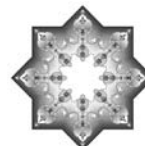


ارائه مدل پیش بینی ورشکستگی  
شرکت های بورس اوراق بهادار تهران  
مبتنی بر مدل ترکیبی شبکه عصبی  
گروهی دستکاری داده ها و الگوریتم ژنتیک



دکتر حسین وظیفه دوست<sup>۱</sup>  
طیبه زنگنه<sup>۲</sup>

از صفحه: ۸۲-۱۰۰  
تاریخ ارایه: ۹۴/۱/۱۷  
تاریخ پذیرش: ۹۴/۲/۴

### چکیده

هدف تحقیق حاضر ارائه مدلی کارا و توانمند جهت پیش بینی ورشکستگی شرکت های تولیدی بورس اوراق بهادار تهران با استفاده از یک مدل جدید ترکیبی الگوریتم ژنتیک- شبکه گروهی دستکاری داده ها (GA-GMDH)، می باشد. هم چنین، با استفاده از تعدادی از پرکاربردترین روش های انتخاب متغیر در ادبیات پیش بینی ورشکستگی، مطالعه جامعی در جهت شناسایی بهترین متغیرهای پیش بینی کننده ورشکستگی شرکت های بورس اوراق بهادار تهران صورت گرفته است. به منظور ساخت مدل های پیش بینی در ابتدا با استفاده از چهار روش انتخاب متغیر ۱-آزمون T نمونه های مستقل (T-test)، ۲- ماتریس همبستگی (CM)، ۳- تحلیل تشخیصی گام به گام (SDA) و ۴- تحلیل مولفه های اصلی (PCA)، نسبت های مالی نهایی از بین ۱۹ نسبت مالی متناسب با بازار سرمایه کشور، انتخاب شده است. با استفاده از نسبت های مالی انتخاب شده و مدل ترکیبی GA-GMDH، شبکه عصبی-فازی تطبیق پذیر (ANFIS) و رگرسیون لجستیک (LR)، ۱۲ مدل جهت پیش بینی ورشکستگی استخراج شد و نتایج حاصل از آن ها مورد مقایسه قرار گرفت. نتایج حاصل از تحقیق، نشان دهنده قابلیت بالای مدل پیشنهادی GA-GMDH در مدل سازی پیش بینی ورشکستگی و برتری آن بر روش های ANFIS و LR می باشد. 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 تطبیق پذیر.

Archive of SID

## مقدمه

مساله ورشکستگی و عدم موفقیت شرکت ها در بازپرداخت تعهدات به عنوان یک پدیده نامطلوب مالی، همواره مشکلی در خور تامل بوده است. پیش بینی ورشکستگی برای اولین بار در سال ۱۹۳۲ در حوزه علمی مطرح شد. روند استفاده از مدل های موجود برای پیش بینی ورشکستگی با یک ترتیب زمانی همراه بوده است به طوری که تحقیقات اولیه با تاکید بر روش های آماری<sup>۱</sup>، مانند تحلیل ممیز چند گانه، تحلیل لوجیت و پروبیت و...، انجام شده و پس از آن با گسترش روز افزون کاربرد کامپیوتر در مسائل پیش بینی، روش های هوش مصنوعی<sup>۲</sup> مانند شبکه های عصبی، الگوریتم ژنتیک، منطق فازی و ... با تاکید بیشتری به کار گرفته شدند پس از آن محققان کوشیده اند که با ترکیب روش های ذکر شده فوق قدرت پیش بینی مدل های خود را افزایش دهند با ترکیب تکنیک های پیش بینی مدل های هوشمند ترکیبی (مدل های محاسبات نرم)<sup>۳</sup> را برای انجام پیش بینی و طبقه بندی ارائه دادند. لازم به ذکر است که این روش ها به علت ترکیب دو روش پیش بینی و مزیت داشتن برتری های دو روش و حذف کاستی ها و محدودیت آن ها دارای عملکرد بهتری نسبت به سایر مدل های آماری یا هوش مصنوعی به تنهایی هستند (Ravi Kumar, 2007, 1-28). اخیرا به منظور تقویت مزایا و کاهش محدودیت های هر یک از مدل ها به تنهایی و دستیابی به مدل های پیش بینی دقیق تر، مدل هایی تحت عنوان مدل های دسته بندی ترکیبی و مونتاژی<sup>۴</sup>، در تحقیقات به کار برده می شوند (Aziz & Dar, 2006, 18-33). یکی از گام های اساسی قبل از شروع فرآیند پیش بینی، مرحله انتخاب متغیرهای ورودی به مدل است. این مرحله یکی از مراحل اصلی در فرآیند داده کاوی<sup>۵</sup> است و هدف آن حذف متغیرهای زائد و نا مرتبط از سری اولیه داده هاست. در تحقیقات انجام شده در حوزه پیش بینی ورشکستگی به تازگی به فرآیند انتخاب متغیر توسط روش های مناسب توجه شده است (Tsai, 2009, 120-127). با توجه به توضیحات ذکر شده، هدف این تحقیق ساخت مدل پیش بینی 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 نتیجه‌گیری از تحقیق ارائه خواهد شد.

### مبانی نظری

آلتمن در سال ۱۹۶۸، 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 ۱۹۸۱ سال 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 می‌رود. تفاوت در تعداد لایه‌ها، تعداد گره‌ها در هر لایه، چگونگی اتصال لایه‌ها و نوع الگوریتم آموزش موجب به وجود آمدن ساختارهای مختلفی از ANN می‌شوند مانند: شبکه پرسپترون چند لایه (MLP)<sup>۱</sup>، شبکه عصبی احتمالی<sup>۲</sup> (PNN)، شبکه مبتنی بر کارکرد شعاعی<sup>۳</sup> (RBFN)، عصبی همبستگی کاسکودی<sup>۴</sup> (cascor)، یادگیری برداری (LVQ)<sup>۵</sup> و نقشه مشخصه خود سازمانده (SOM)<sup>۶</sup>. تمامی ساختارهای ذکر شده فوق در 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

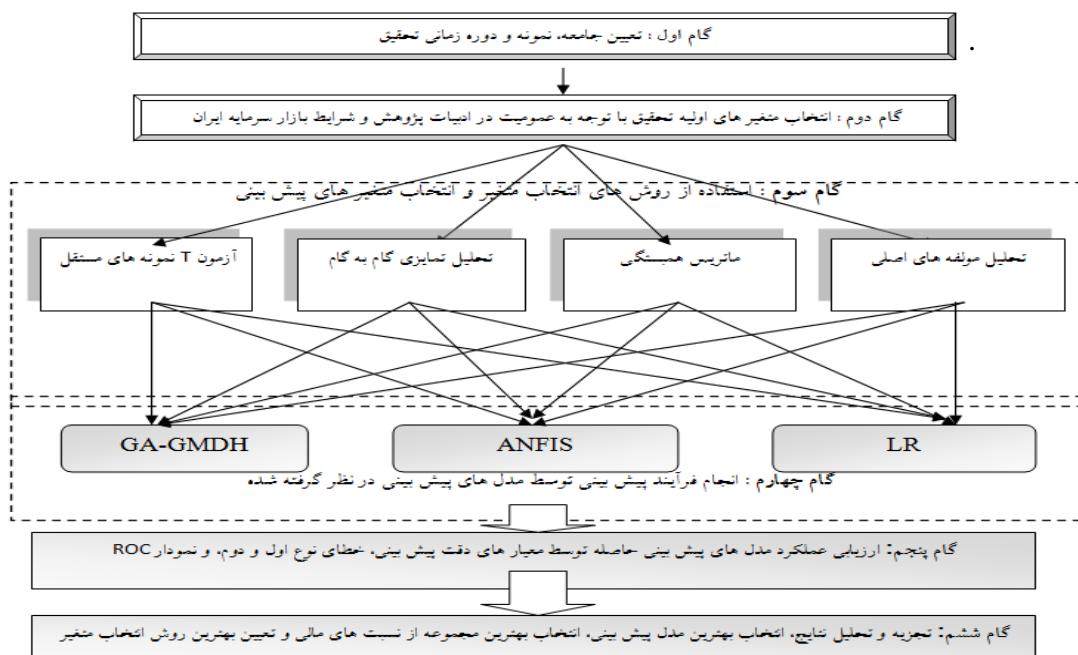
گروهی نظیر پیش‌بینی ورشکستگی است (Leea et al. 2005 1-16).

راوی و کومار (۲۲۲-۲۲۷) یک مدل قوانین فازی برای پیش‌بینی ورشکستگی در سال ۲۰۰۶ پیشنهاد دادند و نتایج آن را با مدل‌های لاجیت، پروبیت، LDA, QDA مقایسه کردند. مدل پیشنهادی آن‌ها 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 سیستم‌های عصبی-فازی تطبیق پذیر<sup>۱</sup> (ANFIS) برای طبقه بندی مشتریان بانک به دو طبقه خوش حساب و بد حساب انجام داده است. آن‌ها 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 می‌دهد 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 شرکت‌های 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 شرکت‌های بورس اوراق تهران به دست 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 شرکت‌ها استفاده نمودند. در این تحقیق از ۴۵ نسبت اولیه استفاده شده است، که اندازه شرکت نیز به عنوان یکی از متغیرها در نظر گرفته شده است، و با استفاده از روش تجزیه تحلیل گام به گام ۵ نسبت از بین این نسبت‌ها انتخاب شده است. نتایج حاکی از برتری مدل پیشنهادی بوده است. با توجه توضیحات و نتیجه گیری‌های ذکر شده در بخش‌هایی 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 گروهی دستگامی داده (GMDH)

یکی از انواع شبکه عصبی مصنوعی که کارایی آنها در مدل‌سازی و پیش‌بینی اثبات شده است، شبکه‌های خود-سازمانده می‌باشد<sup>۱</sup>. 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 الگوریتم های آموزش، الگوریتم GMDH<sup>1</sup> می باشد. نام GMDH، بیانگر ایده اصلی آن یعنی روال تقسیم تصادفی داده های ورودی مدل به دو مجموعه آموزش و تطبیق است. الگوریتم GMDH اولین بار توسط ایواخنکو در سال ۱۹۹۶ برای مدل کردن سیستم های 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 شبکه های عصبی مرسوم، به دست آوردن یک مدل ریاضی برحسب چند جمله ایها، برای فرایند مورد بررسی است. این مدل ریاضی می تواند در شناسایی و یا حتی توصیف کامل فرایند مورد استفاده قرار گیرد. به طور کلی برای مدل کردن سیستم های پیچیده ای که شامل مجموعه ای از داده ها با چندین ورودی و یک خروجی می باشند می توان از چند جمله ای های VKG<sup>2</sup> (رابطه ۱) استفاده نمود.

$$y = a_0 + \sum_{i=1}^n a_i x_i + \sum_{i=1}^n \sum_{j=1}^n a_{ij} x_i x_j + \sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n a_{ijk} x_i x_j x_k + \dots \quad (1)$$

که در آن بردارهای ورودی،  $x = (x_1, x_2, \dots, x_n)$ ، خروجی مدل و  $a_i$  ها ضرایب چند جمله ای می باشند. چند جمله ای های VKG با استفاده از چند جمله ای های درجه دوم تقریب زده می شوند. این چند جمله ای های درجه دوم بر اساس ترکیبات دوتایی ورودی های شبکه ساخته می شوند. الگوریتم GMDH با استفاده از این ایده به عنوان روش یادگیری، برای مدل سازی سیستم های پیچیده معرفی شده است. شبکه عصبی GMDH دارای ساختار یک شبکه چند لایه و رو به جلو می باشد و در برگزیده مجموعه ای از نرون ها است که از پیوند جفت های مختلف ورودی از طریق یک چند جمله ای درجه دوم به وجود می آیند. 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 می باشد. این واحدها عملاً نقش اجزای سازنده مدل را ایفا می نمایند و به شکل یک چند جمله ای درجه دوم فرض می شوند. پارامترهای مجهول الگوریتم GMDH ضرایب چند جمله ای می باشد. جهت محاسبه مقدار خروجی  $\hat{y}_i$  برای هر بردار ورودی  $x = (x_{i1}, x_{i2}, \dots, x_{in})$  براساس رابطه

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

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

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

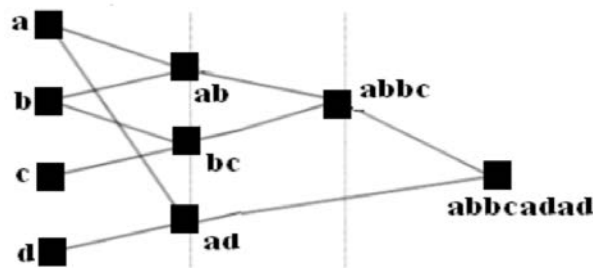
برای یافتن مقدار حداقل خطا، از مشتق جزئی رابطه ۲ استفاده می‌شود. با جایگذاری از رابطه (۲) در این مشتق جزئی، یک معادله ماتریسی ( $Aa = y$ ) به دست می‌آید. در این معادله برای این معادله ماتریسی ( $Aa = y$ )، استفاده از روش تجزیه به مقدار تکین ( $SVD^1$ ) است. هنگام ساختن یک شبکه GMDH همه ترکیبات ورودی‌ها ایجاد شده به اولین لایه شبکه فرستاده می‌شود سپس خروجی‌ها از این لایه برای ورود به لایه بعدی طبقه بندی و انتخاب می‌شود که باهمه ترکیبات خروجی‌های انتخاب شده به لایه ۲ فرستاده می‌شود این روند تا آنجا ادامه می‌یابد که لایه  $(n+1)$  نتیجه بهتری از لایه  $n$  ایجاد کند. هنگامی که لایه  $n+1$  پیدا شد که نسبت به لایه  $n$  بدتر بود فرآیند متوقف می‌شود.

#### 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 نرون‌ها در هر لایه پنهان و نحوه اتصال آن‌ها ارائه می‌دهد که در نتیجه آن سری بهینه از ضرایب مناسب توابع درجه دوم (چند جمله‌ای) پیش‌بینی کننده متغیر وابسته، به دست می‌آید (Varetto, 1998, 1421-1439). در ساختار ترکیبی مورد نظر به طور کلی سه تابع هدف<sup>۲</sup> اصلی به صورت خطای آموزش<sup>۳</sup>، خطای پیش‌بینی<sup>۴</sup> و تعداد نرون‌ها<sup>۵</sup> 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 می گرفته می شود، و کروموزومی (تابع چند جمله ای از متغیرهای پیش بینی) به عنوان جواب بهینه انتخاب می شود که کمترین میزان خطا را در هر دو مجموعه آموزش و تست داشته باشد. معیار میانگین مربع مربعات خطا (MSE) به عنوان معیار خطای پیش بینی در نظر گرفته می شود. در ساختار پیشنهادی GA-GMDH، نمایش ژن ها و کروموزوم ها به صورت رشته نمادین<sup>۱</sup> (غیر دودویی<sup>۲</sup>) 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 شبکه های عصبی GMDH، باید قید استفاده از لایه مجاور در ساختن لایه بعد حذف شود. این نوع از شبکه های عصبی با ساختار عمومی<sup>۳</sup> نامیده می شوند (GS) و برای ساخت لایه جدید از تمام لایه های قبلی (شامل لایه ورودی) استفاده می کنند. در شکل ۲ این امر نشان داده شده است. در واقع برای ساخت کروموزوم نرون خروجی، از تمام لایه ها (لایه دو سه) و نه لزوماً از لایه سه، استفاده شده است. بعد از تعریف کروموزوم ها و ژن ها از اپراتور های ادغام<sup>۴</sup> و جهش<sup>۵</sup> برای تولید دو فرزند از دو والد استفاده می شود. می شود. از روش چرخ رولت<sup>۶</sup> برای انتخاب دو کروموزوم والد استفاده می شود.



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 پیش‌بینی

### 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 شرکت‌های ورشکسته تحقیق، شامل تمام شرکت‌های 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 آن‌ها به طور کامل وجود داشت، به عنوان نمونه شرکت‌های 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 شرکت‌های ورشکسته مطابقت داشته باشند. نمونه شامل ۱۳۶ شرکت به دو بخش نمونه آموزش و نمونه تست (نمونه بسط داده شده) تقسیم می‌شود. 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 مدل‌ها، تقسیم می‌شود.

## فرآیند انتخاب متغیر های پیش بینی

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

شده اند.

جدول ۱- مجموعه نسبت های مالی اولیه تحقیق

نسبت مالی	نوع نسبت های مالی	کد	نسبت مالی	نوع نسبت های مالی	کد	
فروش به سرمایه در گردش	کارایی یا فعالیت	X <sub>11</sub>	سود قبل از بهره و مالیات به کل دارایی ها	سودآوری	X <sub>1</sub>	
فروش به دارایی جاری		X <sub>12</sub>	دارایی جاری به کل دارایی	نقدینگی	X <sub>2</sub>	
فروش به حقوق صاحبان سهام		X <sub>13</sub>	دارایی آنی به کل دارایی		X <sub>3</sub>	
فروش به دارایی آنی		X <sub>14</sub>	سرمایه در گردش به کل دارایی		X <sub>4</sub>	
سود و زیان انباشته به کل دارایی		X <sub>15</sub>	دارایی جاری به بدهی جاری (نسبت جاری)		X <sub>5</sub>	
کل بدهی به حقوق صاحبان سهام	X <sub>16</sub>	دارایی آنی به بدهی جاری (نسبت آنی)	X <sub>6</sub>			
کل بدهی به کل دارایی	X <sub>17</sub>	وجه نقد به کل بدهی	X <sub>7</sub>			
بدهی جاری به حقوق صاحبان سهام	X <sub>18</sub>	هزینه بهره به سود ناخالص	پوششی		X <sub>8</sub>	
بدهی بلند مدت به حقوق صاحبان سهام	سرمایه	X <sub>19</sub>	فروش به کل دارایی		کارایی یا فعالیت	X <sub>9</sub>
			فروش به وجه نقد			X <sub>10</sub>

آزمون T نمونه های مستقل جهت پاسخ گوئی به این سوال که آیا میانگین یک نسبت مالی در گروه شرکت های ورشکسته با میانگین همین نسبت مالی در گروه شرکت های غیر ورشکسته از لحاظ آماری متفاوت است یا خیر؟ سطح اطمینان ۹۹٪ برای تحلیل نتایج حاصل از آزمون در نظر گرفته می شود. در روش ماتریس همبستگی، چنانچه همبستگی یک نسبت مالی با متغیر وابسته (ورشکستگی)، به طور معنی داری 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، حذف متغیرهای مازاد و همبسته از فرآیند مدل‌سازی و جایگزین کردن آن‌ها توسط متغیرهای جدید نا همبسته است. معمولاً 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 و Bartlett's test of sphericity استفاده می‌شود. 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 تقریباً ۸۰٪ سطح تغییرپذیری در داده‌های 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 ۰.۱ در 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 نسبت‌های مالی انتخاب شده توسط روش‌های انتخاب متغیر

روش انتخاب متغیر	نسبت‌های مالی انتخاب شده
آزمون T نمونه‌های مستقل	سود قبل از بهره و مالیات 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 از بهره و مالیات 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 سود و زیان انباشته به کل دارائی ها، کل بدهی ها به کل دارایی ها، سود قبل از بهره و مالیات به کل دارائی گام به گام تجزیه و تحلیل مولفه های اصلی

۷ مولفه ی اصلی جایگزین ۲۰ نسبت مالی اولیه شده اند.

### ساخت مدل های پیش بینی

برای اجرای مدل پیشنهادی GA-GMDH تعداد لایه های پنهان برابر ۲، جمعیت اولیه برابر با ۱۰۰، تعداد نسل ها برابر با ۳۰۰ 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 پیش بینی با استفاده از نسبت های مالی انتخاب شده توسط روش های انتخاب متغیر، در نظر گرفته شد.

جدول ۳- پارامتر های در نظر گرفته شده برای مدل ANFIS پیشنهادی

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

### تجزیه و تحلیل نتایج

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

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

مدل پیش بینی	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

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

نتایج نشان می‌دهد 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 می‌گیرد. بنابراین 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 می‌دهد 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 می‌توان گفت که روش ماتریس همبستگی (CM)، منجر به بالاترین دقت طبقه بندی در ۳ مدل GA-GMDH، ANFIS و LR، نسبت

به سایر روش های انتخاب متغیر می شود. روش های انتخاب متغیر **SDA** و **Ttest** برای مدل های پیش بینی **ANFIS** و **LR** دارای نتایج یکسان می باشند. 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** می باشد. در توجیه نتایج به دست آمده می توان گفت که از آن جاییکه منطق روش **PCA** تنها بر اساس همبستگی بین متغیر های پیش بینی 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 پیش بینی ورشکستگی ایفای نقش کند. بنابراین روش **Ttest** و **PCA** به ترتیب در رتبه های سوم و چهارم قرار می گیرند.

دو مدل **CM-ANFIS** و **CM-GA-GMDH** با عملکرد مشابه با یکدیگر به عنوان بهترین مدل های پیش بینی ورشکستگی انتخاب می شوند. مدل **CM-ANFIS**، ۱۰۰٪ و ۹۰٪ از شرکت ها در نمونه آموزش و تست به درستی طبقه بندی می کند، همچنین دارای خطای نوع اول ۱۵٪ در نمونه تست می باشد. مدل پیش بینی **CM-GA-GMDH** نیز دارای دقت پیش بینی 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 کاملاً بر یکدیگر منطبق می باشد و دارای سطح زیر منحنی برابر با ۰.۹۰۸ می باشد که نشان دهنده توان تشخیصی (قدرت تفکیک کنندگی) بالای دو مدل انتخابی می باشد.

جهت مقایسه دو مدل ساخته شده می توان گفت که مزیت اصلی مدل **GMDH** نسبت به شبکه های عصبی مرسوم، به دست آوردن یک مدل ریاضی برحسب چند جمله ای ها، برای فرآیند مورد بررسی است. این مدل ریاضی می تواند الگوی مناسبی را در جهت شناسایی و یا حتی توصیف کامل فرآیند 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 پیش بینی را نشان می دهد که ممکن است برای استفاده کننده گان ملموس نباشد و در نتیجه درک و استفاده از آن ها 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 شرکت‌ها، 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 متغیرهای  $x_i$  به جدول ۱ مراجعه شود.

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

### نتیجه گیری

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

پیش بینی توجه کافی مبذول نگردیده است. در اغلب تحقیقات انجام گرفته در ایران نیز، شناسایی متغیرهای پیش بینی کننده ورشکستگی، آنچنان که در تحقیق حاضر به آن پرداخته می شود، مورد توجه قرار نگرفته شده است. دلایل بالا انگیزه ای شد تا در این تحقیق با شناسایی متغیرهای تاثیرگذار در زمینه پیش بینی ورشکستگی و مدل سازی یک روش جدید (GA-GMDH) که در مسئله پیش بینی ورشکستگی تا به حال به کار برده نشده است، مدلی جهت پیش بینی ورشکستگی شرکت های بورس اوراق بهادار تهران طراحی شود. با توجه نتایج به دست آمده مدل پیشنهادی تحقیق، دارای قابلیت بالایی در مدل سازی پیش بینی ورشکستگی می باشد. مقایسه نتایج این مدل با مدل های ANFIS و مدل رگرسیون لجستیک گواهی بر این ادعا است. روش ماتریس همبستگی در مقایسه با سایر روش های انتخاب متغیر، توانایی بیشتری در انتخاب متغیرهای موثر بر پیش بینی ورشکستگی شرکت ها دارد. بهترین مدل پیش بینی ورشکستگی ساخته شده در تحقیق حاضر، ترکیب مدل GA-GMDH با روش انتخاب متغیر ماتریس همبستگی (CM-GA-GMDH) می باشد. این مدل مجموعه نسبت های مالی انتخاب شده توسط روش ماتریس همبستگی را دوباره تعدیل کرده و در نهایت نسبت های مالی "سود قبل از بهره و مالیات به کل دارایی"، "کل بدهی به کل دارایی"، "دارایی های سریع به بدهی های جاری"، "هزینه بهره به سود ناخالص" را به عنوان متغیرهای نهایی پیش بینی کنندگی ورشکستگی شرکت های بورس اوراق بهادار تهران معرفی کرده است.

## منابع

- ۱- فلاح پور، سعید، (۱۳۸۳)، پیش بینی در ماندگی 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 مدل بندی پیش بینی 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- Tabachnick, B., Fidell, L. "Using multivariate Statistics" (4th ed.) .Boston: Allyn & Bacon, 2001.
- 11- Kumar, P.R., Ravi, V., "Bankruptcy prediction in banks by fuzzy rule 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.