه شرکت های واسطه گری مالی، بانک ها و بیمه ها (بدلیل نوع فعالیت این شرکت ها) نباشد.
۲. اطلاعات مورد نیاز، از جمله صورت های مالی سالانه و گزارش حسابرس مستقل شرکت حداقل برای یک سال از دوره مورد بررسی موجود باشد.

#### ۱-۷-۴. روش و فرآیند تجزیه و تحلیل داده ها :

پس از جمع آوری و پردازش داده ها، تجزیه و تحلیل آنها به صورت زیر انجام خواهد شد :

الف) توصیفی : داده های جمع آوری شده توسط روش های آمار توصیفی شامل : جداول توزیع فراوانی، نمودارهای آماری، پراکندگی و ... توصیف می شوند. نرم افزار مورد استفاده در این قسمت SPSS 18 و SPSS Clementine 12.0 خواهد بود.

ب) انجام فرآیند داده کاوی مطابق با استاندارد CRISP : داده کاوی یک فرآیند چند مرحله ای است. به عبارتی برای استخراج اطلاعات مفید از داده ها، می بایست مراحل را قبل از استخراج اطلاعات بر روی داده ها پیاده سازی کرد. که مراحل اجرایی آن به صورت زیر خواهد بود :

۱. جمع آوری داده ها از پایگاه داده های موجود

✓ در این مرحله، داده ها از منابع چندگانه جمع آوری و با هم ترکیب می شوند.

۲. آماده سازی داده ها و انتخاب داده های حاوی اطلاعات مورد نظر

✓ در این مرحله داده های غیر معتبر از مجموعه داده ها خارج می شوند. داده های فاقد یکپارچگی، داده های

گم شده<sup>۱۶</sup> و ... ، نمونه هایی از داده هایی هستند که باید پاک سازی<sup>۱۷</sup> در مورد آنها انجام گردد. این مرحله را

می توان بخشی از فرآیند کاهش داده نیز دانست.

۳. تقسیم داده ها به دو مجموعه داده های آموزشی<sup>۱۸</sup> و داده های آزمایشی<sup>۱۹</sup>

<sup>۱۶</sup> Missing Data

<sup>۱۷</sup> Cleaning Data

<sup>۱۸</sup> Training Data Set

<sup>۱۹</sup> Test Data Set

## Prediction of Independent Auditor Opinion: Data Mining Approach

✓ در این مرحله مجموع داده های انتخاب شده را به دو دسته آموزشی به منظور ساخت مدل و آزمایشی به منظور آزمون مدل تقسیم بندی می نماییم. درصد ترکیب داده های آموزشی و آزمایشی به ترتیب ۸۰ و ۲۰ درصد می باشد.

۴. کشف الگوی مناسب در داده های آموزشی برای طبقه بندی گزارش حسابرسان مستقل

✓ در این مرحله داده های آموزشی را به نحوی طبقه بندی و مرتب می نماییم تا یادگیری مدل به بهترین نحو صورت گیرد. به عبارت دیگر در این مرحله، داده ها در قالبی قابل استفاده برای داده کاوی در می آیند.

۵. مدل سازی با استفاده از داده های آموزشی

✓ این مرحله، بخش اصلی فرآیند داده کاوی است که در آن با استفاده از روش ها و تکنیک های خاصی، شروع به استخراج الگوها و روابط موجود در داده ها می نماییم. به عبارتی در این مرحله از تکنیک های شبکه های عصبی مصنوعی، درخت تصمیم C5.0 و رگرسیون لجستیک برای ساخت مدل استفاده می شود.

۶. آزمون مدل به کمک داده های آزمایشی

✓ در این مرحله با استفاده از داده های آزمایشی دقت مدل ساخته شده، تخمین زده می شود.

۷. تعیین مولفه های اثر گذار در پیش بینی گزارش حسابرسان مستقل

✓ در این مرحله متغیرهای با اهمیت به منظور پیش بینی گزارش سال بعد حسابرسان مستقل مشخص می شود.

نرم افزار مورد استفاده، برای اجرای فرآیند داده کاوی SPSS Clementine 12.0 می باشد.

# فصل دوم

## مبانی نظری

### تحقیق

### مقدمه

در این فصل ابتدا در خصوص انواع و محتوای گزارش های حسابرسی مستقل توضیحاتی ارائه و سپس به کاربرد سه تکنیک شبکه های عصبی مصنوعی، درخت تصمیم و رگرسیون لجستیک در تحقیق های مالی اشاره شده است.

### ۱-۲. گزارش حسابرسی مستقل

در این بخش به کلیاتی در خصوص حسابرسی، انواع گزارش های حسابرسی و عواملی که منجر به صدور این نوع گزارش ها می شود، اشاره شده است.

### ۱-۱-۲. تاریخچه حسابرسی

با تداوم انقلاب صنعتی در اروپای قرن نوزدهم، ایجاد کارخانه های بزرگ نیازمند سرمایه های کلان بود. تامین چنین سرمایه هایی از امکانات مالی یک یا چند سرمایه گذار فراتر بوده و از طرفی، یک یا چند سرمایه دار نیز آمادگی پذیرفتن خطر تجارتي چنین فعالیت های بزرگی را نداشتند. از این رو، با بهره گیری از دو دست آورد بزرگ و مفید انقلاب صنعتی، (یعنی سازماندهی و همکاری) شرکت هایی شکل گرفتند که مسئولیت صاحبان سرمایه آن ها محدود به مبلغ سرمایه گذاریشان بود و در قالب چنین مشارکت هایی، سرمایه های کوچک تجهیز و راه حل مناسبی برای تامین سرمایه های کلان و توزیع مخاطرات تجارتي فراهم آمد. این امر باعث جدایی مدیریت از مالکیت شد. شرکت های سهامی توسط هیات مدیره هایی اداره می شوند که معمولا از سهامداران بزرگ تشکیل و یا منتخب آن ها هستند. از آنجایی که هدف های تهیه کنندگان اطلاعات مالی شرکت ها با هدف های استفاده کنندگان از آن ها مغایرت دارد، نیاز به وجود حسابرس مستقل آشکار گردید. به عبارتی افرادی با صلاحیت و صداقت حرفه ای که بتوانند مشخص کنند اطلاعاتی که بر آن اتکا می شود تصویری مطلوب و کامل از واقعیت امر می باشد یا خیر (آقایی و سید نژاد فهیم، ۱۳۸۵).

### ۲-۱-۲. تعریف حسابرسی

حسابرسی فرآیندی منظم و با قاعده (سیستماتیک) جهت جمع آوری و ارزیابی بی طرفانه شواهد درباره ادعاهای مربوط به فعالیت ها و وقایع اقتصادی، به منظور تعیین درجه انطباق این ادعاها یا اظهارات با معیارهای از پیش تعیین شده و گزارش نتایج به افراد ذینفع است. به عبارت دیگر حسابرسی عبارت است از بازرسی جستجو گرانه مدارک حسابداری و سایر شواهد زیربنای صورت های مالی به منظور اعتباردهی به آن است. دلیل اصلی وجود حرفه حسابرسی مستقل وظیفه اعتباردهی است. اعتبار دادن به صورت های مالی به معنی ایجاد اطمینان از مطلوبیت ارائه و قابلیت اتکای آن ها است (آقایی، سید نژاد فهیم، ۱۳۸۵).