تشخیص فرار مالیاتی با استفاده از سیستم هوشمند ترکیبی

اقبال رحیمی کیا[1]

شاپور محمدی[2]

مهدی غضنفری[3]

تاریخ دریافت: ۱۳۹۴/۰۳/۲۲ تاریخ پذیرش: ۱۳۹۴/۰۶/۰۴

چکیده

با توجه به اجرایی شدن سامانه عملیات الکترونیکی مودیان مالیاتی و ایجاد پایگاه داده‌های مالیاتی، امکان پایش اطلاعات موجود با مدل های مختلف فراهم شده است. در این پژوهش، از الگوریتم بهینه سازی جستجوی هارمونی به‌منظور بهینه سازی همزمان پارامترهای شبکه عصبی پرسپترون چند لایه و ترکیب مناسب ورودی ها استفاده شده است. علاوه بر آن نتایج با رگرسیون لجستیک به عنوان هسته سیستم مورد مقایسه قرار گرفته است. متغیرهای ورودی به سیستم ۲۱ مورد بوده که با بررسی پژوهش های مشابه انجام شده طی ۳۰ سال اخیر، اعمال ویژگی های مالیاتی ایران و نظرخواهی از خبرگان انتخاب شده است. مقایسه نتایج حاصل از شبکه عصبی و رگرسیون لجستیک در دو صنعت مواد غذایی و نساجی نشان می دهد، استفاده از شبکه عصبی دارای دقت های بالاتری بوده و این تفاوت از لحاظ آماری معنادار می باشد. در شبکه عصبی به ترتیب در صنعت مواد غذایی و نساجی دقت کلی ٪۸۳/۷۸ و ٪۸۴/۸۵، دقت تشخیص شرکت های فراری ٪۸۰/۳۱ و ٪۸۴/۳۴ و دقت تشخیص شرکت های سالم ٪۸۷/۲۰ و ٪۸۵/۳۶ می باشد. با اعمال مجموعه مدل‌های نهایی سیستم بر روی اطلاعات عملکرد سال ۹۱ اشخاص حقوقی و مقایسه آن با نتایج حسابرسی مالیاتی در دو صنعت مواد غذایی و نساجی، به ترتیب دقت کلی ٪۹۲/۲۲ و ٪۸۲/۳۵، دقت تشخیص شرکت های فراری ٪۸۳/۸۷ و ٪۸۴/۰۵ و دقت تشخیص شرکت های سالم ٪۹۲/۷۱ و ٪۸۲/۲۲ حاصل شده است. نتایج در داده‌های آزمون بر مبنای اعتبارسنجی ضربدری ۱۰ بخشی با تکرار و میانگین گیری بر روی ۸ حلقه موازی ارائه شده است.

واژه های کلیدی: تشخیص فرار مالیاتی، داده کاوی، هوش مصنوعی، سیستم هوشمند ترکیبی

۱- پژوهشگر سیستم های اقتصادی، دانشگاه علم و صنعت (نویسنده مسئول)erahimi@ut.ac.ir

۲- عضو هیات علمی دانشکده مدیریت دانشگاه تهران shmohmad@ut.ac.ir

۳- عضو هیات علمی دانشکده مهندسی صنایع دانشگاه علم و صنعت mehdi@iust.ac.ir

۱- مقدمه

در سال‌های پیشین به دلیل عدم وجود بسترهای نرم افزاری و سخت افزاری برای دریافت اظهارنامه و واریز الکترونیکی مالیات، امکان دسترسی به داده های طبقه بندی شده الکترونیکی شرکت ها برقرار نبود، به این دلیل در دوره فوق حرکت به سمت ایجاد یک سیستم هوشمند نرم افزاری برای کشف فرار مالیاتی و طراحی معیار آن امکان پذیر نبوده است. با پیاده سازی زیر ساخت های نرم افزاری و سخت افزاری مختلف در سازمان امور مالیاتی کشور امکان طراحی و ایجاد ساختارهای هوشمند مختلفی در کنار سیستم های فوق برقرار شده است.

بررسی ها نشان می دهد که تاکنون در داخل کشور مطالعه ای برای ارائه یک مدل هوشمند برای کشف فرار مالیاتی انجام نشده است. همچنین تعداد مطالعات انجام شده در سطح بین المللی با بهره گیری از روش های داده کاوی و هوش مصنوعی محدود می باشد.

لازم به ذکر است که حوزه تشخیص تقلب[1] در حالت جامع دارای زیر مجموعه های متنوعی از مدل ها و روش ها در حوزه های بانکی، مالی و ... می باشد اما در این پژوهش تمرکز اصلی بر روی زیر مجموعه مالیاتی حوزه فوق می باشد. از زیر مجموعه های جامع تشخیص تقلب می توان به تشخیص تروریستی[2]، تشخیص بحران مالی[3] (پیش بینی ورشکستگی)، تشخیص نفوذ و هرزنامه[4] و سایر موارد مشابه اشاره کرد (فوا، لی، اسمیت و گیلر، ۲۰۰۶). از موارد بسیار مهم در طراحی سیستم های نرم افزاری تشخیصی در حوزه مالیاتی انطباق هوشمند سیستم طراحی شده با داده های ورودی مختلف می باشد. به‌طور معمول در پژوهش های انجام شده مشاهده می شود که فرآیند انتخاب متغیرهای ورودی (و در بعضی موارد پارامترهای مدل طبقه بندی) مبتنی بر پژوهش های پیشین انجام شده می باشد. در این پژوهش با مبنا قرار دادن مدل طبقه بندی شبکه عصبی پرسپترون

1. Fraud Detection
2. Terrorist Detection
3. Financial Crime Detection
4. Intrusion and Spam Detection

چندلایه[1] و رگرسیون لجستیک[2] و همچنین بهینه سازی ساختار آن (تعداد لایه ها و نورون های موجود در هر لایه) در کنار بهینه سازی ترکیب مناسب متغیرهای ورودی (از میان ۲۱ متغیر اولیه انتخابی) سیستمی ترکیبی و هوشمند برای تشخیص فرار مالیاتی در کشور ارائه شده است.

علاوه بر آن با بهره گیری از اعتبارسنجی ضربدری ۱۰ بخشی[3] و تکرار آن بر روی هسته های منطقی[4] واحد پردازش مرکزی[5] (CPU) بوسیله استفاده از پردازش موازی[6] از از قرارگیری جواب نهایی الگوریتم بهینه سازی در بهینه محلی[7] جلوگیری به عمل آماده است. همچنین در این پژوهش به دلیل نیاز شبکه عصبی به سه مجموعه داده آموزش، آزمون و اعتبارسنجی از روش جدیدی برای ترکیب اعتبارسنجی ضربدری با شبکه عصبی پرسپترون چندلایه بهره گرفته شده است. جواب نهایی سیستم نیز با تغییراتی، مجموعه ای از مدل ها (تعداد حلقه های سیستم ضرب در تعداد بخش ها در اعتبارسنجی ضربدری) را در برگرفته که در بخش های آتی با جزئیات بیشتری ارائه خواهد شد.

از موارد دیگر ارائه شده در سیستم هوشمند ترکیبی فوق بهره گیری از معیار توقف سیستم می باشد. در این پژوهش از مراحل حرکت الگوریتم بهینه سازی با نام قدم[8] یاد خواهد شد. در صورت عدم بهبود دقت کلی سیستم در تعداد مشخصی از قدم های متوالی فرآیند بهینه سازی مدل طبقه بندی پرسپترون چندلایه و ترکیب ورودی ها متوقف خواهد شد. از ساختار فوق به عنوان شرط توقف سیستم یاد می شود.

در این مقاله ابتدا مروری بر ادبیات استفاده از روش های هوش مصنوعی و داده کاوی به منظور کشف فرار مالیاتی انجام خواهد شد. در بخش بعد به صورت مختصر مدل طبقه بندی شبکه عصبی پرسپترون چندلایه، رگرسیون لجستیک و الگوریتم بهینه سازی

1. Multilayer Perceptron Neural Network
2. Logistic Regression
3. 10-fold Cross Validation
4. Logical Processor
5 . Central Processing Unit
6. Parallel Computing
7. Local Optimum
8. Step

جستجوی هارمونی[1] ارائه خواهد شد. در بخش پایانی نیز ساختار سیستم هوشمند کشف فرار مالیاتی طی بخش های ساختار هسته سیستم، جامعه آماری و ابزار ارائه خواهد شد.

۲- مروری بر ادبیات موضوع

یو، کین و جیا (۲۰۰۳) حوزه کشف فرار مالیاتی را در بخش های انتخاب الگوریتم داده کاوی به عنوان هسته سیستم، طراحی ساختار سیستم، چگونگی بهره گیری از دانش متخصصان، آماده سازی داده ها، تنظیم قواعد و ارزیابی سیستم مورد بررسی قرار داده اند. لازم به ذکر است که بررسی های این پژوهش به تمامی حوزه های داده کاوی قابل تعمیم بوده و تنها حوزه فرار مالیاتی را در بر نمی گیرد.

راویسانکار، راوی، رائو و بس (۲۰۱۱) روش های مختلف داده کاوی از جمله شبکه عصبی، ماشین بردار پشتیبان[2] ، برنامه ریزی ژنتیک[3]، شبکه عصبی گروهی مدیریت داده ها[4]، رگرسیون لجستیک و شبکه عصبی احتمالی[5] را بر روی ۲۰۲ شرکت چینی به منظور کشف فرار مالیاتی با استفاده و عدم استفاده از انتخاب ویژگی مورد بررسی قرار دادند. در شرایط عدم استفاده از انتخاب ویژگی، شبکه عصبی احتمالی بالاترین دقت را نتیجه داده است. در شرایط استفاده از فرآیند انتخاب ویژگی نیز برنامه ریزی ژنتیک و شبکه عصبی احتمالی دقت های بالاتر (البته نزدیک به یکدیگر) را نتیجه داده اند. روش مورد استفاده برای انتخاب ویژگی (ورودی های مدل) روش ساده آماری آزمون تی[6] می باشد. میانگین دقت کلی مجموعه مدل ها با اعمال اعتبارسنجی ضربدری ۱۰ بخشی بین ٪۷۰ تا ٪۹۲ می باشد.

چکینی، آیتوک، کوهلر و پتک (۲۰۱۰) با استفاده از ماشین بردار پشتیبان و اطلاعات مالی، تقلب مدیریتی[7] را مورد بررسی قرار داده اند. از موارد مورد بررسی آنها استفاده از

1. Harmony Search Algorithm
2. Support Vector Machines (SVM)
3. Genetic Programming
4. Group Method of Data Handling (GMDH)
5. Probabilistic Neural Network (PNN)
6. T-statistic
7. Management Fraud

هسته[1] های مختلف در ماشین بردار پشتیبان می باشد. ماشین بردار پشتیبان طراحی شده توسط محققان فوق در مجموعه داده های خارجی[2] ٪۸۰ شرکت های فراری و ٪۹۰/۶ شرکت های سالم را به درستی مورد تشخیص قرار داده است. محققان مدل خود را بر روی چند مجموعه دادهٔ مالی مختلف مورد آزمون قرار داده و برتری این ساختار را نسبت به ساختارهای دیگر موجود در سایر مقالات اثبات کرده اند.

مایکل سررلو و ویرجنیا سررلو (۱۹۹۹) با بررسی روش های مختلف مبتنی بر هوش مصنوعی، استفاده از ساختار شبکه عصبی مصنوعی را برای تعیین وضعیت شرکت ها از لحاظ فرار مالیاتی و سلامت مورد بررسی قرار داده اند.

کرکوس، اسپاتیس و مانولوپولوس (۲۰۰۷) با استفاده از درخت تصمیم گیری، شبکه عصبی مصنوعی و شبکه باور بیزینی[3] و ورودی هایی مبتنی بر نسبت های مالی اقدام به تشخیص فرار مالیاتی کرده اند. طبق نتایج حاصل شده دقت کلی مدل های طبقه بندی فوق به ترتیب در درخت تصمیم گیری، شبکه عصبی و شبکه باور بیزینی برابر با ٪۷۳/۶، ٪۸۰/۰ و ٪۹۰/۳ می باشد که نشان از برتری شبکه باور بیزینی دارد.

کسکیوارا (۲۰۰۰) اثر فرآیند های مختلف پیش پردازش را برای تشخیص فرار مالیاتی با استفاده از شبکه عصبی مورد بررسی قرار داد. نمونه های مورد بررسی وی ۳۱ شرکت تولیدی را طی ۴ سال شامل می شد. چهار روش مورد بررسی این مقاله مقیاس دهی[4] خطی، مقیاس دهی خطی بر پایه سالانه، مقیاس دهی خطی بر پایه شرکت ها و مقیاس دهی خطی بر پایه سالانه و شرکتی (ترکیبی) می باشد. نتیجه حاکی از آن بود که بالاترین دقت شبکه عصبی در حالت استفاده از مقیاس دهی خطی بر پایه سالانه و شرکتی (ترکیبی) حاصل می شود.

کسکیوارا (۲۰۰۳) استفاده از شبکه های عصبی در حسابرسی و سایر حوزه های مرتبط را مورد بررسی قرار داده است. وی بر این موضوع تأکید کرده است که استفاده از

1. Kernel
2. Holdout Set
3. Bayesian Belief Network (BBN)
4. Scaling

سیستم های پشتیبانی فناوری اطلاعات[1] جدید در ساختارهای مالیاتی کنونی از اهمیت خاصی برخوردار می باشد. نتیجه این تحقیق آن است که شبکه عصبی از ساختارهای مفید برای کمک به حسابرسان می باشد.

کوزیاتیز، کوماناکس، زلپیس و تمپکس (۲۰۰۶)، ۱۶۴ شرکت فراری و غیر فراری یونانی از سال ۲۰۰۱ تا ۲۰۰۲ را مورد بررسی قرار داده اند. محققان فوق با بهره گیری از سیستم پشتیبان تصمیم گیری ترکیبی[2] و استفاده از درخت تصمیم گیری (K2، C4.5،3NN)، شبکه عصبی شعاعی پایه[3]، آموزش سست[4] و بهینه سازی حداقلی متوالی[5] اقدام به طراحی یک سیستم ترکیبی کرده اند. روش مورد استفاده برای ترکیب، دسته کردن[6]، رای گیری BestCV[7] (ویتن و فرانک، ۲۰۰۰) و نمره دهی[8] می باشد. روش دسته کردن دارای بالاترین دقت برابر با ۹۳/۹٪ می باشد.

ژو و کاپور (۲۰۱۱) با استفاده از مجموعه مدل های رگرسیونی، درخت تصمیم گیری، شبکه عصبی و شبکه بیزینی اقدام به تشخیص فرار مالیاتی کرده اند. در مرحله بعد محققان مدل سطح پاسخ[9] را نیز با ساختارهای ذکر شده ادغام کرده و نتیجه گرفته اند که این ترکیب نتایج بهتری را در پی خواهد داشت.

نایگ، هو، وانگ، چن و سان (۲۰۱۱) به مرور مقالات و پژوهش های داده کاوی در حوزه کشف تقلب مالی[10] از سال ۱۹۹۷ تا ۲۰۰۸ (۴۹ مورد) که خود از زیربخش های متنوعی تشکیل شده است پرداخته اند.

1. Information Technology Support System
2. Hybrid Decision Support System
3. Radial Basis Function (RBF)
4. Lazy Learning (LL)
5. Sequential Minimal Optimization
6. Stacking
7.Voting
8. Grading
9. Response Surface Method
10. Financial Fraud Detection (FFD)

محققان این پژوهش تقلب مالی را به زیربخش های تقلب بانکی[۱]، تقلب بیمه ای[۲]، تقلب در اوراق بهادار و کالا[۳] و سایر فرارهای مرتبط و همچنین روش داده کاوی مورد استفاده را به زیر بخش های رگرسیون، خوشه بندی، پیش بینی، کشف داده های پرت و تجسم[۴] تقسیم کرده اند. در این پژوهش نتیجه گرفته شده است که روش های داده کاوی در سال های اخیر به صورت گسترده ای در حوزه تقلب بیمه ای و سپس تقلب کارت های اعتباری[۵] مورد استفاده قرار گرفته اند. همچنین مشاهده شده است که استفاده از روش های فوق در حوزه های تقلب در وام مسکن[۶]، پول شویی[۷] و تقلب در اوراق بهادار و کالا بسیار محدود است. مورد توجه ترین روش های داده کاوی در مجموعه حوزه های ذکر شده مدل های لجستیکی، شبکه عصبی، شبکه باور بیزینی و درخت تصمیم گیری می باشد. در پایان نیز با ارائه خلاء های موجود در تحقیقات، پیشنهاداتی برای مطالعات آتی ارائه شده است.

ووا، اوو، لین، چنگ و ین (۲۰۱۳) با اعمال یک سیستم مبتنی بر داده کاوی بر روی پایگاه داده مالیات بر ارزش افزوده[۸] نتیجه گرفته اند که استفاده از سیستم های مبتنی بر داده کاوی موجب بهبود چشمگیری در تشخیص کسب و کارهای فراری در حوزه مالیات بر ارزش افزوده می شود.

لیو، گوانگ زو، کیان زو و ژانگ (۲۰۱۲) با استفاده از یک الگوریتم تشخیص داده های پرت[۹] پیشنهادی اقدام به تشخیص اظهارنامه های غیر عادی کرده اند. محققان فرآیند خود خود را بر روی داده ها اعمال و مناسب بودن این فرآیند را نتیجه گرفته اند.

اسپتیس، داموپاس و زوپنیدس (۲۰۰۲) با استفاده از روش پشتیبانی تصمیم با

1. Bank Fraud
2. Insurance Fraud
3. Securities and Commodities Fraud
4. Visualization
5. Credit Fraud
6. Mortgage Fraud
7. Money Laundering
8. Value-added Tax (VAT)
9. Outlier Detection

معیارهـای چندگانـه[1] و روش UTADIS[2]، ۷۶ شـرکت یونـانی (۳۸ فـراری و ۳۸ سـالم) را بـا با اسـتفاده از ۱۰ نسـبت مـالی ورودی مـورد بررسـی قـرار داده انـد. همچنـین در ایـن پـژوهش از روش جک نایف[3] بـرای اعتبارسـنجی و مقایسـه مـدل هـا بهـره گرفتـه شـده اسـت. نتـایج نشـان می دهد که اسـتفاده از روش فـوق نتـایج بهتـری را نسـبت بـه روش هـای سـنتی ایـن حـوزه در پی دارد. همچنـین نتـایج نشـان دهنـده مناسـب بـودن اسـتفاده از نسـبت هـای مـالی ماننـد کـل بدهی ها به کل دارایی ها، موجودی انبـار بـه فـروش، سـود خـالص بـه فـروش و فـروش بـه کـل دارایی ها برای تشخیص فرار مالیاتی می باشد.

ساختار مدل ها و روش های مورد استفاده

شبکه عصبی پرسپترون چند لایه

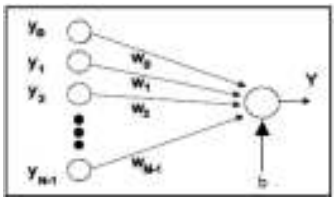
یک شـبکه عصـبی مصـنوعی، مجموعـه‌ای از واحـدهای محاسـباتی مصـنوعی بـه نـام نـورون[4] و نحـوه اتصـالات بـین آن واحـدها مـی‌باشـد. هـر نـورون دارای یـک یـا چنـد ورودی و تنهـا یـک خروجی می‌باشد. در واقع مـی‌تـوان شـبکه عصـبی را بـه صـورت یـک گـراف جهـت‌دار و وزن دار معرفی کرد که در آن نورون‌ها رئوس گـراف و یـال‌هـای وزن‌دار ارتباطـات بـین آن‌هـا مـی‌باشـند (۱۹۹۶).

هـر نـورون یـک واحـد محاسـباتی اسـت کـه تـابعی مشـخص را بـر روی ورودی‌هـای خـود محاسبه کرده و خروجی متناظر آن را به نـورون دیگـری (یـا بـه خروجـی) منتقـل مـی‌کنـد. توابـع در نورون‌ها به صـورت تـابعی یـک متغیـره تعریـف مـی‌شـوند کـه ورودی ایـن تـابع برابـر ضـرب داخلی بردار ورودی در بردار وزن‌ها بـه عـلاوه مقـدار ثـابتی بـه نـام بایـاس[5] مـی‌باشـد کـه رابطـه آن را می‌توان به صورت زیر نمایش داد.

$$\begin{bmatrix} y_0 \\ \vdots \\ y_{n-1} \\ 1 \end{bmatrix} \cdot \begin{bmatrix} w_0 \\ \vdots \\ w_{n-1} \\ b \end{bmatrix} = y_0 . w_0 + \ \dots + y_{n-1} . w_{n-1} + b \qquad (۱)$$

1. Multi-criteria Decision Aid (MCDA)
2. Utilite´s Additives Discriminantes
3. Jackknife
4. Neuron
5. Bias

شکل (۱)- ساختار کلی یک شبکه عصبی مصنوعی (ادوم، ۱۹۹۰)



که در آن $y_0 \dots y_{n-1}$ مقادیر ورودی نورون، $w_0 \dots w_{n-1}$ وزن‌های متناظر با اتصالات هر ورودی و b مقدار بایاس می باشد. از متداول‌ترین توابع فعال‌سازی که در شبکه‌های عصبی استفاده می‌شود می‌توان به تابع خطی[1]، تابع پله‌ای[2]، تابع خطی محدود[3] و تابع سیگموید[4] اشاره کرد.

شبکه‌های عصبی بر اساس معماری شبکه، دارای لایه‌های متفاوتی از نورون ها می‌باشند. هر شبکه عصبی از حداقل یک لایه خروجی تشکیل شده است. در شبکه‌های چند لایه، لایه‌های مختلفی از نورون‌ها (لایه‌های مخفی) بین ورودی و لایه خروجی قرار می‌گیرند. ساده‌ترین شکل شبکه عصبی، از یک لایه تک نورونی به عنوان لایه خروجی تشکیل شده است. یکی از ویژگی‌های جدایی ناپذیر هوشمندی، توانایی یادگیری می‌باشد. به طور کلی فرآیند یادگیری در شبکه عصبی مصنوعی به معنی تنظیم و به روز کردن معماری شبکه و وزن‌های آن است به طوری که شبکه یک مسئله را به نحو کارا حل نماید. در بسیاری از روش‌های یادگیری فقط وزن‌های شبکه تغییر می‌کنند و معماری شبکه ثابت است.

رگرسیون لجستیک

رگرسیون لجستیک نوعی از رگرسیون است که متغیرهای پیش بینی (مستقل) در آن می توانند هم در مقیاس کمّی و هم در مقیاس مقوله ای باشند، ولی متغیر وابسته دو سطحی بوده که به عضویت و عدم عضویت نمونه مورد نظر در یک گروه مشخص اشاره دارد. این مدل رگرسیون، شبیه رگرسیون معمولی است با این تفاوت که روش تخمین ضرایب یکسان نمی باشد و به جای

1. Linear Function
2. Step Function
3. Symmetric Saturating Linear Transfer Function
4. Sigmoid Function

حداقل کردن مجذور خطاها که در رگرسیون معمولی انجام می شود، احتمالی که یک واقعه رخ می دهد را حداکثر می نماید. در رگرسیون لجستیک از مفهومی به نام بخت برای مقدار متغیر وابسته استفاده می شود. در اصطلاح آماری، بخت به معنی احتمال رخداد یک پیشامد (P_i) بر احتمال عدم رخداد $(1 - P_i)$ آن می باشد. احتمال بین ۰ و ۱ تغییر می کند در حالی که بخت ممکن است بیش از یک باشد. ساختار کلیدی در تحلیل رگرسیون لجستیک ساختاری به نام لوجیت است که لگاریتم طبیعی بخت می باشد. رگرسیون لجستیک به شکل زیر تعریف می شود:

$$Z_i = ln\left(\frac{P_i}{1-P_i}\right) = \beta_0 + \sum \beta_i X_i \qquad (۲)$$

در این معادله In بیانگر لگاریتم طبیعی است. در مدل رگرسیون لجستیک احتمال رخداد پیشامد مورد نظر (فرار مالیاتی شرکت مورد نظر) بر اساس رابطه (۳) محاسبه می شود.

$$P_i = \prod i(x_1, x_2, \dots, x_k) = e^{\beta_0} + \frac{\sum \beta_i X_i}{1+e^{\beta_0+\sum \beta_i X_i}} \qquad (۳)$$

یکی از منافع رگرسیون لجستیک بی نیازی آن به مفروضات محدود کننده آماری در رابطه با متغیرهاست.

الگوریتم جستجوی هارمونی

الگوریتم جستجوی هارمونی[1] ابداع شده توسط گیم و همکاران (۲۰۰۱) به دلیل کاربردی بودن برای مسائل بهینه سازی گسسته و پیوسته، محاسبات ریاضیاتی کم، مفهوم ساده، پارامترهای کم و اجرای آسان به یکی از پرکاربردترین الگوریتم های بهینه سازی در سالهای اخیر در مسائل مختلف تبدیل شده است. این الگوریتم در مقایسه با سایر الگوریتم های فرا ابتکاری الزامات ریاضیاتی کمتری دارد و می توان آن را با مسائل مختلف مهندسی با تغییر در پارامترها و عملگرها منطبق نمود. این الگوریتم به طور صعودی در سالیان اخیر توجه زیادی به خود معطوف کرده است. در بعضی از تحقیقات نشان داده شده است که الگوریتم جستجوی هارمونی در بدست آوردن جواب های مناسب، در زمان زودتری نسبت به روش ژنتیک عمل می کند (ناداوی و همکاران، ۲۰۰۷). از مزیت های دیگر این الگوریتم نسبت به سایر الگوریتم ها این است که برای ایجاد حل جدید برخلاف

1. Harmony Search Algorithm

سایر روش ها که از دو بردار حل در هر نسل استفاده می کند، این الگوریتم از همه حل های موجود در حافظه اش استفاده می کند. این ویژگی موجب افزایش انعطاف و جستجوی بهتر فضاهای گسترده تر جواب می شود. از ویژگی های دیگر الگوریتم جستجوی هارمونی این است که در مدت زمان مناسبی فضاهای حل با محدوده عملکرد بهتری را شناسایی می کند.

الگوریتم جستجوی هارمونی از فرآیند بداهه سرایی موزیک گروه ارکستر پیروی می کند. هر نوازنده موزیک، گام هایی از ابزارهای موسیقی خود را می نوازد تا هماهنگی هارمونی در ارکستر به وجود آید. هدف از این فرآیند رسیدن به شرایطی است که هماهنگی کاملی از نواها ایجاد گردد. خروجی این هماهنگی کامل، صدای خوشایندی است که با استانداردهای زیبا مقایسه می گردد. در این الگوریتم هر جواب ممکن یک هارمونی نامیده می شود و با یک بردار N بعدی نمایش داده خواهد شد. الگوریتم فوق سه فاز اصلی دارد: نسل اولیه (مقداردهی اولیه)، بهبود بردار هارمونی جدید و به روز کردن حافظه الگوریتم. لازم به ذکر است که ساختارهای مختلفی از الگوریتم فوق با بهبودهای متفاوت موجود می باشد. ساختار ذکر شده ساختار پایه الگوریتم جستجوی هارمونی می باشد.

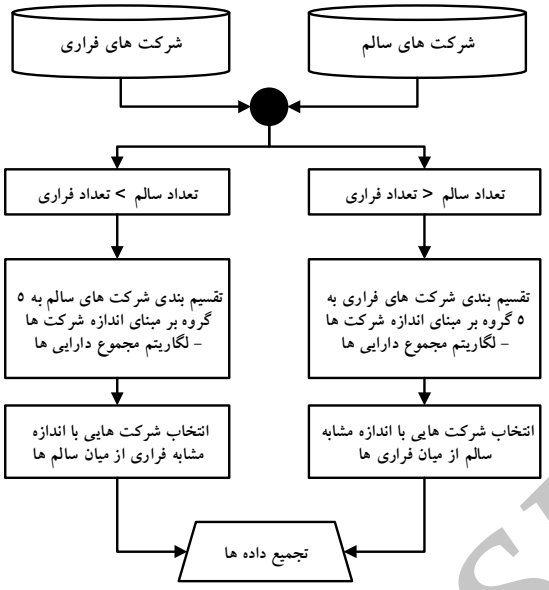
ساختار سیستم هوشمند کشف فرار مالیاتی

ساختار پیش پردازش سیستم و مجموعه داده ها

پس از تعیین شرکت های فراری و سالم بر مبنای معیار فرار مالیاتی طراحی شده با همکاری سازمان امور مالیاتی کشور، مشاهده می شود که در تمامی صنایع مورد بررسی تعداد شرکت های فراری کمتر از تعداد شرکت های سالم بوده که وضعیت فوق منطقی نیز به‌نظر می رسد. در مبانی داده کاوی و هوش مصنوعی از بالانس[1] داده ها برای برابر کردن تعداد نمونه های شرکت های دو گروه مورد بررسی (در اینجا فراری و سالم) استفاده می شود. در پژوهش فوق از ساختار جدیدی مبتنی بر اندازه شرکت ها (الگاریتم مجموع دارایی های شرکت) بهره گرفته شده است. این ساختار در شکل (۲) نمایش داده شده است.

1. Balancing

شکل (۲)- ساختار یکپارچه سازی سیستم



در این ساختار مجموعه با تعداد نمونه های بیشتر (در این پژوهش شرکت های سالم) بر مبنای لگاریتم مجموع دارایی ها (اندازه شرکت) به ۵ بخش تقسیم می شود. سپس برای هر نمونه موجود در مجموعه شرکت های فراری بر مبنای لگاریتم مجموع دارایی های شرکت فراری مورد نظر از میان بخش متناظر از لحاظ اندازه، شرکت سالمی انتخاب خواهد شد. در صورت نبود نمونه در بخش فوق، شرکت سالم از بخش کوچکتر انتخاب شده و در صورت نبود بخش کوچکتر و یا نبود شرکتی در آن به ترتیب اندازه از بخش های بزرگتر بهره گرفته خواهد شد. در پایان تعداد شرکت های فراری و سالم بالانس و برابر خواهند شد. سپس مجموعه داده های حاصل وارد فرآیند نرمالیزه کردن[1] می شوند. به منظور نرمالیزه کردن داده ها از روش استفاده از میانگین و انحراف معیار[2] بهره گرفته شده است. ۲۱ متغیر اولیه ورودی به سیستم به شرح جدول (۱) می باشد.

1. Normalization
2 . Standard Score

جدول (۱)- متغیرهای منتخب سیستم هوشمند کشف فرار مالیاتی

ردیف	متغیرها
۱	فروش خالص به کل دارایی ها
۲	سود (زیان) پس از کسر مالیات به جمع حقوق صاحبان سهام
۳	کل بدهی ها به کل دارایی ها
۴	دارایی های آنی[۱] به کل دارایی ها
۵	کل دارایی ها به حقوق صاحبان سهام
۶	دارایی های ثابت[۲] به کل دارایی ها
۷	کل بدهی ها به حقوق صاحبان سهام
۸	اسناد دریافتنی[۳]
۹	سود (زیان) ناخالص فروش به کل دارایی ها
۱۰	سود (زیان) ناخالص فروش
۱۱	سود (زیان) قبل از بهره و مالیات به کل دارایی ها
۱۲	سود (زیان) پس از کسر مالیات به کل دارایی ها (ROA)
۱۳	سود انباشته به کل دارایی ها
۱۴	دارایی های جاری به کل دارایی ها
۱۵	کل دارایی ها
۱۶	سود (زیان) عملیاتی به کل دارایی ها
۱۷	موجودی انبار به کل دارایی ها
۱۸	کل بدهی ها
۱۹	اسناد دریافتنی به کل دارایی ها
۲۰	کل بدهی ها به (کل بدهی ها + کل حقوق صاحبان سهام)
۲۱	سود (زیان) پس از کسر مالیات

لازم به ذکر است که پس از استخراج موارد فوق نمونه های دارای مقادیر بی نهایت از ورود به سیستم حذف خواهند شد. مجموعه اولیه متغیرهای ورودی به سیستم ۲۱ نسبت و متغیر مالی بوده که با بررسی مجموعه پژوهش های مشابه انجام شده طی ۳۰ سال اخیر،

۱- موجودی نقد بانک + سرمایه گذاری کوتاه مدت + حساب ها و اسناد دریافتنی + جاری شرکا، سهامداران

۲- دارایی ثابت، مشهود + دارایی های نامشهود

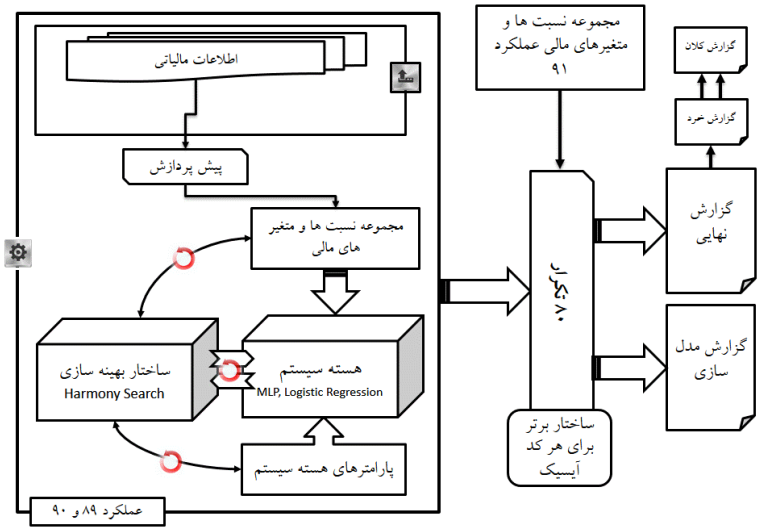
۳ - حساب ها و اسناد دریافتنی تجاری + سایر حساب ها و اسناد دریافتنی

اعمال ویژگی های مالیاتی ایران و نظرخواهی از خبرگان مورد اصلاح قرار گرفته است. نتایج نهایی در جدول (۱) قابل مشاهده می باشد.

ساختار هسته سیستم

ساختار کلی هسته سیستم در شکل (۳) ارائه شده است.

شکل (۳)- ساختار هسته سیستم کشف فرار مالیاتی



مطابق ساختار شکل (۳) الگوریتم بهینه سازی جستجوی هارمونی اقدام به بهینه سازی ترکیب ورودی های سیستم (۲۱ متغیر ورودی پایه) و پارامتر های هسته سیستم شبکه عصبی خواهد کرد. علاوه بر آن در هر قدم برای اندازه گیری تابع هزینه الگوریتم بهینه سازی از ۸۰ بار تکرار ساختار (۸ تکرار موازی اعتبارسنجی ضربدری ۱۰ بخشی) استفاده خواهد شد.

بازه های پارامترهای شبکه عصبی بهینه سازی شده توسط الگوریتم جستجوی هارمونی به این شرح است. تعداد لایه ها می تواند ۱ یا حداکثر ۲ باشد. تعداد نورون ها در هر لایه به ترتیب برای لایه اول برابر با ۴ نورون تا ۲۵ نورون و برای لایه دوم نیز ۰ نورون تا ۲۵ نورون می باشد (در صورت قرار گیری تعداد نورون های لایه دوم بر روی صفر، لایه فوق حذف خواهد شد). بازه های فوق بر پایه نظر خواهی از خبرگان در حوزه

هوش مصنوعی و یادگیری ماشینی و همچنین تکرار و بررسی نتایج سیستم بر مبنای بازه های مختلف پیشنهادی انتخاب شده است. همچنین برای آموزش شبکه عصبی از الگوریتم trainscg[1] با توابع انتقال تانژانت سیگموئید[2] در لایه ورودی و میانی و تابع انتقال Softmax در لایه خروجی بهره گرفته شده است. ساختار شبکه عصبی طبقه بندی دو دویی ذکر شده با برداری سازی[3] خروجی امکان ارائه نتایج احتمالی (احتمال فرار مالیاتی هر شرکت) را دارا می باشد. در شرایط استفاده از رگرسیون لجستیک نیز تنها ترکیب مناسب متغیرها و نسبت ها توسط الگوریتم جستجوی هارمونی از میان ۲۱ متغیر و نسبت مالی ورودی انتخاب خواهد شد. پارامترهای تنظیمی الگوریتم بهینه سازی جستجوی هارمونی به شرح جدول (۲) می باشد. لازم به ذکر است که پارامترهای فوق، پارامترهای پیشنهادی طراحان الگوریتم جستجوی هارمونی می باشند که با تکرار و بررسی بر روی مسائل مختلف حوزه بهینه سازی نتیجه شده است.

جدول (۲)- پارامترهای پیش فرض الگوریتم جستجوی هارمونی

Harmony memory size = 100
Number of new harmonies = 100
Fret width coefficient = 0.01
Pitch adjustment rate = 0.3
Fret width damp ratio = 0.999
Consideration rate = 0.9

همان طوری که ذکر شد از موارد مهم مورد استفاده در سیستم فرآیند، اعتبارسنجی ضربدری بوده که در این پژوهش از نوع ۱۰ بخشی آن استفاده شده است. در روش اعتبارسنجی ضربدری پایه ابتدا مجموعه داده های ورودی (نمونه ها) به ۱۰ بخش تقسیم می شود (به صورت تصادفی). سپس مدل با چهاربخش اول آموزش[4] دیده و با بخش پنجم (آخر) مورد آزمون[5] قرار می گیرد. در مرحله بعد مدل با بخش دوم تا پنجم آموزش دیده و با بخش اول داده مورد آزمون قرار می گیرد. فرآیند فوق ۱۰ بار در کل داده ها تکرار

1. Scaled Conjugate Gradient Backpropagation
2. Tansig
3 .Vectorization
4. Training
5. Testing

خواهد شد. یکی از محدودیت های این روش این مسأله است که برای بخش اعتبارسنجی[1] مورد نیاز، شبکه عصبی داده ای تخصیص داده نمی شود. برای برطرف شدن این مشکل روش جدیدی برای اعتبارسنجی ضربدری در شبکه عصبی طراحی شده است.

در این روش داده ها به ۱۰ بخش (تصادفی) تقسیم می شوند. سپس بخش اول به عنوان داده های آزمون (۲۰٪) و بخش دوم نیز به عنوان داده های اعتبارسنجی (۲۰٪) انتخاب شده مابقی داده ها (۶۰٪) برای آموزش سیستم مورد استفاده قرار خواهند گرفت. فرآیند فوق ۱۰ بار مورد تکرار قرار خواهد گرفت. در واقع در ساختار فوق با یک چرخش ۳ قسمتی داده روبه‌رو خواهیم بود. با توجه به توضیحات فوق مشاهده می شود که ۶۰٪ نمونه ها به منظور آموزش، ۲۰٪ به منظور آزمون و ۲۰٪ نیز برای اعتبارسنجی و جلوگیری از برازش بیش از اندازه[2] استفاده خواهند شد. همچنین برای جلوگیری از برازش بیش از اندازه از روش توقف زود هنگام[3] با توقف در۶ تکرار بهره گرفته شده است.

در بخش آماده سازی مدل های سیستم برای استفاده، به منظور افزایش اطمینان به مدل های اعلامی نهایی (در این پژوهش ۸۰ مدل نهایی) مدل ها وارد یک ساختار ذخیره سازی (به منظور نگهداری ۸۰ مدل) خواهند شد. داده های جدید (عملکرد ۱۳۹۱) وارد تک تک مدل های فوق شده و مجموعه نتایج احتمالی حاصل خواهد شد. سپس بر پایه معادله (۴) دو انحراف معیار از احتمالات پرت حذف خواهد شد. در معادله (۴)، α برابر با ۲ در نظر گرفته شده و X مجموعه احتمالات خروجی را شامل می شود. علاوه بر آن $average$ نشان دهنده اپراتور میانگین و STD نیز نشان دهنده اپراتور انحراف معیار می باشد. احتمالاتی که معادله (۴) برای آنها برقرار می باشد از مجموعه احتمالات حذف شده و برای تعیین احتمال فرار مالیاتی در شرکت مورد نظر، از باقی مانده احتمالات میانگین گرفته خواهد شد.

$$|X - average(X)| \geq \alpha * STD(X) \qquad (۴)$$

1. Validation Data
2. Over Fitting
3. Early Stopping

۳- جامعه آماری

جامعه آماری این پژوهش صنایع مواد غذایی و نساجی را با مجموعه کدهای آیسیک[1] جدول (۳) در بر می گیرد.

جدول (۳)- کدهای آیسیک و صنایع مورد استفاده

صنعت	کدهای آیسیک (سطح ۲)
مواد غذایی	۹۱۵۱،۹۱۵۲،۹۱۵۳،۹۱۵۴،۹۱۵۵
نساجی	۹۱۷۱،۹۱۷۲،۹۱۷۳،۹۱۸۱،۹۱۸۲،۹۱۹۱،۹۱۹۲

همچنین حوزه مورد بررسی، اشخاص حقوقی و مجموعه داده ورودی به سیستم سال های ۱۳۸۹ و ۱۳۹۰ را در بر می گیرد. برای آزمون سیستم بر روی داده های خارجی از عملکرد ۱۳۹۱ و مقایسه با نتایج حسابرسی آن سال استفاده شده است. علاوه بر آن خروجی سیستم (متغیر وابسته – شرکت های فراری و سالم) بر پایه معیاری تعیین شده است. معیار فوق نتایج دو دویی (۱ به معنای فرار مالیاتی و ۰ به معنای سلامت) را برای مجموعه شرکت ها در سال های ۱۳۸۹، ۱۳۹۰ و ۱۳۹۱ ارائه می دهد.

برای طراحی ساختار اصلی سیستم، از نرم افزار MATLAB R2014a بهره گرفته شده است. همچنین برای گزارش گیری مناسب و استاندارد از سیستم، از VBA[2] در محیط Microsoft Excel 2013 استفاده شده و ارتباطات بین دو نرم افزار کدنویسی شده است. علاوه بر موارد فوق برای بهبود عملکرد و کارایی سیستم در بخش هایی از زبان جاوا[3] در محیط MATLAB R2014a استفاده شده است. در پایان به منظور ارتباط با پایگاه داده و تبدیل و انتقال داده ها از SQL server 2014 بهره گرفته شده است. واحد پردازش مرکزی مورد استفاده Intel Core i7–4702MQ دارای ۴ هسته فیزیکی[4] و ۸ هسته منطقی[5] می باشد. علاوه بر آن هشت گیگابایت حافظه

1. International Standard Industrial Classification (ISIC)
2. Visual Basic for Applications (VBA)
3. Java
4. Physical Processor
5. Logical Processor

دسترسی تصادفی[1] و سیستم عامل Windows 8.1 Enterprise edition 64bit مورد استفاده قرار گرفته است.

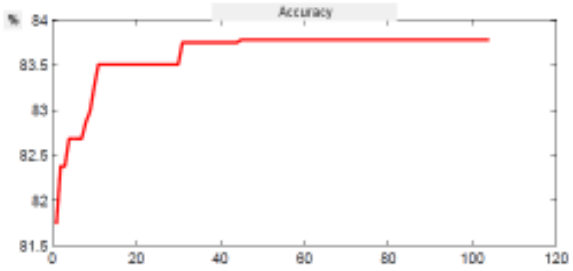
۴- جمع بندی و نتیجه گیری

در این بخش نتایج سیستم در دو صنعت مواد غذایی و نساجی به صورت جداگانه بر روی داده های آزمون (عملکرد ۱۳۸۹و ۱۳۹۰) و خارجی (عملکرد ۱۳۹۱) ارائه خواهد شد. نمودارهای مرتبط با دقت کلی تشخیص، دقت تشخیص شرکت های فراری و دقت تشخیص شرکت های سالم در هر بخش ارائه شده است.

صنعت مواد غذایی

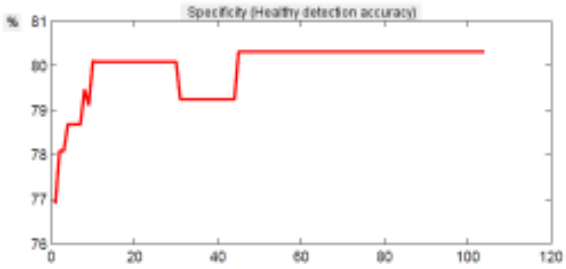
نمودارهای فرآیند بهینه سازی الگوریتم جستجوی هارمونی در ۳۰۰ قدم با توقف یک پنجم (عدم بهبود در۶۰ قدم) در صنعت مواد غذایی به ترتیب در ادغام با شبکه عصبی و رگرسیون لجستیک در زیر ارائه شده است. لازم به ذکر است که نتایج این بخش و بخش قبل بر مبنای میانگین دقت های حاصل از داده های آزمون با ۸۰ تکرار (۸ تکرار اعتبارسنجی ضربدری ۱۰ بخشی با حذف ۲ انحراف معیار ذکر شده) ارائه شده و از ارائه دقت های حاصل از داده های آموزش صرف نظر شده است. همچنین داده های خارجی، مجموعه داده های مربوط به عملکرد حسابرسی شده سال ۱۳۹۱ بوده که در آموزش و ساخت مدل مورد استفاده قرار نگرفته و بمنظور اعتبارسنجی مجموعه مدل های طراحی شده مورد استفاده قرار گرفته است.

شکل (۴)- دقت کلی تشخیص در داده های آزمون (شبکه عصبی)–صنعت مواد غذایی

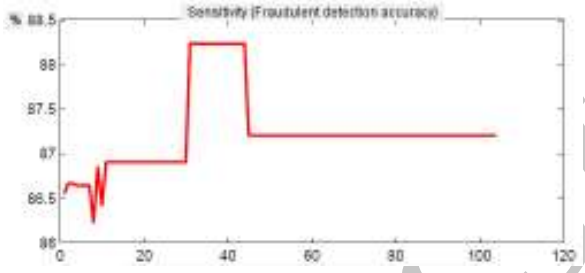


1. Random Access Memory (RAM)

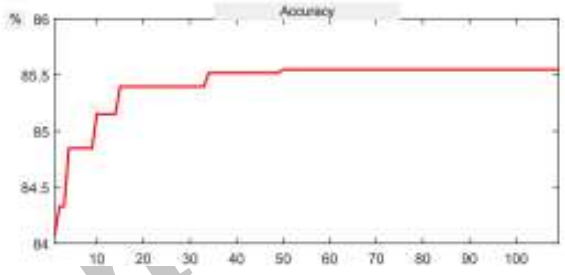
شکل (۵)- دقت تشخیص سالم در داده های آزمون (شبکه عصبی)–صنعت مواد غذایی



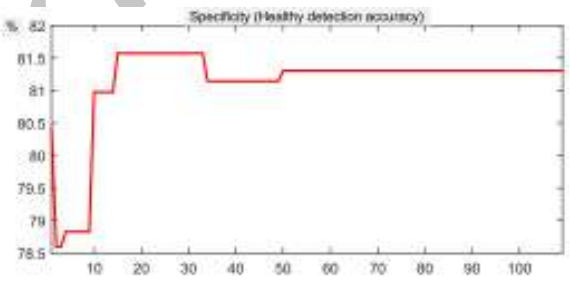
شکل (۶)- دقت تشخیص فراری در داده های آزمون (شبکه عصبی)–صنعت مواد غذایی



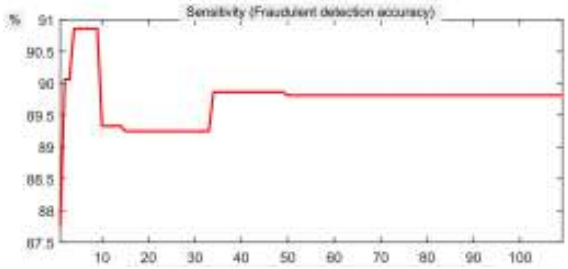
شکل (۷)- دقت کلی تشخیص در داده های آزمون (رگرسیون لجستیک)–صنعت مواد غذایی



شکل (۸)- دقت تشخیص سالم در داده های آزمون (رگرسیون لجستیک)–صنعت مواد غذایی



شکل (۹)- دقت تشخیص فراری در داده های آزمون (رگرسیون لجستیک)–صنعت مواد غذایی



تعداد شرکت ها و دقت های حاصل بر روی داده هـای آزمـون (عملکـرد ۱۳۸۹ و ۱۳۹۰) و خـارجی (عملکرد ۱۳۹۱) به شرح زیر می باشد:

جدول (۴)- تعداد شرکت های فراری و سالم در داده های آزمون و خارجی - صنعت مواد غذایی

داده های آزمون (عملکرد ۱۳۸۹ و ۱۳۹۰)			داده های خارجی (عملکرد ۱۳۹۱)		
تعداد کل شرکت ها	تعداد شرکت های فراری	تعداد شرکت های سالم	تعداد کل شرکت ها	تعداد شرکت های فراری	تعداد شرکت های سالم
۳۰۹۷	۲۸۸۸	۲۰۹	۱۱۳۵	۶۲	۱۰۵۷

جدول (۵)- دقت تشخیص شرکت های فراری و سالم در داده های آزمون و خارجی - صنعت مواد غذایی - شبکه عصبی

داده های آزمون (عملکرد ۱۳۸۹ و ۱۳۹۰)			داده های خارجی (عملکرد ۱۳۹۱)		
دقت کل	شرکت های فراری	شرکت های سالم	دقت کل	شرکت های فراری	شرکت های سالم
٪۷۸/۸۳	٪۳۱/۸۰	٪۲۰/۸۷	۹۲/۲۲٪	۸۳/۸۷٪	۹۲/۷۱٪

جدول (۶)- دقت تشخیص شرکت های فراری و سالم در داده های آزمون و خارجی - صنعت مواد غذایی - رگرسیون لجستیک

داده های آزمون (عملکرد ۱۳۸۹ و ۱۳۹۰)			داده های خارجی (عملکرد ۱۳۹۱)		
دقت کل	شرکت های فراری	شرکت های سالم	دقت کل	شرکت های فراری	شرکت های سالم
٪۸۵/۵۴	٪۸۱/۳۰	٪۸۹/۸۱	٪۸۶/۵۸	٪۷۹/۰۳	٪۸۷/۰۲

توجه بـه ایـن نکتـه ضـروری بـه نظـر مـی رسـد کـه در جـدول (۴)، مجمـوع تعـداد شـرکت های فراری و سـالم بـه میـزان ۱۶ نمونـه کمتـر از تعـداد کـل شـرکت هـا مـی باشـد. در شـرکت

هـای فـوق بـه دلیـل ناکـافی و یـا نامناسـب بـودن داده هـا، امکـان محاسـبه متغیرهـای ورودی برقـرار نبـوده و ایـن مجموعـه از شـرکت هـا از گـزارش نهـایی حـذف شـده اسـت. تحلیـل ذکـر شده قابل تعمیم به جدول (۷) نیز می باشد.

با مقایسـه دقـت هـای کـل حاصـل از اعمـال سیسـتم بـر روی داده هـای خـارجی (عملکـرد ۱۳۹۱) بـا دقـت هـای کـل آزمـون در شـرایط اسـتفاده از شـبکه عصـبی و رگرسـیون لجسـتیک نتیجـه مـی شـود کـه در هـر دو حالـت دقـت هـای اعمـال سیسـتم بـر روی داده هـای خـارجی بـالاتر از داده هـای آزمـون بـوده کـه ایـن مسـاله نشـان دهنـده مناسـب بـودن اعمـال سیسـتم طراحی شده بر روی داده های خـارجی یـا جدیـد مـی باشـد. در واقـع مـدل طراحـی شـده قابلیـت تعمیم دهی[1] مناسبی را در صنعت مـواد غـذایی دارا مـی باشـد کـه دقـت ایـن تعمـیم دهـی حتـی از دقـت اعمـال سیسـتم بـر روی داده هـای آزمـون نیـز بـالاتر اسـت. عـلاوه بـر مـوارد فـوق بـا مقایسه جداول (۵) و (۶) مـی تـوان نتیجـه گرفـت کـه دقـت سیسـتم بـر روی داده هـای آزمـون در حالـت اسـتفاده از رگرسـیون لجسـتیک بـه عنـوان هسـته سیسـتم نسـبت بـه شـبکه عصـبی بالاتر بوده اما شـرایط در اعمـال سیسـتم بـر روی داده هـای خـارجی معکـوس مـی باشـد. لـذا بـا توجه به نقش محـوری داده هـای خـارجی بـرای انتخـاب سـاختار بهینـه، تفـاوت نتـایج حاصـل از سـاختار شـبکه عصـبی و رگرسـیون لجسـتیک در ترکیـب بـا الگـوریتم جسـتجوی هـارمونی بـر روی داده های خارجی (عملکرد ۱۳۹۱) مورد آزمون آماری قرار خواهد گرفت.

به‌منظور آزمون ادعـای تفـاوت یـا عـدم تفـاوت آمـاری از آزمـون مکنمـار[2] بهـره گرفتـه شـده است (مکنمار، ژوئـن ۱۹۴۷). آزمـون مکنمـار یـک آزمـون ناپارامتریـک اسـت کـه اغلـب در مـورد داده های اسمی دو مقوله ای یا دو پاسخی مربـوط بـه دو نمونـهٔ مـرتبط یـا همبسـته بـه کـار مـی رود. ایـن آزمـون بـه‌طـور ویـژه در مـواردی بـه کـار مـی رود کـه مـی خـواهیم نظرهـا یـا عملکردهـای قبلـی یـا بعـدی موردهـا (در اینجـا مقایسـه نتـایج حاصـل از دو مـدل) را بـا هـم مقایسـه کنـیم. آزمـون مـورد اسـتفاده در ایـن پـژوهش از نـوع اصـلاح شـده ییـتس[3] مـی باشـد (ییـتس، ۱۹۳۴). همچنـین آمـاره آزمـون فـوق چـی اسـکوئر[4] اسـت. مقـدار حاصـل شـده از ایـن

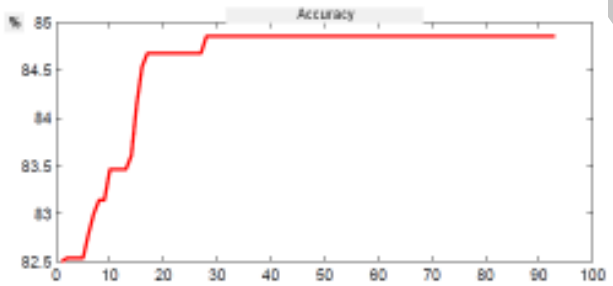
1. Generalization
2. Mc Nemar's Test
3. Yates's Correction
4. Chi-square Statistics

آزمون ۰۱۲۵/۷۸ و p-value متناظر با آن نیز برابر با ۰/۰۰۰ می باشد. نتیجه می شود که تفاوت های دو ساختار طبقه بندی شبکه عصبی و رگرسیون لجستیک در سطح اطمینان ۹۵٪ معنادار بوده و با توجه به بالاتر بودن تمامی دقت ها در حالت استفاده از شبکه عصبی به عنوان هسته سیستم در صنعت مواد غذایی در مجموعه داده های خارجی، ساختار فوق در ترکیب با الگوریتم جستجوی هارمونی به عنوان ساختار برتر در این صنعت انتخاب می شود.

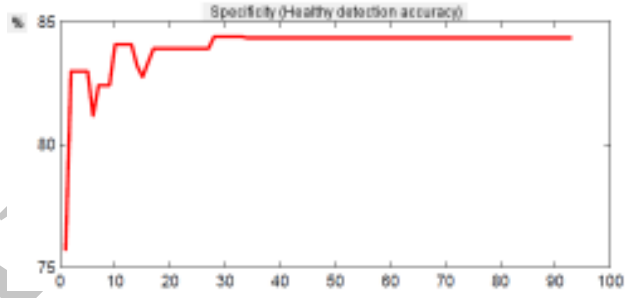
صنعت نساجی

نمودارهای فرآیند بهینه سازی الگوریتم جستجوی هارمونی در ۳۰۰ قدم با توقف یک پنجم (عدم بهبود در۶۰ قدم) در صنعت نساجی به ترتیب در ادغام با شبکه عصبی و رگرسیون لجستیک در زیر ارائه شده است:

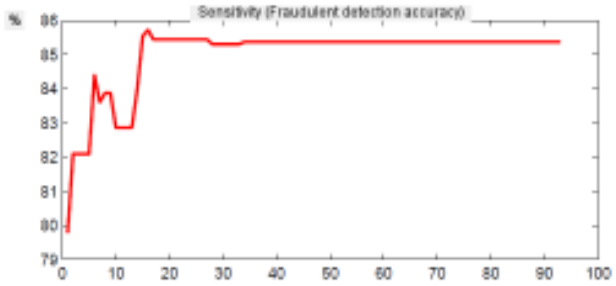
شکل (۱۰)- دقت کلی تشخیص در داده های آزمون (شبکه عصبی)–صنعت نساجی



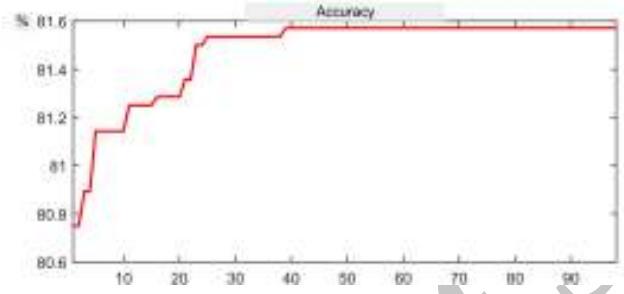
شکل (۱۱)- دقت تشخیص سالم در داده های آزمون (شبکه عصبی)–صنعت نساجی



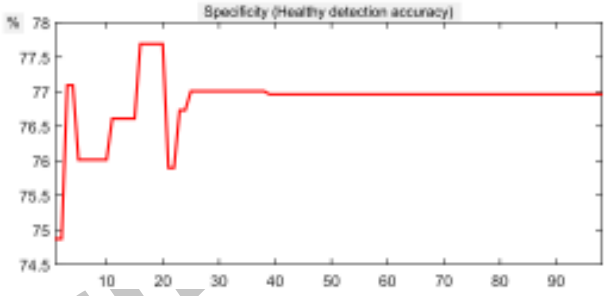
شکل (۱۲)- دقت تشخیص فراری در داده های آزمون (شبکه عصبی)–صنعت نساجی



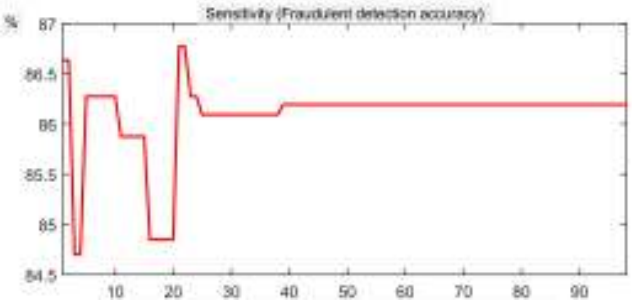
شکل (۱۳)- دقت کلی تشخیص در داده های آزمون (رگرسیون لجستیک)–صنعت نساجی



شکل (۱۴)- دقت تشخیص سالم در داده های آزمون (رگرسیون لجستیک)–صنعت نساجی



شکل (۱۵)- دقت تشخیص فراری در داده های آزمون (رگرسیون لجستیک)–صنعت نساجی



تعداد شرکت ها و دقت های حاصل در شرایط بهره گیری از شبکه عصبی و رگرسیون لجستیک بر روی داده های آزمون (عملکرد ۱۳۸۹ و ۱۳۹۰) و خارجی (عملکرد ۱۳۹۱) به شرح زیر می باشد:

جدول (۷)- تعداد شرکت های فراری و سالم در داده های آزمون و خارجی – صنعت نساجی

داده های آزمون (عملکرد ۱۳۸۹ و ۱۳۹۰)			داده های خارجی (عملکرد ۱۳۹۱)		
تعداد کل شرکت ها	تعداد شرکت های فراری	تعداد شرکت های سالم	تعداد کل شرکت ها	تعداد شرکت های فراری	تعداد شرکت های سالم
۲۳۵۶	۱۷۸	۲۱۷۸	۹۳۶	۶۹	۸۵۵

جدول (۸)- تعداد شرکت های فراری و سالم در داده های آزمون و خارجی - صنعت نساجی - شبکه عصبی

داده های آزمون (عملکرد ۱۳۸۹ و ۱۳۹۰)			داده های خارجی (عملکرد ۱۳۹۱)		
دقت کل	شرکت های فراری	شرکت های سالم	دقت کل	شرکت های فراری	شرکت های سالم
۸۴/٪۸۵	۸۴/۳۴٪	۸۵/۳۶٪	۸۲/۳۵٪	۸۴/۰۵٪	۸۲/۲۲٪

جدول (۹)- تعداد شرکت های فراری و سالم در داده های آزمون و خارجی - صنعت نساجی - رگرسیون لجستیک

داده های آزمون (عملکرد ۱۳۸۹ و ۱۳۹۰)			داده های خارجی (عملکرد ۱۳۹۱)		
دقت کل	شرکت های فراری	شرکت های سالم	دقت کل	شرکت های فراری	شرکت های سالم
۸۱/۵۷٪	٪۹۶/۷۶	۸۶/۱۹٪	۷۹/۲۶٪	۷۳/۹۱٪	۷۹/۶۹٪

مشاهده می شود که در حالت استفاده از شبکه عصبی و همچنین رگرسیون لجستیک دقت طبقه بندی در داده های آزمون قابل قبول بوده و با وجود کمتر بودن دقت تشخیص بر روی داده های خارجی (عملکرد ۱۳۹۱) نسبت به آزمون، مجموعه دقت های اعمال مدل بر روی داده های خارجی نیز مناسب می باشد. همچنین با مقایسه جداول (۸) و (۹) می توان نتیجه گرفت که در حالت استفاده از رگرسیون لجستیک در صنعت نساجی، تمامی دقت های حاصل از اعمال سیستم بر داده های آزمون و داده های خارجی نسبت به استفاده از شبکه عصبی کمتر می باشد. با اعمال آزمون مکنمار بر داده های خارجی، خروجی آزمون برابر با ۰۲۵۰/۳۸ و p-value متناظر با آن نیز ۰/۰۰۰ می باشد که در نتیجه تفاوت دو ساختار شبکه عصبی و رگرسیون لجستیک در صنعت نساجی در سطح

اطمینان ۹۵٪ معنادار بوده و با توجه به بالاتر بودن دقت شبکه عصبی، این ساختار در ترکیب با جستجوی هارمونی به عنوان ترکیب برتر انتخاب می شود.

متغیرهای مالی منتخب در صنایع مورد بررسی

در بخش قبل مشاهده شد که در دو صنعت مواد غذایی و نساجی دقت های حاصل از استفاده از شبکه عصبی به عنوان هسته سیستم در ترکیب با الگوریتم بهینه سازی جستجوی هارمونی از ترکیب رگرسیون لجستیک با این الگوریتم بالاتر بوده و این تفاوت ها از لحاظ آماری معنادار نیز می باشند. به این دلیل استفاده از شبکه عصبی در ترکیب با الگوریتم بهینه سازی جستجوی هارمونی در دو صنعت مورد بررسی به عنوان ترکیب برتر برگزیده می شود. جدول (۱۰) متغیرهای منتخب موثر در تشخیص فرار مالیاتی (انتخاب شده توسط سیستم) را در دو صنعت مواد غذایی و نساجی با استفاده از ترکیب ذکر شده ارائه می دهد.

جدول (۱۰)- متغیرهای ورودی منتخب سیستم در صنعت مواد غذایی و نساجی

صنعت مواد غذایی	صنعت نساجی
سود (زیان) قبل از بهره و مالیات به کل دارایی ها	فروش خالص به کل دارایی ها
سود (زیان) پس از کسر مالیات به کل دارایی ها	سود (زیان) پس از کسر مالیات به کل دارایی ها
کل دارایی ها	سود انباشته به کل دارایی ها
کل دارایی ها به حقوق صاحبان سهام	کل دارایی ها
کل بدهی ها به حقوق صاحبان سهام	سود (زیان) ناخالص فروش به کل دارایی ها
کل بدهی ها	سود (زیان) ناخالص فروش
اسناد دریافتنی	سود (زیان) پس از کسر مالیات
سود (زیان) ناخالص فروش	

"سود (زیان) پسر از کسر مالیات به کل دارایی ها"، "کل دارایی ها" و "سود (زیان) ناخالص فروش" از ورودی های تکراری در هر دو صنعت بوده اما برای بررسی بیشتر لازم است تا با تعداد تکرار بالا نسبت های مالی انتخابی سیستم مورد ارزیابی قرار گیرد. به طور کلی از موارد مهم موجود در سیستم فوق، گزارش نسبت های مالی انتخابی بوده که در بلندمدت اثر زیادی در شناخت ساختار فرار مالیاتی در کشور و در صنایع مختلف خواهد داشت. علاوه بر آن در جدول (۱۰) مشاهده می شود که به غیر از متغیرهای مالی ذکر شده، سایر متغیرهای انتخابی سیستم در دو صنعت مورد بررسی متفاوت می باشند. این

مسأله نشان از متفاوت بودن متغیرهای مالی تأثیرگذار به منظور پیش بینی فرار مالیاتی در صنایع مختلف دارد.

در مجموع نتیجه گیری می شود که اعمال ساختار فوق بر اطلاعات مالیاتی صنایع مواد غذایی و نساجی از توانایی بالایی برای تشخیص شرکت های فراری و سالم در داده های آزمون و به طور ویژه در داده های خارجی برخوردار می باشد. به قطع بررسی اعمال سیستم بر مجموعه داده های دیگر مانند داده های صنایع یا مناطق مختلف مالیاتی از اهمیت بالایی برخوردار می باشد. لازم به ذکر است که زمان مورد نیاز اجرای سیستم برای هر یک از صنایع مواد غذایی و نساجی در حدود ۱۰ ساعت در شرایط استفاده از پردازش موازی بوده است. توسعه این سیستم با بهره گیری از مدل های طبقه بندی و الگوریتم های بهینه سازی دیگر از موارد قابل بررسی آتی می باشد.

فهرست منابع

1. Cecchini, M., Aytug, H., Koehler, G. J., & Pathak, P. (2010). Detecting Management Fraud in Public *Management Science, 56*(7), 1146-1160 .

2. Cerullo, M. J., & Cerullo, V. (1999). Using neural networks to predict financial reporting fraud: Part 1. *omputer Fraud & Security*(5), 14-17 .

3. Geem, Z. W., Kim, J.-H., & Loganathan, G. V. (2001). A New Heuristic Optimization Algorithm: Harmony Search. *SIMULATION: Transactions of The Society for Modeling and Simulation International*, 60-68 .

4. González, P. C., & Velásquez, J. D. (2013). Using data mining technique to enhance tax evasion detection performance. *Expert Systems with Applications, 40*(5), 1427–1436 .

5. Jain, A. K., Mao, J., & Mohiuddin , K. M. (March 1996). Artificial Neural Networks: A Tutorial. *Computer - Special issue: neural computing, 29*(3), 31-44 .

6. Kirkos, E., Spathis, C., & Manolopoulos, Y. (2007). Data Mining techniques for the detection of fraudulent. *Expert Systems with Applications, 32*(4), 995–1003 .

7. Koskivaara, E. (2000). Different Pre-Processing Models for Financial Accounts when Using Neural networks for auditing .

8. Koskivaara, E. (2003). Artificial neural networks in auditing: state of the art. *Turku Centre for Computer Science* .

9. Kotsiantis, S., Koumanakos, E., Tzelepis, D., & Tampakas, V. (2006). Forecasting fraudulent financial statements using data mining. *International Journal of Computational Intelligence, 3*(2), 104-110 .

10. Liu, B., Xu, G., Xu, Q., & Zhang, N. (2012). Outlier Detection Data Mining of Tax Based on Cluster. Physics Procedia, 33, 168,

1694-9 .

11. Mc Nemar, Q. (June 1947). Note on the sampling error of the difference between correlated proportions or percentages. Psychometrika, 12(2), 153-157 .

12. Nahdavi, M., Fesanghary, M., & Damangir, E. (2007). An improvd harmony search algorithm for solving optimization problems. Appl Math Comput., 1567-1579 .

13. Ngai, E. W. T., Hu, Y., Wong, Y. H., Chen, Y., & Sun, X. (2011). The application of data mining techniques in financial fraud detection: A classification framework and an academic review of literature. Decision Support Systems, 50(3), 559–569 .

14. Odom, M. D., & Sharda, R. (1990). A neural network model for bankruptcy prediction. In Neural Networks. Paper presented at the IJCNN International Joint Conference on IEEE.

15. Phua, C., Lee, V., Smith, K., & Gayler, R. (2006). A Comprehensive Survey of Data Mining-based Fraud Detection Research. Paper presented at the Communications and Electronics, 2006. ICCE '06. First International Conference on Hanoi.

16. Ravisankar, P., Ravi, V., Rao, R., & Bose, I. (2011). Detection of financial statement fraud and feature selection using data mining techniques. Decision Support Systems, 50(2), 491–500 .

17. Spathis, C., Doumpos, M., & Zopounidis, C. (2002). Detecting falsified financial statements: a comparative study using multicriteria analysis and multivariate statistical techniques. European Accounting Review, 11(3), 509-535 .

18. Witten , I., & Frank, E. (2000). Data Mining: Practical Machine Learning Tools.

19. Yates, F. (1934). Contingency Tables Involving Small Numbers and the χ2 Test. Supplement to the Journal of the Royal Statistical Society, 1(2), 217-235 .

20. Yu, F., Qin, Z., & Jia, X.-l. (2003). Data mining application issues in fraudulent tax declaration detection. Paper presented at the Machine Learning and Cybernetics, 2003 International Conference on (Volume 4).

21. Zhou, W., & Kapoor, G. (2011). Detecting evolutionary financial statement fraud. Decision Support Systems, 50(3), 570-575 .