



بررسی تاثیر روش های مختلف خوشه بندی داده ها بر روی دقت مدل های مربوط به برآوردی های حسابداری از طریق مقایسه روش های خوشه بندی سنتی و کلاسیک

سید محسن صالحی وزیری^۱
جمال برزگری خانقاه^۲

تاریخ پذیرش: ۹۷/۰۷/۰۹

تاریخ دریافت: ۹۷/۰۶/۰۳

چکیده

امروزه استفاده از برآورد اطلاعات حسابداری همانند سایر رشته ها بدلیل عدم دسترسی به تمام اطلاعات، لازم و ضروری می باشد. به همین دلیل در این پژوهش سعی شده با استفاده از روش های مختلف خوشه بندی دقت مدل های مربوط به برآورد های حسابداری مورد بررسی قرار گیرد تا مشخص گردد روش های مختلف خوشه بندی به چه میزان دقت مدل های مدنظر را افزایش میدهند و روش ارجح تر در میان روش های مختلف خوشه بندی در افزایش دقت مدل ها کدام روش می تواند باشد. نمونه آماری پژوهش بدین منظور شامل ۹۹ شرکت پذیرفته شده در بورس اوراق بهادار تهران بوده و به منظور جمع آوری داده های مورد نیاز، از اطلاعات صورت های مالی و یادداشت های دوره ۹ ساله (۱۳۸۷-۱۳۹۵) شرکت های مورد نظر استفاده شده است. نتایج حاصل از تحقیق نشان داد که استفاده از روش های مختلف خوشه بندی دقت مدل های مربوط به برآوردهای حسابداری را در اکثر موارد افزایش می دهد ولی از میان روش های خوشه بندی مورد استفاده در تحقیق، روش خوشه بندی کلاسیک روش مناسب تری نسبت به روش سنتی در افزایش دقت مدل های مربوط به برآورد های حسابداری می باشد.

واژه های کلیدی: داده کاوی، خوشه بندی، خوشه بندی سنتی، خوشه بندی کلاسیک.

۱- کارشناسی ارشد حسابداری، دانشگاه یزد، یزد، ایران.

۲- استادیار و عضو هیات علمی دانشگاه یزد، یزد، ایران. (نویسنده مسئول) barzegari@yazd.ac.ir

۱- مقدمه

برآورد های حسابداری و همچنین با امکان دستکاری آن توسط مدیران روبرو می باشد. با توجه به چنین محدودیت هایی نگرانی هایی درباره قابل اتکا بودن برآورد های حسابداری وجود دارد که می تواند سودمند بودن برآورد ها را تحت تاثیر قرار دهد (لو و لی، ۲۰۱۰) با توجه به مباحثی که مطرح شد خوشه بندی به عنوان یکی از شاخه های داده کاوی می تواند با توجه به روش های مختلفی که وجود دارد نقش بسزایی در افزایش دقت مدل های مربوط به برآورد های حسابداری داشته باشد و شرکت ها و مدیران آنها با استفاده از روش های مختلف خوشه بندی دقت برآورد ها و پیش بینی های خود از داده های مالی را افزایش دهند. از آنجایی که تاکنون تحقیقی در این زمینه انجام نشده است با استفاده از روش های مختلف خوشه بندی دقت مدل های مربوط به برآورد های حسابداری را مورد بررسی قرار دهیم تا ببینیم کدام روش خوشه بندی، دقت مدل های مربوط به برآورد های حسابداری را افزایش می دهد.

۲- مبانی نظری و پیشینه پژوهش

امروزه با حجم زیادی از داده ها روبرو هستیم که برای استفاده از آنها به ابزار های کشف دانش نیاز داریم. داده کاوی به عنوان یک توانایی پیشرفته، در تحلیل داده و کشف دانش مورد استفاده قرار می گیرد. برای این کار روش های متعددی وجود دارد که هر یک از آنها برای اهداف خاصی مورد استفاده قرار می گیرند. یکی از مهم ترین روش های داده کاوی، خوشه بندی است که کاربرد بسیاری در کشف دانش دارد. خوشه بندی یکی از روش های اصلی تحلیل داده است. از این روش به صورت گسترده برای تشخیص الگو، استخراج ویژگی ها، مجزا سازی تصاویر، تخمین تابع و داده کاوی استفاده می شود. خوشه بندی به عنوان یک روش دسته بندی غیر نظارتی، برخی از ساختار های ذاتی موجود در مجموعه ای از اشیا را بر اساس تشابه در سنجش مشخص می کند. تحلیل خوشه ای ابزاری برای تجسس در داده های چند متغیره است که با سازمان بندی در خوشه ها به

پیچیدگی و گستردگی سازمان ها و وجود رقیبان در عصر حاضر ایجاب می کند مدیریت هر سازمان به اطلاعات بهینه و به هنگام مجهز باشد و امروزه آمار و اطلاعات نقش محوری در تصمیم گیری و اداره مطلوب سازمان ها دارد و این اطلاعات زمانی مفید خواهند بود که به سرعت و به سهولت قابل دسترسی باشد. بطور کلی هدف از گزارشگری این صورت های مالی فراهم آوردن اطلاعاتی است که آثار مالی معاملات، عملیات و رویداد های مالی موثر بر وضعیت مالی و نتایج عملیات یک واحد انتفاعی را بیان می کند و از این طریق سرمایه گذاران، اعطا کنندگان تسهیلات مالی و سایر استفاده کنندگان برون سازمانی را در قضاوت و تصمیم گیری نسبت به امور یک واحد انتفاعی یاری دهد. با توجه به ابهامات موجود در محیط اقتصادی، انجام برآورد های حسابداری و تجدید نظر در آنها برای تهیه صورت های مالی امری اجتناب ناپذیر است. پیش بینی رویدادهای آتی و برآورد اثرات آنها مستلزم اعمال قضاوت و تجدید نظر در این برآورد ها به هنگام وقوع رویداد های جدید، حصول تجربه بیشتر و دستیابی به اطلاعات اضافی است. برآورد حسابداری تاثیر بسزایی بر اکثر بازیگران صحنه گزارشگری مالی از قبیل استاندارد گذاران و واحد های گزارش حسابرسان دارد (لو، ۱۹۸۹). بر اساس چارچوب نظری، یکی از اهداف انتشار صورت های مالی ارائه اطلاعات سودمند به استفاده کنندگان است، در نتیجه تلاش و هزینه صرف شده در ارائه برآورد های حسابداری به شرطی قابل توجیه است که اطلاعات سودمند در اختیار استفاده کنندگان قرار دهد. سودمندی برآورد های حسابداری به عنوان یکی از پایه ای ترین مسائل در حسابداری با چالش مواجه است چون برآورد های حسابداری، از یکسو بطور بالقوه مربوط بودن اطلاعات حسابداری را بطور کلی افزایش می دهد و بطور بالقوه می تواند برای سرمایه گذاران سودمند باشد و آنها را در پیش بینی جریان نقد آتی یاری دهد. ولی از سوی دیگر برآورد های حسابداری با محدودیت هایی مانند دشواری ذاتی انجام دقیق

داده کاوی

تعریف متنوعی از داده کاوی در مراجع مختلف و توسط افراد مختلف بیان شده، از جمله موارد زیر را می توان اشاره کرد:

داده کاوی: عبارت است از فرآیند استخراج اطلاعات معتبر، از پیش ناشناخته های قابل فهم و قابل اعتماد از پایگاه داده های بزرگ و استفاده از آن در تصمیم گیری در فعالیت های تجاری مهم (دیوید^۳، ۲۰۰۲). تعریفی که در اکثر مراجع به اشتراک گذاشته شده عبارت است از: استخراج اطلاعات و دانش و کشف الگوهای پنهان از پایگاه داده های بزرگ و پیچیده. داده کاوی یک راهبرد بسیار قوی با پتانسیل بالا می باشد که به سازمان ها کمک می کند بر روی مهم ترین اطلاعات از مخزن داده های خود تمرکز نمایند. ویژگی های داده کاوی سبب شده که امروزه در تجارت و کسب و کار از آن بسیار استفاده شود. در دنیای تجارت نیز داده های مالی به عنوان سرمایه راهبرد مطرح هستند. داده های مالی توسط موسساتی مثل بانک ها، بورس اوراق بهادار، سازمان های مالیاتی، پایگاه داده ویژه حسابرسی و حسابداران جمع آوری و نگهداری می شوند. روش های داده کاوی در داده های مالی، می تواند در حل مشکلات طبقه بندی و پیش بینی و تسهیل فرآیند تصمیم گیری به کار رود.

خوشه بندی

فرآیند دسته بندی مجموعه ای از داده های مادی یا انتزاعی در درون طبقه هایی از داده های مشابه را فرآیند خوشه بندی می نامند. یک خوشه مجموعه ای از داده ها است که در درون همان خوشه مشابه یکدیگر بوده ولی با داده های خوشه های دیگر متفاوت هستند. یک خوشه از داده ها را به طور کلی می توان به عنوان مثال گروه و در نتیجه به عنوان یک تراکم داده ای در نظر گرفت. (هان و کامبر^۴، 2006). ایده اصلی خوشه بندی اطلاعات، جداکردن نمونه ها از یکدیگر و قراردادن آنها در گروه های شبیه به هم می باشد. به این معنی که نمونه های شبیه به هم باید

محقق در یافتن ویژگی ساختار داده ها برای دست یابی به هدف معین یاری میدهد. به عبارت دیگر تحلیل خوشه ای برای دسته بندی یک نمونه به خوشه های مجزا طراحی شده است به طوری که نمونه های داخل هر خوشه دارای خاصیت مشابه و خوشه ها دارای خاصیت های به طور کامل متفاوت از یکدیگر باشند. با استفاده از خوشه بندی داده ها می توانیم تحلیل دقیق تری از داده ها داشت و به روابط پنهان بین داده ها پی برد که یکی از موارد استفاده در برآورد داده های حسابداری می باشد که میتوان برآورد دقیق تری از داده ها حسابداری انجام داد و داده های برآوردی را به داده های واقعی نزدیک نمود.

برآورد های حسابداری

برآورد های حسابداری دارای نقش عمده در تهیه صورت های مالی بر مبنای بهای تاریخی دارند. این برآورد ها در نتیجه بکارگیری مفهوم حسابداری تعهدی بطور عمده در تخصیص هزینه ها و بعضا در تخصیص دارایی ها به دوره های مختلف به صورت مستقیم و غیر مستقیم ضرورت می یابند و از جمله موارد مندرج در صورت های مالی اند که دارای ریسک ذاتی بالایی اند، زیرا چنانچه مدیریت واحد تجاری تمایل به بازی کردن با اعداد و ارقام و در نتیجه دستکاری عملکرد واقعی واحد تجاری داشته باشد، مطمئنا روش های مورد عمل در زمینه برآورد های حسابداری را مورد استفاده قرار خواهد داد و در واقع می تواند از این برآورد ها برای کم یا زیاد کردن سود و ایجاد تغییر در وضعیت مالی استفاده کند. (عرب مازار یزدی، ۱۳۸۸). با توجه به اینکه تمامی اطلاعات حسابداری در تمامی زمان ها در دسترس نمی باشند نیاز می شود در جاهای مختلف از اطلاعات برآوردی حسابداری استفاده شود. در یکی از مواردی که برآورد ها مورد استفاده قرار می گیرند در مدل های تخمین می باشد. در این مدل ها از اطلاعات حسابداری برای تخمین یا برآورد سایر اطلاعات حسابداری مورد استفاده قرار می گیرد که در این پژوهش دو مدل تخمین مورد استفاده قرار گرفته شده است.

جهت رفع معایب الگوریتم های قبلی معرفی شده که برخی پارامترها را بهبود داده اما باز خود دارای معایب دیگری است. در کل روش هایی که نیازی به تعیین تعداد خوشه نمی باشد خوشه بندی بهتری را انجام می دهند (شاکری و عبداللهی، ۱۳۹۴).

در تحقیقی به مقایسه الگوریتم های خوشه بندی فازی و غیر فازی پرداخته شده است. نتایج حاصل از تحقیق این شد که استفاده از الگوریتم k-means اگرچه سرعت همگرایی بالایی دارد و زمان اجرای الگوریتم کم می باشد ولی خوشه بندی مطلوبی را به دست نمی دهد و نسبت به انتخاب مراکز اولیه خوشه ها بسیار حساس می باشد. سپس به بررسی عملکرد الگوریتم fcm پرداخته شد. این الگوریتم نسبت به نقاط اولیه حساس نمی باشد و نتیجه مطلوب را به دست می دهد و نسبت به انتخاب مراکز اولیه خوشه از حساسیت کمی برخوردار می باشد (معتمنی و همکاران، ۱۳۹۰).

در مقاله ای با عنوان یادگیری پیشرفته با استفاده از بهینه سازی کیفیت خوشه به بررسی چگونگی تاثیر گذاری فشرده سازی و جداسازی خوشه ها و استفاده از الگوریتم خوشه بندی در بهینه سازی کیفیت خوشه پرداختند. به این نتیجه رسیدند که الگوریتم استاندارد k-means به عنوان یک مورد خاص می تواند باعث بهینه سازی کیفیت خوشه شود (لوپز و همکاران، ۲۰۱۸).

در مقاله ای به بررسی تاثیر روش سلسله مراتبی به عنوان یکی از روش های سنتی در تجزیه و تحلیل خوشه پرداخته شده است. نتایج نشان داد این روش می تواند برای تشخیص ساختار خوشه ای واقعی و شبیه سازی مفید باشد. (پاتریک و همکاران، ۲۰۱۷)

در مقاله ای با عنوان خوشه بندی مبتنی بر مدل به بررسی نحوه خوشه بندی داده ها با توجه به مدل های مختلف و الگوریتم مختلف پرداخته شده است. نتایج تحقیقات نشان می دهد که روش های مختلف خوشه بندی مبتنی بر مدل مزایا و معایبی دارند که روش های جدید بر طرف کننده نقاط ضعف روش های گذشته می باشند که با توجه به کاربرد آنها می

در یک گروه قرار گیرند و با نمونه های گروه های دیگر حداکثر تفاوت را داشته باشند (جین و مارتین^۵، ۲۰۰۵). از آنجا که اکثر روش های خوشه بندی پایه، روی جنبه های خاصی از داده ها تاکید می کنند، در نتیجه روی مجموعه داده های خاصی کارآمد می باشند. به همین دلیل نیاز به روش هایی است که بتواند با استفاده از ترکیب این الگوریتم ها و گرفتن نقاط قوت هر یک، نتایج بهینه تری را تولید کند. (استرل و قوش^۶، ۲۰۰۲؛ فرد و لورنسو^۷، ۲۰۰۸).

با توجه به موضوعات مطرح شده می توان روش های خوشه بندی را به دو روش خوشه بندی کلاسیک و خوشه بندی سنتی تقسیم نمود. در خوشه بندی سنتی تعداد خوشه ها را می توان تعیین نمود که می تواند تاثیر گذار در نتیجه کار باشد ولی در خوشه بندی به روش سنتی نمی توان تعداد خوشه ها را مشخص نمود و خود سیستم با توجه به ویژگی داده ها تعداد خوشه ها را مشخص می کند.

۲-۱- پیشینه تحقیق

در مقاله ای با استفاده از تکنیک خوشه بندی فازی و خوشه بندی سنتی و بر اساس معیارهای منطقه بهینه پولی، درجه تشابه اقتصادی بین کشورهای اسلامی را بررسی کرده اند. بر اساس نتایج حاصل، از بین کشورهای اسلامی کشورهای عضو اتحادیه اقتصادی عرب از درجه تشابه بالایی برخوردار می باشند. به این ترتیب اگر کشورهای اسلامی به ضرورت و اهمیت همگرایی پولی بین خود پی برده و جدیت کافی در ایجاد این همگرایی را داشته باشند آنگاه می توانند اتحادیه پولی کشورهای اسلامی را با ورود کشورهای عضو اتحادیه اقتصادی عرب شروع و اصلاحات اقتصادی را انجام دهند که با کمک روش خوشه بندی ممکن می باشد (رزاقی و همکاران، ۱۳۹۶).

در تحقیقی به خوشه بندی داده ها و روش های موجود و مقایسه عملکرد آنها پرداخته شده است. نتایج نشان می دهد که هر کدام از روش ها دارای مزایا و معایبی می باشند. بعضی از این الگوریتم ها

توانند در موضوعات مختلف مورد استفاده قرار بگیرند(نیکولاس^۸، ۲۰۱۶).

در مطالعه ای در مورد نحوه ی خوشه بندی داده های حسابداری به عنوان یک روش کنترل داخلی مورد توجه قرار گرفته است. در این مطالعه ابزاری تحت نظارت به عنوان نقشه های خود سازمان ده مطرح می شود. این مطالعه در یک شرکت حمل و نقل از طریق یک سری از آزمایش ها بر روی یک سری پایگاه داده های حسابداری انجام شده است. نتایج تجربی نشان می دهد که خوشه بندی داده های حسابداری در خوشه های همگن و خوشه های جداگانه می تواند ابزاری مناسب برای تفسیر های حسابداری باشد(ناواس و همکاران^۹، ۲۰۱۴).

در مقاله ای در چین در سال ۲۰۱۲ با عنوان خوشه بندی داده های دسته بندی شده با استفاده از مدل های چند بعدی به بررسی تاثیر گذاری خوشه بندی داده ها با توجه به مدل های چند بعدی پرداختند که نتایج تحقیقات نشان می دهد که خوشه بندی داده ها با توجه به مدل ها کار تجزیه و تحلیل داده ها را راحت تر و دقیق تر می کند(چن و همکاران، ۲۰۱۲).

در مقاله ای با عنوان انتخاب متغیر ها برای خوشه بندی مبتنی بر مدل به بررسی تاثیر متغیر های مختلف بر خوشه بندی مبتنی بر مدل پرداخته است. نتایج تحقیقات نشان می دهد که حذف متغیر های نامناسب اغلب عملکرد مدل های مختلف را بهبود می بخشد و انتخاب متغیر های مناسب باعث می شود که برآورد دقیق تری با توجه به مدل های مختلف حاصل شود(رافتری و دین^{۱۰}، ۲۰۰۶).

در پژوهشی تعدادی از واحد های خروجی استفاده شده در شبکه های خود سازمان ده را که بر روی قابلیت اجرایی آنها برای خوشه بندی و تصور تاثیر گذار است بررسی شده است و اثبات شده که نقشه های خود سازمان ده می توانند بصورت جداگانه برای خوشه بندی یا تصور و یا به طور همزمان برای خوشه بندی و تصویر و یا حتی برای خوشه بندی همراه با تصویر مورد استفاده قرار گیرند. (فلکسر^{۱۱}، ۲۰۰۱).

در مقاله ای با عنوان خوشه بندی مجموعه داده های همگن توزیع شده به بررسی اثرگذاری داده های همگن بر نتیجه حاصل خوشه بندی پرداخته شد که نتایج تحقیق نشان می دهد که این روش می تواند با محدودیت زمانی سازگار تر باشد و برآورد و تخمین دقیق تری انجام می گیرد(پارساساراسی و اگیهارا^{۱۲}، ۲۰۰۰).

۳- سوالات تحقیق

- ۱) آیا استفاده از روش های مختلف خوشه بندی دقت مدل های مربوط به برآورد های حسابداری را افزایش می دهد؟
- ۲) روش ارجح خوشه بندی در افزایش دقت مدل های مربوط به برآورد های حسابداری کدام روش می باشد؟

۴- روش شناسی پژوهش

تحقیق حاضر از حیث هدف کاربردی است. از جهت ماهیت و روش توصیفی همبستگی است و از جنبه روش گردآوری داده ها، کتابخانه ای و اسنادی است. برای پاسخ به سوالات پژوهش نیز روشی که در نظر گرفته شده این است که داده های دو مدل پژوهش را به روش های مختلف در نرم افزار Rapidminer خوشه بندی کرده، که در این پژوهش در خوشه بندی به سمت خوشه های همگن تر رفته و خوشه هایی را مد نظر قرار می دهیم که بعد از چند مرحله خوشه بندی به بهترین خوشه ممکن در مدل مربوط به آن رسیده ایم. به عنوان مثال در ابتدا داده های ۹۹ شرکت را خوشه بندی کرده، که بعد از خوشه بندی اولیه به عنوان مثال به پنج خوشه میرسیم که خوشه اول دارای ۵ داده، خوشه دوم ۷ داده، خوشه سوم ۶ داده، خوشه چهارم ۹ داده و خوشه پنجم ۷۲ داده می باشند. در مرحله بعد دوباره خوشه پنجم را که شامل ۷۲ داده است خوشه بندی کرده و داده های چهار خوشه دیگر را مد نظر قرار نمی دهیم. در مرحله بعد که ۷۲ داده را خوشه بندی کردیم شبیه مرحله قبل عمل می کنیم و خوشه هایی

مدل دوم مخارج اختیاری طبیعی را به عنوان یک تابع خطی از فروش باقی مانده بیان می کند.

Disexp در این مدل به عنوان مجموع هزینه های تبلیغات، مخارج تحقیق و توسعه و مجموع هزینه اداری و فروش می باشد. A_{t-1} : باقی مانده کل دارایی ها در پایان دوره، S_{t-1} : فروش دوره قبل در این مدل می توان گفت با توجه به فروش دوره قبل برآوردی از مجموع هزینه های تبلیغات، مخارج تحقیق و توسعه و هزینه اداری و فروش این دوره را داشت.

۶- متغیر های پژوهش

اولین گام برای پاسخ به سوالات پژوهش ارایه تعریف دقیق و مناسبی از متغیر هایی است که امکان اندازه گیری خصوصیات مورد توجه در این پژوهش را فراهم می کند که بر دو نوع است: متغیر وابسته، متغیر مستقل. در این پژوهش از دو مدل برای مقایسه ی روش های مختلف برآورد های حسابداری استفاده می شود که در جدول زیر توضیح داده می شود:

جدول ۱- تعریف متغیر های تحقیق

مدل اول		
$CFOt \div At - 1 = \alpha + \alpha 1(1 \div At - 1) + \beta 1(S_t \div At - 1) + \beta 2(\Delta St \div At - 1) + \epsilon_{t \square \square}$		
تعریف	نماد متغیر	نوع متغیر
در این مدل جریان نقد حاصل از عملیات مد نظر است.	Cfo	وابسته
درآمد حاصل از فروش کالا و خدمات این دوره	S_t	مستقل
تغییرات درآمد حاصل از فروش کالا و خدمات این دوره	ΔS_t	مستقل
مدل دوم		
$DISXPt \div At - 1 = \alpha + \alpha 1(1 \div At - 1) + \beta 1(S_t - 1 \div At - 1) + \epsilon_{t \square \square}$		
تعریف	نماد متغیر	نوع متغیر
در این مدل به عنوان مجموع هزینه های تبلیغات، مخارج تحقیق و توسعه و مجموع هزینه تبلیغات تعریف می شود.	DISXP _t	وابسته
درآمد حاصل از فروش دوره قبل مد نظر است.	S_{t-1}	مستقل

که دارای داده کمی می باشند حذف کرده تا به همگن ترین و بهترین خوشه ممکن رسیده و رگرسیون را بر روی خوشه مد نظر در نرم افزار spss انجام می دهیم و با مقایسه میزان R و R^2 اولیه مدل (قبل از خوشه بندی) و بعد از خوشه بندی، میزان تاثیر گذاری خوشه بندی را بر روی دقت مدل میسنجیم. اگر میزان R و R^2 با خوشه بندی افزایش یابد میزان دقت مدل افزایش یافته و اگر میزان R و R^2 کاهش یابد میزان دقت مدل کاهش یافته است.

۵- مدل های پژوهش

در این پژوهش دو مدل تخمین مورد استفاده قرار گرفته شده است که در مدل اول به برآورد جریان نقد حاصل از عملیات از طریق فروش این دوره و تغییرات فروش این دوره (رویچادهاری^{۱۳}، ۲۰۰۶)، در مدل دوم به برآورد مخارج اختیاری طبیعی (مجموع هزینه های تبلیغات، مخارج تحقیق و توسعه و اداری و فروش) از طریق فروش دوره قبل (رویچادهاری، ۲۰۰۶) پرداخته شده است. در این مطالعه به مقایسه ی روش های مختلف خوشه بندی با توجه به مدل های برآورد مبتنی بر رگرسیون خطی در حسابداری می پردازیم.

مدل اول:

$$CFOt \div At - 1 = \alpha + \alpha 1(1 \div At - 1) + \beta 1(S_t \div At - 1) + \beta 2(\Delta St \div At - 1) + \epsilon_{t \square \square}$$

اولین مدل از جریان نقدی نرمال از عملیات به عنوان یک تابع خطی فروش و تغییرات در فروش در دوره جاری است.

A_{t-1} : باقی مانده کل دارایی ها در پایان دوره، S_t :

فروش طی دوره، ΔSt : تغییر در فروش طی دوره با توجه به مدل می توان با استفاده از فروش و تغییر فروش طی دوره برآوردی از جریان نقد عملیات طی دوره و حتی برآوردی از جریان نقد عملیاتی دوره های آینده داشت.

مدل دوم:

$$DISXPt \div At - 1 = \alpha + \alpha 1(1 \div At - 1) + \beta 1(S_t - 1 \div At - 1) + \epsilon_{t \square \square}$$

۷- روش ها و ابزار تجزیه و تحلیل داده ها

داده های پژوهش به روش کتابخانه ای و با مراجعه به صورت های مالی شرکت ها و گزارش های هیات مدیره و همچنین اطلاعات مالی منتشره توسط سازمان بورس اوراق بهادار تهران و وب سایت www.codal.ir گردآوری شده است. در تجزیه و تحلیل اطلاعات از آمار توصیفی و روش های آماری مربوطه و مدل های رگرسیونی چند گانه و با استفاده از نرم افزار SPSS و rapidminer انجام شده است.

۹- یافته های پژوهش

۹-۱- آمار توصیفی

مهمترین شاخص مرکزی میانگین می باشد که نشان دهنده ی نقطه تعادل و مرکز ثقل توزیع است. همانگونه که در جدول مشاهده می شود که در میان متغیر های وابسته، متغیر Cfo برابر ۱۰۳۷,۷ و متغیر DISEXP_t برابر ۴۵۶,۵ به ترتیب دارای بیشترین و کمترین مقدار میانگین در میان متغیر های وابسته می باشند. در میان متغیر های مستقل نیز متغیر S_t برابر ۷۲۸۲,۰۳ و متغیر ΔS_t نیز برابر ۰,۸۹ می باشد که به ترتیب بیشترین و کمترین مقدار میانگین در میان متغیر های مستقل می باشند.

میانه یکی دیگر از شاخص های مرکزی است که مرکزیت داده ها را نشان می دهد. در میان متغیر های وابسته متغیر Cfo برابر ۹۲,۱۲ و متغیر DISEXP_t برابر ۷۴,۲ به ترتیب بیشترین و کمترین مقادیر میانه را در میان متغیر های وابسته را دارا می باشند. در میان متغیر های مستقل نیز متغیر S_{t-1} برابر ۶۵۶,۳۳

متغیر ΔS_t برابر ۰,۱۸ به ترتیبی بیشترین و کمترین مقدار میانه را در میان متغیر های مستقل دارا می باشند.

میزان عدم تقارن منحنی فراوانی را چولگی می نامند. اگر ضریب چولگی صفر باشد، جامعه کاملا متقارن است و چنانچه ضریب مثبت باشد، چولگی به راست و اگر منفی باشد چولگی به چپ وجود خواهد داشت. در میان متغیر های وابسته متغیر DISEXP_t برابر ۰,۰۵۳ و متغیر Cfo برابر ۰,۰۳۸ می باشند که به ترتیب بیشترین و کمترین مقدار چولگی در میان متغیر های وابسته را دارند. در میان متغیر های مستقل ΔS_t برابر ۰,۰۹۹ و متغیر S_{t-1} برابر ۰,۰۵۵ می باشند که به ترتیب بیشترین و کمترین میزان چولگی در میان متغیر های مستقل را دارا می باشند.

میزان کشیدگی منحنی فراوانی نسبت به منحنی نرمال استاندارد را برجستگی یا کشیدگی می نامند. اگر کشیدگی حدود صفر باشد منحنی فراوانی از لحاظ کشیدگی وضعیت متعادل و نرمالی خواهد داشت، اگر این مقدار مثبت باشد منحنی برجسته و اگر منفی باشد منحنی پهن می باشد. در میان متغیر های وابسته متغیر DISEXP_t برابر ۰,۰۳۶ و متغیر Cfo برابر ۰,۱۵۲ می باشند که به ترتیب بیشترین و کمترین مقدار کشیدگی در میان متغیر های وابسته را دارا می باشند. در میان متغیر های مستقل متغیر ΔS_t برابر ۰,۰۹۸ و متغیر S_t برابر ۰,۰۳۳ که به ترتیب بیشترین و کمترین مقدار کشیدگی در میان متغیر های مستقل دارا می باشند.

جدول ۲- آمار توصیفی

نماد متغیر	میانگین	ماکزیمم	مینیمم	واریانس	دامنه	میانه	کشیدگی	چولگی
Cfo	۱۰۳۷,۷۶	۱۷۲۵۲,۲	-۳۴,۲۰	۸۳۲۱۱,۶	۱۷۲۸۶,۴	۹۲,۱۲	۰,۰۱۵	۰,۰۰۳۸
St	۷۲۸۲,۰۳	۱۸۸۳۹۲,۱	۲۰,۸۸	۶۳۷۹۶,۸	۱۸۸۳۷۱,۲	۶۵۶,۳	۰,۰۰۳	۰,۰۰۰۵
ΔS_t	۰,۸۹	۸۶,۲	-۰,۷۰	۷۵۱۴۳,۱	۸۶,۹۷	۰,۰۱۸	۰,۰۰۹	۰,۰۰۰۹
DISEXP _t	۴۵۶,۵	۹۰۸۱,۲۸	۱,۹۶	۱۲۵۰۶,۰۸	۹۰۷۹,۳	۷۴,۲۲	۰,۰۰۳۶	۰,۰۰۰۵۳
St-1	۷۲۸۱,۹	۱۸۸۳۹۲,۱	۲۰,۸۸	۶۳۷۹۷	۱۸۸۳۷۱,۲	۶۵۶,۳۳	۰,۰۰۳۳	۰,۰۰۰۵۵

۹-۲- آمار استنباطی

تجزیه و تحلیل مدل تحقیق

برای تجزیه و تحلیل مدل اول جدولی ارایه می گردد که به شرح زیر است:

جدول ۳- خلاصه مدل تحقیق مدل اول

آماره دوربین واتسون	ضریب تعیین چندگانه تعدیل شده	ضریب تعیین چندگانه (R^2)	ضریب همبستگی چند (R) گانه	مدل
۱,۷۱۹	۰,۳۰۳	۰,۳۱۷	۰,۵۶۳	۱

مقدار R در جدول ۳ در مدل اول برابر ۰,۵۶۳ می باشد که نشان می دهد مدل رگرسیون ارائه شده قدرت توضیح دهندگی متوسطی را دارا می باشد. در جدول فوق مقدار مجذور R برای مدل اول برابر ۰,۳۱۷ می باشد که نشان می دهد متغیرها به تنهایی حدود ۳۲٪ از کل تغییرات را توجیه می کند و مابقی سهم سایر متغیرها می باشد. برای تجزیه و تحلیل مدل دوم جدولی ارایه می گردد که به شرح زیر است:

جدول ۴- خلاصه مدل تحقیق مدل دوم

آماره دوربین واتسون	ضریب تعیین چندگانه تعدیل شده	ضریب تعیین چندگانه (R^2)	ضریب همبستگی چند (R) گانه	مدل
۱,۵۸۳	۰,۱۶۶	۰,۱۷۵	۰,۴۱۸	۲

مقدار R در جدول ۴ در مدل دوم برابر ۰,۴۱۸ می باشد که نشان می دهد مدل رگرسیون ارائه شده قدرت توضیح دهندگی متوسطی را دارا می باشد. در جدول فوق مقدار مجذور R برای مدل دوم برابر ۰,۱۷۵ می باشد که نشان می دهد متغیرها به تنهایی حدود ۱۷,۵٪ از کل تغییرات را توجیه می کند و مابقی سهم سایر متغیرها می باشد.

بررسی میزان دقت مدل ها با توجه به روش های مختلف خوشه بندی

در این قسمت معیاری که برای بررسی میزان دقت مدل های تحقیق ملاک قرار میگیرد میزان R^2 می باشد که میزان دقت مدل های تحقیق را قبل و بعد از خوشه بندی، با توجه به میزان آنها میسنجیم که طبق جدول های زیر ارایه می گردد:

خوشه بندی به روش مدرن

در اینجا روشی که مورد استفاده قرار می گیرد روش som (نقشه خود سازمان ده) می باشد که در این روش تعداد ابعاد شبکه، اندازه شبکه، تعداد دفعات تکرار آزمون، حداقل و حداقل تعداد خوشه را می توان مشخص نمود.

مدل اول تحقیق خوشه بندی به روش SOM

برای تجزیه و تحلیل مدل اول جدولی ارایه می گردد که به شرح زیر است: (عدد داخل پرانتز نشان دهنده تعداد داده های موجود در هر خوشه می باشد).

جدول ۶- خلاصه مدل تحقیق مدل اول خوشه بندی به روش SOM

آماره دوربین واتسون	ضریب تعیین چندگانه تعدیل شده	ضریب تعیین چندگانه (R^2)	ضریب همبستگی چند (R) گانه	خوشه (تعداد داده موجود در هر خوشه)	مدل
۱,۷۶۲	۰,۴۱۷	۰,۴۴۱	۰,۶۶۴	۱(۳۵)	۱

مقدار R در جدول ۶ در مدل اول و خوشه اول برابر ۰,۶۶۴ می باشد که نشان می دهد در این خوشه مدل رگرسیون ارائه شده قدرت توضیح دهندگی بالایی را دارا می باشد. در جدول فوق مقدار مجذور R برای مدل اول و خوشه اول برابر ۰,۴۴۱ می باشد که نشان می دهد متغیرها به تنهایی حدود ۴۴٪ از کل تغییرات را توجیه می کند و مابقی سهم سایر متغیرها می باشد. با توجه به نتایج جدول میزان R و R^2 کلی برابر ۰,۵۶۳ و ۰,۳۱۷ می باشد که در خوشه اول

مدل اول بر اساس خوشه بندی به روش K-MEANS برای تجزیه و تحلیل مدل اول جدولی ارایه می گردد که به شرح زیر است: (عدد داخل پرانتز نشان دهنده تعداد داده های موجود در هر خوشه می باشد).

جدول ۹- خلاصه مدل تحقیق مدل اول خوشه بندی

به روش K-MEANS

مدل	خوشه (تعداد داده موجود در هر خوشه)	ضریب ضریب همبستگی چند گانه (R)	ضریب تعیین چند گانه (R ²)	ضریب تعیین چندگانه تعدیل شده	آماره دوربین واتسون
۱	۱(۲۷)	۰,۵۰۳	۰,۲۵۳	۰,۱۹۱	۱,۶۶۷
۱	۲(۳۱)	۰,۶۵۱	۰,۴۲۴	۰,۳۹۹	۱,۷۲۵

مقدار R در جدول ۹ در مدل اول و خوشه اول برابر 0.503 می باشد که نشان می دهد در این خوشه مدل رگرسیون ارائه شده قدرت توضیح دهنده متوسطی را دارا می باشد. در جدول فوق مقدار مجذور R برای مدل اول و خوشه اول برابر ۰,۲۵۳ می باشد که نشان می دهد متغیرها به تنهایی حدود ۲۵٪ از کل تغییرات را توجیه می کند و مابقی سهم سایر متغیرها می باشد. مقدار R در جدول ۹ در مدل اول و خوشه دوم نیز برابر ۰,۶۵۱ می باشد که نشان می دهد در این خوشه مدل رگرسیون ارائه شده قدرت توضیح دهنده بالایی را دارا می باشد. در جدول فوق مقدار مجذور R برای مدل اول و خوشه دوم نیز برابر ۰,۴۲۴ می باشد که نشان می دهد متغیرها به تنهایی حدود ۴۲٪ از کل تغییرات را توجیه می کند و مابقی سهم سایر متغیرها می باشد. با توجه به نتایج جدول میزان R و R² کلی برابر ۰,۵۶۳ و ۰,۳۱۷ می باشد که در خوشه دوم به مقدار ۰,۶۵۱ و ۰,۴۲۴ رسیده است که نشان می دهد دقت مدل در این خوشه افزایش یافته است ولی در خوشه اول بهبودی در دقت مدل حاصل نشده است.

به مقدار ۰,۶۶۴ و ۰,۴۴۱ رسیده است که نشان می دهد دقت مدل در این خوشه افزایش یافته است.

مدل دوم بر اساس خوشه بندی به روش SOM برای تجزیه و تحلیل مدل دوم جدولی ارایه می گردد که به شرح زیر است:

جدول ۷- خلاصه مدل تحقیق دوم خوشه بندی به

روش SOM

مدل	خوشه (تعداد داده موجود در هر خوشه)	ضریب ضریب همبستگی چند گانه (R)	ضریب تعیین چند گانه (R ²)	ضریب تعیین چندگانه تعدیل شده	آماره دوربین واتسون
۲	۱(۳۳)	۰,۷۱۹	۰,۵۱۷	۰,۴۷۷	۱,۷۸۲

مقدار R در جدول ۷ در مدل دوم و خوشه اول برابر ۰,۷۱۹ می باشد که نشان می دهد در این خوشه مدل رگرسیون ارائه شده قدرت توضیح دهنده بالایی را دارا می باشد. در جدول فوق مقدار مجذور R برای مدل دوم و خوشه اول برابر ۰,۵۱۷ می باشد که نشان می دهد متغیرها به تنهایی حدود ۵۲٪ از کل تغییرات را توجیه می کند و مابقی سهم سایر متغیرها می باشد. با توجه به نتایج جدول میزان R و R² کلی برابر ۰,۴۱۸ و ۰,۱۷۵ می باشد که در خوشه اول به مقدار ۰,۷۱۹ و ۰,۵۱۷ رسیده است که نشان می دهد دقت مدل در این خوشه بسیار افزایش یافته است.

خوشه بندی به روش سنتی

در اینجا دو روش مورد استفاده روش k-means و x-means می باشند که در روش k-means الگوریتم از k نقطه تصادفی در فضای حالت شروع می کند و به صورت تکراری داده ها را به نزدیک ترین مراکز خوشه ها انتساب می دهد و مراکز خوشه ها را با آن بروز رسانی می نماید. آن عمل را تا جایی که هیچ بهبودی امکان پذیر نباشد ادامه می دهد و در روش x-means تفاوتی که وجود دارد این می باشد که میتواند حداقل و حداکثر تعداد خوشه را تعیین نمود.

و خوشه دوم نیز برابر ۰,۶۵۱ می باشد که نشان می دهد در این خوشه مدل رگرسیون ارائه شده قدرت توضیح دهندگی بالایی را دارا می باشد. در جدول فوق مقدار مجذور R برای مدل اول و خوشه دوم برابر ۰,۴۲۱ می باشد که نشان می دهد متغیرها به تنهایی حدود ۴۲٪ از کل تغییرات را توجیه می کند و مابقی سهم سایر متغیرها می باشد. با توجه به نتایج جدول میزان R و R^2 کلی برابر ۰,۵۶۳ و ۰,۳۱۷ می باشد که در خوشه دوم به مقدار ۰,۶۵۱ و ۰,۴۲۴ رسیده است که نشان می دهد دقت مدل در این خوشه افزایش یافته است ولی در خوشه اول بهبودی در دقت مدل حاصل نشده است.

مدل دوم بر اساس خوشه بندی به روش k-means برای تجزیه و تحلیل مدل دوم جدولی ارائه می گردد که به شرح زیر است:

جدول ۱۰- خلاصه مدل تحقیق مدل دوم خوشه بندی

به روش K-MEANS

آماره دوربین واتسون	ضرب تعیین چندگانه تعدیل شده	ضرب تعیین چند گانه (R^2)	ضرب همبستگی چند (R) گانه	خوشه (تعداد داده موجود در هر خوشه)	مدل
۱,۶۰۰	-۰,۰۱۲	۰,۰۰۱	۰,۰۳۴	۱(۳۴)	۲

جدول ۱۲- خلاصه مدل تحقیق مدل اول خوشه بندی

به روش x-means

آماره دوربین واتسون	ضرب تعیین چندگانه تعدیل شده (R^2)	ضرب همبستگی چند (R) گانه	خوشه (تعداد داده موجود در هر خوشه)	مدل
۱,۶۶۷	۰,۱۹۱	۰,۲۵۰	۰,۵۰۳	۱(۲۲)
۱,۷۳۵	۰,۳۹۹	۰,۴۲۱	۰,۶۵۱	۱(۳۱)

مقدار R در جدول ۱۰ در مدل دوم و خوشه اول برابر ۰,۰۳۴ می باشد که نشان می دهد در این خوشه مدل رگرسیون ارائه شده قدرت توضیح دهندگی پایینی را دارا می باشد. در جدول فوق مقدار مجذور R برای مدل دوم و خوشه اول برابر ۰,۰۰۱ می باشد که نشان می دهد متغیرها به تنهایی حدود ۱٪ از کل تغییرات را توجیه می کند و مابقی سهم سایر متغیرها می باشد. با توجه به نتایج جدول میزان R و R^2 کلی برابر ۰,۴۱۸ و ۰,۱۷۵ می باشد که در خوشه اول به مقدار ۰,۰۳۴ و ۰,۰۰۱ رسیده است که نشان می دهد دقت مدل در این خوشه افزایش نیافته است.

مدل دوم بر اساس خوشه بندی به روش x-means

برای تجزیه و تحلیل مدل دوم جدولی ارائه می گردد که به شرح زیر است:

مقدار R در جدول ۱۳ در مدل دوم و خوشه اول برابر ۰,۰۸۱ می باشد که نشان می دهد در این خوشه مدل رگرسیون ارائه شده قدرت توضیح دهندگی پایینی را دارا می باشد. در جدول فوق مقدار مجذور R برای مدل دوم و خوشه اول برابر ۰,۰۰۷ می باشد که نشان می دهد متغیرها به تنهایی حدود ۱٪ از کل تغییرات را توجیه می کند و مابقی سهم سایر متغیرها می باشد. مقدار R در جدول ۱۳ در مدل دوم و خوشه دوم برابر ۰,۷۳۷ می باشد که

مدل اول بر اساس خوشه بندی به روش x-means برای تجزیه و تحلیل مدل اول جدولی ارائه می گردد که به شرح زیر است: (عدد داخل پرانتز نشان دهنده تعداد داده های موجود در هر خوشه می باشد).

مقدار R در جدول ۱۲ در مدل اول و خوشه اول برابر ۰,۵۰۳ می باشد که نشان می دهد در این خوشه مدل رگرسیون ارائه شده قدرت توضیح دهندگی متوسطی را دارا می باشد. در جدول فوق مقدار مجذور R برای مدل اول و خوشه اول برابر ۰,۲۵۰ می باشد که نشان می دهد متغیرها به تنهایی حدود ۲۵٪ از کل تغییرات را توجیه می کند و مابقی سهم سایر متغیرها می باشد. مقدار R در جدول ۱۲ در مدل اول

نشان می دهد در این خوشه مدل رگرسیون ارائه شده قدرت توضیح دهندگی بالایی را دارا می باشد. در جدول فوق مقدار مجذور R برای مدل دوم و خوشه دوم برابر $0,543$ می باشد که نشان می دهد متغیرها به تنهایی حدود 54% از کل تغییرات را توجیه می کند و مابقی سهم سایر متغیرها می باشد. با توجه به میزان R و R^2 که برابر $0,418$ و $0,175$ می باشد در خوشه دوم میزان R و R^2 برابر $0,737$ و $0,543$ می باشد که نشان می دهد در این خوشه میزان دقت مدل افزایش یافته است.

برآورد های حسابداری به شرطی قابل توجیه است که اطلاعات سودمند در اختیار استفاده کنندگان قرار دهد. سودمندی برآورد های حسابداری به عنوان یکی از پایه ای ترین مسائل در حسابداری با چالش مواجه است. چون برآورد های حسابداری، از یکسو بطور بالقوه مربوط بودن اطلاعات حسابداری را بطور کلی افزایش می دهد و بطور بالقوه می تواند برای سرمایه گذاران سودمند باشد و آنها را در پیش بینی جریان نقد آتی یاری دهد. ولی از سوی دیگر برآورد های حسابداری با محدودیت هایی مانند دشواری ذاتی انجام دقیق برآورد های حسابداری و همچنین امکان دستکاری آن توسط مدیران می باشد. با توجه به چنین محدودیت هایی نگرانی هایی درباره قابل اتکا بودن برآورد های حسابداری وجود دارد که می تواند سودمند بودن برآورد ها را تحت تاثیر قرار دهد. مدیران نیز می توانند با توجه به موارد مطرح شده دستکاری در برآوردها انجام داده و داده های برآوردی را مطابق میل خود طراحی نمایند. در این راستا روش هایی باید وجود داشته باشد که دستکاری مدیران را با محدودیت هایی روبرو کند که یکی از این روش ها خوشه بندی داده ها می باشد. در این پژوهش نیز دو مدل که با توجه به آن مدل ها می توان متغیرهایی را مورد برآورد قرار داد مطرح شده است که با استفاده روش های مختلف خوشه بندی داده ها می خواهیم دقت مدل ها را افزایش داده تا مدل ها برآورد دقیق تری از اطلاعات داشته باشند. بر اساس سوال اول تحقیق که آیا روش های خوشه بندی بر دقت مدل ها تاثیر دارد یا خیر. در روش خوشه بندی مدرن که از روش som استفاده می شود با توجه به میزان R و R^2 میزان دقت مدل ها به میزان زیادی افزایش یافته است و در روش خوشه بندی سنتی در روش k -means دقت مدل اول در یک خوشه افزایش یافته و در یک خوشه افزایش نیافته ولی در مدل دوم دقت مدل در تنها خوشه ممکن افزایش نیافته است. در روش x -means دقت مدل اول و دوم در یک خوشه افزایش یافته ولی در خوشه دیگر افزایش نیافته است. بر اساس سوال دوم تحقیق که کدام روش ارجح

جدول ۱۳- خلاصه مدل تحقیق مدل دوم خوشه

بندی به روش X-MEANS

آماره دوربین واتسون	ضریب تعیین چندگانه تعدیل شده	ضریب تعیین چندگانه (R^2)	ضریب همبستگی چندگانه (R)	خوشه (تعداد داده موجود در هر خوشه)	مدل
۲,۱۶۸	-۰,۱۱	۰,۰۰۷	۰,۰۸۱	۱ (۳۲)	۲
۱,۱۰۲	۰,۵۱۵	۰,۵۴۳	۰,۷۳۷	۲ (۲۹)	۲

۱۰- نتیجه گیری و بحث

با توجه به ابهامات موجود در محیط اقتصادی، انجام برآورد های حسابداری و تجدید نظر در آنها برای تهیه صورت های مالی امری اجتناب ناپذیر است. پیش بینی رویدادهای آتی و برآورد اثرات آنها مستلزم اعمال قضاوت و تجدید نظر در این برآوردها به هنگام وقوع رویداد های جدید، حصول تجربه بیشتر و دستیابی به اطلاعات اضافی است. حرکت حسابداری به سمت حسابداری ارزش منصفانه، موجب افزایش برآورد مدیر در شناسایی و اندازه گیری اقلام صورت های مالی شده است. برآورد حسابداری تاثیر بسزایی بر اکثر بازیگران صحنه گزارشگری مالی از قبیل استاندارد گذاران و واحد های گزارش حسابرسان دارد. بر اساس چارچوب نظری، یکی از اهداف انتشار صورت های مالی ارائه اطلاعات سودمند به استفاده کنندگان است، در نتیجه تلاش و هزینه صرف شده در ارائه

دارای ویژگی متفاوتی با نمونه مورد نظر هستند باید با احتیاط صورت گیرد.

پیشنهاد های کاربردی

- (۱) با توجه به نتایج تحقیق در مورد مدل ها که نشان می دهد خوشه بندی داده های موجود موجب بهبود و افزایش کارایی مدل ها شده و اطلاعات را به اطلاعات واقعی تر نزدیک می کند به مدیران و تهیه کنندگان صورت های مالی توصیه می شود که با استفاده از روش مختلف خوشه بندی اطلاعات، دقت داده های برآوردی خود را افزایش داده تا اطلاعات کامل تری در اختیار استفاده کنندگان از این اطلاعات قرار گیرد تا بتوانند تصمیمات بهتری برای سرمایه گذاری اتخاذ کنند.
- (۲) به سرمایه گذاران نیز پیشنهاد می شود در جهت گرفتن تصمیمات در مورد تخصیص منابع و سرمایه گذاری در شرکت ها، به صورت های مالی شرکت ها توجه بیشتری داشته باشند که شرکت ها با استفاده از روش های مختلف که یکی از این روش ها خوشه بندی اطلاعات موجود می باشد، اطلاعات برآوردی خود را به اطلاعات واقعی نزدیک کرده باشند.

پیشنهاد برای پژوهش های آتی

- (۱) استفاده از متغیر های کنترلی مختلف مانند سن شرکتها، دارایی های مشهود، دارایی های نامشهود و ... برای بررسی میزان تاثیر گذاری متغیر های کنترلی بر نتایج تحقیق را مورد توجه قرار دهند.
- (۲) همچنین خوشه بندی داده های تحقیق بر اساس صنایع و ویژگی های مختلف صنایع و بررسی میزان تاثیر گذاری این روش بر نتایج تحقیق را مورد توجه قرار دهند.

فهرست منابع

* رزاقی، سمیه؛ سلمانی، بهزاد، کازرونی، سید علیرضا و فلاحی، فیروز. (۲۰۱۷). "بررسی امکان تشکیل منطقه بهینه پولی در کشورهای عضو سازمان همکاری اسلامی (OIC) با استفاده از تکنیک خوشه بندی فازی و خوشه بندی سنتی".

تر می باشد. با توجه به نتایج تحقیق خوشه بندی به روش های مختلف تاثیر زیادی بر افزایش دقت مدل های تحقیق داشت ولی از بین دو دسته روش مورد بررسی، روش خوشه بندی SOM که در دسته بندی خوشه بندی به روش مدرن قرار دارند تاثیر بیشتری بر افزایش دقت مدل ها داشته اند و بهتر از خوشه بندی به روش سنتی عمل کرده اند. در این راستا شرکت ها و مدیران آنها می توانند با استفاده از روش های مختلف خوشه بندی مخصوصا روش های کلاسیک دقت برآورد های خود از داده های حسابداری را افزایش داده و پیش بینی های دقیق تری را انجام دهند. نتایج این تحقیق با نتایج تحقیقات ایگر و همکاران (۲۰۱۳) و دفوند و جیامبالوو (۱۹۹۴) مطابقت دارد که آنها به این نتیجه رسیدند که برآورد مبتنی بر رگرسیون با استفاده از روش (SOM) مدل بهتری از رگرسیون در هنگام برآورد اقلام تعهدی اختیاری و متغیر های خاص خوشه بندی ارائه می کند. مزیت قابل توجهی که این رویکرد نسبت به رویکرد روش سنتی دارد این است که می توان با استفاده از رگرسیون غیر خطی نیز اجرا کرد.

محدودیت های پژوهش

(۱) به دلیل عدم الزام شرکت ها به تهیه صورت های مالی تعدیل شده بر اساس سطح عمومی قیمت ها، در این تحقیق اطلاعات مالی بدون تعدیل آثار ناشی از تورم استفاده شده است که این موضوع می تواند نتایج حاصل را تحت تاثیر قرار داده باشد.

(۲) در تعمیم نتایج این تحقیق به واحد های اقتصادی خارج از بورس باید احتیاط لازم صورت گیرد زیرا ویژگی های شرکت های خارج از بورس ممکن است با شرکت های پذیرفته شده در بورس متفاوت باشند.

(۳) از آنجایی که برای انتخاب نمونه از میان شرکت های جامعه آماری، ویژگی های خاص که بیشتر به آن اشاره شد مبنا قرار داده شده است، لذا تعمیم نتایج تحقیق به آن گروه از شرکت های پذیرفته شده در بورس اوراق بهادار تهران که

- * Jain.A.K.law.martin.H.C. (2005).data clustering a users dilemma.1st international conference.premi Kolkata india december20-22.
- * Lev, B. (1989), "On the Usefulness of Earnings and Earnings Research: Lessons and Directions from Two Decades of Empirical Research" *Journal of Accounting Research*, Vol. 27: 153-192
- * Lev, B., S. Li., T. Sougiannis (2010), "The Usefulness of Accounting Estimates for Predicting Cash Flows and Earnings" *Review of Accounting Studies*, Vol. 15., 4., 779-807
- * López-Rubio, E., Palomo, E. J., & Ortega-Zamorano, F. (2018). Unsupervised learning by cluster quality optimization. *Information Sciences*, 436, 31-55.*McNicholas, P. D. (2016). Model-based clustering. *Journal of Classification*, 33(3), 331-373.
- * Nawaz, T., Cavallaro, A., & Rinner, B. (2014, October). Trajectory clustering for motion pattern extraction in aerial videos. In *Image Processing (ICIP), 2014 IEEE International Conference on* (pp. 1016-1020).
- * Patrick, K., Hayes, N., & Stephen, J. (2017). Statistical significance for hierarchical clustering. *Biometrics*.
- * Parthasarathy, S., & Ogihara, M. (2000, September). Clustering distributed homogeneous datasets. In *European Conference on Principles of Data Mining and Knowledge Discovery* (pp. 566-574). Springer, Berlin, Heidelberg.
- * Raftery, A. E., & Dean, N. (2006). Variable selection for model-based clustering. *Journal of the American Statistical Association*, 101(473), 168-178.
- * Roychowdhury, S. (2006). Earnings management through real activities manipulation. *Journal of Accounting and Economics*, 42, 335-370
- * Sterehl.a.ghosh.j.2002.cluster ensembles –a knowledge reuse framework for combining multiple partitions.jornal of machine learning research 3(dec):583-617
- * فصلنامه علمی پژوهشی نظریه های کاربردی اقتصاد.
- * شاکری، محمود و عبداللهی، محمود (۱۳۹۴). خوشه بندی داده ها، مروری بر روش های موجود و مقایسه عملکرد آنها. کنفرانس بین المللی پژوهش های کاربردی در فناوری اطلاعات، کامپیوتر و مخابرات. دانشگاه آزاد اسلامی واحد تربت حیدریه.
- * عرب مازار یزدی، محمد و قاسمی، مهسا (۱۳۸۸). "برآورد قیمت عرضه های عمومی اولیه با استفاده از شبکه عصبی مصنوعی". بررسی های حسابداری و حسابرسی. ۱۳۸۸. ص ۷۴-۹۵.
- * معتمنی، همایون؛ بیکی، منظر و جعفر نژاد، عین الله (۱۳۹۰). "مقایسه الگوریتم های خوشه بندی فازی". کنفرانس بین المللی مدل سازی غیر خطی و بهینه سازی، آمل، ۱۳۹۰.
- * Chen, T., Zhang, N. L., Liu, T., Poon, K. M., & Wang, Y. (2012). Model-based multidimensional clustering of categorical data. *Artificial Intelligence*, 176(1), 2246-2269.
- * David.j.HAND.Data mining:statistics and more December 2002.
- * Dechow, P. M., & Dichev, I. D. (2002). The quality of accruals and earnings: The role of accrual estimation errors. *The Accounting Review*, 77, 35-59.
- * DeFond, M. L., & Jiambalvo, J. (1994). Debt covenant violation and manipulation of accruals. *Journal of Accounting and Economics*, 17, 145-176.
- * Ecker, F., Francis, J., Olsson, P., & Schipper, K. (2013). Estimation sample selection for discretionary accruals models. *Journal of Accounting and Economics*, 56, 190-211
- * Flexer, A. (2001). On the use of self-organizing maps for clustering and visualization. *Intelligent Data Analysis*, 5(5), 373-384.
- * Fred.a.lourenco.a.2008.cluster ensemble methods:from single clustering to combined solution studies in computerings intelligence (sci)126.3-30.
- * Han.j.m.kamber.data mining:concepts and techniques morgan kauifmann san Francisco 2000

یادداشت ها

1. Lev
2. Lev&Li
3. David
4. han &kamber
5. jain&martin
6. strehl and ghosh
7. fered and lourenco

8. McNicholas
9. Nawaz et al
10. Raftery & Dean
11. Flexer
12. Parthasarathy & Ogihara
13. Roychowdhury