فصلنامه علمی ـ پژوهشی مطالعات تجربی حسابداری مالی ـ سال ۱۵ ـ شماره ۵۹ ـ پائیز ۱۳۹۷
صفحات ۱۳۵ تا ۱۵۷

خوشه‌بندی و پیش‌بینی سودآوری شرکت‌های پذیرفته شده در بورس اوراق بهادار تهران با رویکرد درخت‌تصمیم C5

محمدرضا مهربان‌پور *
ملیحه حبیب زاده **

چکیده

امروزه سرمایه‌گذاران با توجه به فضای رقابتی حاکم باید محتاط‌تر از قبل تصمیم بگیرند. به این منظور آنها می‌توانند از بانک‌های اطلاعاتی بورس اوراق بهادار استفاده کنند. اما این اطلاعات به تنهایی مثمرثمر نیست، بنابراین لازم است با استفاده از فنون داده‌کاوی تجزیه، تحلیل و تفسیرداده‌ها انجام شود تا اطلاعات قابل‌اتکاتری در اختیار استفاده‌کنندگان قرارگیرد. هدف این پژوهش خوشه‌بندی و پیش‌بینی سودآوری شرکت‌ها و تعیین عوامل موثر بر سودآوری شرکت‌های عضو بورس اوراق بهادار تهران است. جهت این کار، ۸۸۸ شرکت-سال در محدوده زمانی ۱۳۹۵-۱۳۸۷ انتخاب شدند. پس‌از پیش‌پردازش اولیه داده‌ها، با نرم‌افزارهای متلب و Clementine و با استفاده از معیار SSE و روش K-Means شرکت‌ها به ۳ خوشه تبدیل شدند و نتایج این خوشه‌بندی‌ها بوسیله معیار سنجش کیفیت، مورد سنجش قرار گرفت. در ادامه با استفاده از درخت‌تصمیم C5 خوشه‌ها تحلیل و متغیرهای تاثیرگذار بر سودآوری، شناسایی شد. از۳۲ متغیر تحلیل شده تنها ۸ متغیر شامل: سودخالص به کل دارایی ، فروش به کل دارایی، سودخالص به حقوق صاحبان سهام، سود عملیاتی به فروش خالص، سود و زیان انباشته به حقوق صاحبان سهام، سودخالص به فروش خالص، کل بدهی‌ها به کل دارایی‌ها و دارایی‌های جاری به کل دارایی‌ها بر سودآوری شرکت‌ها تاثیر می گذارند. در نهایت با درنظرگرفتن این متغیرها، پیش‌بینی سودآوری شرکت ها طبق هرخوشه انجام شد که دقت پیش‌بینی خوشه‌ها به ترتیب ۸۶/۳۴ درصد، ۸۸/۱۵ درصد و ۸۱/۶۸ درصد است.

واژگان کلیدی: پیش‌بینی، سودآوری، خوشه‌بندی، درخت‌تصمیم C5.

---

* استادیار حسابداری دانشگاه تهران (نویسنده مسئول) mehrabanpour@ut.ac.ir
** کارشناس ارشد حسابداری، مدرس دانشگاه قم

تاریخ دریافت: ۱۳۹۷/۰۱/۲۷ **تاریخ پذیرش: ۱۳۹۷/۰۷/۱۷**

مقدمه

پیش‌بینی آینده در زمینه های مختلف، همواره برای انسان جالب و جذاب بوده است. بااطمینان می‌توان گفت که پیش بینی آینده و روند تغییرات در همه حوزه ها، از دغدغه‌های اصلی و همیشگی مدیران سطح بالا و میانی می‌باشد. از آنجایی که شناخت فرصت‌های مناسب سرمایه‌گذاری در بازار اوراق بهادار، نیازمند تحلیل دقیق صورت‌های مالی شرکت‌ها و پیش‌بینی مناسب سودآتی می باشد، بنابراین تحلیل و پیش‌بینی متغیرهای موثر برسودآوری شرکت‌ها همواره مدنظر سرمایه‌گذاران، کارگزاران و سایر افراد ذی‌نفع در بازار اوراق بهادار می‌باشد (Finger,1994). امروزه فضای رقابتی شدیدی بین شرکت‌های مختلف حاکم شده است؛ بنابراین مدیران باید سریع‌تر و درست‌تر از قبل تصمیم بگیرند. لازمه چنین امری، پیش‌بینی دقیق از سودآوری آتی شرکت‌ها و شناسایی عوامل موفقیت در سودآوری خود می‌باشد. پیش‌بینی سود باید اطلاعاتی را فراهم کند که منطقی و به موقع باشد تا بتواند نیازهای استفاده‌کنندگان را برطرف نماید و به تبع آن، تاثیر خود بر شرکت را به نحو مناسبی نشان دهد (Ying Chan, 2016). میزان اطلاعات افشا شده برای مشارکت کنندگان در بازار سرمایه، بر دقت پیش‌بینی سود تحلیلگران تاثیر می‌گذارد، یافته ها بیانگر آن است که شرکت هایی که اطلاعات بیشتری برای استفاده کنندگان خارجی منتشر می‌کنند، دقت پیش‌بینی سود بیشتری دارند (کردستانی،۱۳۸۹).

اما افزایش حجم اطلاعات درباره شرکت‌ها، عملا کار شناسایی عوامل تاثیرگذار در سودآوری شرکت‌ها به روش رایج و سنتی را سخت و دشوار کرده و مسلما با خطاهای انسانی و مشکلاتی همراه است. از جمله این مشکلات، طولانی بودن فرآیند شناسایی و یا سطحی بودن امر شناسایی یا اعمال سلیقه های کارشناسان می باشد. افزایش این حجم اطلاعات نیز این خطاها را بیش‌از پیش افزایش می‌دهد. بنابراین، استفاده از روش‌هایی که به کمک آن بتوان شناخت بیشتر و ارزیابی دقیق‌تری از سودآوری شرکت‌ها بدست آورد، مهم می‌باشد (افسر و همکاران،۱۳۹۳). برای اینکه شناسایی بهتری در این حجم انبوه از اطلاعات داشته باشیم و نیز جهت کاهش زمان فرآیند، می‌توان با خوشه‌بندی؛ شرکت‌های همگن با ارزش یکسان را در خوشه‌های کوچک‌تر مشابه قرارداد تا حل مساله راحت‌تر شود. خوشه‌بندی درواقع یک روش برای دسته‌بندی داده‌ها است که داده‌ها را با توجه به میزان شباهتشان در دسته‌هایی قرار می‌دهد.

باتوجه به مطالب ذکر شده، مساله مورد نظر این پژوهش این است که آیا می‌توان

بااستفاده‌از اطلاعات مالی، عملکرد و وضعیت سودآوری شرکت‌ها، و به شکل قابل‌اتکایی نتیجه عملکرد آتی آنها را پیش‌بینی نمود؟ در این صورت خوشه بهینه سودآوری شرکت‌های پذیرفته شده در بورس اوراق بهادار چه تعداد خواهد بود؟

مبانی نظری، ادبیات و سوالات

پژوهش هائی که مستقیما حوزه این تحقیق را پوشش دهد، به ندرت یافت می شوند. با این حال در این بخش، جهت آشنائی با موضوع پیش بینی سود و عوامل تاثیرگذار بر آن و روش هائی که جهت این امر پیشنهاد شده است، نزدیک ترین پژوهش های انجام شده در خارج و داخل ارائه و مرور می شوند.

پژوهش‌های خارجی

گراهام و همکاران (۱۹۶۲) در پژوهش خود جهت پیش‌بینی سودآتی، از مطالعه و بررسی متوسط سودهای گذشته در طول زمان استفاده کردند و اعتقاد داشتند که تنها راه پیش‌بینی سود، استفاده از متوسط سودهای گذشته است و این نظر توسط پژوهشگران دیگر نیز مورد توجه قرارگرفت و الگوهای قابل قبولی نیز دراین زمینه ارائه شد. اما محقق دیگری به نام لیو (۱۹۹۳) در بررسی‌های خود دریافت که غیر از سری‌های زمانی سودهای گذشته، اطلاعات دیگری نیز میتواند در پیش‌بینی سودهای آتی موثر باشد، ولی در صورت عدم دسترسی به سایر اطلاعات، سودهای گذشته بهترین پارامتر در پیش‌بینی سودآتی واحد تجاری است. در راستای تحقیق لیو، محقق دیگری به نام فینگر (۱۹۹۴)، به بررسی سود در پیش‌بینی سودها و جریان‌های نقدی آتی پرداخت. در این تحقیق که دوره زمانی سال‌های ۱۹۳۵-۱۹۸۷ را پوشش می‌داد و نمونه‌ای متشکل از ۵۰ شرکت پذیرفته شده در بورس نیویورک را دربرمی‌گرفت، این نتیجه حاصل شد که برای ۸۸ درصد شرکت‌های نمونه، سودهای گذشته را می‌توان به عنوان یک پیش‌بینی‌کننده خوب برای سودهای آتی دانست. از طرف دیگر چاریتو و همکارانش (۲۰۰۰) پژوهشی با عنوان "مربوط بودن ارزش سودها و گردش وجوه نقد" در ژاپن انجام دادند. نتایج پژوهش آنها نشان داد که سودها و گردش وجوه نقد هر دو در پیش‌بینی سودهای آتی موثرند.

توماس و ژانگ (۲۰۰۲) در پژوهشی با ۳۹۳۱۵ مشاهده بین سال های ۱۹۷۰ تا ۱۹۹۷، به بررسی ارتباط بین تغییرات موجودی کالا و بازده های آتی شرکت‌ها پرداختند. نتایج پژوهش آنان نشان داد که ارتباط معکوس بین اقلام تعهدی و بازده‌های غیرعادی آتی

شرکت‌ها، کاملا از تغییرات در موجودی کالا ناشی می‌شود. آنها بیان کردند که موجودی کالا عامل تعیین‌کننده بسیار مهمی برای عملکرد و ارزش شرکت است. در راستای تحقیقات لیو، کاسکی و هانلن (۲۰۰۵) با بررسی رابطه بازده سال جاری سهام و سود آتی برای واحدهای تجاری که در سال جاری سود پرداخت می‌کنند، درمقایسه با واحدهای تجاری که سود پرداخت نمی‌کنند، به این نتیجه رسیدند که رابطه میان سودسال آتی و بازده سال جاری در شرکت‌های توزیع‌کننده سود، رابطه مستقیم و معناداری است.

بینون و همکاران (۲۰۰۸) در کار مطالعاتی خود بااستفاده از متغیرهای حسابداری در شرکت‌های انگلستان و با بکارگیری تکنیک داده‌کاوی به پیش‌بینی طبقه سودآور و غیرسودآور پرداختند که نتایج ارائه‌شده، صحت الگوی اندرس و همکاران را نشان می‌داد. میوجنگ و همکاران (۲۰۱۰)، در بعد دیگری فعالیت کرده و نتایج تحقیق آنها درخصوص شناسایی عوامل تعیین‌کننده برتری صحت نسبی پیش‌بینی سود مدیریت بر پیش‌بینی سود توسط تحلیل‌گران عبارتنداز: ۱.زمانی که کیفیت سود افزایش یابد، برتری صحت نسبی پیش‌بینی سود مدیریت بر پیش‌بینی سود توسط تحلیل‌گران کاهش می‌یابد. ۲.زمانی که پیش‌بینی سود با دشواری همراه باشد، برتری صحت پیش‌بینی سود مدیریت بر پیش‌بینی سود توسط تحلیل‌گران کاهش می‌یابد. در راستای استفاده از نرم‌افزارها برای پیش‌بینی، اگنس و همکاران (۲۰۱۱) با بکارگیری یک سیستم هوشمند نرم‌افزاری، اقدام به طراحی الگو برای پیش‌بینی عملکرد شرکت‌ها نمودند. نتایج کار آنها انعکاس گویای میزان خطای پایین این الگوریتم در این پیش‌بینی‌ها بود. در ادامه تحقیقات اگنس در زمینه پیش‌بینی، گوارا و همکاران (۲۰۱۴) توانایی پیش‌بینی مدل‌های تحلیل تمایزی چندگانه و شبکه‌های عصبی را در پیش‌بینی ورشکستگی شرکت‌ها بررسی کردند. نتایج این پژوهش نشان‌داد هر دو مدل قادر به پیش‌بینی ورشکستگی شرکت‌ها می‌باشد اما شبکه‌های عصبی به طور معنی‌داری از توانایی بالاتری برخورداراست.

چوی، کلی و سدکا (۲۰۱۶) در پژوهشی باعنوان «اخبار سود، سودهای موردانتظار و بازده انباشته سهام» به بررسی رابطه بین تغیرات سود و بازده سهام انباشته پرداختند. نتایج آنان نشان داد بین اخبار سود و بازده سهام همزمان رابطه مستقیمی وجود دارد. همچنین مورامیا و تاکادا (۲۰۱۷) در پژوهشی، عنوان می کنند که عدم دقت در پیش‌بین سود انجام شده توسط مدیریت، می‌تواند ناشی از در دسترس نبودن اطلاعات اولیه موردنیاز پیش‌بینی (ضعیف بودن اطلاعات ورودی مورد نیاز پیش‌بینی) و یا بکارگیری یا تفسیر نادرست این

اطلاعات توسط مدیریت باشد.

پژوهش‌های داخلی

در داخل کشور نیز مطالعاتی در خصوص کمک به پیش بینی بهتر سود شرکت ها انجام شده است. خالقی مقدم و آزاد (۱۳۷۷) دقت پیش‌بینی سود شرکت‌ها که توسط مدیریت اعلام می‌شود را در ۴۵ شرکت بورس اوراق بهادار تهران مورد بررسی قرار دادند. ارتباط چهار متغیر قیمت سهام، اندازه، عمر شرکت و درجه اهرم مالی با دقت پیش‌بینی سود، بااستفاده از رگرسیون های یک و چند متغیره مورد آزمون قرار گرفت. نتایج آزمون فقط ارتباط بین دقت پیش‌بینی و متغیرهای قیمت سهام و اندازه شرکت را تایید نمود. در مطالعه ای دیگر، جنت رستمی (۱۳۷۸) نقش سود در پیش‌بینی جریان های نقدی و سودهای آتی را بررسی کرد. نمونه این تحقیق ۵۱ شرکت پذیرفته شده در بورس اوراق بهادار تهران و روش شناسی آن برگرفته از روش تحقیق فینگر در سال ۱۹۹۴ بود و نتایج نشان داد که برای ۹۲/۱۶ درصد شرکت‌های نمونه، سودهای تاریخی را می‌توان به عنوان یک پیش‌بینی کننده خوب برای سودهای آتی دانست. برخلاف تحقیقات محققینی چون لیو، نوروش و غلام زاده (۱۳۸۲) در بررسی رفتار سود حسابداری با استفاده از سری های زمانی باکس- جنکیز نشان دادند که سودهای گذشته در مورد سودهای آینده اطلاعات چندانی ارائه نمی کند.

در یکی از پژوهش ها با استفاده از اطلاعات دوساله ۸۱ و ۸۲، ۱۲۰ شرکت و ۶ متغیر اقدام به بررسی و پیش‌بینی شرکت‌های موفق و ناموفق کردند. متغیرهای مورد بررسی آنها متشکل از ۳ متغیر مالی و ۳ متغیر غیرمالی بود. در مدل نهایی آنها، تنها دو متغیر بازده حقوق صاحبان سهام و رشد فروش به عنوان متغیرهای توضیح‌دهنده شرکت‌های موفق و ناموفق استفاده شد و سایر متغیرها رابطه مهمی با طبقه‌بندی انجام شده نداشتند (مهرانی و همکاران،۱۳۸۳). در این میان بهرامفر و ساعی (۱۳۸۵) در پژوهشی نشان‌دادند که نسبت های مرتبط با سنجش فعالیت، و نسبت‌های مرتبط با اندازه‌گیری وضعیت بدهی‌ها، اندازه شرکت و نوع صنعت، در پیش‌بینی رتبه عملکرد شرکت‌ها مفید می‌باشند. برخلاف نتایج حاصل از تحقیق نوروش و همکاران، مدرس و عباس زاده (۱۳۸۷) در پژوهشی به این نتیجه رسیدند که سودهای گذشته می‌توانند سودآتی را باکمترین خطای ممکن پیش‌بینی کنند و همچنین ورود یکی از اجزای سود (نقدی و تعهدی) به مدل ها، پیش‌بینی را بهبود می بخشد و در ضمن مشخص شد تاثیر جزء نقدی در مدل های پیش‌بینی از اقلام تعهدی

بیشتر است. در ادامه تحقیقات خالقی مقدم، نمازی و شمس الدینی (۱۳۸۷) در بررسی سازه های موثر بر دقت پیش‌بینی سود، بااستفاده از نمونه ۷۷تایی از شرکت‌ها در طی سال های ۱۳۷۹ تا ۱۳۸۳ به این نتیجه رسیدند که بین رشد سود، رشد فروش، رشد دارایی ها، سود پیش‌بینی شده گذشته، اهرم مالی، قیمت سهام و دقت پیش‌بینی سود، رابطه وجود دارد؛ اما بین سود سهام پرداختی و اندازه شرکت با دقت پیش‌بینی سود، رابطه‌ای نیست. همچنین یکی از پژوهش ها برای بررسی عوامل موثر بر میزان دقت پیش‌بینی سود شرکت، ۵ عامل افق زمانی پیش‌بینی، دفعات تجدیدنظر، نوع اظهارنظر حسابرس، نوع صنعت و اندازه شرکت را مورد توجه قرار داد. نتایج تحقیق حاکی از آن بود که از بین عوامل گفته شده، افق زمانی و نوع صنعت بر دقت پیش‌بینی سود شرکت‌ها موثرند و بین سایر عوامل با دقت پیش‌بینی سود رابطه معناداری مشاهده نشد(حقیقت وهمکاران،۱۳۹۰).

اعتمادی و همکاران (۱۳۹۱) با بکارگیری شبکه عصبی مصنوعی بااستفاده از اطلاعات هفت ساله ۸۰ - ۸۶، ۹۰ شرکت و ۴۲ متغیر اقدام به بررسی و پیش‌بینی سودآوری شرکت‌های بورس اوراق بهادار تهران کردند. در مدل نهایی آنها، تنها ۹ متغیر به عنوان توانمندترین متغیرها در تشخیص و تفکیک دوگروه شرکت‌های سودآور و زیان آور استفاده شده است. نتایج کار آنها با ۸۶ درصد صحت نشان از کارایی مناسب این تکنیک داشت. در پژوهش دیگری حاکی از تاثیرات پیش‌بینی سود، حسینی نسب و همکاران (۱۳۹۴) به بررسی واکنش بازار پرداختند. هدف آن ها بررسی واکنش بازار به اخبار خوب یا بد شرکت‌ها با توجه به نوع پیش‌بینی سود سال قبل آن ها، تمایل مدیران به حفظ یا اصلاح شهرت خود در پیش‌بینی و مقایسه تمایل مدیران به مدیریت افزایشی سود از طریق انواع مختلف مدیریت سود با توجه به نوع پیش‌بینی سال قبل آن ها بود. آن ها صورت های مالی ۸۶ شرکت را طی سال های ۱۳۸۵-۱۳۹۱ مورد تجزیه و تحلیل قرار دادند. نتایج پژوهش بیانگر آن بود که واکنش بازار به پیش‌بینی سود با محتوای اخبار خوب (بد)، زمانی که پیش‌بینی دوره قبل بدبینانه (خوشبینانه) است، مثبت تر (منفی‌تر) از حالتی است که پیش‌بینی دوره قبل خوشبینانه (بدبینانه) است. همچنین مدیران در پیش‌بینی های خود از ثبات رفتاری برخوردار می باشند. علاوه براین، نتایج حاکی از آن بود که پیش‌بینی خوشبینانه در دوره قبل می تواند عاملی جهت مدیریت افزایشی سود دوره جاری باشد. رحمانی و حیاتی (۱۳۹۵) با بررسی تاثیر دقت برآوردی پیش‌بینی سود مدیران بر رانش پس از اعلان سود نشان دادند که ویژگی های پیش بینی سود به طور معناداری می‌تواند

دقت واقعی پیش‌بینی سود مدیران را توضیح دهد. از طرفی سودغیرمنتظره نمی‌تواند بازده‌های غیرعادی تعدیل شده را تبیین کند. همچنین مهربان‌پور و همکاران (۱۳۹۶) نیز بادرنظرگرفتن ۱۱متغیر، اقدام به شناسایی عوامل موثر برسودآوری کردند. در پژوهش آنها، متغیر بازده حقوق صاحبان سهام به عنوان معیار سودآوری مطرح شد و سپس بااستفاده از اطلاعات مالی ۱۰ ساله بانک‌ها؛ ساختاردارایی، تنوع درآمدی، رشد اقتصادی و تورم به عنوان متغیرهای موثر بر سودآوری شناسایی شدند.

علاوه بر موارد بیان شده در فوق، تحقیقات انجام شده در زمینه خوشه‌بندی را می توان به صورت خلاصه در جدول ۱ مشاهده نمود:

جدول۱. تحقیقات انجام گرفته در زمینه خوشه‌بندی

محقق	سال	پژوهش
فیشر	۱۹۳۶	ایده جداسازی گروه ها در یک جمعیت
آلتمن	۱۹۳۸	توسعه اولین سیستم ارزیابی تقاضانامه های اعتباری را با بکارگیری ۵ معیار
مایرزو هنرجی	۱۹۶۳	ارائه آنالیز ممیزی چند متغیره برای رتبه بندی اعتباری مشتریان
مور و کلن	۱۹۶۸	رتبه بندی شرکت‌ها بااستفاده از روش آنالیز ممیزی چند متغیره
دسایی و همکاران	۱۹۹۸	طبقه بندی مشتریان وام های بین المملی در سه کشور آمریکا، المان و استرالیا
وست	۲۰۰۰	تقسیم مشتریان را به دودسته خوب و بد بااستفاده از شبکه عصبی
هسیه	۲۰۰۴	طبقه بندی مشتریان بانک به سه گروه عمده سودآور بااستفاده از شبکه عصبی

از جمله مقالات معدودی که در حوزه بورس برای تقسیم بندی انجام شد، تحقیقی است که شین و سون (۲۰۰۴) انجام داده‌اند. در این تحقیق بااستفاده از دو روش K-means و SOM توانستند نرخ کمیسیون کارگزاری را برای مشتریان پیشنهاد دهند. نتایج حاصل از تجزیه و تحلیل آنها نشان داد که روش k-means قویترین روش برای تقسیم بندی است. به همین منظور در این پژوهش نیز از این روش برای خوشه‌بندی شرکت‌های مذکور استفاده شده است.

پژوهش حاضر، فاقد فرضیه می باشد. در عوض سوالاتی به شرح زیر طرح شده که در بخش های بعدی به آنها جواب داده شده است: ۱. آیا با خوشه‌بندی و پیش‌بینی سودآوری شرکت‌ها می توان به شکل قابل‌اتکایی، نتیجه عملکرد آتی آنها را پیش‌بینی نمود؟ ۲. خوشه بهینه سودآوری شرکت‌های پذیرفته شده در بورس اوراق بهادار چه تعداد است؟

۳. عوامل موثر بر سودآوری شرکت‌های عضو بورس اوراق بهادار تهران کدام هستند؟

روش تحقیق

داده‌کاوی به بررسی و تجزیه تحلیل مقدار عظیمی از داده ها به منظور کشف الگوها و قوانین معنی دار اطلاق می گردد. داده‌کاوی در دو نوع ظاهر می شود: داده‌کاوی هدایت شده و داده‌کاوی هدایت نشده. در داده‌کاوی هدایت شده هدف دسته بندی اطلاعات براساس برخی پارامترها مشخص می باشد اما در داده‌کاوی هدایت نشده هدف یافتن الگوها یا تشابهات بین گروه هایی از اطلاعات، بدون استفاده از هیچگونه پیش زمینه ای در مورد اطلاعات می باشد. از نمونه روش‌های داده‌کاوی هدایت نشده و هدایت شده می توان به خوشه‌بندی و دسته بندی اشاره نمود. خوشه‌بندی، به عمل تقسیم جمعیت ناهمگن به تعدادی از زیر مجموعه ها یا گروه های همگن گفته می شود. در دسته بندی هر داده به دسته ای از پیش تعیین شده براساس دانش قبلی اختصاص می یابد؛ اما در خوشه‌بندی هیچ دسته از پیش تعیین شده ای وجود ندارد. در واقع خوشه‌بندی راهی برای یافتن ساختار داده های پیچیده فراهم می کند (باقرزاده وهمکاران،۱۳۸۷). خوشه‌بندی در واقع تقسیم کردن داده ها به گروه های مشابه است و هر گروه را اصطلاحا یک خوشه می نامند. داده‌ها در هر خوشه به یکدیگر شبیه و در عین حال با داده های موجود در خوشه‌های دیگر متفاوت اند (Abraham,2006). به همین دلیل در این پژوهش در مرحله اول، خوشه‌بندی داده ها انجام می‌گیرد.

روش‌های متعددی برای خوشه‌بندی وجود دارد. یکی از الگوریتم های رایج خوشه‌بندی، الگوریتم K-means است. روش K-Means با وجود سادگی آن، یک روش پایه برای بسیاری از روش‌های خوشه‌بندی دیگر محسوب می شود (Alpaydin, 2004). ولی تعیین تعداد خوشه بهینه در ابتدای امر از معایب آن می باشد و نتایج به انتخاب اولیه خوشه‌ها وابسته است. بنابراین می توان گفت که تعیین تعداد خوشه‌ها از اهمیت زیادی برخوردار بوده و برنتیجه کار تاثیر خواهد گذاشت. از این رو، برای تعیین تعداد خوشه بهینه از روشی به نام SSE استفاده شده است. دراین روش، نخست مراکز خوشه درنظر گرفته می شود و سپس فاصله نقطه مورد نظر از مراکز خوشه محاسبه می گردد. خوشه ای که SSE پایین تر دارد، نشان دهنده بهترین خوشه‌بندی (تعداد خوشه‌ها) است (مطیعیان ونعیمی،۱۳۹۲). برای این امر از رابطه ۱ استفاده می شود:

$$SSE = \sum_{i=1}^{k} \sum_{p \in c_i} d(P, m_i)$$ رابطه ۱

الگوریتم K-Means به شرح زیر می باشد:

۱. انتخاب K داده به عنوان مراکز خوشه

۲. تعیین فواصل بقیه داده ها با مراکز خوشه‌ها

۳. قرار گیری داده هایی که به مرکز هر خوشه نزدیک ترند در آن خوشه

۴. محاسبه میانگین هر خوشه به عنوان مرکز جدید خوشه

۵. تکرار مرحله دوم تا چهارم تا رسیدن با عدم تغییر در خوشه‌ها (دهقان وهمکاران،۱۳۹۲).

معیار سنجش کیفیت خوشه ها

همانطور که اشاره شد، گونه ای از خوشه‌بندی بهتر است که فاصله درون خوشه‌ها در آن کمترین و فاصله بین خوشه‌ها بیشترین باشد. برای مقایسه خوشه‌بندی‌های انجام شده به کمک الگوریتم‌های خوشه‌بندی، این معیارسنجش کیفیت مطرح شده است:

اگر $O = \{c^n | n = 1, \dots, k\}$ مجموعه مراکز خوشه‌ها و c^n مراکز خوشه ها باشد و $O^n = \{c_i | i = 1, \dots., |T^c - O|\}$ مجموعه باقیمانده شرکت ها که به عنوان مرکز خوشه انتخاب نشده‌اند و T^c مجموعه کلیه شرکت‌هایی که خوشه‌بندی روی آنها صورت گرفته‌است؛ باشد، کیفیت نتایج خوشه‌بندی با K خوشه می‌تواند بصورت رابطه ۲ تعریف شود:

(۲)
$$\rho(k) = \frac{1}{k} \sum_{n=1}^{k} \left(\max \left\{ \frac{\gamma_n + \gamma_m}{\delta_{nm}} \right\}\right)$$

(۳)
$$\gamma_m = \frac{1}{||O^m||} \sum_{c_j \in o^m} dist(c_j, c^m)$$

(۴)
$$\gamma_n = \frac{1}{||O^n||} \sum_{c_i \in o^n} dist(c_i, c^n)$$

(۵)
$$\delta_{nm} = dist(c^n, c^m)$$

معادله (3)، γ_m را به عنوان میانگین فاصله بین مرکز خوشه c^m و همه شرکت ها خوشه O^m را تعریف می‌کند. معادله (4) نیز توضیحاتی مشابه معادله مذکور دارد. معادله (۵) δ_{nm} را به عنوان فاصله c^m و c^n تعریف می‌کند. معیار فوق هرچه کمتر باشد تراکم

درون خوشه‌ای بیشتر و فاصله بین خوشه‌ای بیشتر است، درنتیجه کیفیت خوشه بندی صورت گرفته بهتر است (Tsai and chiu, 2004).

برای اجرای این تحقیق، پس از پالایش و آماده سازی داده ها، برای آن که بتوان سودآوری شرکت‌ها را به صورت دقیق تری پیش‌بینی کرد، از سیستم های هوشمند و تکنیک درخت‌تصمیمC5 استفاده شد. این روش به خاطر قابلیت برتر این روش پیش‌بینی، نسبت به انواع مسائل مدل سازی پیش‌بینی و محبوبیت در انتشارات اخیر ادبیات داده‌کاوی انتخاب شده است.

درخت‌تصمیم به صورت بازگشتی، مشاهدات مجزا را در یک شاخه قرار می‌دهد تا یک درخت به منظور دستیابی به بالاترین دقت پیش‌بینی ممکن ساخته شود. در انجام این کار، الگوریتم های ریاضی مختلفی استفاده می شود تا یک ویژگی و حد آستانه مربوط برای آن ویژگی شناسایی شود، تا مخزن مشاهدات به دو یا چند زیرگروه تقسیم شود. این مرحله تا گره برگ تکرار می شود تا درخت کامل ساخته شود. تعداد شاخه برای هر گره ایجاد شده به الگوریتم خاص استفاده شده و تعداد مقادیر ویژگی انتخاب شده وابسته است. در نهایت باتوجه به امکانات موجود در نرم‌افزارهای Excel و Clementine به پیش‌بینی سودآوری شرکت‌های مذکور پرداخته می شود.

متغیرها

با توجه به تحقیقات پیشین، نسبت ها و متغیرهای مربوطه و تاثیرگذار بر موفقیت شرکت ها و سودآوری یا ورشکستگی آنها در جدول ۲ از کار تحقیقاتی فرج زاده دهکردی (۱۳۸۶) مستخرج شده است.

جدول۲. متغیرها و نسبت های مورد استفاده در پژوهش های پیشین

محقق/ محققان	سال	نسبت ها و متغیرهای مورد استفاده
بیور	۱۹۶۶	گردش وجوه نقد به کل بدهی ها، سودخالص به کل دارایی ها، کل بدهی ها به کل دارایی‌ها، سرمایه در گردش به کل دارایی ها، دارایی‌های جاری به بدهی های جاری
بیور	۱۹۶۸	فروش، سودخالص، گردش وجوه نقد، اوراق قرضه قابل معامله، حساب های دریافتنی، دارایی های آنی، موجودی کالا، دارایی های جاری، سرمایه در گردش، کل دارایی‌ها، بدهی‌های جاری، جمع بدهی ها و سهام ممتاز، ارزش ویژه، وجوه نقد به کل بدهی ها، سودخالص به کل دارایی ها، کل بدهی ها به کل دارایی ها، دارایی های جاری به کل دارایی ها، دارایی های آنی به کل دارایی ها، سرمایه در گردش به کل دارایی‌ها، دارایی‌های آنی به بدهی‌های جاری، وجوه نقد به بدهی های جاری، دارایی های جاری به فروش، دارایی های آنی به فروش، سرمایه در گردش به فروش، وجوه نقد به فروش
آلتمن	۱۹۶۸	سرمایه درگردش به کل دارایی ها، سود (زیان) انباشته به کل دارایی ها، سود قبل از بهره و مالیات به کل دارایی‌ها، ارزش بازار حقوق صاحبان سهام به ارزش دفتری بدهی‌ها، فروش به کل دارایی‌ها
ویلکاکس	۱۹۷۱	الگوی تصادفی از توابعی بر مبنای اقلام صورت های مالی
ادمیستر	۱۹۷۲	جریان وجوه سالانه به بدهی های جاری، حقوق صاحبان سهام به فروش، سرمایه در گردش به فروش، بدهی های جاری به حقوق صاحبان سهام، موجودی کالا به فروش
دیکن	۱۹۷۲	نگاه کنید به بیور۱۹۶۸
منساح	۱۹۸۴	گردش وجوه نقد، نقدینگی، اهرم مالی، گردش حقوق صاحبان سهام
کیسی و بارتجک	۱۹۸۴	گردش وجوه نقد عملیاتی
کیسی و بارتجک	۱۹۸۵	گردش وجوه نقد عملیاتی، گردش وجوه نقد عملیاتی به بدهی های جاری، گردش وجوه نقد عملیاتی به کل بدهی ها
گنتری، نیوبلد، ویتنفورد	۱۹۸۵	وجوه حاصل از عملیات، سرمایه در گردش، تامین مالی، هزینه های ثابت، مخارج سرمایه‌ای، سود سهام، سایر جریان های دارایی‌ها و بدهی‌ها، تغیر در اوراق بهادار نقدی و قابل معامله
عزیز، امانوئل، لاوسن	۱۹۸۸	گردش وجوه نقد عملیاتی، خالص سرمایه‌گذاری ثابت، مالیات پرداختنی، تغیرات نقدینگی، بهره پرداختنی، بدهی های میان مدت یا بلندمدت افزایش یافته یا بازپرداخت شده، سودسهام نقدی، حقوق صاحبان سهام افزایش یافته یا بازپرداخت شده
عزیز و لاوسن	۱۹۸۹	گردش وجوه نقد عملیاتی، خالص سرمایه‌گذاری ثابت، مالیات پرداختنی، تغیرات نقدینگی، بهره پرداختنی، بدهی های میان مدت یا بلندمدت افزایش یافته یا بازپرداخت شده، سودسهام نقدی، حقوق صاحبان سهام افزایش یافته یا بازپرداخت شده
گیلبرت، منون و	۱۹۹۰	وجوه نقد به کل دارایی ها، گردش وجوه نقد عملیاتی به بدهی های جاری، گردش وجوه نقد عملیاتی به کل بدهی ها، گردش وجوه نقد عملیاتی به کل دارایی ها، دارایی‌های جاری به بدهی‌های جاری، دارایی های جاری به کل دارایی‌ها، سودقبل از بهره و مالیات به کل

اسچوارتز		دارایی‌ها، حقوق صاحبان سهام به کل بدهی ها، سودخالص به کل دارایی ها، سودانباشته به کل دارایی ها، فروش به دارایی های جاری، فروش به کل دارایی ها، سرمایه در گردش به کل دارایی‌ها
پلات و پلات	۱۹۹۰	گردش وجوه نقد به کل دارایی ها، سودخالص به کل دارایی ها، سودخالص به ارزش ویژه، سود عملیاتی به کل دارایی ها، گردش وجوه نقد به فروش، سود خالص به فروش، دارایی‌های جاری به کل دارایی ها، فروش به کل دارایی ها، کل بدهی ها به کل دارایی‌ها، کل بدهی‌ها به ارزش ویژه، بدهی های جاری به کل بدهی ها، سرمایه سهامداران عادی به ارزش ویژه، فروش به سرمایه در گردش، دارایی های جاری به فروش، حساب های دریافتنی به موجودی کالا، حساب های دریافتنی به فروش، حساب های دریافتنی و موجودی کالا به کل دارایی‌ها، دارایی های جاری به بدهی‌های جاری، خالص دارایی‌های ثابت به کل دارایی‌ها، وجه نقد به بدهی های جاری، وجه نقد به کل دارایی‌ها، درصد تغییرات در فروش، سود قبل از بهره و مالیات به کل هزینه بهره، گردش وجوه نقد به کل هزینه بهره، درصد تغییر در بازده دارایی‌ها، گردش وجوه نقد به بدهی ها
فوستر و وارد	۱۹۹۷	گردش وجوه نقد حاصل از فعالیت های عملیاتی، گردش وجوه نقد حاصل از فعالیت های تامین مالی، گردش وجوه نقد حاصل از سرمایه‌گذاری
سانگ، چانگ و لی	۱۹۹۹	نرخ رشد کل دارایی ها، نرخ رشد اموال و ماشین آلات و تجهیزات، نرخ رشد دارایی های جاری، نرخ رشد فروش، نرخ رشد سودخالص، سوداناخالص به فروش خالص، سود عملیاتی به فروش خالص، سود عملیاتی به کل دارایی ها، سود عادی به کل دارایی‌ها، سود خالص به کل دارایی ها، سود عادی به حقوق صاحبان سهام، سودخالص به حقوق صاحبان سهام، نسبت سود سهام، سود سهام به سود خالص، سود هر سهم، گردش وجوه نقد هر سهم، حقوق صاحبان سهام به کل دارایی ها، نسبت دارایی های ثابت به حقوق صاحبان سهام و بدهی های بلند مدت، نسبت جاری، نسبت آنی، نسبت بدهی، نسبت پوشش بدهی، گردش وجوه نقد به بدهی ها، گردش وجوه نقد به کل دارایی‌ها، نسبت گردش کل دارایی ها، نسبت گردش حقوق صاحبان سهام، نسبت گردش موجودی کالا، میانگین دوره گردش موجودی کالا، میانگین دوره جمع آوری حساب های دریافتنی، ارزش افزوده ناخالص، بهره وری نیروی کار، سود عادی هریک از کارکنان، کل دارایی‌های هر ریال سرمایه، بهره وری سرمایه، ارزش افزوده ناخالص به فروش خالص
شاه و مورتازا	۲۰۰۰	دارایی های جاری به بدهی های جاری، فروش به وجه نقد، گردش حساب های دریافتنی، بهره کسب شده، بدهی به حقوق صاحبان سهام، کل دارایی ها به حقوق صاحبان سهام، سود خالص به فروش خالص، فروش خالص به حقوق صاحبان سهام عادی
گریس و اینگرام	**2001**	نگاه کنید به آلتمن ۱۹۹۶
شین و لی	۲۰۰۲	نسبت سریع، بدهی جاری به کل دارایی، هزینه های مالی به فروش، نسبت‌های‌نقدینگی، درآمد خالص به حقوق صاحبان سهام، درآمد عملیاتی به هزینه عملیاتی، سودانباشته به کل دارایی، حقوق صاحبان سهام به کل دارایی، ارزش افزوده به

		کل هزینه
سیلن	۲۰۰۴	جریان نقد به دارایی های در گردش محدود شده، نسبت سرمایه (سرمایه به کل دارایی ها)، مالیات منقضی شده، سود انباشته به کل دارایی ها، موجودی کالا، درآمد ناخالص، پوشش بدهی، درآمد خالص، نسبت سریع، نسبت بدهی
اندرس	۲۰۰۵	هزینه بدهی، وضعیت بدهی، رشد، بدهکاری، سهم هزینه نیروی انسانی، نقدینگی کوتاه مدت و گردش دارایی ها

متغیروابسته در این پژوهش، سودآوری (زیان آوری) شرکت‌ها در سال آینده است. منظور از این دو واژه مورد تاکید در این تحقیق، موارد زیر است:

۱. سودآوری: شرکتی که خالص سود و زیان پس از کسر مالیات آن در سال مالی آتی مثبت باشد.

۲. زیان آوری: شرکتی که خالص سود و زیان پس از کسر مالیات آن در سال مالی آتی منفی باشد (اعتمادی وهمکاران،۱۳۹۲).

متغیرهای مستقل با توجه به کارهای تحقیقاتی قبلی (متغیرهایی که قابل محاسبه در محیط ایران بوده و برای پیش‌بینی موفقیت شرکت‌ها طی تحقیقات پیشین به کار رفته اند) در پیش‌بینی سودآوری (یا زیان آوری) شرکت‌ها استفاده شده است. فرج زاده دهکردی (۱۳۸۶) پس از بررسی مبسوط تحقیقات پیشین، متغیرهائی را که قابل محاسبه در محیط ایران هستند، برای پیش‌بینی ورشکستگی به بوته آزمایش گذاشت. این پژوهش نیز توانمندی هر یک از ۳۲ متغیر را که طی پژوهش‌های گذشته در پیش‌بینی های عملکرد از جمله ورشکستگی و موفقیت شرکت‌ها به کار رفته بودند، بررسی نمود. این متغیرها در جدول شماره ۳ آمده است.

جدول۳. متغیرهای مستقل مورد استفاده در پژوهش

ردیف	متغیر	ردیف	متغیر
۱	حقوق صاحبان سهام به بدهی بلند مدت	۱۷	بدهی جاری به کل دارایی ها
۲	سود عملیاتی به فروش خالص	۱۸	فروش به دارایی‌های جاری
۳	کل بدهی به کل حقوق صاحبان سهام	۱۹	سودخالص به کل دارایی‌ها
۴	کل بدهی به کل دارایی	۲۰	سود خالص به حقوق صاحبان سهام
۵	دارایی‌های جاری به کل دارایی‌ها	۲۱	فروش به حقوق صاحبان سهام
۶	سودعملیاتی به کل دارایی‌ها	۲۲	فروش به کل دارایی‌ها

۷	سود خالص به فروش خالص	۲۳	دارایی های آنی به بدهی جاری
۸	موجودی کالابه کل دارایی‌ها	۲۴	دارایی ثابت به کل دارایی ها
۹	بدهی به ارزش ویژه	۲۵	فروش به وجه نقد
۱۰	وجه نقد به بدهی جاری	۲۶	سود خالص به دارایی ثابت
۱۱	دارایی های آنی به کل دارایی ها	۲۷	سود و زیان انباشته به کل دارایی‌ها
۱۲	حساب های دریافتنی به موجودی کالا	۲۸	فروش به حساب های دریافتنی
۱۳	وجه نقد به کل دارایی ها	۲۹	سود و زیان انباشته به حقوق صاحبان سهام
۱۴	حقوق صاحبان سهام به کل دارایی	۳۰	دارایی جاری به بدهی جاری
۱۵	هزینه بهره به سود ناخالص	۳۱	بدهی جاری به حقوق صاحبان سهام
۱۶	سود ناخالص به فروش	۳۲	بدهی جاری به کل بدهی

جامعه و نمونه آماری

جامعه آماری مورد استفاده در این تحقیق، داده های مربوط به اطلاعات شرکت‌های پذیرفته شده در بورس اوراق بهادار تهران در قلمرو زمانی ۱۳۸۷ تا ۱۳۹۵ می باشد. جامعه آماری شامل تمام شرکت‌های پذیرفته شده در بورس اوراق بهادار تهران می باشند که حائز معیارهای زیر باشند:

۱. شرکت‌هایی که طی سالهای ۸۷ تا۹۵ عضو بورس اوراق بهادار بوده اند.

۲. در تمامی۹ سال مذکور به بورس اوراق بهادار تهران صورت مالی ارائه کرده باشند.

۳. اطلاعات آنها در دسترس باشد.

۴. سال مالی آنها منتهی به پایان اسفندماه باشد.

پس از تشکیل جامعه آماری و تعیین اعضای واجد شرایط، ۱۸۶ شرکت شرایط ذکر شده را داشتند. در تحقیق حاضر به دلیل استفاده از داده‌کاوی و لزوم داشتن هرچه بیش‌تر تعداد نمونه‌ها در این روش، از نمونه‌گیری استفاده نشد و تمام داده‌های موجود به شرح فوق به عنوان داده‌های این تحقیق به کار برده شده‌اند.

یافته های تحقیق

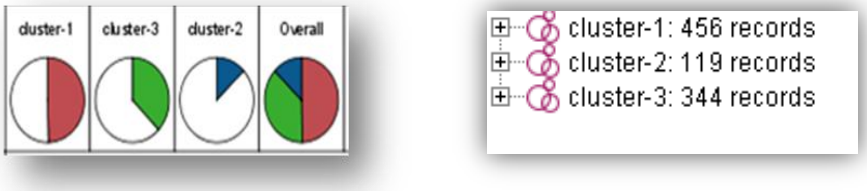
آماده سازی داده ها یکی از مراحل اصلی است که داده ها را برای مراحل دیگر آماده می کند. بنابراین ابتدا متغیرهای تحقیق از بانک اطلاعاتی بورس اوراق بهادار تهران استخراج شد. روشهای متعددی برای خوشه‌بندی وجود دارد. در این مطالعه روش K-Means

جهت خوشه‌بندی شرکت‌ها بکار گرفته شد. روش K-Means باوجود سادگی آن یک روش پایه برای بسیاری از روش‌های خوشه‌بندی دیگر است (Alpaydin, 2004). ولی تعیین تعداد خوشه بهینه در ابتدای امر از معایب آن می باشد و نتایج به انتخاب اولیه خوشه‌ها وابسته است، بنابراین می توان گفت که تعیین تعداد خوشه‌ها از اهمیت زیادی برخوردار بوده و برنتیجه کار تاثیر خواهد گذاشت. از این رو برای تعیین تعداد خوشه بهینه از روشی به نام SSE استفاده شد. خوشه ای که SSE پایین تر دارد، نشان دهنده بهترین خوشه‌بندی است. برای محاسبه SSE از نرم‌افزار متلب نسخه ۲۰۱۵ استفاده شد که نتایج طبق جدول ۴ ارائه شده است.

جدول۴. تعداد خوشه‌ها در K-Means

تعداد خوشه‌ها	SSE
۱	۷,۸۹۲۹
۲	۲,۴۵۳۹
۳	۱,۴۳۵۹
۴	۹,۹۱۲۳

با توجه به نتیجه داده های جدول بالا، تعداد خوشه بهینه K=3 می باشد و همچنین در شکل ۱ تعداد ثبت های موجود در هر خوشه مشاهده می‌شوند که بیشترین شرکت‌ها در خوشه ۱ قرار گرفته اند.



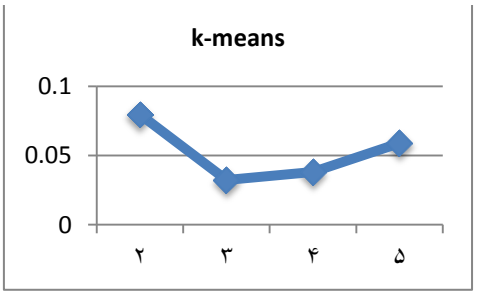
شکل۱. مقادیر ثبت ها در هر خوشه

پس از محاسبه متغیرها، بااستفاده از الگوریتم K-means شرکت ها خوشه‌بندی شدند. برای سنجش کیفیت خوشه‌بندی از معیار مذکور برای خوشه‌بندی بین ۲ تا ۵ خوشه محاسبه شده‌است. اکنون نتایج حاصله بررسی می شوند:

معیار: باتوجه به رابطه(۲)، کیفیت نتایج خوشه‌بندی بااستفاده از الگوریتم K-mean در جدول (۵) نشان داده شده است. همانطور که پیشتر اشاره شد، کمتر بودن معیار ذکرشده نشان می‌دهد که خوشه‌های تشکیل شده، کیفیت بهتری دارند. همانطور که در نمودار (۱) مشاهده می‌شود، کیفیت خوشه‌های تشکیل شده بااستفاده از این الگوریتم مطلوب است و همچنین خوشه بندی با سه خوشه بهترین کیفیت را دارد.

جدول۵. الگوریتم خوشه بندی براساس معیار کیفیت

تعداد خوشه	k-means
۲	0/079767389
۳	0/032394225
۴	0/038239037
۵	0/058896799



نمودار۱. کیفیت خوشه بندی الگوریتمk-means بااستفاده از معیارکیفیت

در این مرحله بااستفاده از درخت تصمیم به تحلیل بخش های تشکیل شده توسط الگوریتم خوشه بندیk-means پرداخته شده است، که وضعیت خوشه‌ها به شرح زیر است:

خوشه ۱: شامل شرکت‌هایی می باشند که دارای خصوصیات زیر هستند:

نسبت دارایی جاری به کل دارایی آنها قابل ملاحظه و نسبت بدهی جاری به کل دارایی آنها اندک بوده و از طرفی نسبت فروش به کل دارایی آنها زیاد بوده است.

خوشه ۲: شامل شرکت‌هایی می باشند که همگی نسبت کل بدهی به کل حقوق صاحبان

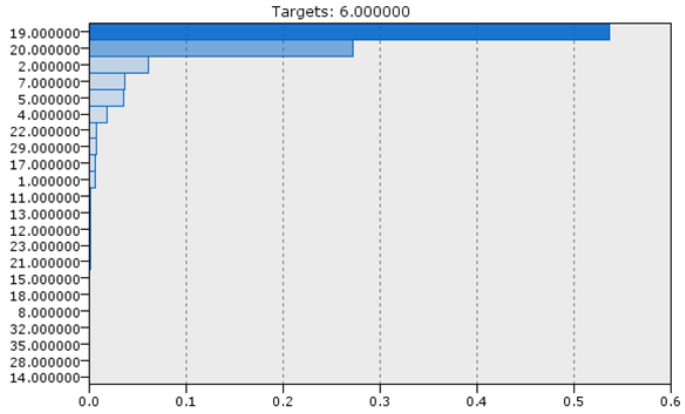
سهام آنها در حد متوسط می باشد.

خوشه ۳: شامل شرکت‌هایی هستند که نسبت سودخالص به فروش خالص آنها بسیار بالا بوده و همچنین نسبت کل بدهی به کل حقوق صاحبان سهام آنها نیز بسیار ناچیز است، ولی در عین حال نسبت بدهی جاری به کل دارایی‌های آنها بالا است.

در نهایت بااجرای درخت‌تصمیم، طبق شکل ۲ از بین ۳۲ متغیر، ۸ متغیر طبق جدول ۶ به عنوان متغیرهای تاثیرگذار بر سودآوری شرکت‌ها شناخته شدند که بدین شرح هستند:

جدول۶. متغیرهای تاثیرگذار بر سودآوری

۱. سودخالص به کل دارایی‌ها	۵. دارایی جاری به کل دارایی‌ها
۲. فروش به کل دارایی‌ها	۶. سودخالص به حقوق صاحبان سهام
۳. سود عملیاتی به فروش خالص	۷. سود و زیان انباشته به حقوق صاحبان سهام
۴. سودخالص به فروش خالص	۸. کل بدهی ها به کل دارایی ها



شکل۲. متغیرهای تاثیر گذار بر سودآوری

پس از شناسایی متغیرهای تاثیرگذار بر سودآوری شرکت‌ها به عنوان ورودی درخت‌تصمیم C5 به پیش‌بینی سودآوری آنها در سال ۹۶ در هر خوشه پرداخته شد. اما از آنجایی که یکی از محدودیت های درخت‌تصمیم در نرم‌افزار Clementine این است که نمی تواند مستقیما از داده های استخراج شده از پایگاه داده بورس اوراق بهادار تهران استفاده کند، بنابراین نیاز به تبدیل این داده ها وجود دارد که براساس طیف لیکرت داده ها