تحقیقات مالی
Journal of Financial Research
DOI: 10.22059/jfr.2017.102470.1005775
دانشکدهٔ مدیریت دانشگاه تهران
دورهٔ ۱۹، شمارهٔ ۱
بهار ۱۳۹۶
صص. ۱۵۶- ۱۳۹

کاربرد روش ترکیبی ماشین بردار پشتیبان و انتخاب ویژگی به‌منظور پیش‌بینی درماندگی مالی شرکت‌های پذیرفته شده در بورس اوراق بهادار تهران

سعید فلاح‌پور[۱]، عیسی نوروزی‌یان لکوان[۲]، محمد هندیجانی‌زاده[۳]

چکیده: پیش‌بینی درماندگی مالی از مسائل مهمی است که همواره پژوهشگران، مؤسسه‌های اعتباری و بانک‌ها به آن توجه کرده‌اند. تاکنون تحقیقات بسیاری در این زمینه صورت گرفته است، ولی استفاده از مدل‌های ترکیب‌شدهٔ انتخاب ویژگی و مدل طبقه‌بندی‌کننده، از مسائلی است که فقط در سال‌های اخیر توجه پژوهشگران را به خود جلب کرده است. در این مقاله ماشین بردار پشتیبان با چهار تابع کرنل خطی، چندجمله‌ای، شعاعی و سیگموید به‌عنوال مدل طبقه‌بندی‌کننده و ترکیب آن با روش‌های انتخاب ویژگی فیلترکننده و پوشش‌دهنده استفاده شده است. همچنین از الگوریتم ژنتیک که یکی از انواع روش‌های پوشش‌دهندهٔ انتخاب ویژگی است و روش‌های آنالیز اجزای اساسی، زنجیرهٔ اطلاعات و رلیف که جزء روش‌های فیلترکنندهٔ انتخاب ویژگی هستند، استفاده شده است. نتایج به‌دست آمده نشان داد که روش الگوریتم ژنتیک نسبت به روش‌های فیلترکننده، عملکرد بهتری دارد. همچنین دقت ماشین بردار پشتیبان با توابع کرنل خطی، چندجمله‌ای، شعاعی و سیگموید در ترکیب با الگوریتم ژنتیک، با سطح اطمینان ۹۵ درصد تفاوت معناداری با هم ندارند.

واژه‌های کلیدی: *الگوریتم ژنتیک، پوشش‌دهنده، درماندگی مالی، فیلترکننده، ماشین بردار پشتیبان.*

۱. استادیار مدیریت مالی، دانشکدهٔ مدیریت، دانشگاه تهران، تهران، ایران
۲. دانشجوی کارشناسی ارشد مهندسی مالی، دانشکدهٔ مدیریت دانشگاه تهران، تهران، ایران
۳. دانشجوی کارشناسی ارشد مهندسی مالی، دانشکدهٔ مدیریت دانشگاه تهران، تهران، ایران

تاریخ دریافت مقاله: ۱۳۹۳/۰۳/۳۰
تاریخ پذیرش نهایی مقاله: ۱۳۹۳/۱۱/۲۵
نویسندهٔ مسئول مقاله: عیسی نوروزی‌یان لکوان
E-mail: Norouzian@ut.ac.ir

مقدمه

از آنجا که درماندگی مالی وضعیت بحرانی‌ای است که می‌تواند زیان زیادی را به مدیران، سهامداران، بانک‌ها و مؤسسه‌های اعتباردهنده وارد کند، پیش‌بینی به موقع آن اهمیت شایان توجهی دارد. در ادبیات مالی تعاریف مختلفی از درماندگی مالی ارائه است. گوردن (۱۳۷۱) در یکی از اولین مطالعات خود روی تئوری درماندگی مالی، آن را کاهش سودآوری شرکت تعریف می‌کند که احتمال ناتوانایی بازپرداخت بهره و اصل بدهی را افزایش می‌دهد. ویتاکر (۱۹۹۱) درماندگی مالی را وضعیتی در نظر می‌گیرد که در آن، جریان نقدی شرکت از مجموع هزینه‌های بهرهٔ مربوط به بدهی بلندمدت کمتر است. یکی از راه‌هایی که با استفاده از آن می‌توان به بهره‌مندی مناسب از فرصت‌های سرمایه‌گذاری و تخصیص بهتر منابع کمک کرد، پیش‌بینی درماندگی مالی یا ورشکستگی است، یعنی با ارائهٔ هشدارهای لازم می‌توان شرکت‌ها را نسبت به وقوع درماندگی مالی هوشیار کرد تا آنها با توجه به این هشدارها دست به اقدام مقتضی بزنند و دوم اینکه سرمایه‌گذاران و اعتباردهندگان فرصت‌های مطلوب سرمایه‌گذاری را از فرصت‌های نامطلوب تشخیص دهند و منابعشان را در فرصت‌های مناسب سرمایه‌گذاری کنند. ارزش نسبت‌های مالی در ماجرای ورشکستگی شرکت‌ها مشخص شده است. در این پژوهش ملاک درماندگی مالی شرکت‌ها، مشمول مادهٔ ۱۴۱ بودن قانون تجارت است. باید به این نکته توجه کرد که درماندگی مالی لزوماً موجب ورشکستگی نمی‌شود، بلکه یکی از پیامدهای آن بوده و معمولاً آخرین مرحلهٔ آن است (راعی و فلاح‌پور، ۱۳۸۷).

مدل‌های ارائه شده در این پژوهش، از نوع مدل‌های ترکیبی انتخاب ویژگی با مدل طبقه‌بندی کننده است. از نسبت‌های مالی که ویژگی‌های ورودی مدل هستند، برای پیش‌بینی درماندگی مالی استفاده شده است. در این مقاله عملکرد روش انتخاب ویژگی الگوریتم ژنتیک که از نوع روش‌های پوشش‌دهنده[1] محسوب می‌شود و روش‌های انتخاب ویژگی آنالیز اجزای اساسی، زنجیرهٔ اطلاعات و رلیف که از روش‌های فیلترکننده[2] به‌شمار می‌روند، در ترکیب با ماشین بردار پشتیبان مقایسه می‌شود. ماشین بردار پشتیبان دارای توابع کرنل مختلفی است که در این پژوهش از چهار نوع تابع خطی، چندجمله‌ای، شعاعی و سیگموید[3] برای پیش‌بینی درماندگی مالی استفاده شده است.

---

1. Wrapper
2. Filter
3. Sigmoid

پیشینهٔ پژوهش

تحقیقات اولیه براساس مدل‌های تک‌عاملی بودند و پژوهشگران از یک نسبت مالی برای پیش‌بینی درماندگی مالی استفاده می‌کردند. یکی از قدیمی‌ترین نسبت‌های مالی که در سال ۱۸۷۰ برای ارزیابی وضعیت اعتباری استفاده شد، نسبت جاری بود (راعی و فلاح‌پور، ۱۳۸۷). فیتزپارتریک (۱۹۳۲) یکی از پیشگامان تحقیق در زمینهٔ درماندگی مالی بر مبنای نسبت‌های مالی بود که بعد از آن توسط بیور (۱۹۶۶) که اساس تحقیقات او مدل‌های تک‌عاملی بود، دنبال شد. در این پژوهش بیور ناتوانی شرکت برای انجام تعهدات مالی را درماندگی مالی تعریف کرد. مدل آلتمن (۱۹۶۸) که ساختار آن بر اساس توزیع z-score است، نقطهٔ عطفی در پیش‌بینی درماندگی مالی بود (سان و هیو، ۲۰۱۲).

در اوایل سال ۱۹۹۰ شبکهٔ عصبی که یکی از روش‌های هوش مصنوعی است، در مسائل پیش‌بینی درماندگی مالی استفاده شد و بعد از آن ماشین بردار پشتیبان به‌عنوان روش جدید هوش مصنوعی، عملکرد بسیار خوبی را نشان داد. شین و همکاران (۲۰۰۵) و مین و لی (۲۰۰۵) هر دو از ماشین بردار پشتیبان برای پیش‌بینی شرکت‌های کره‌ای استفاده کردند و به این نتیجه رسیدند که دقت آن بالاتر از دقتی است که از مدل‌های آماری نظیر آلتمن و لوجیت به‌دست می‌آید و همچنین از روش شبکهٔ عصبی نیز دقت بیشتری دارد. هیو و سان (۲۰۰۶) و دینگ، سانگ و زن (۲۰۰۸) این بار از ماشین بردار پشتیبان برای پیش‌بینی شرکت‌های چینی استفاده کردند و به نتایج مشابهی رسیدند. ماشین بردار پشتیبان اغلب نتایج بهتری نسبت به سایر مدل‌های طبقه‌بندی‌کننده برای پیش‌بینی درماندگی مالی روی داده‌هایی که تعداد نمونه‌ها در آن نسبتاً کم است، نشان می‌دهد (سان، ۲۰۱۲).

به‌دلیل اهمیت تشخیص درماندگی مالی، پژوهشگران برای افزایش دقت پیش‌بینی سعی کردند از میان نسبت‌های مالی، آنهایی را که موجب افزایش دقت می‌شوند، انتخاب کنند. مین، لی و هان (۲۰۰۶) و وو و همکاران (۲۰۰۷) از الگوریتم ژنتیک برای افزایش دقت ماشین بردار پشتیبان استفاده کردند.

در ایران نیز تحقیقات مشابهی انجام شده است. سلیمانی در سال ۱۳۸۲، قدرت نسبت‌های مالی را در پیش‌بینی بحران مالی شرکت‌ها بررسی کرد و برای این کار از مدل رگرسیون چندگانه برای پیش‌بینی بحران مالی شرکت‌ها بهره برد. تحقیق صورت گرفته در سال ۱۳۸۳ توسط راعی و فلاح‌پور نشان داد شبکهٔ عصبی از عملکرد بهتری نسبت به مدل تحلیل ممیز چندگانه در پیش‌بینی درماندگی مالی برخوردار است. در سال ۱۳۸۷ راعی و فلاح‌پور نشان دادند ماشین بردار پشتیبان عملکرد بهتری نسبت به مدل رگرسیون لجستیک دارد. خوانساری و میرفیض

(۱۳۸۸) برای پیش‌بینی شرکت‌های ورشکسته، مدل KMV را به‌کار بردند که در این مدل علاوه‌بر داده‌های تاریخی، از داده‌های بازار نیز برای هشدار در مورد وضعیت فعلی مشتریان حقوقی بانک‌ها و حتی انتظارات نسبت به وضعیت آینده آنها استفاده شده است. قدیری مقدم، غلام‌پور فرد و نصیرزاده (۱۳۸۸) دو الگوی آلتمن و اولسون را با استفاده از دو روش ممیزی چندگانه و لوجیت مقایسه کردند. نتایج این پژوهش نشان داد اگر الگوی آلتمن با روش لوجیت برآورد شود، به دقت بالاتری خواهد رسید. نبوی چاشمی، احمدی و مهدی فرح‌آبادی (۱۳۸۹) از مدل لوجیت برای پیش‌بینی ورشکستگی شرکت‌ها استفاده کردند. آنها ۹ نسبت مالی را برای پیش‌بینی ورشکستگی به‌کار بردند که در نهایت مشخص شد سه مورد از آنها تأثیر معناداری در پیش‌بینی درماندگی مالی دارد. فدایی نژاد و اسکندری (۱۳۹۰) به بررسی تأثیر داده‌های بازار و نسبت‌های مالی در پیش‌بینی درماندگی مالی پرداختند و همچنین قدرت پیش‌بینی روش الگوریتم ژنتیک و روش بهینه‌سازی تجمعی ذرات در پیش‌بینی ورشکستگی شرکت‌ها را با هم مقایسه کردند. در پژوهش دیگری که پناهی، اسدزاده و جلیلی مرند (۱۳۹۳) انجام دادند، از نسبت‌های به‌کار رفته در مدل آلتمن و نسبت‌های جاری، به‌منظور پیش‌بینی درماندگی مالی برای فاصلهٔ زمانی ۵ ساله استفاده کردند و همچنین از مدل‌های احتمال خطی، لوجیت و پروبیت، به‌عنوان مدل‌های پیش‌بین بهره بردند.

روش‌شناسی پژوهش

در این پژوهش از چهار نوع کرنل ماشین بردار پشتیبان (SVM) برای پیش‌بینی درماندگی مالی استفاده شده است؛ دلیل نخست این است که SVM از نظریه‌های یادگیری آماری مختلفی ساخته شده است (کریستیانینی و شاو تیلور، ۲۰۰۰ و واپنیک، ۱۹۷۸) و دلیل دوم اینکه نتایج تحقیقات گذشته نشان داده است SVM دقت بالایی در پیش‌بینی درماندگی مالی دارد (دینگ و همکاران، ۲۰۰۸؛ هیو و همکاران، ۲۰۰۶؛ سان و همکاران، ۲۰۰۹؛ مین و همکاران، ۲۰۰۶؛ مین و همکاران ۲۰۰۵؛ پرماچاندرا، بهابرا و سیوشی ۲۰۰۹؛ شین و همکاران، ۲۰۰۵؛ تان، کویک و سی، ۲۰۰۷ و هیو، وانگ، ایکسو، زهانگ و لیانگ، ۲۰۰۷).

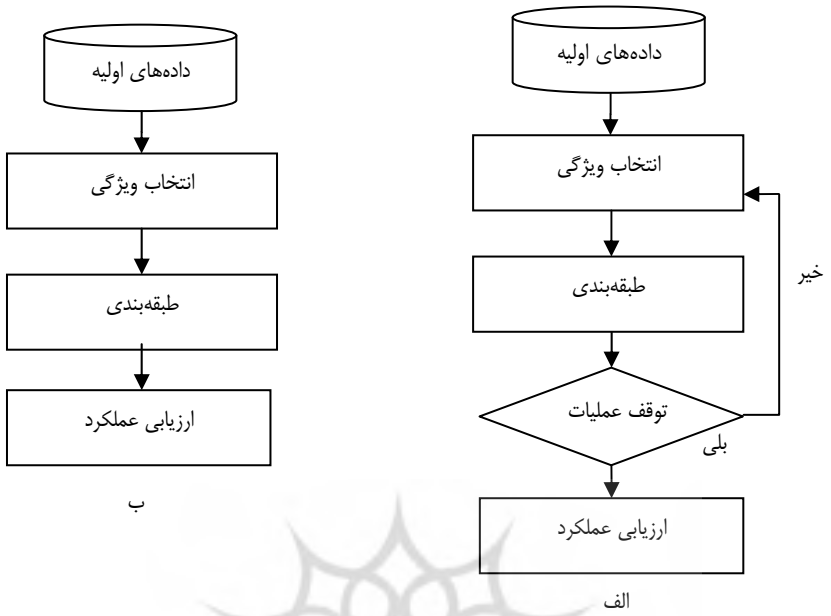
انتخاب ویژگی

فرایند انتخاب ویژگی معمولاً برای مسائلی استفاده می‌شود که داده‌ها شامل ویژگی‌های زیادی هستند. این فرایند باعث کاهش ابعاد بردار ویژگی‌ها می‌شود که با حذف ویژگی‌های غیرضروری و انتخاب ویژگی‌های ضروری برای یادگیری مدل، بهبود دقت پیش‌بینی و افزایش قابلیت توضیحی مدل‌های پیش‌بینی‌کننده را به همراه دارد (ارسکی و ارسکی، ۲۰۱۴). انتخاب ویژگی در

حقیقت مسئلهٔ انتخاب یک زیرمجموعه از ویژگی‌هاست که برای توضیح هدف مد نظر کافی و ضروری است (ارسکی و ارسکی، ۲۰۱۴). وقتی که بردار ویژگی داده‌ها شامل تعداد زیادی ویژگی باشد، به‌دلیل وجود ویژگی‌های اضافی و غیرضروری، دقت به‌دست آمده از مدل طبقه‌بندی‌کننده کمتر از دقتی است که یک زیرمجموعهٔ مناسب برای یادگیری مدل انتخاب می‌شود. می‌توان مشکل انتخاب یک زیرمجموعه از ویژگی‌ها را برای رسیدن به دقت بهینه یا دقت نزدیک به دقت بهینه با استفاده از یک روش مناسب انتخاب ویژگی حل کرد.

به‌طور کلی می‌توان روش‌های انتخاب ویژگی را به دو گروه عمده دسته‌بندی کرد: روش‌های فیلترکننده و روش‌های پوشش‌دهنده (گایون و الیسف، ۲۰۰۳ و کوهاوی جان، ۱۹۹۷). روش‌های فیلترکننده انتخاب ویژگی، اشاره به این دارند که یک الگوریتم به جست‌وجو در میان فضای ممکن ویژگی‌ها می‌پردازد و پس از انتخاب هر زیرمجموعه از ویژگی‌ها و اجرای یک تابع فیلترکننده، ویژگی‌ها را رتبه‌بندی می‌کند. این رویکرد در انتخاب کردن بهترین زیرمجموعه از ویژگی‌ها توجهی به مدل طبقه‌بندی‌کننده ندارد (لی، لی، وو و سان، ۲۰۱۴). در واقع، روش‌های فیلترکننده به ویژگی کلی بردار ویژگی‌ها و ارزیابی و انتخاب زیرمجموعه‌ها بدون ترکیب شدن با مدل طبقه‌بندی‌کننده متکی هستند. روش‌های فیلترکننده به‌دلیل استفاده‌نکردن از مدل طبقه‌بندی‌کننده، سرعت بالایی دارند، بنابراین این روش برای حل مسائلی که داده‌های آن ویژگی‌های زیادی دارند، بسیار مناسب است. همچنین این روش‌ها به راحتی با انواع مختلف مدل‌های طبقه‌بندی‌کننده اجرا می‌شوند. روش‌های پوشش‌دهنده ابتدا از یک الگوریتم بهینه‌سازی استفاده می‌کنند، این الگوریتم ویژگی‌های مختلف را اضافه یا کم می‌کند تا زیرمجموعه‌های متفاوتی را تشکیل دهد. سپس از یک مدل طبقه‌بندی‌کننده برای ارزیابی این زیرمجموعه‌های ساخته شده، استفاده می‌کنند و چون به‌طور مکرر مدل طبقه‌بندی‌کننده را فراخوانی می‌کنند، سرعت کمتری نسبت به روش‌های فیلترکننده دارند (ارسکی و ارسکی، ۲۰۱۴).

از محدودیت‌های روش فیلترکننده این است که کاربران باید تعداد ویژگی‌هایی را که می‌خواهند انتخاب کرده یا معیار توقف را مشخص کنند (نجی، یانگ، فیرد، تسانگ و وانگ، ۲۰۰۸)، ولی در روش‌های پوشش‌دهنده تعداد ویژگی‌های انتخاب شده بر اساس دقتی است که مدل طبقه‌بندی‌کننده به آن دست می‌یابد و زیرمجموعه‌ای از ویژگی‌ها را که بیشترین دقت را نتیجه می‌دهد، انتخاب می‌کنند. شایان ذکر است که تعیین پارامترهای جست‌وجوی مناسب برای روش‌های پوشش‌دهنده اهمیت بسیاری دارد؛ به‌خصوص در مسائلی که تعداد ویژگی‌ها زیاد است، در صورت عدم تنظیم صحیح پارامترها، به‌دلیل تشکیل زیرمجموعه‌های زیاد و محاسبهٔ دقت آنها، زمان زیادی صرف محاسبه خواهد شد.



شکل ۱. مقایسهٔ روش‌های پوشش‌دهنده (الف) و فیلترکننده برای انتخاب ویژگی(ب)

آنالیز اجزای اساسی (PCA)[1]

یک روش ریاضی است که با استفاده از تبدیل متعامد، مجموعه‌ای از مشاهدات احتمالاً همبسته را به مجموعهٔ ارزش از مشاهدات ناهمبسته که به آن مؤلفه‌های اصلی می‌گویند، تبدیل می‌کند. تعداد مؤلفه‌های اساسی کوچک‌تر یا مساوی تعداد ویژگی‌های ورودی است. اولین جزء اصلی به‌گونه‌ای انتخاب می‌شود که بالاترین واریانس را داشته باشد. هر جزئی که واریانس بالایی داشته باشد، می‌توان آن را ناهمبسته از اجزای دیگر دانست و ویژگی‌ها بر اساس واریانس خود مرتب می‌شوند.

زنجیرهٔ اطلاعات[2]

این روش ارتباط بین ویژگی‌ها را بر اساس زنجیرهٔ اطلاعات در نظر گرفته و با توجه به طبقهٔ هر ویژگی، به ویژگی‌ها وزن می‌دهد. X و Y نشان دهندهٔ ویژگی است.

$$InfoGain = H(Y) - H(Y|X)$$ رابطهٔ ۱)

1. Principal Component Analysis
2. Information Gain

$$H(Y) = -\sum_{y\epsilon Y} P(y) log_2(p(y))$$ رابطهٔ ۲)

$$H(Y|X) = -\sum_{x\epsilon X} P(x) \sum_{y\epsilon Y} (y|x) \log\,(p(y|x))$$ رابطهٔ ۳)

رلیف[1]

اهمیت ویژگی‌ها را بر اساس نمونه‌برداری و جاگذاری با در نظرگرفتن ارزش داده شده به ویژگی با نزدیک‌ترین جایگزین با طبقهٔ یکسان یا متفاوت مشخص می‌کند. اهمیت داده شده به ویژگی‌ها در این موارد باید مرتب به‌روزآوری شده و بار دیگر امتیازدهی شود (نی، نی و گائو ۲۰۱۱).

این سه روش تعریف شده از نوع روش‌های فیلترکننده هستند. در این پژوهش معیار توقف برای این سه روش، عدد ۱۰ در نظر گرفته شده است که دلیل آن نتایج بهتر حاصل شده از این عدد در مقایسه با اعداد دیگر است. این الگوریتم‌ها بعد از وزن‌دهی و رتبه‌بندی ویژگی‌ها با توجه به معیار توقف تعیین شده، فقط از ۱۰ ویژگی برتر برای پیش‌بینی استفاده می‌کنند.

الگوریتم ژنتیک (GA)

یکی از مشهورترین الگوریتم‌های بهینه‌سازی برای حل مسائل پیچیده با فضای حل وسیع است. این الگوریتم برگرفته از مفاهیم زیستی است (شوف و فوستر، ۱۹۹۶). اصلی‌ترین ویژگی‌های الگوریتم ژنتیک عبارت‌اند از: الگوریتم ژنتیک روی مجموعه جواب‌هایی از فضای قابل قبول کار می‌کند نه روی یک جواب؛ وسعت جست‌وجوی عملگرهای الگوریتم ژنتیک کمک می‌کند تا به‌طور مؤثری جواب‌های کشف نشده در فضای جست‌وجو شناسایی و آزمایش شوند؛ احتمالی بودن ساختار مسئله با عملگرهای الگوریتم ژنتیک باعث می‌شود یک جواب بهتر کشف و ارائه شود؛ وجود جمعیت‌های مختلف باعث می‌شود که احتمال گیرافتادن الگوریتم در یک نقطهٔ بهینهٔ محلی کاهش یابد؛ گام‌های مختلف الگوریتم ژنتیک به‌گونه‌ای است که بعد از اجرای تمام گام‌ها دوباره به نخستین گام برگشته و تا رسیدن به یک جواب مطلوب این عملیات تکرار می‌شود؛ گام اول در الگوریتم ژنتیک، تولید یک سری جواب‌های تصادفی است که جمعیت یا نسل اولیه نام دارد (تان، کوییک و سی، ۲۰۰۷ و ارسکی، ارسکی و ارسکی، ۲۰۱۲). کیفیت نسل اولیهٔ انتخاب شده، نقش بسیار مهمی در کیفیت جواب نهایی مسئله دارد (ییو و لیو، ۲۰۰۳)؛ الگوریتم ژنتیک از

1. Relief

نوع روش‌های پوشش‌دهنده است که با استفاده از آن می‌توان به یک جواب بهینه یا یک جواب نزدیک به جواب بهینه رسید.

ماشین بردار پشتیبان (SVM)

این الگوریتم نوع خاصی از مدل‌های خطی را می‌یابد که حداکثر حاشیهٔ ابر صفحه را حاصل می‌کنند. حداکثر کردن حاشیهٔ ابر صفحه، به حداکثر شدن تفکیک بین طبقات منجر می‌شود. به نزدیک‌ترین نقاط آموزشی به حداکثر حاشیهٔ ابرصفحه، بردار پشتیبان اطلاق می‌گردد. این بردارها فقط برای مشخص کردن مرز بین طبقات به‌کار می‌روند (شین و همکاران، ۲۰۰۵).

اگر داده‌ها به صورت خطی مجزا از هم باشند، SVM به ماشین‌های خطی برای تولید سطح بهینه‌ای که داده‌ها را بدون خطا و با حداکثر فاصله میان صفحه و نزدیک‌ترین نقاط آموزشی (بردار پشتیبان) تفکیک کند، آموزش می‌دهد. اگر نقاط آموزشی را به‌صورت $[x_i, y_i]$ و بردار ورودی $x_i \in R^n$ و ارزش طبقه y_i را $i = 1, \dots, L = \{-1,1\}$ تعریف کنیم، آنگاه در حالتی که داده‌ها به‌صورت خطی قابل تفکیک‌اند، قواعدی که تصمیم‌گیری باینری را تفکیک می‌کند، به‌صورت رابطهٔ زیر است:

$$Y = sig\, n \left(\sum y_i a_i (X.X) + b \right)$$ رابطهٔ ۴)

که در آن Y خروجی معادله؛ y_i ارزش طبقهٔ نمونهٔ آموزشی و x_i نشان‌دهندهٔ ضرب داخلی است. بردار $x = (x_1, x_2, \dots, x_3)$ نشان‌دهندهٔ یک دادهٔ ورودی و بردارهای $X_i, i = 1, \dots, N$ بردارهای پشتیبان هستند. در رابطهٔ ۴، پارامترهای a_i و b تعیین‌کنندهٔ ابرصفحه هستند. اگر داده‌ها به‌صورت خطی قابل تفکیک نباشند، رابطهٔ ۴ به رابطهٔ زیر تغییر می‌یابد:

$$Y = sig\, n \left(\sum y_i a_i K(X.X_i) + b \right)$$ رابطهٔ ۵)

تابع $K(X.X_i)$ تابع کرنلی است که برای ایجاد ماشین‌هایی با انواع مختلف از سطوح تصمیم‌گیری غیرخطی در فضای داده‌ها، ضرب‌های داخلی تولید می‌کند (راعی و فلاح‌پور، ۱۳۸۷). چهار نوع تابع کرنل به‌کار برده شده در این پژوهش عبارت‌اند از:

۱. ماشین بردار پشتیبان با تابع کرنل خطی (liner):

$$K(x, y) = x \cdot y$$ رابطهٔ ۶)

۲. ماشین بردار پشتیبان با تابع کرنل چندجمله‌ای (polynomial):

رابطهٔ ۷) $$K(x,y) = [g(x,y) + c]^d$$

d نشان‌دهندهٔ درجهٔ چندجمله‌ای است.

۳. ماشین بردار پشتیبان با تابع کرنل سیگموید (sigmoid):

رابطهٔ ۸) $$K(x,y) = Tanh[g(x,y) + c]$$

۴. ماشین بردار پشتیبان با تابع کرنل شعاعی (radial):

رابطهٔ ۹) $$K(x,y) = exp[-g\|x - y\|^2]$$

فرایند یادگیری برای ایجاد توابع تصمیم‌گیری دارای ساختاری دو لایه است. SVM از تئوری بهینه‌سازی برای طبقه‌بندی استفاده می‌کند؛ تئوری بهینه‌سازی بر اساس تئوری یادگیری آماری، خطای طبقه‌بندی را به حداقل می‌رساند.

به‌طور معمول چهار نوع کرنل وجود دارد که در تحقیقات از آنها استفاده می‌شود ۱. کرنل خطی؛ ۲. کرنل چندجمله‌ای؛ ۳. کرنل شعاعی؛ ۴. کرنل سیگموید (زاوگرن، ۱۹۸۵) که در این پژوهش نیز از این چهار نوع تابع کرنل استفاده می‌شود. یکی از نکات بسیار مهم در استفاده از SVM، تعیین پارامترهای ضریب جریمه و ضریب گاماست. به‌طور کلی دو روش عمده برای تعیین پارامترهای بهینه وجود دارد: تعیین پارامترها بر اساس تجربه و تعیین پارامترها بر اساس جست‌وجوی شبکه‌ای. در پژوهش حاضر برای یافتن پارامترهای بهینه از روش جست‌وجوی شبکه‌ای استفاده شده است.

روش اعتبارسنجی مقطعی[1]

این عملگر داده‌ها را به دو گروه، داده‌های آموزشی و داده‌های آزمایشی دسته‌بندی می‌کند؛ به‌طوری که ابتدا داده‌ها را به K زیرمجموعهٔ مساوی تقسیم می‌کند، پس از آن هر بار K-1 زیرمجموعه از داده‌ها را برای داده‌های آموزشی و فقط یک زیرمجموعه را برای داده‌های آزمایشی در نظر می‌گیرد و این کار را K بار انجام می‌دهد؛ به طوری که هر زیرمجموعه یک بار به‌عنوان دادهٔ آزمایشی قرار می‌گیرد و دقت پیش‌بینی برای آن زیرمجموعه به‌دست می‌آید؛ دقت کلی مدل نیز از میانگین دقت‌ها محاسبه می‌شود.

پس از مطالعه و بررسی نسبت‌های مالی مهم استفاده شده در تحقیقات گذشته و اطلاعات در دسترس از شرکت‌های منتخب، در نهایت ۲۹ متغیر به‌عنوان متغیرهای ورودی مدل انتخاب شد.

---

1. Cross Validation

در جدول ۱ این ویژگی‌ها درج شده است. برای افزایش دقت پیش‌بینی، تمام این متغیرها قبل از ورود به مدل نرمال استاندارد شده‌اند.

$$V_i = \frac{v_i - \bar{v}_i}{\sigma_i}$$ رابطهٔ ۱۰)

جدول ۱. متغیرهای استفاده شده برای پیش‌بینی درماندگی مالی

نام متغیر	شمارهٔ متغیر	نام متغیر	شمارهٔ متغیر
سود خالص به فروش	V1	سرمایه در گردش خالص	V16
سود عملیاتی به فروش	V2	گردش موجودی کالا	V17
سود ناخالص به فروش	V3	دوره وصول مطالبات	V18
سود خالص به سود ناخالص	V4	نسبت موجودی کالا به سرمایه در گردش	V19
بازده دارایی‌ها (ROA)	V5	گردش سرمایه جاری	V20
درصد بازده سرمایه	V6	گردش دارایی ثابت	V21
گردش مجموع دارایی	V7	گردش مجموع دارایی	V22
بازدهی سرمایه (ROE)	V8	نسبت بدهی	V23
بازده سرمایهٔ در گردش	V9	نسبت بدهی به ارزش ویژه	V24
بازده دارایی ثابت	V10	نسبت دارایی ثابت به ارزش ویژه	V25
سنجش سودمندی وام	V11	نسبت بدهی بلند مدت به ارزش ویژه	V26
نسبت جاری	V12	نسبت بدهی جاری به ارزش ویژه	V27
نسبت آنی	V13	نسبت مالکانه	V28
نسبت نقدینگی	V14	نسبت پوشش بدهی	V29
نسبت دارایی‌های جاری	V15		

جامعهٔ آماری این پژوهش، شرکت‌های پذیرفته شده در بورس اوراق بهادار تهران از سال ۱۳۷۵ تا ۱۳۹۱ در نظر گرفته شده است. نمونهٔ مورد استفاده شامل ۱۸۰ شرکت است که با توجه به تعداد شرکت‌های بورسی که مشمول مادهٔ ۱۴۱ قانون تجارت بودند، فهرستی از تمام شرکت‌های درماندهٔ مالی (حدود ۱۰۳ شرکت) تهیه شد. از میان آنها ۹۰ شرکت که دسترسی به اطلاعات آنها به‌صورت کامل وجود داشت، انتخاب گردید و همچنین ۹۰ شرکت سالم با استفاده از نمونه‌گیری تصادفی انتخاب شد. در مجموع ۱۸۰ شرکت، شامل ۹۰ شرکت درماندهٔ مالی و ۹۰ شرکت سالم بررسی شدند. برای آزمودن مدل‌های بیان شده، عملکرد این مدل‌ها در سال درماندگی مالی (t)، یک سال قبل از درماندگی مالی (t−۱) و دو سال قبل از درماندگی مالی (t−۲) بررسی شده است. بنابراین به اطلاعات سه سال مالی هر شرکت نیاز داریم.

با توجه به نتایج به‌دست آمده از تحقیقات گذشته، برای این پژوهش دو فرضیه مطرح شده است:

فرضیهٔ اول: از لحاظ آماری، تفاوت معناداری بین روش الگوریتم ژنتیک و مدل‌های فیلترکنندهٔ انتخاب ویژگی در ترکیب با SVM وجود دارد.

فرضیهٔ دوم: از لحاظ آماری، تفاوت معناداری بین توابع کرنل خطی، چندجمله‌ای، شعاعی و سیگموید SVM در ترکیب با مدل انتخاب ویژگی الگوریتم ژنتیک وجود دارد.

کلیهٔ مدل‌سازی‌های این تحقیق با نرم‌افزار Rapid Miner و آزمون مقایسات زوجی در نرم‌افزار متلب انجام شده است.

یافته‌های پژوهش

در این قسمت یافته‌های پژوهش و تحلیل آنها ارائه می‌شود. جدول ۲ نتایج به‌دست آمده برای دقت پیش‌بینی مدل‌های به‌کار برده شده را برای سال t نشان می‌دهد.

جدول ۲. عملکرد چهار روش انتخاب ویژگی در ترکیب با توابع کرنل مختلف SVM برای سال t

نام روش	تعداد ویژگی‌های انتخاب شده	نوع روش	میانگین دقت کلی پیش‌بینی
PCA+ SVM(liner)	۱۰	فیلترکننده	۹۵/۰۰
PCA+ SVM(polynomial)	۱۰	فیلترکننده	۹۳/۳۳
PCA+ SVM(radial)	۱۰	فیلترکننده	۹۴/۴۴
PCA+ SVM(sigmoid)	۱۰	فیلترکننده	۹۵/۵۶
Information gain + SVM(liner)	۱۰	فیلترکننده	۹۵/۰۰
Information gain + SVM(polynomial)	۱۰	فیلترکننده	۹۲/۲۲
Information gain + SVM(radial)	۱۰	فیلترکننده	۹۵/۵۶
Information gain + SVM(sigmoid)	۱۰	فیلترکننده	۹۵/۵۶
Relief + SVM(liner)	۱۰	فیلترکننده	۹۵/۰۰
Relief + SVM(polynomial)	۱۰	فیلترکننده	۹۱/۶۷
Relief + SVM(radial)	۱۰	فیلترکننده	۹۵/۵۶
Relief + SVM(sigmoid)	۱۰	فیلترکننده	۹۵/۵۶
GA + SVM(liner)	۶	پوشش‌دهنده	۹۷/۷۸
GA + SVM(polynomial)	۳	پوشش‌دهنده	۹۷/۷۸
GA + SVM(radial)	۷	پوشش‌دهنده	۹۷/۷۸
GA + SVM(sigmoid)	۱۱	پوشش‌دهنده	۹۷/۷۸

برای بررسی قسمت اول فرضیهٔ پژوهشی مبنی بر عملکرد بهتر روش الگوریتم ژنتیک نسبت به روش‌های فیلترکننده، آزمون مقایسات زوجی اجرا شده است. در این آزمون عملکرد بهتر الگوریتم ژنتیک از نظر آماری بررسی شده است. در جدول ۳ نتایج آزمون مقایسهٔ زوجی برای سال t مشاهده می‌شود. میانگین دقت کلی روش الگوریتم ژنتیک، بیان‌کنندهٔ میانگین دقت‌های به‌دست آمده از توابع کرنل مختلف SVM است. همچنین میانگین دقت کلی روش‌های فیلترکننده، نشان‌دهندهٔ میانگین هر سه روش فیلترکنندهٔ انتخاب ویژگی با توابع کرنل مختلف SVM است.

جدول ۳. نتایج آزمون مقایسهٔ زوجی برای سال t

	الگوریتم ژنتیک	روش‌های فیلترکننده
میانگین دقت کلی	۹۷/۷۸	۹۴/۴۹۶
p-value	۰/۰۰۰۴۱۹	
بررسی فرضیهٔ اول پژوهش	فرضیه پذیرفته می‌شود	

جدول ٤. عملکرد چهار روش انتخاب ویژگی در ترکیب با توابع کرنل مختلف SVM برای سال ۱- t

نام روش	تعداد ویژگی های انتخاب شده	نوع روش	میانگین دقت کلی پیش‌بینی
PCA+ SVM(liner)	۱۰	فیلترکننده	۹۰/۰۰
PCA+ SVM(polynomial)	۱۰	فیلترکننده	۷۶/۶۷
PCA+ SVM(radial)	۱۰	فیلترکننده	۹۰/۵۶
PCA+ SVM(sigmoid)	۱۰	فیلترکننده	۸۵/۰۰
Information gain + SVM(liner)	۱۰	فیلترکننده	۹۱/۶۷
Information gain + SVM(polynomial)	۱۰	فیلترکننده	۷۸/۳۳
Information gain + SVM(radial)	۱۰	فیلترکننده	۹۰/۵۶
Information gain + SVM(sigmoid)	۱۰	فیلترکننده	۸۹/۴۴
Relief + SVM(liner)	۱۰	فیلترکننده	۹۰/۵۶
Relief + SVM(polynomial)	۱۰	فیلترکننده	۷۶/۱۱
Relief + SVM(radial)	۱۰	فیلترکننده	۹۱/۱۱
Relief + SVM(sigmoid)	۱۰	فیلترکننده	۸۷/۷۸
GA + SVM(liner)	۵	پوشش‌دهنده	۹۳/۸۹
GA + SVM(polynomial)	۴	پوشش‌دهنده	۹۲/۷۸
GA + SVM(radial)	۵	پوشش‌دهنده	۹۴/۴۴
GA + SVM(sigmoid)	۱۰	پوشش‌دهنده	۹۳/۹۸

الگوریتم ژنتیک با SVM در سطح اطمینان ۰/۹۵، از دقت کلی مدل‌های ترکیبی نوع فیلترکننده بیشتر است. بنابراین قسمت اول فرض پژوهش برای سال t پذیرفته می‌شود. جدول ۴ کلیهٔ نتایج به‌دست آمده برای دقت پیش‌بینی مدل‌های به‌کار برده شده برای یک سال قبل از درماندگی را نشان می‌دهد.

برای بررسی قسمت اول فرض پژوهش مبنی بر عملکرد بهتر الگوریتم ژنتیک نسبت به روش‌های فیلترکننده نیز از آزمون مقایسات زوجی استفاده شده است. جدول ۵ نتایج آزمون مقایسات زوجی را برای یک سال قبل از درماندگی مالی نشان می‌دهد.

جدول ۵. نتایج آزمون مقایسهٔ زوجی برای سال t-۱

	الگوریتم ژنتیک	روش‌های فیلترکننده
میانگین دقت کلی	۹۳/۷۵	۸۷/۸۶
p-value	۰/۰۳۲۷	
بررسی فرضیهٔ اول پژوهش	فرضیه پذیرفته می‌شود	

جدول ٦. عملکرد چهار روش انتخاب ویژگی در ترکیب با توابع کرنل مختلف SVM برای سال t-۲

نام روش	تعداد ویژگی های انتخاب شده	نوع روش	میانگین دقت کلی پیش‌بینی
PCA+ SVM(liner)	۱۰	فیلترکننده	۸۲/۲۲
PCA+ SVM(polynomial)	۱۰	فیلترکننده	۷۵/۵۶
PCA+ SVM(radial)	۱۰	فیلترکننده	۸۱/۶۷
PCA+ SVM(sigmoid)	۱۰	فیلترکننده	۸۰/۰۰
Information gain + SVM(liner)	۱۰	فیلترکننده	۷۹/۴۴
Information gain + SVM(polynomial)	۱۰	فیلترکننده	۷۰/۵۶
Information gain + SVM(radial)	۱۰	فیلترکننده	۷۹/۴۴
Information gain + SVM(sigmoid)	۱۰	فیلترکننده	۸۰/۰۰
Relief + SVM(liner)	۱۰	فیلترکننده	۸۱/۱۱
Relief + SVM(polynomial)	۱۰	فیلترکننده	۶۹/۴۴
Relief + SVM(radial)	۱۰	فیلترکننده	۸۳/۸۹
Relief + SVM(sigmoid)	۱۰	فیلترکننده	۸۲/۲۲
GA + SVM(liner)	۹	پوشش‌دهنده	۸۷/۲۲
GA + SVM(polynomial)	۵	پوشش‌دهنده	۸۷/۲۲
GA + SVM(radial)	۱۰	پوشش‌دهنده	۸۸/۸۹
GA + SVM(sigmoid)	۸	پوشش‌دهنده	۸۷/۷۸

همان‌طور که جدول ۵ نشان می‌دهد، می‌توان گفت که دقت کلی مدل ترکیبی الگوریتم ژنتیک با SVM در سطح اطمینان ۰/۹۵، از دقت کلی مدل‌های ترکیبی نوع فیلترکننده بیشتر است. بنابراین قسمت اول فرض پژوهشی برای سال t −۱ پذیرفته می‌شود. جدول ۶ نتایج به‌دست آمده برای دقت پیش‌بینی مدل‌های به‌کار برده شده برای دو سال قبل از درماندگی مالی را نشان می‌دهد.

برای بررسی قسمت اول فرض پژوهشی مبنی بر عملکرد بهتر الگوریتم ژنتیک نسبت به روش‌های فیلترکننده نیز از آزمون مقایسات زوجی استفاده شده است. جدول ۷ نتایج آزمون مقایسات زوجی را برای دو سال قبل از درماندگی مالی نشان می‌دهد.

جدول ۷. نتایج آزمون مقایسهٔ زوجی برای سال t −۲

	الگوریتم ژنتیک	روش‌های فیلترکننده
میانگین دقت کلی	۸۷/۷۷۷	۷۸/۴۸۴
p-value	۰/۰۰۱۹	
بررسی فرضیهٔ اول پژوهش	فرضیه پذیرفته می‌شود	

همان‌طور که جدول ۷ نشان می‌دهد، می‌توان گفت که دقت کلی مدل ترکیبی الگوریتم ژنتیک با SVM در سطح اطمینان ۰/۹۵، از دقت کلی مدل‌های ترکیبی نوع فیلترکننده بیشتر است. بنابراین قسمت اول فرضیهٔ پژوهش برای سال t −۲ پذیرفته می‌شود.

برای بررسی قسمت دوم فرضیهٔ پژوهش، مبنی بر اینکه تفاوت معناداری بین توابع کرنل مختلف SVM در ترکیب با روش انتخاب ویژگی الگوریتم ژنتیک وجود دارد نیز، از آزمون مقایسهٔ زوجی استفاده شده است. جدول ۸ نشان‌دهندهٔ نتایج به‌دست آمده از آزمون مقایسهٔ زوجی است؛ به‌طوری که نتایج حاصل شده از هر تابع کرنل طی سال‌های t، t −۱ و t −۲ دو به دو با هم مقایسه شده‌اند.

جدول ۸. نتایج آزمون مقایسهٔ زوجی برای کرنل‌های مختلف SVM در ترکیب با الگوریتم ژنتیک

نوع کرنل	چندجمله‌ای		شعاعی		سیگموید	
	بررسی فرضیهٔ دوم	p-value	بررسی فرضیهٔ دوم	p-value	بررسی فرضیهٔ دوم	p-value
خطی	تأیید نمی‌شود	۰/۹۶۷	تأیید نمی‌شود	۰/۸۶۳۲	تأیید نمی‌شود	۰/۹۳۶۱
چندجمله‌ای	تأیید نمی‌شود	۰/۸۷۲۱	تأیید نمی‌شود	۰/۷۶۵۷	—	—
شعاعی	تأیید نمی‌شود	۰/۸۹۴	—	—	—	—

همان‌طور که جدول ۸ نشان می‌دهد، می‌توان گفت که دقت کلی توابع کرنل خطی، چندجمله‌ای، شعاعی و سیگموییدSVM در سطح اطمینان ۰/۹۵ با هم تفاوت معناداری ندارند. بنابراین قسمت دوم فرضیهٔ پژوهش پذیرفته نمی‌شود.

نتیجه‌گیری و پیشنهادها

پیش‌بینی درماندگی مالی و ورشکستگی شرکت‌ها یکی از مسائل مهم در حوزهٔ مالی است. بانک‌ها و مؤسسه‌های رتبه‌بندی اعتباری، از مدل‌های پیش‌بینی درماندگی مالی برای تخصیص اعتبار و رتبه‌بندی استفاده می‌کنند. همچنین این مدل‌ها می‌توانند به مدیران، سرمایه‌گذاران و کارکنان در گرفتن تصمیم مناسب کمک کنند؛ به طوری که با اقدامات پیشگیرانه از وقوع درماندگی مالی و ورشکستگی جلوگیری نمایند و میزان خسارات آتی را کاهش دهند. تاکنون تحقیقات بسیار زیادی در زمینهٔ مقایسهٔ انواع مدل‌های طبقه‌بندی‌کننده بدون ترکیب با مدل‌های انتخاب ویژگی انجام شده است و نتایج حاصل نشان‌دهندهٔ عملکرد خوب مدل طبقه‌بندی‌کنندهٔ SVM در مسائل پیش‌بینی درماندگی مالی و ورشکستگی بوده است. اخیراً مطالعهٔ مدل‌های ترکیبی انتخاب ویژگی و مدل‌های طبقه‌بندی‌کننده به‌عنوان روشی با عملکرد بالا افزایش یافته است، ولی همچنان نیاز به مطالعات بیشتری دارد. مدل‌های ارائه شده در این مقاله برای پیش‌بینی درماندگی مالی، از نوع مدل‌های ترکیبی انتخاب ویژگی و مدل SVM هستند. روش‌های انتخاب ویژگی استفاده شده، شامل روش‌های پوشش‌دهنده و فیلترکننده است که با توابع کرنل خطی، چندجمله‌ای، شعاعی و سیگموییدSVM ترکیب شده‌اند. نتایج نشان داد، الگوریتم ژنتیک که جزء روش‌های پوشش‌دهنده است، از عملکرد بهتری نسبت به روش‌های تجزیه و تحلیل اجزای اساسی، زنجیرهٔ اطلاعات و رلیف که از خانوادهٔ روش‌های فیلترکننده‌اند در ترکیب با توابع کرنل مختلف SVM برخوردار است. همچنین بررسی‌های انجام شده روی دقت‌های به‌دست آمده از ترکیب الگوریتم ژنتیک با توابع کرنل خطی، چندجمله‌ای، شعاعی و سیگموییدنشان داد از لحاظ آماری تفاوت معناداری بین توابع کرنل مختلف SVM در ترکیب با روش الگوریتم ژنتیک وجود ندارد. برای تحقیقات آتی می‌توان از روش‌های دیگر انتخاب ویژگی مانند الگوریتم بهینه‌سازی کلونی زنبور عسل و سایر الگوریتم‌های جدید بهینه‌سازی استفاده کرد و می‌توان رگرسیون بردار پشتیبان (SVR) را به‌عنوان مدل طبقه‌بندی‌کننده به کار برد.

فهرست منابع

پناهی، ح.، اسدزاده، ا.، جلیلی مرند، ع. (۱۳۹۳). پیش‌بینی پنج سالهٔ ورشکستگی مالی برای شرکت‌های پذیرفته شده در بورس اوراق بهادار تهران. *نشریهٔ تحقیقات مالی*، ۱۶ (۱)، ۷۶-۵۷.

خوانساری، ر.، فلاح شمس، م. (۱۳۸۸). ارزیابی کاربرد مدل ساختاری KMV در پیش‌بینی نکول شرکت‌های پذیرفته شده در بورس اوراق بهادار تهران. *نشریهٔ تحقیقات مالی*، ۱۱ (۲۸)، ۶۸-۴۹.

راعی، ر.، فلاح‌پور، س. (۱۳۸۳). پیش‌بینی درماندگی مالی شرکت‌ها با استفاده از شبکهٔ عصبی مصنوعی. *نشریهٔ تحقیقات مالی*، ۶ (۱)، ۶۹-۳۹.

راعی، ر.، فلاح‌پور، س. (۱۳۸۳). کاربرد ماشین بردار پشتیبان در پیش‌بینی درماندگی مالی شرکت‌ها با استفاده از نسبت‌های مالی . *فصلنامهٔ علمی پژوهشی بررسی‌های حسابداری و حسابرسی*، ۱۵ (۵۳)، ۳۴-۱۷.

سلیمانی امیری، ق. (۱۳۸۲). نسبت‌های مالی و پیش‌بینی بحران مالی شرکت‌ها در بورس اوراق بهادار تهران. *نشریهٔ تحقیقات مالی*، ۵ (۱)، ۱۳۶-۱۲۱.

فدائی‌نژاد، ا.، اسکندری، ر. (۱۳۹۰). طراحی و تبیین مدل پیش‌بینی ورشکستگی شرکت‌ها در بورس اوراق بهادار تهران. *تحقیقات حسابداری و حسابرسی*، ۳ (۹)، ۲۴-۱.

قدری مقدم، ا.، غلام‌پور فرد، م.، نصیرزاده، ف. (۱۳۸۷). بررسی توانایی مدل‌های پیش‌بینی ورشکستگی آلتمن و اهلسون در پیش‌بینی ورشکستگی شرکت‌های پذیرفته‌شده در بورس اوراق بهادار. *مجلهٔ دانش و توسعه*، ۱۶ (۲۸)، ۲۲۰-۱۹۳.

Altman, E. I. (1968). Financial ratios discriminant analysis and the prediction of corporate bankruptcy. *Journal of Finance*, 23(4), 589 - 609.

Beaver, W. (1966). Financial ratios as predictors of failure. *Empirical Research in Accounting: Selected Studies*, (4), 71–111.

Cristianini, N. & Shawe-Taylor, J. (2000). *An Introduction to Support Vector Machines*. Cambridge University Press, Cambridge.

Ding, Y., Song, X. & Zen, Y. (2008). Forecasting financial condition of Chinese listed companies based on support vector machine. *Expert Systems with Applications*, 34(4), 3081–3089.

Fadayi-Nejad, M. & Eskandari, R. (2011). Design and explained corporate bankruptcy prediction model in Tehran Stock Exchange. *Iranian Accounting Association*, 3 (9), 1-24. (*in Persian*)

Fitzpartrick, P. J. (1932). A comparison of ratios of successful industrial enterprises with those of failed companies. *Journal of Accounting Research*, 10, 598–605.

Ghadri Moghadam, A., Gholampour- Fard, M., Nasirzadeh, F. (2009). Evaluate of ability of altman and ohlsoon models in bankruptcy prediction of companies listed in the stock exchange. *Journal of knowledge and development*, 16 (28), 193- 220. (*in Persian*)

Gordon, M. J. (1971). Towards a theory of financial distress. *The Journal of Finance*, 26(2), 347-356.

Guyon, I. & Elisseeff, A. (2003). An introduction to variable and feature selection. *Journal of Machine Learning Research*, 3, 1157–1182.

Hua, Z., Wang, Y., Xu, X., Zhang, B., Liang, L. (2007). Predicting corporate financial distress based on integration of support vector machine and logistic regression. *Expert Systems with Applications*, 33 (2), 434-440.

Hui, X. & Sun, J. (2006). An application of support vector machine to companies'cial distress prediction. *Lecture Notes in Artificial Intelligence*, 3885, 274–282.

Khansari, R., Mirfeyz, F. (2009). Assessment of the structural model in predicting default KMV companies listed in Tehran Stock Exchange. *Financial reaserch*, 11 (28), 49-68. (*in Persian*)

Kohavi, K. & John, G. (1997). Wrappers for feature subset selection. *Artificial Intelligence*, 97 (1-2), 273–324.

Li, H., Li, C. J., Wu, X. J., Sun, J. (2014). Statistics-based Wrapper for feature selection: An implementation on financial distress identification with support vector machine. *Applied soft computing*, 19, 57–67.

Li, H., Sun, J. (2009). Predicting business failure using multiple case-based reasoning combined with support vector machine. *Expert Systems with Applications*, 36 (6), 10085–10096.

Min, J. H., Lee, Y. C. (2005). Bankruptcy prediction using support vector machine with optimal choice of kernel function parameters. *Expert Systems with Applications*, 28(4), 603–614.

Min, S. H., Lee, J., Han, I. (2006). Hybrid genetic algorithms and support vector machines for bankruptcy prediction. *Expert Systems with Applications*, 31(3), 652–660.

Nabavi Chashmi, A., Ahmadi, M., Mahdavi Farahabadi, S. (2010). Bankuraptcy prediction of companies with using of logit model. *Journal of financial engineering and portfolio management*, 1 (5), 55-81.

Ng, W.W.Y., Yeung, D. S., Firth, M., Tsang, C.C., Wang, X. (2008). Feature selection using localized generalization error for supervised classification problems using RBFNN, *Pattern Recognition*, 41(12), 3706–3719.

Ni, L. G., Ni, Z., Gao, W. Y. (2011). Stock trend prediction based on fractal feature selection and support vector machine. *Expert Systems with Applications*, 38(5), 5569-5576.

Oreski, S. & Oreski, G. (2014). Genetic algorithm-based heuristic for feature selection in credit risk assessment. *Expert Systems with Applications*, 41 (4), 2052-2064.

Oreski, S., Oreski, D. & Oreski, G. (2012). Hybrid system with genetic algorithm and artificial neural networks and its application to retail credit risk assessment. *Expert systems with applicars*, *39* (16), 12605-12617.

Panahi, H., Asadzadeh, A., Jalili Marand, A. (2014). Bankruptcy prediction prediction of listed companies in Tehran stock exchange market. *Financial research*, 1 (16), 57-76. (*in Persian*)

Premachandra, I. M., Bhabra, G. & Sueyoshi, T. (2009). DEA as a tool for bankruptcy assessment: a comparative study with logistic regression technique. *European Journal of Operational Research*, 192 (2), 412 - 424.

Raei, R. & Fallahpur, S. (2004). Use of neural network for financial distress prediction. *Financial research*, 6(1), 39-69. (*in Persian*)

Raei, R., Fallahpur, S. (2008). Application of support vector machine in financial distress prediction with using of financial ratios, *Financial research*, 15(53), 17-34. (*in Persian*)

Shin, K.S., Lee, T.S., Kim, H.J. (2005). An application of support vector machines in bankruptcy prediction model. *Expert Systems with Applications*, 28(1), 127-135.

Shoaf, J. S. & Foster, J. A. (1996). A genetic algorithm solution to the efficient set problem: A technique for portfolio selection based on the Markowitz model. *In Proceedings of the decision sciences institute annual meeting*, Orlando, Florida, 571-573.

Soleymani Amiri, GH. (2003). Financial ratios and financial crisis of companies in Tehran stock exchange market. *Financial research*, (15), 121-126. (*in Persian*)

Sun, J., Hu, L. (2012). Financial distress prediction using support vector machines: ensemble vs. *individual. Applied Soft*, 12(8), 2254–2265.

Tan, T. Z., Quek, C., See Ng. G. (2007). Biological brain-inspired genetic complementary learning for stock market and bank failure prediction. *Computational Intelligence*, 23 (2), 236 -261.

Vapnik, V. N. (1998). *Statistical Learning theory*, Springer, NewYork.

Whitaker, R. (1999). The Early Stage of Financial Distress. *Journal of Economics and Finance*, 23 (2), 123-133.

Wu, C.H., Tzeng, G.H., Goo, Y.J., Fang, W.C. (2007). A real-valued genetic algorithm to optimize the parameters of support vector machine for predicting bankruptcy. *Expert Systems with Applications*, 32 (2), 397-408.

Yu, L. & Liu, H. (2003). Feature selection for high-dimensional data: A fast correlation-based filter solution. *Proceedings, Twentieth International Conference on Machine Learning*, Washington, DC, United States, 2, 856-863.

Zavgren, C.V. (1985). Assessing the vulner ability to failure of American industrial firm: a logistic analysis. *Journal of Business Finance and Accounting*, 12, 19-45.