ارائه مدل پیش‌بینی ورشکستگی شرکت های بورس اوراق بهادار تهران مبتنی بر مدل ترکیبی شبکه عصبی گروهی دستکاری داده ها و الگوریتم ژنتیک

دکتر حسین وظیفه دوست [۱]
طیبه زنگنه [۲]

از صفحه: ۸۳- ۱۰۰
تاریخ ارایه: ۹۴/۱/۱۷
تاریخ پذیرش: ۹۴/۲/۴

چکیده

هدف تحقیق حاضر ارائه مدلی کارا و توانمند جهت پیش‌بینی ورشکستگی شرکت‌های تولیدی بورس اوراق بهادار تهران با استفاده از یک مدل جدید ترکیبی الگوریتم ژنتیک- شبکه گروهی دستکاری داده ها (GA-GMDH)، می‌باشد. هم چنین، با استفاده از تعدادی از پر کاربردترین روش‌های انتخاب متغیر در ادبیات پیش‌بینی ورشکستگی، مطالعه جامعی در جهت شناسائی بهترین متغیرهای پیش‌بینی کننده ورشکستگی شرکت‌های بورس اوراق بهادار تهران صورت گرفته است. به منظور ساخت مدل‌های پیش‌بینی در ابتدا با استفاده از چهار روش انتخاب متغیر ۱-آزمون T نمونه‌های مستقل (T-test)، ۲- ماتریس همبستگی(CM) ، ۳- تحلیل تشخیصی گام به گام (SDA) و ۴- تحلیل مولفه‌های اصلی (PCA)، نسبت‌های مالی نهایی از بین ۱۹ نسبت مالی متناسب با بازار سرمایه کشور، انتخاب شده است. با استفاده از نسبت‌های مالی انتخاب شده و مدل ترکیبی GA-GMDH،شبکه‌عصبی-فازی تطبیق‌پذیر(ANFIS) و رگرسیون لجستیک(LR)، ۱۲ مدل جهت پیش‌بینی ورشکستگی استخراج شد و نتایج حاصل از آن‌ها مورد مقایسه قرار گرفت. نتایج حاصل از تحقیق، نشان دهنده قابلیت بالای مدل پیشنهادی GA-GMDH در مدل‌سازی پیش‌بینی ورشکستگی و برتری آن بر روش‌های ANFIS و LR می‌باشد. همچنین،

۱.دانشیار مدیریت بازرگانی، دانشگاه آزاد اسلامی واحد علوم و تحقیقات vazifehdust@yahoo.com
۲.دانشجوی دکتری مدیریت بازرگانی، دانشگاه آزاد اسلامی واحد علوم و تحقیقات t-zangene@yahoo.com

نتایج تحقیق نشان می‌دهد که روش ماتریس همبستگی در مقایسه با سایر روش‌های انتخاب متغیر، توانایی بیشتری در انتخاب متغیرهای موثر بر پیش‌بینی ورشکستگی شرکت‌ها دارد. بنابراین، مدل CM-GA-GMDH به عنوان بهترین مدل پیش‌بینی کننده ورشکستگی شرکت‌های بورس اوراق بهادار تهران شناخته می‌شود.

واژگان کلیدی: پیش‌بینی ورشکستگی، بورس اوراق بهادار تهران، انتخاب متغیرهای پیش‌بینی، نسبت‌های مالی ، شبکه گروهی دستکاری داده‌ها، شبکه‌های عصبی- فازی تطبیق‌پذیر.

مقدمه

مساله ورشکستگی و عدم موفقیت شرکت‌ها در بازپرداخت تعهدات به عنوان یک پدیده نامطلوب مالی، همواره مشکلی در خور تامل بوده است. پیش‌بینی ورشکستگی برای اولین بار در سال ۱۹۳۲ در حوزه علمی مطرح شد. روند استفاده از مدل‌های موجود برای پیش‌بینی ورشکستگی با یک ترتیب زمانی همراه بوده است به طوری که تحقیقات اولیه با تاکید بر روش‌های آماری[1]، مانند تحلیل ممیز چند گانه، تحلیل لوجیت و پروبیت و...، انجام شده و پس از آن با گسترش روز افزون کاربرد کامپیوتر در مسائل پیش‌بینی، روش‌های هوش مصنوعی[2] مانند شبکه‌های عصبی، الگوریتم ژنتیک، منطق فازی و ... با تاکید بیشتری به کار گرفته شدند پس از آن محققان کوشیده اند که با ترکیب روش‌های ذکر شده فوق قدرت پیش‌بینی مدل‌های خود را افزایش دهند با ترکیب تکنیک‌های پیش‌بینی مدل‌های هوشمند ترکیبی (مدل‌های محاسبات نرم)[3] را برای انجام پیش‌بینی و طبقه بندی ارائه دادند. لازم به ذکر است که این روش‌ها به علت ترکیب دو روش پیش‌بینی و مزیت داشتن برتری‌های دو روش و حذف کاستی ها و محدویت آن‌ها دارای عملکرد بهتری نسبت به سایر مدل‌های آماری یا هوش مصنوعی به تنهائی هستند (Ravi Kumar,2007,1-28). اخیرا به منظور تقویت مزایا و کاهش محدویت های هر یک از مدل‌ها به تنهائی و دستیابی به مدل‌های پیش‌بینی دقیق تر، مدل‌هایی تحت عنوان مدل‌های دسته بندی ترکیبی و مونتاژی[4]، در تحقیقات به کار برده می شوند(Aziz & Dar,2006,18-33).یکی از گام‌های اساسی قبل از شروع فرآیند پیش‌بینی، مرحله انتخاب متغیرهای ورودی به مدل است. این مرحله یکی از مراحل اصلی در فرآیند داده کاوی[5] است و هدف آن حذف متغیرهای زائد و نا مرتبط از سری اولیه داده هاست. در تحقیقات انجام شده در حوزه پیش‌بینی ورشکستگی به تازگی به فرآیند انتخاب متغیر توسط روش‌های مناسب توجه شده است (Tsai ,2009,120-127). با توجه به توضیحات ذکر شده، هدف این تحقیق ساخت مدل پیش‌بینی ورشکستگی مناسب با شرایط محیطی کشور می باشد. در ادامه

1. Statistical models
2. artificial intelligences techniques
3. Hybrid Intelligent (Soft computing) models
4. hybrid and ensemble classification models
5. Data Mining

تحقیق در ابتدا مروری بر مبانی نظری و ادبیات تحقیق خواهیم داشت، سپس روش انجام تحقیق به همراه معرفی مفاهیم مربوط به هریک از روش‌های انتخاب متغیر و مدل‌های پیش‌بینی مورد استفاده در تحقیق مسئله بیان خواهد شد. پس از آن نحوه ساخت مدل‌های تحقیق و نتایج به دست آمده تشریح خواهد شد. در بخش آخر نیز نتیجه گیری از تحقیق ارائه خواهد شد.

مبانی نظری

آلتمن در سال۱۹۶۸، مدل MDA را برای نخستین بار پیشنهاد کرد. مدل او به مدل "Z-Score" معروف است. (Altman,1968,589-609). تحلیل لوجیت نخستین بار توسط مارتین برای پیش‌بینی ورشکستگی بانک‌ها پیشنهاد شد. نتایج تحقیق نشانده دقت پیش‌بینی درست ۹۶٪ و ۸۹٪ به ترتیب برای شرکت‌های سالم و ورشکسته بود (Martin,1997,249-276). پس از تحقیق مارتین، مدل لاجیت توسط اولسن برای پیش‌بینی ورشکستگی شرکت‌های تجاری به کار رفت. پس از ۱۹۸۱سال به دلیل محدودیت‌های موجود در روش‌های MDA, مطالعات ورشکستگی اغلب بر استفاده از لوجیت تمرکز یافتند. شبکه‌های عصبی مصنوعی (ANN) از پرکاربردترین مدل‌های به کار گرفته شده در مسئله پیش‌بینی ورشکستگی به شمار می رود. تفاوت در تعداد لایه‌ها، تعداد گره‌ها در هر لایه، چگونگی اتصال لایه ها و نوع الگوریتم آموزش موجب به وجود آمدن ساختارهای مختلفی از ANN می‌شوند مانند: شبکه پرسپترون چند لایه (MLP)[1]، شبکه عصبی احتمالی[2] (PNN)، شبکه مبتنی بر کارکرد شعاعی[3] (RBFN)، عصبی همبستگی کاسکودی[4] (cascor)، یادگیری برداری (LVQ)[5] و نقشه مشخصه خود سازمانده (SOM)[6]. تمامی ساختارهای ذکر شده فوق در تحقیقات تحقیقات ورشکستگی به کار گرفته شده‌اند و نتایج مطلوبی نیز داشته‌اند. از میان ساختارهای فوق پرسپترون چند لایه‌ای (MLP) یکی از رایج‌ترین ساختار ANNها برای مسائل طبقه بندی دو

1. Multi-layer perception
2. Probabilistic neural network
3. Radial basis function network
4. Cascade Correlation Neural Network
5. Learning vector quantization
6. Self organizing feature map

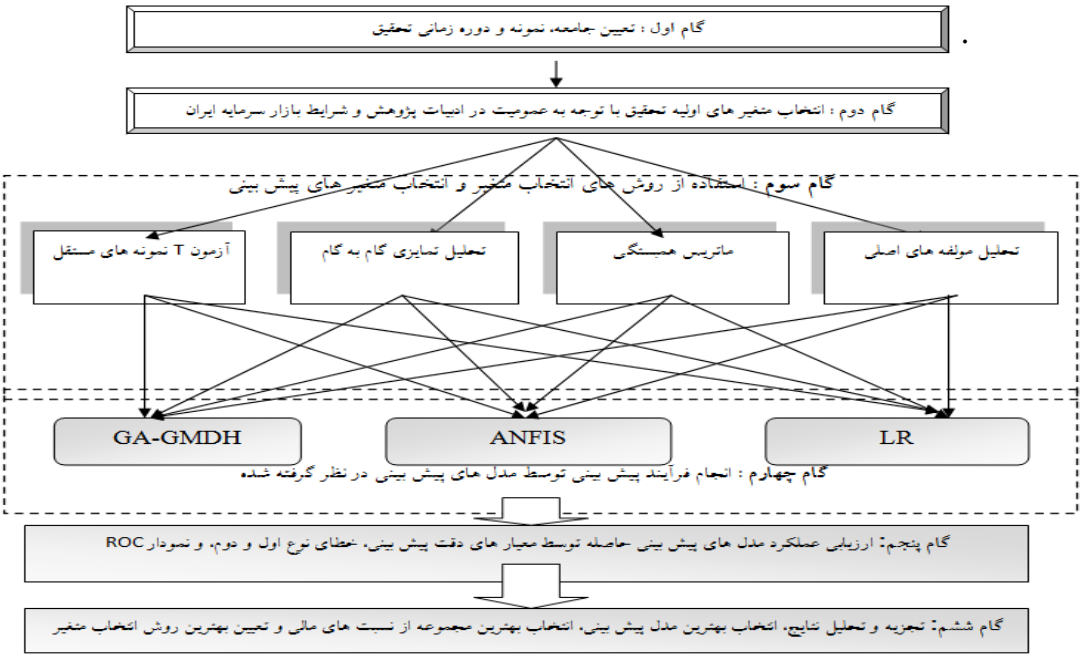
گروهی نظیر پیش‌بینی ورشکستگی است (Leea et al. 2005 1-16).

راوی و کومار (۲۲۲-۲۲۷) یک مدل قوانین فازی برای پیش‌بینی ورشکستگی در سال ۲۰۰۶ پیشنهاد دادند ونتایج آن را با مدل‌های لاجیت، پروبیت، LDA, QDA مقایسه کردند. مدل پیشنهادی آن ها با خطای نوع اول و دوم برابر با ۷/۴۸ و ۴۴/۸۳ در مجموع خطایی برابر با ۲۶/۱۶ نسبت به مدل‌های آماری ذکر شده بهتر عمل کرد. مالهاترا یک مطالعه جامع بر روی استفاده از مدل سیستم های عصبی-فازی تطبیق پذیر [1] (ANFIS) برای طبقه بندی مشتریان بانک به دو طبقه خوش حساب و بد حساب انجام داده است. آن ها نشان دادند که انفیس از مدل تحلیل ممیزی بهتر عمل کرده است(Malhotra,2002,190-211). کومار و واری نیز در سال ۲۰۰۶ در تحقیقی یک مدل مونتاژی شامل هفت مدل طبقه بندی برای پیش‌بینی ورشکستگی استفاده کردند. نتایج نشان می دهد که مدل‌های انفیس، مدل semi- online BF2 و MLP از میان هفت مدل به کار رفته مهمترین مدل‌های طبقه بندی شناخته شده اند. سعید فلاح پور(۱۳۸۳) در مقایسه‌ای بین شبکه‌های عصبی مصنوعی (ANN) و مدل تحلیل تشخیصی چندگانه (MDA) بر روی شرکت های پذیرفته شده در بورس اوراق بهادار تهران، برتری شبکه عصبی مصنوعی را در مدل‌سازی پیش‌بینی ورشکستگی نشان داد. وحید روشن (۱۳۸۶) در تحقیق خود با استفاده از مدل رگرسیون لجستیک و ۱۷ نسبت مالی، مدل‌های پیش‌بینی ورشکستگی را برای یک تا سه سال قبل از ورشکستگی برای شرکت های بورس اوراق تهران به‌دست آورد. مدل‌های برآورد شده دارای ۷۸/۲، ۷۲/۵ و ۷۶/۵ دقت پیش‌بینی به ترتیب برای یک، دو و سه سال قبل از موقع ورشکستگی بودند. آقای فرج زاده (۱۳۸۶)از برنامه‌ریزی ژنتیک جهت پیش‌بینی ورشکستگی شرکت ها استفاده نمودند. در این تحقیق از ۴۵ نسبت اولیه استفاده شده است، که اندازه شرکت نیز به عنوان یکی از متغیرها درنظر گرفته شده است، و با استفاده از روش تجزیه تحلیل گام به گام ۵ نسبت از بین این نسبت ها انتخاب شده است. نتایج حاکی از برتری مدل پیشنهادی بوده است. با توجه توضیحات و نتیجه گیری های ذکر شده در بخش هایی که از نظر گذشت، تا به حال از مدل گروهی دستکاری داده‌ها (GMDH) و ترکیب آن با الگوریتم ژنتیک جهت مدل بندی پیش‌بینی ورشکستگی استفاده نشده است.

1. Adaptive-Network-Based Fuzzy Inference System (ANFIS)

روش اجرای تحقیق

در شکل ۱ روش تحقیق این مطالعه به صورت شماتیک نشان داده شده است. در نتیجه این فرآیند تعداد ۱۲ مدل پیش‌بینی به دست خواهد آمد.



شکل ۱- فرآیند انجام تحقیق

مدل پیشنهادی تحقیق

شبکه‌های روش گروهی‌دستگاهی داده (GMDH)

یکی از انواع شبکه عصبی مصنوعی که کارائی آنها در مدل‌سازی و پیش‌بینی اثبات شده است، شبکه‌های خود-سازمانده می باشد[1]. یکی از انواع شبکه‌های خود- سازمانده، شبکه‌های چند جمله‌ای است. شبکه‌های چند جمله‌ای، نتیجه ترکیب تکنیک‌های رگرسیون خطی و شبکه‌های

1. Self-Organizing Modeling

عصبی مصنوعی هستند. یکی از الگوریتم‌های آموزش، الگوریتم GMDH[1] می‌باشد. نام GMDH، بیانگر ایده اصلی آن یعنی روال تقسیم تصادفی داده‌های ورودی مدل به دو مجموعه آموزش و تطبیق است. الگوریتم GMDH اولین بار توسط ایواختنکو در سال ۱۹۹۶ برای مدل کردن سیستم‌های پیچیده ای که شامل یک سری داده با چندین ورودی و یک خروجی بودند، مورد استفاده قرار گرفت. در واقع هدف اصلی این روش، ساختن تابعی در یک شبکه بر اساس تابع انتقال درجه دوم است. مزیت اصلی GMDH نسبت به شبکه‌های عصبی مرسوم، به دست آوردن یک مدل ریاضی برحسب چندجمله ایها، برای فرایند مورد بررسی است. این مدل ریاضی می‌تواند در شناسائی و یا حتی توصیف کامل فرایند مورد استفاده قرار گیرد. به‌طور کلی برای مدل کردن سیستم‌های پیچیده‌ای که شامل مجموعه‌ای از داده‌ها با چندین ورودی و یک خروجی می‌باشند می‌توان از چند جمله‌ای‌های VKG[2] (رابطه۱) استفاده نمود.

$$y = a_{\circ} + \sum_{i=1}^{n} a_i x_i + \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij} x_i x_j + \sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n} a_{ij} x_i x_j x_k + \ldots \qquad (۱)$$

که در آن $x = (x_1, x_2, \ldots, x_n)$ بردارهای ورودی، y خروجی مدل و a_i ها ضرایب چند جمله‌ای می‌باشند. چند جمله‌ای های VKG با استفاده از چند جمله‌ای های درجه دوم تقریب زده می‌شوند. این چند جمله‌ای های درجه دوم بر اساس ترکیبات دوتایی ورودی های شبکه ساخته می‌شوند. الگوریتم GMDH با استفاده از این ایده به عنوان روش یادگیری، برای مدل‌سازی سیستمهای پیچیده معرفی شده است.شبکه عصبی GMDH دارای ساختار یک شبکه چند لایه و رو به جلو می‌باشد و در برگیرنده مجموعه ای از نرون‌ها است که از پیوند جفت‌های مختلف ورودی از طریق یک چند جمله‌ای درجه دوم به وجود می‌آیند. هر لایه در این شبکه از یک یا چند واحد پردازشگر تشکیل شده است که هر یک از آنها دارای دو ورودی و یک خروجی می‌باشد. این واحدها عملاً نقش اجزای سازنده مدل را ایفا می‌نمایند و به شکل یک چند جمله‌ای درجه دوم فرض می‌شوند. پارامترهای مجهول الگوریتم GMDH ضرایب چند جمله‌ای می‌باشد. جهت محاسبه مقدار خروجی $\hat{y}_i$ برای هر بردار ورودی $x = (x_{i1}, x_{i2}, \ldots, x_{in})$ براساس رابطه

1. Group Method of Data Handling
2. Volterra-Kolmogorov-Gabor

(۳)، باید میانگین مربعات خطا (رابطه ۲) به حداقل برسد.

$$e = \sum_{i+1}^{n} \left(\hat{y}_i - y_i\right)^2 \qquad (۲)$$

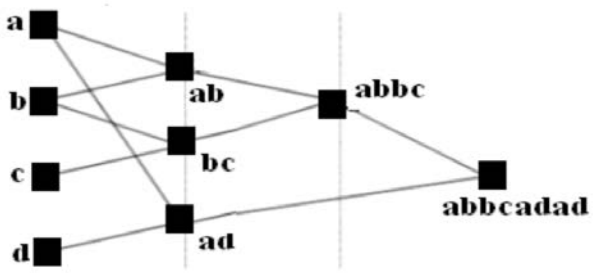
برای یافتن مقدار حداقل خطا، از مشتق جزیی رابطه ۲ استفاده می‌شود. با جایگذاری از رابطه (۲) در این مشتق جزئی، یک معادله ماتریسی($Aa = y$) به‌دست می‌آید. در این معادله $a = \{a_\circ, a_1, a_2, a_3, a_4, a_5\}$ ، $Y = \{y_1, ..., y_m\}^T$ و ماتریس A می‌باشد. یک روش حل برای این معادله ماتریسی($Aa = y$)، استفاده از روش تجزیه به مقدارتکین (SVD[1]) است. هنگام ساختن یک شبکه GMDH همه ترکیبات ورودی‌ها ایجاد شده به اولین لایه شبکه فرستاده می‌شود سپس خروجی‌ها ازاین لایه برای ورود به لایه بعدی طبقه بندی و انتخاب می شود که باهمه ترکیبات خروجی های انتخاب شده به لایه ۲ فرستاده می شود این روند تا آنجا ادامه می‌یابد که لایه (n+1) نتیجه بهتری ازلایه n ایجاد کند. هنگامی که لایه n+1 پیدا شد که نسبت به لایه n بدتر بود فرآیند متوقف می شود.

مدل ترکیبی شبکه‌های GMDH با الگوریتم ژنتیک

روش‌های تکاملی مانند الگوریتم ژنتیک کاربرد وسیعی در مراحل مختلف طراحی شبکه‌های عصبی به دلیل قابلیت‌های منحصر به فرد خود در پیدا کردن مقادیر بهینه و امکان جستجو در فضاهای غیر قابل پیش‌بینی، دارند. در این ساختار ترکیبی، الگوریتم ژنتیک تعداد بهینه نرون ها در هر لایه پنهان و نحوه اتصال آن‌ها ارائه می‌دهد که در نتیجه آن سری بهینه از ضرایب مناسب توابع درجه دوم (چند جمله ای) پیش‌بینی کننده متغیر وابسته، به‌دست می‌آید (Varetto, 1998,1421-1439). در ساختار ترکیبی مورد نظر به طور کلی سه تابع هدف[2] اصلی به صورت خطای آموزش [3]، خطای پیش‌بینی[4] و تعداد نرون ها[5] وجود دارد. بنابراین ساختار بهینه بر مبنای توازن (Trade-offs) بین این توابع هدف به دست می‌آید. در تحقیق حاضر تابع هدف، به صورت مینیمم کردن خطای پیش‌بینی دو مجموعه

1. Singular Value Decomposition
2. Objective function
3. Training error
4. Prediction Error
5. Number of Neurons

آموزش و تست در نظر می‌گرفته می‌شود، و کروموزومی(تابع چند جمله‌ای از متغیرهای پیش‌بینی) به عنوان جواب بهینه انتخاب می‌شود که کمترین میزان خطا را در هر دو مجموعه آموزش و تست داشته باشد. معیار میانگین مربع مربعات خطا (MSE) به عنوان معیار خطای پیش‌بینی درنظر گرفته می‌شود. در ساختار پیشنهادی GA-GMDH ، نمایش ژن ها و کروموزوم‌ها به صورت رشته نمادین[1] (غیر دودویی[2]) در نظر گرفته می شوند که در آن هر یک از متغیر ها به صورت حروف الفبایی (a,b,…) نمایش داده می شوند. برای عمومیت بخشیدن به شبکه‌های عصبی GMDH، باید قید استفاده از لایه مجاور در ساختن لایه بعد حذف شود. این نوع از شبکه‌های عصبی با ساختار عمومی[3] نامیده می شوند (GS)و برای ساخت لایه جدید از تمام لایه‌های قبلی (شامل لایه ورودی) استفاده می‌کنند. در شکل۲ این امر نشان داده شده است. در واقع برای ساخت کروموزوم نرون خروجی، از تمام لایه ها (لایه دو سه) و نه لزوما از لایه سه، استفاده شده است. بعد از تعریف کروموزوم‌ها و ژن‌ها از اپراتور های ادغام[4] و جهش[5] برای تولید دو فرزند از دو والد استفاده می‌شود. می‌شود. از روش چرخ رولت [6] برای انتخاب دو کروموزوم والد استفاده می‌شود.



شکل ۲- نمایش کروموزوم یک شبکه GMDH عمومی

1. Symbolic String
2. Not binary String
3. Generalized Structure
4. Crossover
5. Mutation
6. Roulette Wheel

ساخت مدل‌های پیش بینی

جامعه، نمونه آماری و دوره زمانی تحقیق

بازه زمانی مورد نظر برای تحقیق یک دوره ۱۱ ساله از سال ۱۳۸۲ تا سال ۱۳۹۲ در نظر گرفته شده است. نمونه شرکت های ورشکسته تحقیق، شامل تمام شرکت های پذیرفته شده در بورس اوراق بهادار تهران می‌باشند که در طی دوره تحت بررسی مشمول ماده ۱۴۱ قانون تجارت شده‌اند. تعداد ۶۸ شرکت ورشکسته در طی دوره مورد بررسی که اطلاعات آن ها به طور کامل وجود داشت، به عنوان نمونه شرکت های ورشکسته در نظر گرفته شد. برای انطباق با شرکت‌های ورشکسته ، ۶۸ شرکت غیر ورشکسته نیز، به صورت نمونه گیری تصادفی انتخاب شدند که سعی شد از نظر سال مالی با شرکت های ورشکسته مطابقت داشته باشند. نمونه شامل ۱۳۶ شرکت به دو بخش نمونه آموزش و نمونه تست (نمونه بسط داده شده) تقسیم می شود. برخی از محققان نسبت نمونه آزمایش به نمونه تست را ۶۰ به ۴۰ یا ۷۰ به ۳۰ در نظر می‌گیرند (Zadeh, 1994,48-56). در تحقیق حاضر نمونه به صورت تصادفی به تعداد ۹۶ شرکت (۴۸ شرکت سالم و ۴۸ شرکت ورشکسته) جهت آموزش و ۴۰ شرکت(۲۰ شرکت سالم و ۲۰ شرکت ورشکسته) به منظور تست مدل ها، تقسیم می‌شود.

فرآیند انتخاب متغیر های پیش‌بینی

۲۰ نسبت مالی به عنوان نسبت‌های مالی اولیه تحقیق در نظر گرفته شد، که در جدول ۱ معرفی شده‌اند.

جدول ۱- مجموعه نسبت های مالی اولیه تحقیق

کد	نوع نسبت‌های مالی	نسبت مالی	کد	نوع نسبت‌های مالی	نسبت مالی
X_1	**سودآوری**	سود قبل از بهره و مالیات به کل دارایی ها	X_{11}	**کارائی یا فعالیت**	فروش به سرمایه در گردش
X_2	**نقدینگی**	دارائی جاری به کل دارائی	X_{12}		فروش به دارائی جاری
X_3		دارائی آنی به کل دارائی	X_{13}		فروش به حقوق صاحبان سهام
X_4		سرمایه در گردش به کل دارائی	X_{14}		فروش به دارایی آنی
X_5		دارائی جاری به بدهی جاری (نسبت جاری)	X_{15}		سود و زیان انباشته به کل دارائی
X_6		دارائی آنی به بدهی جاری (نسبت آنی)	X_{16}	**نسبت‌های توانائی پرداخت بدهی(اهرمی یا ساختار سرمایه)**	کل بدهی به حقوق صاحبان سهام
X_7		وجه نقد به کل بدهی	X_{17}		کل بدهی به کل دارائی
X_8	**پوششی**	هزینه بهره به سود ناخالص	X_{18}		بدهی جاری به حقوق صاحبان سهام
X_9	**کارائی یا فعالیت**	فروش به کل دارائی	X_{19}		بدهی بلند مدت به حقوق صاحبان سهام
X_{10}		فروش به وجه نقد			

آزمون T نمونه های مستقل جهت پاسخ‌گوئی به این سوال که آیا میانگین یک نسبت مالی در گروه شرکت‌های ورشکسته با میانگین همین نسبت مالی در گروه شرکت‌های غیر ورشکسته از لحاظ آماری متفاوت است یا خیر؟ سطح اطمینان ۹۹٪ برای تحلیل نتایج حاصل از آزمون در نظر گرفته می‌شود. در روش ماتریس همبستگی، چنانچه همبستگی یک نسبت مالی با متغیر وابسته (ورشکستگی)، به طور معنی‌داری از مخالف صفر باشد، آن نسبت مالی به عنوان یکی از متغیرهای

پیش‌بینی کننده ورشکستگی انتخاب می‌شود. آزمون فرض فوق برای هر یک نسبت‌های مالی اولیه در سطح اطمینان ۹۹٪ تست می‌شود. هدف روش تجزیه و تحلیل مولفه‌های اصلی کاهش داده‌ها[1]، حذف متغیرهای مازاد و همبسته از فرآیند مدل‌سازی و جایگزین کردن آن‌ها توسط متغیرهای جدید نا همبسته است. معمولا تعداد مشخصی (و نه همه مولفه‌ها) حداکثر تغییرپذیری در نمونه را شامل می‌شوند، که می‌توانند جایگزین متغیرهای اولیه گشته و وارد مدل شوند. برای پاسخگویی به این سوال که انجام روش PCA بر روی داده‌های تحقیق امکان‌پذیر می‌باشد و نتایج مطلوب خواهد داشت یا خیر، از دو آزمون KMO وBartlett's test of sphericity استفاده می شود. نتایج به دست آمده از این دو آزمون(KMO=0.647; Bartlett's test=2.374, df=190, Sig.= 0.00) نشانگر قابلیت مطلوب روش PCA بر روی داده‌های تحقیق می‌باشد. مشاهده می‌شود که ۷ مولفه‌ی اولیه تقریبا ۸۰٪ سطح تغییرپذیری در داده های اولیه را پوشش می‌دهند. برای انجام فرآیند روش تحلیل تشخیصی گام به گام مقدار ۰.۰۵ برای آماره F را، حداکثر سطح اهمیت که یک متغیر می‌تواند وارد مدل شود و حد آستانه خروج یک متغیر از مدل نیز ۰.۱در نظر گرفته می‌شود. در نتیجه اعمال مدل SDA بر روی ۱۹ متغیر اولیه، ۵ متغیر انتخاب شدند. در جدول ۲ نسبت‌های مالی انتخاب شده توسط چهار روش انتخاب متغیر ذکر شده است.

جدول ۲- مجموعه نسبت های مالی انتخاب شده توسط روش‌های انتخاب متغیر

روش انتخاب متغیر	نسبت های مالی انتخاب شده
آزمون T نمونه‌های مستقل	سود قبل ازبهره ومالیات به کل دارائی، کل بدهی به کل دارایی، وجه نقد به کل بدهی، سودوزیان انباشته به کل دارایی، دارایی‌های سریع به بدهی‌های جاری، کل بدهی به حقوق صاحبان سهام، فروش به دارایی جاری
ماتریس همبستگی	سودوزیان انباشته به کل دارایی، سود قبل ازبهره ومالیات به کل دارایی، کل بدهی به کل دارایی، دارایی‌های سریع به بدهی‌های جاری، کل بدهی به حقوق صاحبان سهام، وجه نقد به کل بدهی، هزینه بهره به سود ناخالص، دارایی جاری به بدهی جاری، فروش به حقوق صاحبان

1. Data reduction

تحلیل تشخیصی گام به گام	سود و زیان انباشته به کل دارائی ها، کل بدهی ها به کل دارایی ها، سود قبل ازبهره ومالیات به کل دارائی
تجزیه و تحلیل مولفه‌های اصلی	۷ مولفه ی اصلی جایگزین ۲۰ نسبت مالی اولیه شده اند.

ساخت مدل‌های پیش بینی

برای اجرای مدل پیشنهادی **GA-GMDH** تعداد لایه های پنهان برابر ۲، جمعیت اولیه برابر با ۱۰۰، تعداد نسل ها برابر با۳۰۰ و عملگرهای تقاطع و جهش به ترتیب برابر با ۰.۹ و ۰.۱ در نظر گرفته شد. در نتیجه تحلیل نتایج به دست آمده ساختار ANFIS با پارامترهایی که در جدول ۳ نشان داده شده است، جهت انجام فرآیند پیش‌بینی با استفاده از نسبت های مالی انتخاب شده توسط روش‌های انتخاب متغیر، در نظر گرفته شد.

جدول ۳-پارامتر های درنظر گرفته شده برای مدل ANFIS پیشنهادی

تلرانس خطا	تعداد تکرار	الگوریتم آموزش	تابع عضویت خروجی	نوع توابع عضویت متغیرهای پیش‌بینی	نرخ رد	نرخ پذیرش	شعاع خوشه‌ها	روش خوشه‌بندی
0	5	ترکیبی	خطی	گوسین	۰/۱۵	۰/۵	۰/۰۱	خوشه بندی کاهشی

تجزیه و تحلیل نتایج

نتایج حاصل از ۱۲ مدل پیش‌بینی ساخته شده بر روی نمونه تست در جدول ۴ ارائه شده است.

جدول ۴- عملکرد پیش‌بینی مدل ها بر روی نمونه آزمایشی (ارقام به درصد)

مدل پیش‌بینی	مدل انتخاب متغیر	دقت طبقه بندی	خطای نوع اول	خطای نوع دوم	AUC	رتبه بندی مدل ها
T-test	GA-GMDH	۸۷.۵	۲۰	۵	۰.۸۸۴	۱
	ANFIS	۸۲.۵	۳۰	۵	۰.۸۳۶	۳

	LR	۸۲.۵	۲۰	۱۵	۰.۸۴۲	۲
CM	GA-GMDH	۹۰	۱۵	۵	۰.۹۰۸	۲
	ANFIS	۹۰	۱۵	۵	۰.۹۰۸	۱
	LR	۸۵	۱۵	۱۵	۰.۸۶۶	۳
SDA	GA-GMDH	۸۷.۵	۲۰	۵	۰.۸۶۳	۱
	ANFIS	۸۲.۵	۲۰	۱۵	۰.۸۴۸	۲
	LR	۸۲.۵	۲۵	۱۰	۰.۸۴۵	۳
PCA	GA-GMDH	۸۲.۵	۲۰	۱۵	۰.۸۴۲	۱
	ANFIS	۷۵	۲۵	۲۵	۰.۷۵۶	۲
	LR	۷۲.۵	۳۰	۲۵	۰.۷۵۳	۳

نتایج نشان می دهد که مدل GA-GMDH در سه حالت رتبه اول و تنها با ترکیب با روش ماتریس همبستگی(CM) در رتبه دوم و بعد از مدل ANFIS قرار می گیرد. بنابرای مدل ترکیبی الگوریتم ژنتیک با شبکه گروهی دستکاری داده (GA-GMDH) به عنوان مدل برتر نسبت به مدل‌های ANFIS و LR در مدل بندی پیش‌بینی ورشکستگی شناخته می‌شود. همان‌طور که انتظار داشتیم، نتایج نشان می دهد دقت مدل‌های پیش‌بینی به وضوح متاثر از روش‌های انتخاب متغیر می‌باشند. به طور مثال دقت مدل ترکیبی GA-GMDH با روش PCA برابر با ۸۲.۵ ٪ می‌باشد در حالی که دقت همین مدل با متغیرهای انتخاب شده توسط روش CM برابر ۹۰٪ در نمونه آزمایشی است. با توجه به نتایج به دست آمده می توان گفت که روش ماتریس همبستگی (CM)، منجر به بالاترین دقت طبقه بندی در ۳ مدل GA-GMDH ، ANFIS و LR، نسبت

به سایر روش‌های انتخاب متغیر می‌شود. روش‌های انتخاب متغیر Ttest و SDA برای مدل‌های پیش‌بینی ANFIS و LR، دارای نتایج یکسان می‌باشند. اما از آن جا که دقت طبقه بندی مدل SDA-GA-GMDH نسبت به مدل T-test بیشتر و خطای نوع اول آن کمتر است، می توان گفت که روش انتخاب متغیر SDA در رتبه دوم قرار خواهد گرفت. مقایسه نتایج Ttets و PCA نیز، به وضوح نشان دهنده برتری روش انتخاب متغیر Ttest به PCA می‌باشد. در توجیه نتایج به دست آمده می توان گفت که از آن جائیکه منطق روشPCA تنها بر اساس همبستگی بین متغیر های پیش‌بینی کننده (نسبت های مالی) بدون توجه به متغیر وابسته(ورشکستگی) استوار است، این روش نمی تواند به عنوان یک روش مناسب انتخاب متغیر در مسئله پیش‌بینی ورشکستگی ایفای نقش کند. بنابراین روش Ttest و PCA به ترتیب در رتبه‌های سوم و چهارم قرار می‌گیرند.

دو مدل CM-ANFIS و CM-GA-GMDH با عملکرد مشابه با یکدیگر به عنوان بهترین مدل‌های پیش‌بینی ورشکستگی انتخاب می‌شوند. مدل CM-ANFIS، ۱۰۰ ٪ و ۹۰٪ از شرکت‌ها در نمونه آموزش و تست به درستی طبقه بندی می کند، همچنین دارای خطای نوع اول ۱۵٪ در نمونه تست می‌باشد. مدل پیش‌بینی CM-GA-GMDH نیز دارای دقت پیش‌بینی ۹۱.۲۷ و ۹۰ درصد به ترتیب در نمونه های آموزش و تست می باشد. در منحنی ROC هر چه نمودار مربوط به یک مدل به نقطه چپ و بالای محور ها نزدیکتر باشد، نشان دهنده بالاتر بودن قدرت تفکیک کنندگی آن مدل است. منحنی ROC این دو مدل نیز کاملا بر یکدیگر منطبق می باشد و دارای سطح زیر منحنی برابر با ۰.۹۰۸ می‌باشد که نشان دهنده توان تشخیصی (قدرت تفکیک کنندگی) بالای دو مدل انتخابی می‌باشد.

جهت مقایسه دو مدل ساخته شده می توان گفت که مزیت اصلی مدل GMDH نسبت به شبکه‌های عصبی مرسوم، به دست آوردن یک مدل ریاضی برحسب چندجمله ای‌ها، برای فرآیند مورد بررسی است. این مدل ریاضی می‌تواند الگوی مناسبی را در جهت شناسائی و یا حتی توصیف کامل فرایند مورد استفاده در اختیار مدیران، تحلیل گران و سرمایه گذاران قرار دهد، در حالی که خروجی مدل ANFIS تنها مجموعه ای از قواعد بین متغیرهای پیش‌بینی را نشان می‌دهد که ممکن است برای استفاده کننده گان ملموس نباشد و در نتیجه درک و استفاده از آن‌ها با مشکلاتی همراه باشد. از طرفی نتایج مدل ANFIS در نمونه آزمایشی و آموزشی نشان دهنده

وقوع پدیده Overfitting می‌باشد که در فرآیند مدل‌سازی و پیش‌بینی امر مطلوبی به شمار نمی‌آید. با توجه به توضیحات فوق می‌توان گفت که علیرغم عملکرد مشابه دو مدل ساخته شده، مدل GA-GMDH دارای برتری نسبی بر مدل ANFIS می‌باشد. از این رو مدل CM-GA-GMDH به عنوان بهترین مدل پیش‌بینی ورشکستگی تحقیق در نظر گرفته می‌شود. مدل ورشکستگی ساخته شده توسط مدل CM-GA-GMDH و ضریب تاثیرگذاری نسبت‌های مالی بر متغیر ورشکستگی شرکت ها،در جدول شماره ۳ نشان داده شده است.

جدول ۳- مدل پیش‌بینی ورشکستگی CM-GAGMDH

$$Y = -0.107 + 0.866z_3 + 0.317z_4 + 0.163z_3^2 - 0.265z_4^2 - 0.430z_3z_4$$

Where:

$$z_1 = 0.725 - 4.108x_1 + 0.258x_6 + 1.321x_1^2 - 0.382x_6^2 + 1.95x_1x_6$$

$$z_2 = -0.33 + 2.171x_6 - 1.003x_{17} - 0.693x_6^2 + 0.436x_{17}^2 + 0.099x_6x_{17}$$

$$z_3 = -0.347 + 1.219x_1 + 3.193z_1 - 0.444x_1^2 - 2.247z_1^2 - 7.051x_1z_1$$

$$z_4 = 0.072 + 0.463x_8 + 0.787z_2 - 0.028x_8^2 - 0.109z_2^2 - 1.089x_8z_2$$

*:به منظور یافتن تعریف متغیر های x_i به جدول ۱ مراجعه شود.

مدل پیش‌بینی به دست آمده از مدل CM-GMDH (جدول ۳) تنها شامل نسبت های مالی سود قبل ازبهره ومالیات به کل دارایی (X_1)، کل بدهی به کل دارایی (X_{17})، دارایی‌های سریع به بدهی‌های جاری (X_6)، هزینه بهره به سود ناخالص (X_8)، می باشد، و نشان دهنده این امر است که متغیر های دیگر انتخاب شده توسط روش ماتریس همبستگی تاثیر شایانی بر انجام فرآیند پیش‌بینی ندارند.

نتیجه گیری

از ورشکستگی به عنوان مهم ترین ریسکی که بر شرکت و سرمایه گذاران وارد می‌شود، یاد می‌کنند. در اغلب تحقیقات انجام گرفته در زمینه پیش‌بینی ورشکستگی به فرآیند انتخاب متغیرهای

پیش‌بینی توجه کافی مبذول نگردیده است. در اغلب تحقیقات انجام گرفته در ایران نیز، شناسایی متغیرهای پیش‌بینی کننده ورشکستگی، آنچنان‌که در تحقیق حاضر به آن پرداخته می‌شود، مورد توجه قرار نگرفته شده است. دلایل بالا انگیزه‌ای شد تا در این تحقیق با شناسائی متغیر های تاثیرگذار در زمینه پیش‌بینی ورشکستگی و مدل‌سازی یک روش جدید (GA-GMDH) که در مسئله پیش‌بینی ورشکستگی تا به حال به کار برده نشده است، مدلی جهت پیش‌بینی ورشکستگی شرکت‌های بورس اوراق بهادار تهران طراحی شود. با توجه نتایج به دست آمده مدل پیشنهادی تحقیق، دارای قابلیت بالایی در مدل‌سازی پیش‌بینی ورشکستگی می‌باشد. مقایسه نتایج این مدل با مدل‌های ANFIS و مدل رگرسیون لجستیک گواهی بر این ادعا است. روش ماتریس همبستگی در مقایسه با سایر روش‌های انتخاب متغیر، توانایی بیشتری در انتخاب متغیر های موثر بر پیش‌بینی ورشکستگی شرکت‌ها دارد. بهترین مدل پیش‌بینی ورشکستگی ساخته شده در تحقیق حاضر، ترکیب مدل GA-GMDH با روش انتخاب متغیر ماتریس همبستگی (CM-GA-GMDH) می‌باشد. این مدل مجموعه نسبت‌های مالی انتخاب شده توسط روش ماتریس همبستگی را دوباره تعدیل کرده و در نهایت نسبت‌های مالی "سود قبل ازبهره ومالیات به کل دارایی"، "کل بدهی به کل دارایی"، "دارایی‌های سریع به بدهی‌های جاری"، "هزینه بهره به سود ناخالص" را به عنوان متغیرهای نهایی پیش‌بینی کنندگی ورشکستگی شرکت های بورس اوراق بهادار تهران معرفی کرده است.

منابع

۱- فلاح پور، سعید،(۱۳۸۳)، پیش بینی درماندگی مالی شرکت ها با استفاده از شبکه های عصبی مصنوعی، دانشگاه تهران، دانشکده مدیریت، ۱۳۸۳.

۲- روشن قلب، وحید، بررسی و شناخت نسبت های موثر بر پیش بینی ورشکستگی در شرکت های پذیرفته شده در بورس اوراق بهادار تهران، دانشگاه تهران، دانشکده مدیریت، ۱۳۸۶.

۳- فرج زاده دهکردی، حسن، (۱۳۸۶)، کاربرد الگوریتم ژنتیک در مدلبندی پیش‌بینی ورشکستگی، پایان نامه کارشناسی ارشد، دانشگاه تربیت مدرس، ۱۳۸۶.

4- McKee, T.E., Lensberg, T., "Genetic programming and rough sets: A hybrid approach to bankruptcy classification." European Journal of Operational Research, 138, 2002, 436-451.

5- Haber, J., " Theoretical development of bankruptcy prediction variables", the Journal of Theoretical Accounting Research, 2, 2006, 82-101.

6- Tsai, C. F., "Feature selection bankruptcy prediction", Knowledge-Based Systems, 22(2), 2009, 120-127.

7- Aziz, M. A., Dar, H. A., "Predicting corporate bankruptcy: Where we stand?", Corporate Governance, 6(1), 2006, 18–33.

8- Ravi Kumar, P., Ravi, V., "Bankruptcy prediction in banks and firms via statistical and intelligent techniques – A review", European Journal of Operational Research, 180 (1), 2007, 1–28.

9- Martin,D., "Early warning of bank failure:A logit regression approach", Journal of banking and Finance,1, 1997, 249-276.

10- Tabachnich, B., Fidell, L. "Using multivariate Statistics" (4th ed.) .Boston: Allyn & Bacon, 2001.

11- Kumar, P.R., Ravi, V., "Bankruptcy prediction in banks byfuzzyrule based classifier" , Proceedings of the 2006 first international conference on digital information management, 2006, 222–227.

12- Verikas, A., Kalsyte, Z., Bacauskie., Gelzinis, A., "Hybrid and ensemble-based soft computing techniques in bankruptcy prediction: a survey" , Soft Comput, 2009.

13- Zadeh, L.A. , "Soft computing and fuzzy logic", IEEE Software 11 (6),1994, pp. 48–56.

14- Malhotra, R., Malhotra, D.K., "Differentiating between good credits and bad credits using neuro-fuzzy systems", European Journal of Operational Research, 136, 2002, 190-211.

15- Ravi kumar, P., Ravi, V., "Bankruptcy prediction in banks by an ensemble classifier", Industrial technology. IEEE International Conference on ICIT,2006, 2032–2036.

16- Karles, G.V., Prakash, A.J., " Multivariate normality and forecasting of business bankruptcy"., Journal of Business Finance & Accounting, 14, 1987, 573-593.

17- Jang, J.-S. R., " Neuro-Fuzzy and Soft Computing", Prentice- Hall: New Jersey, 1997.

18- Ivakhnenko, A.G. "Group Method of Data Handling — a rival of the method of stochastic approximation", Soviet Automatic Control, 13, 1966, 43–71.

Varetto, F., " Genetic algorithm application in the analysis of insolvency risk", Journal of Banking and Finance, 22, 1998, 1421-1439.