فصلنامه پژوهش در علوم، مهندسی و فناوری
دوره ۷، شماره ۱، بهار ۱۴۰۰
صفحات ۸۴-۹۸
ISSN: 2476-5082
www.uctjournals.com

یک روش هوشمند هدایت تحصیلی دانش آموزان با استفاده از داده‌کاوی

ام البنین نوری۱ امین شهرکی مقدم۲

۱- گروه مهندسی کامپیوتر واحد زاهدان، دانشگاه آزاد اسلامی، زاهدان، ایران
۲- گروه مهندسی کامپیوتر، واحد زاهدان، دانشگاه آزاد اسلامی، زاهدان، ایران

چکیده

اجرای طرح هدایت تحصیلی بر اساس استعداد و عملکرد تحصیلی دانش‌آموزان ، نقطه آغازی است برای کمک به کسب شغل آنان در آینده که تنها نباید بر دوش آموزش و پرورش باشد و نیازمند آموزش خانواده‌ها و هم‌افزایی جامعه دارد.در پایان نامه حاضر به بررسی و هدایت تحصیلی دانش آموزان و ارائه یک روش هوشمند پرداخته شده است. برای این امر ابتدا تحلیل دادها با روش های مرسوم بررسی شده و درنهایت ارائه یک روش عصبی – ژنتیک برای بهبود نتایج ارائه داده شده است. به این منظور به روش های رگرسیون خطی به تحلیل داده ها پرداخته و سپس یک شبکه عصبی آموزش داده شده و در نهایت به ارائه یک روش هوشمند عصبی-ژنتیک به بهبود نتایج پرداخته شده است. برای این منظور از پرونده هدایت تحصیلی ۳۵۰ دانش آموز و داده های آموزشی موجود استفاده شده و یک روش جامع ارائه داده شده و مدلی طراحی شده است که رشته‌ای را با بیشترین شانس موفقیت در آینده به دانش آموزان جدید پیشنهاد می دهد. نتایج کلی تحقیق نشان داد که مدلهای خطی MLR و MQR مقدار خطای RSE بالا و قابل بحث نمی باشد و مقدار RSE در مدل MLR برابر ۴۴٪ و در مدل MQR برابر ۳۵٪ به دست آمده است. سایر مقادیر خطای محاسبه شده یعنی MPE و MSE در محدوده زیادی می باشند و مقدار خطا RSE در مدل شبکه عصبی برابر ۱۴٪ و در مدل عصبی ژنتیک برابر ۹٪ به دست آمده است همانطور که مشاهده شد مدل ارائه شده عصبی ژنتیک دارای کمترین خطا است. نتایج کلی تحقیق نشان داد که یک شبکه عصبی بهینه شده میتواند نتایج و تخمین خوبی برای هدایت تحصیلی دانش آموزان رقم بزند که مدیران بتوانند به آن اتکا کنند.

کلیدواژه: داده‌کاوی، هدایت تحصیلی، شبکه عصبی، انتخاب رشته.

مقدمه

در سال‌های اخیر اهتمام مردم به علم‌آموزی و تحصیلات، روند رو به رشد ادامه تحصیل را به همراه داشته است. با افزایش دانش‌آموزان در مقطع متوسطه و رشته‌های مختلف، لزوم هدایت صحیح دانش‌آموزان در زمینه‌های گوناگون به ویژه بحث انتخاب رشته بیش از پیش احساس می‌شود. نظر به این که دانش‌آموزان باید در مقطع دبیرستان از میان رشته‌های موجود، مناسب‌ترین رشته را با توجه به معیارهایی از قبیل: در نظر گرفتن وضعیت تحصیلی مقاطع قبلی، فرم‌های مشاوره‌ای ، تجربه انتخاب رشته‌های قبلی اخذ نمایند و این امر نیازمند بررسی حجم وسیعی از داده‌های آموزشی سال‌های گذشته دانش‌آموزان است؛ لذا این فرآیند طولانی، انتخاب رشته‌ای مناسب را برای دانش‌آموز و برای مشاور مدرسه، به امری دشوار تبدیل نموده است.

از کارهایی که بر اساس داده‌کاوی در حوزه آموزش انجام گرفته است می‌توان به نمونه‌های زیر اشاره نمود:

مینایی و همکاران در سال ۱۳۹۱ شناسایی عوامل مؤثر بر افت تحصیلی با استفاده از قواعد انجمنی و تحلیل خوشه‌ای را مورد بررسی قرار داده‌اند. در این پژوهش تلاش برای پیاده‌سازی مدل‌های داده‌کاوی پیش بینی کننده به منظور پیش‌بینی وضعیت تحصیلی دانشجویان بر اساس مشخصات فردی و گذشته تحصیلی آن‌ها بوده است. با توجه به نتایج آماری که از ساخت مدل-های پیش‌بینی کننده وضعیت دانشجو در این پژوهش بدست آمده است می‌توان با اطمینان بالایی از آینده تحصیلی دانشجویان بر مبنای داده‌های گذشته اطلاع حاصل نمود (مینایی، بهروز، سادات میرافضل، هانی، ۱۳۹۱).

صفاریان و همکارانش در سال ۱۳۸۹ پیش بینی رتبه داوطلبان کنکور سراسری دانشگاه‌ها را با استفاده از روش‌های داده‌کاوی انجام داده‌اند که نتایج بکارگیری الگوریتم عصبی را نشان می‌دهد که با دقت بالای ۹۰ درصد می‌توان رتبه داوطلب در کنکور سراسری را تخمین زد و نتیجه بدست آمده نشان می‌دهد که رتبه داوطلب در کنکور سراسری تا حد بسیار بالایی تابعی از وضعیت تحصیلی و درسی آن‌ها در طی دوران تحصیل است (صفاریان، محسن، خیرآبادی، اعظم، ۱۳۸۹).

نوری و همکاران در سال ۱۳۸۹ پیش بینی نمره از طریق داده کاوی را مورد بررسی قرار داده‌اند در این تحقیق سعی شده بر اساس مسائل ذکر شده و عوامل مؤثر و تأثیرگذار از قبیل ترم (نیمسال اول یا دوم) و جنسیت و سن و وضعیت اشتغال نمره دانشجو پیش گویی شود در این مقاله وابستگی بین دروس از طریق الگوریتم های سبد خرید و پیش گویی نمره از طریق داده کاوی انجام می‌گیرد (نوری، بهاره، مقصودی، بهروز، شیخ احمدی، سید امیر، ۱۳۸۹).

طاهری و فرهادی در سال ۱۳۸۹ پیش‌بینی موفقیت یا عدم موفقیت دانشجویان در درس مبانی کامپیوتر را بررسی کرده‌اند. در این مقاله تکنیک‌های داده‌کاوی بر روی داده‌های مربوط به درس مبانی کامپیوتر اعمال و پیش‌بینی‌های مناسبی در مورد موفقیت و یا عدم موفقیت دانشجویان این درس ارائه شده است در حقیقت عوامل مؤثر بر نمره دانشجو در این درس توصیف و معرفی شد تا به کمک این آگاهی دانشجو بتواند نقاط ضعف خود را برطرف و یا حداقل نماید. (طاهری ، فرهادی ، ۱۳۸۹).

حاتملو و هاشمی‌نژاد در سال ۱۳۸۷ تحلیل رفتار آموزشی دانشجویان با استفاده از تکنیک‌های داده‌کاوی را مورد بررسی قرار داده اند در این مقاله نشان می‌دهد که چگونه می‌توان از نتایج الگوریتم‌های داده‌کاوی در جهت شناخت رفتار و وضعیت دانشجویان و تصمیم‌گیری برای تغییر برنامه‌های آموزشی در جهت افزایش بازدهی آن‌ها و رفع مشکلات آموزشی و همچنین پیش‌بینی مشکلات احتمالی آینده و جلوگیری از وقوع آن‌ها استفاده کرد (حاتملو، عبدالرضا، هاشمی نژاد، سیدجواد، ۱۳۸۷).

خیرآبادی و مینایی در سال ۱۳۸۹ به بررسی و خوشه‌بندی نتایج ارزشیابی اساتید دانشگاه با استفاده از روش‌های داده‌کاوی پرداخته است .در این تحقیق داده‌های مربوط به اطلاعات شخصی و نمرات ارزشیابی و کلاس‌های اساتید مورد بررسی قرار گرفته و دو روش ارزشیابی تلفیقی و ارزشیابی الکترونیکی با هم مقایسه شده‌اند. همچنین با استفاده از روش خوشه‌بندی تعداد خوشه‌های موجود در هر یک از روش‌ها تعیین و مورد تحلیل قرار گرفته است (خیرآبادی، اعظم، مینایی بیدگل، بهروز، ۱۳۸۹).

بدیهی است دانش‌آموز باید بر اساس ملاک‌های مناسب علمی تصمیماتی اتخاذ نماید و احتمال شکست در آینده را کاهش دهد. بنابراین با توجه به مشکلات موجود در زمینه استخراج دانش از پرونده‌های مشاوره‌ای، تحصیلی و استفاده از آن‌ها به

منظور تصمیم‌گیری‌های مناسب، تکنیک‌های داده کاوی در امر آموزش مورد توجه قرار گرفته است، این زمینه تحقیقاتی جدید داده کاوی آموزشی نامیده می‌شود که به توسعه روش‌های کشف دانش از داده‌های محیط آموزشی می‌پردازد .بر همین اساس، ما به کمک حجم عظیمی از داده‌های انتخاب رشته سال‌های قبل و به منظور استفاده از دانش نهفته موجود در آن ها، از روش‌های داده کاوی و تکنیک داده کاوی برای تصمیم گیری و هدایت علمی دانش آموزان استفاده نموده‌ایم.

در شیوه کنونی(سال ۱۳۹۷) انتخاب رشته در مدارس با توجه به فرم مشاوره که امتیازات مشاوره‌ای در آن ثبت می‌شود و فرم هدایت تحصیلی که خلاصه‌ای از نمرات متوسطه اول و میانگین نمرات برای هر رشته در آن درج شده، انتخاب رشته صورت می‌گیرد. یعنی امتیاز مشاوره‌ای برای هر رشته با امتیاز موجود در هدایت تحصیلی آن رشته جمع شده و به صورت صعودی مرتب می‌شود و در نهایت براساس آن اولویت‌ها مشخص می گردد. ما در این پایان نامه با تحلیل انتخاب رشته دانش‌آموزان در سالهای ۹۱و۹۰ و کشف دانش نهفته در آن‌ها مدلی را طراحی می‌نماییم که رشته‌ای را با بیشترین شانس موفقیت در آینده به دانش آموزان جدید پیشنهاد می‌دهد .این امر می‌تواند بخشی از وظایف مشاور تحصیلی مدارس را پوشش دهد .نکته قابل توجه آن است که قوانین بدست آمده از تکنیک‌های داده‌کاوی وسیع تر از روش‌های مرسوم است و ممکن است روش‌هایی استخراج گردد که در سیستم فعلی وجود ندارد و این مسئله می‌تواند شانس انتخاب رشته را برای دانش آموز بسیار ارتقاء دهد.

نگاه گسترده به نقش آموزش و پرورش، به عنوان عاملی مؤثر در فعلیت بخشیدن به امکانات بالقوه و ذاتی افراد، انتقال اندوخته تجارب گذشتگان، ارائه ارزش‌های مطلوب، افزایش معلومات و ایجاد مهارت‌های لازم در افراد برای زندگی و بالاخره تسهیل سیر حرکت وجودی آدمی به سوی کمال، اهمیت امروزه‌ی این نهاد اجتماعی را بیش از پیش نشان می‌دهد .گسترش کمی آموزش و پرورش نشانگر گرایش و استقبال همگان به اثربخشی این نهاد و گسترش کیفی آن ناشی از توسعه معارف، علوم و رشد بی وقفه‌ی تکنولوژی است؛ که این دو، بر پیچیدگی و ظرافت جریان آموزش و پرورش افزوده است (صافی، احمد ۱۳۹۲). نقش راهنمایی و مشاوره در آموزش و پرورش افراد، به حدی است که عده‌ای از دانشمندان آن را مترادف با معنای تعلیم و تربیت شمرده اند و بعضی، راهنمایی و مشاوره را تسهیل کننده‌ی جریان تعلیم و تربیت می‌دانند (قاضی، قاسم، ۱۳۸۵). اما مشکل اساسی که در مسیر هدایت محصلین وجود دارد، وجود روابط پیچیده‌ی تحصیلی بر سر راه انتخاب رشته‌ی دانش آموزان است (اعلایی، غلامحسین، ۱۳۷۱). کشف این روابط پیچیده و مولفه‌های مرتبط با آن، نیازمند تحلیل و بررسی عوامل متعدد تحصیلی موجود در نظام‌های آموزشی است.

روش پیشنهادی

هدف از این پژوهش ارائه یک روش جدید برای راهنمایی و انتخاب رشته دانش آموزان است تا بتوانند بهترین رشته را انتخاب کنند که ابتدا به بررسی رگرسیون خطی پرداخته می شود و بعد نتایج با شبکه عصبی مقایسه شده و در نهایت به ارائه یک روش عصبی-ژنتیک پرداخته و این روش کاملا پویا بر اساس مشخصات خود دانش آموز ارائه می شود تا بهترین رشته برای شخص مورد نظر پیشنهاد شود.ضمنا جامعه آماری تحقیق دبیرستانهای دخترانه ناحیه ۲ شهر زاهدان می باشد.

در روش پیشنهادی ما یک مجموعه داده بزرگ به عنوان مجموعه داده آموزش داریم که متعلق به دانش آموزانی است که در گذشته انتخاب رشته کرده اند و مشخص است که رشته انتخاب شده توسط آنها نتیجه بخش بوده یا خیر؟ و یک مجموعه داده هم وجود دارد که وضعیت دانش آموزان در آن مشخص شده است و می خواهیم با استفاده از این دو مجموعه داده و روش پیشنهادی بهترین رشته را برای دانش آموزان جدید پیشنهاد دهیم. هر یک از این موارد در ادامه توضیح داده شده اند.

در این پژوهش، رگرسیون های خطی وظیفه مدل سازی و ارائه مدل خطی را برعهده دارند و همچنین وظیفه الگوریتم ژنتیک بهینه سازی شبکه عصبی(غیر خطی) می باشد. در این حالت می‌توان گفت که خود شبکه عصبی ارائه شده نقش تابع شایستگی الگوریتم ژنتیک را دارد. در روش پیشنهادی یک مجموعه از ترکیب‌های مختلف اطلاعات دانش آموزان است که مقادیر آن مشخص شده است.با توجه به این اطلاعات و سه ابزار رگریسون خطی ، الگوریتم ژنتیک و شبکه عصبی وضعیت ترکیب‌های جدید پیش‌بینی خواهد شد تا بهترین گزینه پیشنهاد شود.

مدلسازی

بعد از انجام تقسیم‌بندی، دیتا ها در سه دسته آموزشی ، تست، و ارزیابی از نرم‌افزار مینی تب برای ایجاد مدل ریاضی، ما بین این توصیف‌گرها و فعالیت آموزشی و هدایت تحصیلی استفاده شده است.

در این تحقیق ابتدا از مدل های رگرسیون خطی استفاده شده تا میزان موفقیت دانش آموزان مشخص شود و درنهایت از شبکه عصبی استفاده شده است که بعد از مشخص شدن میزان موفقیت هر یک از روش ها در پیش بینی هدایت تحصیلی به ارائه یک روش ترکیبی عصبی-ژنتیک پرداخته شده است. چند روش مدل‌سازی که در این تحقیق استفاده شده است عبارتند از:

۱-مدلهای رگرسیونی خطی،۲-مدلهای شبکه عصبی،۳-مدلهای بهینه سازی

مدل سازی با رگرسیون خطی- چندگانه (MLR)

برای انجام این آنالیز ابتدا مقادیر۸ توصیف‌گر باقی‌مانده به عنوان متغیرهای مستقل انتخاب و میزان هدایت تحصیلی (R٪) به عنوان متغیر وابسته وارد نرم‌افزار مینی تب می‌شود. متغیرهای مستقل ده‌ستون از صفحه کاری نرم‌افزار مینی تب را تشکیل می‌دهند. این ده‌ستون به عنوان متغیرهای x (متغیرهای مستقل) در قسمت آنالیز رگرسیونی وارد خواهند شد. یک ستون دیگر که حاوی مقادیر R ٪ واقعی می‌باشد، در صفحه کاری مینی تب تشکیل می‌شود. این ستون نیز به عنوان متغیرهای y (متغیرهای وابسته) در قسمت آنالیز رگرسیونی وارد می‌شوند سطح اطمینان **۹۵**٪ انتخاب شده برنامه اجرا می‌شود.

مدل سازی با رگرسیون درجه دوم چندگانه (MQR)

همانند مدل‌سازی MLR، برای انجام این آنالیز هم ابتدا مقادیر ۸ توصیف‌گر باقی‌مانده به عنوان متغیرهای مستقل انتخاب و میزان تاثیر هدایت تحصیلی یا موفقیت دانش آموز (R٪) به عنوان متغیر وابسته وارد می‌شود. به این صورت مقادیر ۸ توصیف-گر باقی‌مانده در ده ستون از اول و ستون‌های بعدی نیز با توجه به معادله درجه دوم فرض شده، در صفحه کاری مینی تب تشکیل داده می‌شود. این بیست ستون به عنوان متغیرهای x در قسمت آنالیز رگرسیونی مینی تب وارد خواهند شد

مراحل الگوریتم پیاده سازی شده

هدف ایجاد مدلی برای تشخیص سطح موفقیت در رشته انتخابی بر اساس پارامترهای اصلی تعیین کننده است، مدلی که از جامعیت بالایی برخوردار باشد و سرعت اجرای بالایی داشته و نسبت به مدل‌های دیگر از نرخ تشخیص بالایی برخوردار باشد. به این منظور از برنامه‌های موجود در یک دیتاست آموزش استفاده می‌کنیم. این مجموعه داده از دانش آموزان است. ابتدا با استفاده از روش‌های استاندارد پایگاه داده مورد نظر را نرمال کرده و داده‌هایی که اطلاعات آن‌ها ناقص یا گم شده است را حذف می‌کنیم. در ادامه دو مرحله داده کاوی با دو الگوریتم متفاوت روی این مجموعه داده به انجام می‌رسد. گام اول ایجاد مدل های خطی با اعمال رگرسیون های خطی که مدل های خطی را ارائه میدهد و همچنین متغییر هایی که تاثیر کمی بر خروجی دارد را نشان میدهد.

الگوریتم دوم مبتنی بر شبکه‌های عصبی بوده که در ادامه توضیح داده شده است.

شبکه عصبی طراحی شده در این پژوهش یک شبکه رو به جلو با هشت لایه پنهان است. در لایه اول از هشت نرون استفاده شده است در قسمت آموزش قسمتی از داده‌ها به شبکه داده می‌شود و پس از بهینه سازی توسط الگوریتم ژنتیک وزن‌های شبکه به نحوی تنظیم خواهند شد که خطای کمتری بین هدف و خروجی پیش‌بینی شده ایجاد گردد.

جزئیات شبکه عصبی به شرح زیر است:

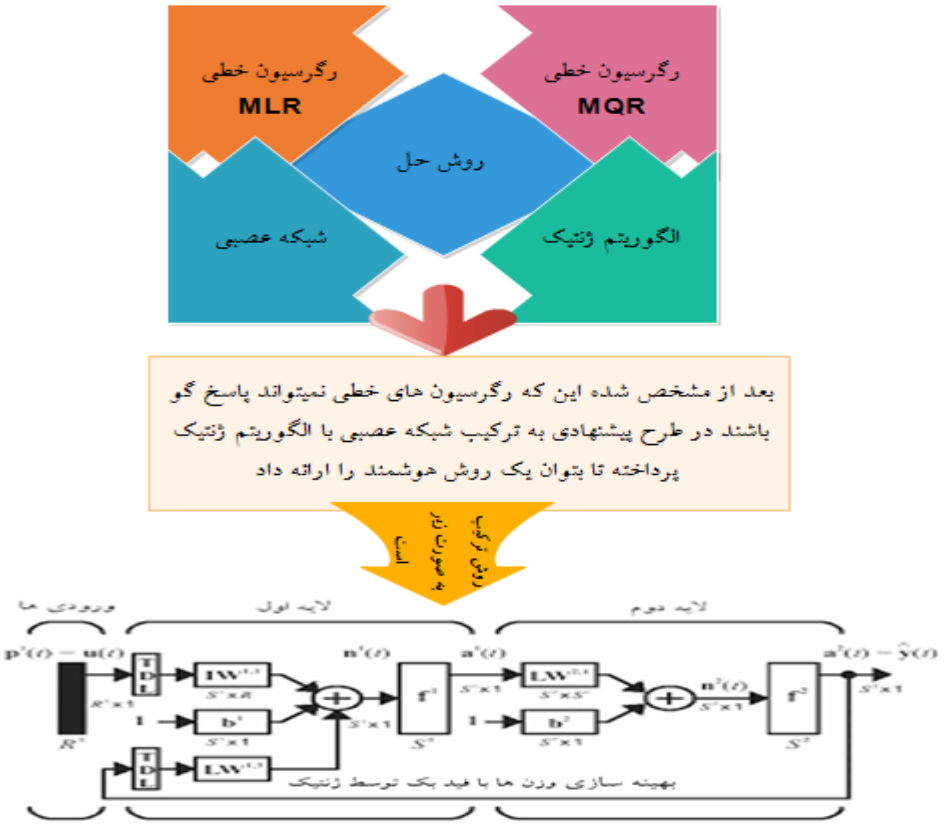
این شبکه عصبی دارای هشت ورودی،هشت لایه پنهان و یک خروجی است. ورودی‌ها همان ویژگی‌های دانش آموزان می‌باشند، در هر لایه پنهان یک ویژگی با چهار مقدار قرار دارد و خروجی سطح موفقیت یا متناسب بودن رشته تحصیلی است. برای هر خروجی رابطه‌ای به شکل زیر وجود دارد:

$$Result=w_1F_1+w_2F_2+w_3F_3+w_4F_4+w_5F_5+ w_6F_6+w_7F_7+w_8F_8$$

F2،F1و... همان هشت ویژگی انتخاب شده می‌باشند و w ها وزن هر ویژگی است. مقدار w ها بر اساس مقادیر داده آموزش به دست می‌آیند در واقع شبکه عصبی یاد می‌گیرد که برای هر مجموع داده چه ضریبی را برای هر ویژگی انتخاب کند. در

نهایت با شبکه عصبی بهینه شده و الگوریتم ژنتیک به بررسی و ارائه یک روش ترکیبی عصبی-ژنتیک پرداخته تا بتوان با دقت بالا تری به پیشبینی و هدایت تحصیلی دانش اموزان پرداخت. در این مدل از ۸ وردی که در ابتدا فصل مشخص شده بود استفاده کردیم که هر ورودی متعلق به یک کروموزم هست و خروجی تحقیق میزان موفقیت دانش اموزان است مدل استفاده شده در روش پیشنهادی در شکل1- مشخص شده است.

در نهایت همان‌طور که مشاهده می‌شود، این الگوریتم از دو بخش اصلی تشکیل شده است. در ابتدا مدل خطی ارائه میشود تا بهترین متغیرها مشخص شود. دقت شود در این حالت تابع شایستگی الگوریتم ژنتیک خود یک شبکه عصبی است که کارایی وزن‌های ایجاد شده را بررسی می‌نماید. سپس وزن‌های ایجاد شده به‌عنوان ورودی و همراه با داده‌های اصلی مسئله به شبکه عصبی اصلی داده می‌شود. این شبکه وظیفه رده‌بندی داده‌ها را بر عهده دارد. در بخش‌های بعد نتایج حاصل از پیاده‌سازی این الگوریتم روی داده‌های مختلف را مورد بررسی قرار خواهیم داد.

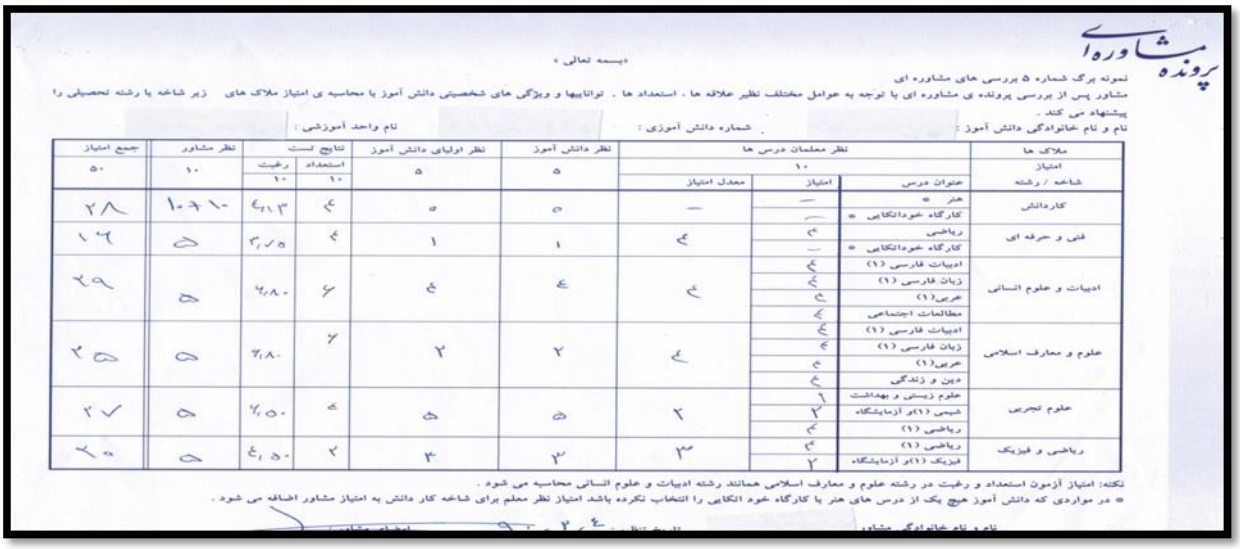


شکل-۱: فلوچارت مسئله

داده های موجود در مدارس:

در مدارس برای هر دانش آموز پرونده مشاوره ای و تحصیلی وجود دارد وانتخاب رشته برای هر دانش آموز بر اساس فرم ها و اطلاعات موجود در این پرونده صورت می گیرد که نمونه ای از این فرم ها درذیل بررسی می شود.

فرم مشاوره ای شماره پنج که نمونه آن در شکل2- آمده و برای هررشته امتیازاتی بدین شرح داده شده است



پرونده مشاوره ای

«بسمه تعالی»

نمونه برگ شماره ۵ بررسی های مشاوره ای

مشاور پس از بررسی پرونده ی مشاوره ای با توجه به عوامل مختلف نظیر علاقه ها ، استعداد ها ، توانایی ها و ویژگی های شخصیتی دانش آموز با محاسبه ی امتیاز ملاک های زیر شاخه یا رشته تحصیلی را پیشنهاد می کند .

نام و نام خانوادگی دانش آموز : شماره دانش آموزی : نام واحد آموزشی :

ملاک ها / امتیاز / شاخه / رشته	نظر معلمان درس ها ۱۰		نظر دانش آموز ۵	نظر اولیای دانش آموز ۵	نتایج تست: استعداد ۱۰	نتایج تست: رغبت ۱۰	نظر مشاور ۱۰	جمع امتیاز ۵۰
	عنوان درس	امتیاز / معدل امتیاز						
کاردانش	هنر *	—	۵	۵	۴	۴٫۱۳	۱۰٫۲۰	۲۸
	کارگاه خوداتکایی *	—						
فنی و حرفه ای	ریاضی	۴	۱	۱	۴	۳٫۷۵	۵	۱۰۶
	کارگاه خوداتکایی *	—						
ادبیات و علوم انسانی	ادبیات فارسی (۱)	۴	۴	۴	۶	۶٫۸۰	۵	۲۹
	زبان فارسی (۱)	۴						
	عربی (۱)	۴						
	مطالعات اجتماعی	۴						
علوم و معارف اسلامی	ادبیات فارسی (۱)	۴	۲	۲	۷	۷٫۸۰	۵	۲۵
	زبان فارسی (۱)	۴						
	عربی (۱)	۴						
	دین و زندگی	۴						
علوم تجربی	علوم زیستی و بهداشت	۱	۵	۵	۴	۷٫۵۰	۵	۲۷
	شیمی (۱) و آزمایشگاه	۲						
	ریاضی (۱)	۴						
ریاضی و فیزیک	ریاضی (۱)	۳	۳	۳	۲	۴٫۵۰	۵	۲۰
	فیزیک (۱) و آزمایشگاه	۲						

نکته: امتیاز آزمون استعداد و رغبت در رشته علوم و معارف اسلامی همانند رشته ادبیات و علوم انسانی محاسبه می شود .

* در مواردی که دانش آموز هیچ یک از درس های هنر یا کارگاه خود اتکایی را انتخاب نکرده باشد امتیاز نظر معلم برای شاخه کار دانش به امتیاز مشاور اضافه می شود .

نام و نام خانوادگی مشاور

شکل-۲فرم مشاوره ای شماره پنج

نظر معلمان درس ها برای هر رشته که محدوده آن از ۰ تا ۱۰ امتیاز است شامل: نظر معلمان درس ها برای کاردانش ، فنی و حرفه ای، ادبیات و علوم انسانی، علوم و معارف اسلامی ، علوم تجربی و ریاضی فیزیک می باشد.

نظر دانش آموز برای هر رشته که محدوده آن از ۰ تا ۵ امتیاز است شامل : نظر دانش آموز برای کاردانش ، فنی و حرفه ای، ادبیات و علوم انسانی، علوم و معارف اسلامی ، علوم تجربی و ریاضی فیزیک می باشد.

نظر اولیای دانش آموز برای هر رشته که محدوده آن از ۰ تا ۵ امتیاز است شامل : نظراولیای دانش آموز برای کاردانش ، فنی و حرفه ای، ادبیات و علوم انسانی، علوم و معارف اسلامی ، علوم تجربی و ریاضی فیزیک می باشد.

استعداد دانش آموز برای هر رشته که محدوده آن از ۰ تا ۱۰ امتیاز است شامل : استعداد دانش آموز برای کاردانش ، فنی و حرفه ای، ادبیات و علوم انسانی، علوم و معارف اسلامی ، علوم تجربی و ریاضی فیزیک می باشد. (استعداد بر اساس نتایج آزمون می باشد)

رغبت دانش آموزبرای هر رشته که محدوده آن از ۰ تا ۱۰ امتیاز است شامل : رغبت دانش آموز برای کاردانش ، فنی و حرفه ای، ادبیات و علوم انسانی، علوم و معارف اسلامی ، علوم تجربی و ریاضی فیزیک می باشد. (رغبت بر اساس نتایج آزمون می باشد).

نظر مشاور برای هر رشته که محدوده آن از ۰ تا ۱۰ امتیاز است شامل : نظر مشاور برای کاردانش ، فنی و حرفه ای، ادبیات و علوم انسانی، علوم و معارف اسلامی ، علوم تجربی و ریاضی فیزیک می باشد.

مجموع امتیاز ها برای هر رشته حداکثر ۵۰ امتیاز است. در فرم شماره ۶ هدایت تحصیلی سوابق تحصیلی دانش آموز وجود دارد که در شکل3- مشاهده می کنید.

شکل-۳فرم مشاوره ای شماره شش

مدل

در این تحقیق به میزان موفقیت دانش آموزان پرداخته شده است، شبیه سازی به چهار روش رگرسیون خطی چندگانه، رگرسیون درجه دوم چندگانه و شبکه عصبی svm و همچنین الگوریتم ژنتیک صورت گرفت. در هرکدام از مدل‌سازی‌ها ابتدا با استفاده از داده‌هایی که از پرونده دانش آموزان استخراج شده است به این صورت که ابتدا از مدل رگرسیون (یعنی MLR و MQR) که قبلا در فصل سه توضیح کامل داده شده است استفاده شده و ضرایب معادلات از طریق رگرسیون به دست آمد است. برای مدل‌سازی شبکه عصبی مصنوعی از شبکه بردار پشتیبان (SVM)، استفاده شده است برای بهینه کردن شبکه عصبی از سه نوع خطا به نامهای خطای انحراف معیار (ریشه دوم میانگین مربعات خطا (RMSE))، میانگین خطای پیش‌بینی (MPE) و خطای استاندارد نسبی (RSE) برای محاسبه مقادیر خطای مدل‌ها استفاده شد. از مقدار RMSE برای مقایسه نهایی استفاده شده است.

در نهایت شبکه عصبی را با الگوریتم ژنتیک ترکیب کرده تا بتوانیم یک شبکه عصبی بهینه شده را آموزش دهیم و نتایج دقیق تری نسبت به قبل دست یابیم.

نتایج

در این بخش به ترتیب نتایج مربوط به قسمت مدل MLR، مدل MQR و شبکه عصبی مصنوعی و الگوریتم ژنتیک ارائه می‌شود. موارد بحث شده در این فصل به قرار زیر است:

- ✓ ایجاد مدل MLR ارائه و بررسی نتایج
- ✓ مدلسازی با روش رگرسیون درجه دوم چندگانه (MQR) و ارائه و بررسی نتایج
- ✓ طراحی و بهینه‌سازی شبکه عصبی مصنوعی
- ✓ استفاده از الگوریتم ژنتیک برای بهینه سازی نتایج شبکه عصبی

ایجاد مدل ***MLR*** و نتایج حاصل از آن

نتایج حاصل از مدل سازی MLR در جدول-۱نشان داده شده وهمانطور که مشخص است با این اهم که سطح اطمینان بالای ۹۰ درصد انتخاب شده است و در مواردی که P_Value ها بزرگتر از ۰/۰۵ باشد، این ضرایب معنی دار نبوده و بایستی از معادله حذف شوند.

جدول۱: مقادیر ضرایب معادله در شبیه سازی MLR

	ضریب	P_Value
Constant	30.5516145522467	0.00694828622975817
X_1	2.54535752644853	0.00634198960121721
X_2	17.1201506539206	0.0627146552233314
X_3	-1.99063899226310	6.88921610058175e-05
X_4	-2.68292175929888	0.00219372179828199
X_5	-3.11733259444988	0.00839428605604618
X_6	69.0036832837511	0.00688965109334102
X_7	-18.4247817093458	0.00238559970341119
R		۴۵,۶

معادله خط

$$\%R = ۳۰,۵+ ۲,۵X_1 + ۱۷,۱X_2 -۱,۹X_3 -۲,۶X_4 -۳,۱X_5 + ۶۹X_6 -۱۸,۴X_7$$

پارامتر های موجود در جدول-۲ :فیلد نوع جنسیت ، جهت ورود نظر معلمان درس ها برای رشته های مختلف، جهت ورود نظر دانش آموز برای رشته های مختلف، جهت ورود نظر والدین برای رشته های مختلف، جهت ورود استعداد دانش آموز برای رشته های مختلف، جهت ورود علاقه دانش آموز برای رشته های مختلف، جهت ورود میانگین نمرات اصلی برای رشته های مختلف است.

جدول۲- پارامتر های اصلی حل

X1	sex
X2	Parent
X3	Teacher
X4	Student
X5	Talent
X6	interest
X7	MMS

که مجددا برای بررسی دقیق نتایج آنالیز رگرسیونی با حذف متغیر X_2 انجام شده که نتایج آن در جدول-۳ آمده است.همانطور مشخص است r تغییر آنچنانی نکرده است و مشخص می شود تاثیر کمی بر نتایج خروجی داشته است.

جدول۳- آنالیز رگرسیونی مجدد با سطح اطمینان ۹۵٪ پس از حذف ضرایب فاقد معنی دار

	ضریب	P_Value
Constant	142.620946546431	0
X_1	7.63034981530772	0.00756645530385296
X_3	-14.4940842013773	0.000762846536914593
X_4	-4.82219473997953	0.00488570895641290
X_5	-2.63494016160438	0.000243524968724989
X_6	95.3052771940821	0.00278779086956168
X_7	-8.43110892257404	0.000699967531969618
R		۴۵/۳

معادله نهایی بدست آمده برای هر کدام از متغیرهابه شرح زیر است:

$$\%R = 142 + 7.6X_1 - 14.5\ X_3 - 4.8\ X_4 - 2.63X_5 + 95.3\ X_6 - 8.43\ X_7$$

محاسبات خطا برای مدل *MLR*

برای صحت سنجی مدل فرض شده با معادله بدست آمده که بفهمیم تا چه اندازه به داده های مستخرج از پرونده دانش آموزان ما نزدیک است، خطای بین پاسخ حاصل از معادله رگرسیونی (مقادیر پیش‌بینی) و داده های تجربی (مقادیر واقعی) به سه صورت RSE، MPE و RMSE محاسبه می‌شوند و در جدول (۳-۳) خطاهای داده‌های اصلی که برای یافتن و پردازش معادله به کار رفته است، مشاهده می‌گردد. همانگونه که مشاهده می‌شود مقادیر خطاها در شبیه سازی با روش رگرسیون خطی چندگانه (MLR) بسیار بالا و حدود ۴۸ ٪ می‌باشد. مقادیر خطای MSE و MPE نیز این ادعا را تایید می کنند. از این رو مشخص میشود هیچ رابطه خطی بین هدایت تحصیلی دانش آموزان و میزان موفقیت آنها وجود ندارد. که در جدول-۴ نشان داده شده است.

جدول۴- خطاهای نسبی در مدل MLR و ضریب همبستگی مدل

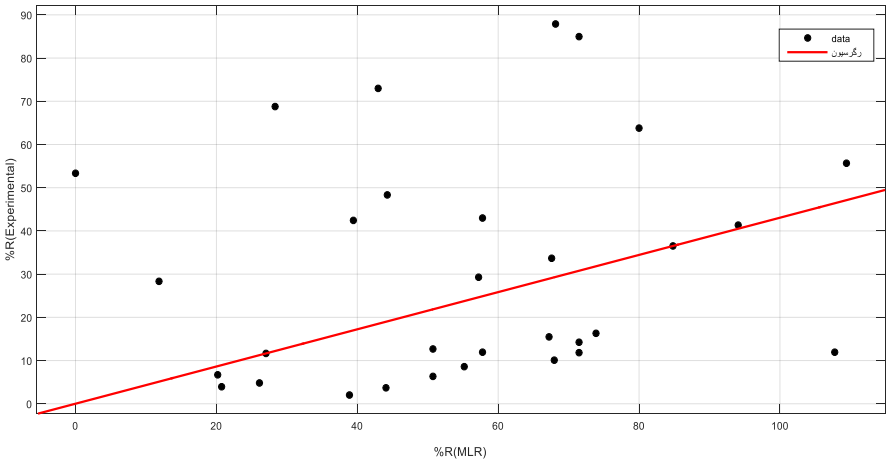
R	RMSE	MPE	RSE(%)
0.51700	33.880000	26.0700	50.160

ارزیابی وتست مدل ریاضی (*MLR*) بدست آمده با استفاده از مجموعه ارزیابی و شبیه سازی

تست مدل با گروه ارزیابی

جدول۵- خطاهای نسبی در مدل MLR و ضریب همبستگی مدل برای گروه ارزیابی

R	RMSE	MPE	RSE(%)
0.409238722	36.278459680	29.531551020	50.878327



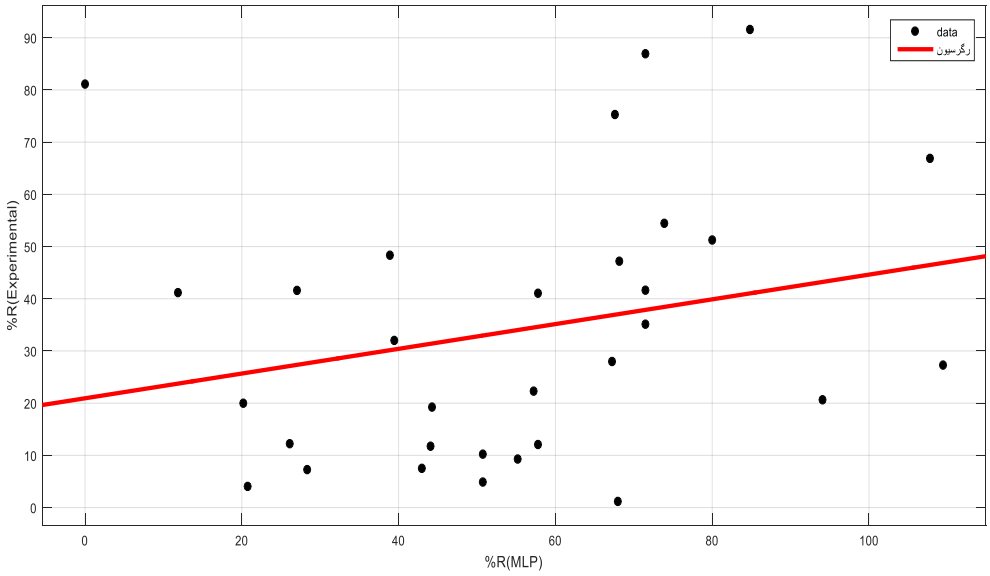
شکل ۴: همبستگی نتایج بدست آمده از مدل خطی با مجموعه ارزیابی

همانطور که در جدول و نمودار شکل 4- و جدول-۵ مشاهده می‌شود همبستگی بین نتایج مدل ریاضی بدست آمده و داده-های واقعی خیلی کم است و از مقدار خطای نسبی بالایی برخوردار است و بنابراین این روش مدل‌سازی مناسب نیست.

تست مدل با مجموعه شبیه سازی

جدول۶- خطاهای نسبی بین پاسخهای واقعی (مجموعه شبیه سازی) و پاسخهای مدل MLR

R	RMSE	MPE	RSE(%)
0.47560175800000	35.3936192000000	27.6512650000000	51.9843782000000



شکل۵: همبستگی نتایج بدست آمده از مدل خطی با مجموعه شبیه سازی

همانطور که نتایج با توجه به شکل 5- و جدول-۶ نشان داده شده است از مدلهای خطی MLR و جواب های واقعی همبستگی کمی وجود داشته و خطای زیادی در حد ۵۰ درصد را نشان می‌دهد.

مدلسازی با روش رگرسیون درجه دوم چندگانه *(MQR)*

نتایج مدلسازی MQR در جدول -۷ مشخص شده است با توجه اینکه سطح اطمینان ۹۵٪ انتخاب شده است و در مواردی P_Value ها بزرگتر از ۰/۰۵ می‌باشد، لذا این ضرایب معنی‌دار نبوده و بایستی از معادله حذف شوند. مقدار P_Value ضرایب X2 ، X4، X7)بزرگتر از۰/۰۵ می‌باشد که حذف خواهند شد و مجدداً آنالیز رگرسیونی بعد از حذف متغیر های فوق، انجام شده که در فصل سه آمده است.

جدول۷- مقادیر معادله در شبیه سازی MQR

	ضریب	P_Value
Constant	21.1948836037512	0.0166811583918412
X_1	0.603211572783927	0.277794803255756
X_2	-9.96751347310817	0.232258885373378
X_3	-11.9659820182551	0.0239319640365103
X_4	-4.01666953625178	0.0577396245836193
X_5	-7.14337924228782	0.0732196372334501
X_6	111.109269479894	0.000703223224556291
X_7	-3.89016559903571	0.412357553497785
R	۵۷/۱	

معادله درجه دوم بدست آمده در اولین اجرا به شرح زیر است:

$$\%R = 21 + 0.6\ X_1 - 10\ X_2 - 12\ X_3 - 4.0\ X_4 - 7.15\ X_5 + 111\ X_6 - 38\ X_7$$

جدول۸- آنالیز رگرسیونی مجدد با سطح اطمینان ۹۵٪ پس از حذف ضرائب عدم معنی دار

	ضریب	P_Value
Constant	57.876948935894	0.0694859308490859
X_1	-8.94975948594748	0.0304985948594584
X_3	19.7469264019855	0.0647581757826370
X_5	11.0729198795853	0
X_6	0.00746674374515697	0.00162320516199065

معادله درجه دوم بدست آمده بعد از حذف ضرایب فاقد معنی با توجه به جدول-۸ به شرح زیر است:

$$\%R = 57.87+Constant\ -8.9\ X_1+19.7\ X_3 + 11.0741\ X_5 + 0.0074\ X_6$$

محاسبات خطا برای مدل *MQR*

برای آگاهی از اینکه مدل ارائه شده در معادله به دست‌آمده چه اندازه به واقعیت نزدیکتر است، خطای بین پاسخ حاصل از معادله رگرسیونی (مقادیر پیش‌بینی) و داده‌های تجربی مجموعه آموزشی (مقادیر واقعی) به سه صورت RSE، MPE و RMSEمحاسبه می شوند که در جدول-۹ آمده است

جدول ۹- خطاهای نسبی در مدل MQR و ضریب همبستگی مدل

RSE(%)	MPE	RMSE	R
56.06570491400	30.8366907280000	37.8490515320000	18.9980188۰0

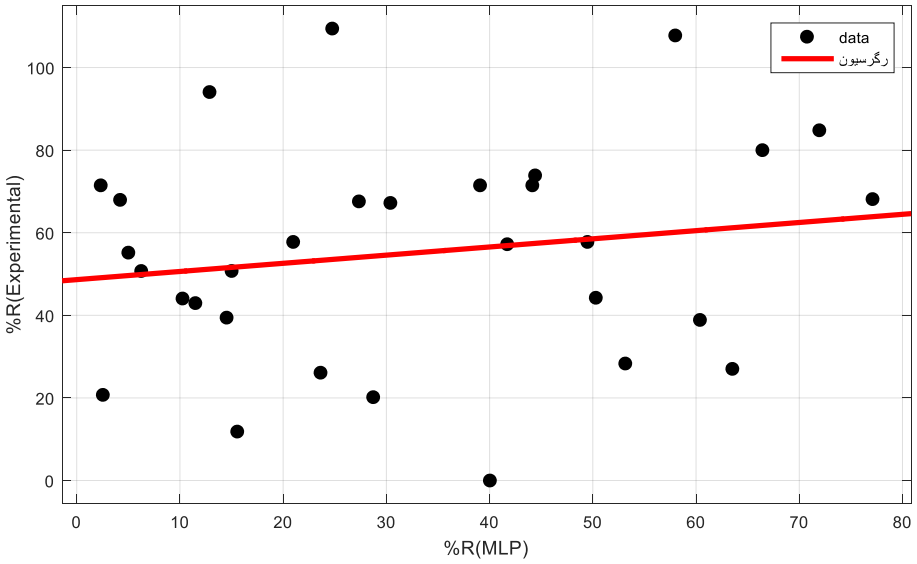
همانگونه که مشاهده می‌شود مقادیر خطاها در شبیه سازی با روش رگرسیون خطی چندگانه (MQR) بسیار بالا و مقدار RSE در حدود ۵۶,۰۶٪ می‌باشد. مقادیر خطای RMSE و MPE نیز این اهم را تایید می‌کنندو جدول۱۰ نشان داده شده است. بنابراین می‌توان ادعا کرد که هیچ رابطه خطی بین میزان موفقیت دانش آموزان و مقادیر متغیرهای مستقل آنها وجود ندارد.

مدل ریاضی درجه دوم چندگانه (MQR) بدست آمده با استفاده از مجموعه ارزیابی و شبیه سازی

تست مدل با مجموعه ارزیابی

جدول۱۰- خطاهای نسبی بین پاسخهای واقعی (ست ارزیابی) و پاسخهای مدل MQR

RSE(%)	MPE	RMSE	R
58.974617992	33.2146995180	41.3331109220	0.10839295880



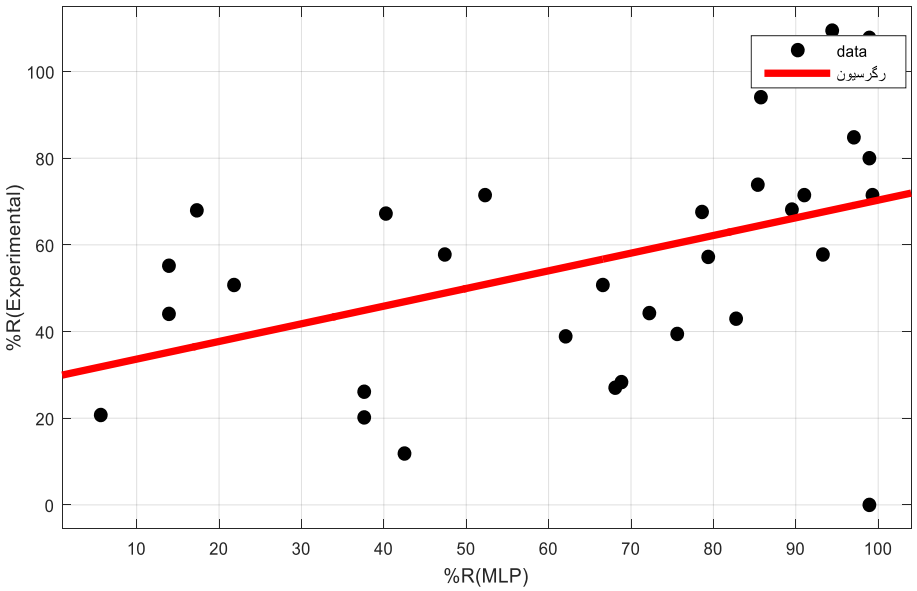
شکل-۶: همبستگی نتایج بدست آمده از مدل خطی با مجموعه ارزیابی

همانطور که مشاهده شد مشخص است با توجه به شکل ۶ و جدول-۱۱ .بین داده های واقعی و تخمین خطی فاصله زیادی وجود دارد و هیچ رابطه ای وجود ندارد که بتوان با فرض دقیق موفقیت دانش آموزان را با آن معادله تخمین و تایید نمود.

تست مدل *MQR* با مجموعه شبیه سازی

جدول-۱۱- خطاهای نسبی و ضریب همبستگی در مدل MQR برای مجموعه شبیه سازی

R	RMSE	MPE	RSE(%)
0.0276512650	44.2088424820000	37.3734497740000	64.869867690



شکل-۷: همبستگی نتایج بدست آمده از مدل خطی با مجموعه شبیه سازی

همانطور که در شکل6-و جدول-۱۱ مشاهده می‌شود نتایج بدست آمده از مدلهای خطی و پاسخ‌های واقعی همبستگی کمی داشته و خطای زیادی در حد ۶۰٪ نشان می دهد. بنابراین می‌توان نتیجه گرفت وجود یک رابطه خطی میزان موفقیت دانش آموزان و متغیر های مستقل را نمی‌تواند تایید کند.

ورودی ها و خروجی های شبکه عصبی بردار پشتیبان

ورودی های شبکه عصبی ۸ توصیف‌گر باقی‌مانده از حذف توصیف‌گرها توسط برنامه Mini-Tab به روش رگرسیون چندگانه مرحله‌ای هستند. در طراحی شبکه مورد نظر مقادیر اولیه وزن ها به طور تصادفی در محدوده [۱و۱-] بوده و قبل از عمل آموزش مقادیر ورودی بین [۱- تا۱] وخروجی در محدوده[۰/۲ تا ۰/۸] تنظیم می‌شود. چراکه تابع انتقال انتخاب شده تابع سیگموید است. و بعد برای ورود دیتاها به شبکه عصبی متلب ابتدا نرمالایز کرده در بازه ۰ تا ۱ و بعد دیتا های نرمال شده را به شبکه عصبی تعریف میکنیم. که شبکه استفاده شده. تعداد متغیرها در لایه ورودی با تعداد توصیف‌گرهای انتخاب شده توسط نرم‌افزار Mini-Tab برابر است. تعداد توصیف‌گرهای انتخاب شده توسط Mini-Tab (۷) مورد بوده، در ضمن تعداد خروجی نیز یک مورد می‌باشد، در نتیجه شبکه های مورد نظر برای مدل‌سازی بدین ترتیب نشان داده می‌شود، (۱-X-۷)، که X تعداد نورون‌ها در لایه‌ی پنهان شبکه‌ی عصبی می‌باشد.

بررسی مقادیر خطا روی پاسخ‌های حاصل از داده‌های آموزشی شبکه عصبی

خطای بین مقادیر پیشگویی شده و مقادیر واقعی برای داده‌های آموزشی و ارزیابی در شبکه عصبی طراحی شده، به سه صورت RMSE، MPE و RSE محاسبه شدند ملاحظه می شود خطای بسیار کمتری نسبت به روشهای رگرسیونی وجود دارد. به طوریکه درصد خطا برای داده‌های آموزشی و ارزیابی به ترتیب برابر ۱۱٪ و ۱۹٪ بوده است. مقادیر پایین سایر پارامترهای خطا و همچنین مقادیر بسیار خوب ضرایب همبستگی نیز مؤید این ادعا می باشد که در جدول-۱۲ مشخص شده است.

جدول۱۲- خطاهای نسبی و ضریب همبستگی برای دو مجموعه آموزشی و ارزیابی در شبکه عصبی

	RSE(%)	MPE	RMSE	R
داده‌های آموزش	۷,۹۲۳۰۶۵۶۸۰۷۲	۴,۹۹۲۰۵۳۰۵۹	۱۱,۷۳۶۵۸۳۲۶۹۴۳	۰,۹۴۳۵۹۸۶۶۵۷۳
داده‌های ارزیابی	۱۳/۴۰۵۴۵۹۶۸	۸/۴۷۵۹۹۸۳۴۵	۱۹/۱۵۵۴۴۲۹۴	۰/۹۱۲۳۳۲۷۳

ارزیابی نهایی مدلهای طراحی شده و مقایسه مدلها با هم

تعدادی از داده‌ها برای تست نهایی مدلها (دسته شبیه‌سازی) به طور تصادفی جدا شده است و متغیرهای مستقل این داده‌ها به عنوان ورودی به مدلهای به دست آمده وارد شده و پاسخ های حاصل با متغیرهای وابسته واقعی مقایسه و نتیجه آن ارزیابی می‌گردد.

ارزیابی شبکه عصبی مصنوعی توسط داده های شبیه سازی شده

داده‌ها به صورت تصادفی به سه مجموعه تقسیم‌بندی شده است:مجموعه آموزشی، مجموعه ارزیابی و مجموعه شبیه‌سازی. از مجموعه آموزشی برای ایجاد مدل، از مجموعه ارزیابی برای ارزیابی مدل ساخته شده و از مجموعه شبیه‌سازی برای عمومی بودن مدل استفاده می‌شود. برای این منظور داده‌های مجموعه شبیه‌سازی و داده‌های حاصل از مدلسازی شبکه عصبی به نرم-افزار Matlab منتقل می‌شوند و مقادیر خطا محاسبه می شود.

جدول ۱۳-خطاهای نسبی و ضریب همبستگی در مدل SVM

	RMSE	MPE	RSE(%)	R
داده‌های شبیه‌سازی	۱۸/۰۰۳۸۲۳۳۰۰	۱۰/۹۹۸۸۶۵۸۶۷۳۴	۱۴/۹۸۸۹۷۹۸۳۹۴۸۹۳۸۴	۰/۸۸۹۸۳۸۴۷۹

با توجه به مقادیر ثبت شده در جدول۱۳ ملاحظه می‌شود مقادیر خطا برای مدلهای SVM بسیار کم می باشد به طوریکه مقدار درصد خطای استاندارد نسبی (RSE) برابر ۱۴٪ می‌باشد. نتایج نشان دهنده همبستگی مطلوب مدل شبکه عصبی و داده های واقعی است. بنابراین ارتباط بین داده‌ها (متغیر وابسته و متغیرهای مستقل) یک رابطه غیرخطی است.

در صورتی که طبق جدول۱۴ ، در مدلهای خطی MLR و MQR مقدار خطای RSE بالا و قابل بحث نمی باشد. مقدار RSE در مدل MLR برابر ۴۴٪ و در مدل MQR برابر ۳۵٪ به دست آمده است. سایر مقادیر خطای محاسبه شده یعنی MPE و MSE در جدول تحلیل و نتیجه فوق را کاملاً تائید می کنند.

جدول ۱۴- خطاهای نسبی و ضرایب همبستگی در داده‌های شبیه‌سازی

مدل	RMSE	MPE	RSE(%)	R
MLR	51.98437820	27.651265000	35.39361920000	0.3756017580
MQR	64.869867690	37.3734497740000	44.2088424820000	0.276512650
SVM	۱۸/۰۰۳۸۲۳۳۰۰	۱۰/۹۹۸۸۶۵۸۶۷۳۴	۱۴/۹۸۸۹۷۹۸۳۹۴۸۹۳۸۴	۰/۸۸۹۸۳۸۴۷۹
GA	۱۳,۴۵۵۴۰۰۰۳۴	۴,۹۸۴۹۵۴۷۸۳۴۷۸۷	۹,۸۴۷۸۳۶۴۷۲۴۷۲۸۴۶۶	۰,۹۴۵۷۳۷۵۸۷۳

همانطور که در نمودار مشاهده می‌شود همبستگی نتایج حاصل از مدل‌سازی شبکه عصبی و مقادیر واقعی نشان دهنده وجود رابطه غیرخطی بین موفقیت دانش آموزان و متغیرهای مستقل است.

نتیجه گیری

طبق نتایج بدست آمده از مدل بهینه سازی شده و پاسخ‌های واقعی همبستگی زیادی وجود دارد و خطای حدود ۹ درصد را نشان می دهد. بنابراین می توان از بهینه سازی برای بررسی پیش بینی میزان موفقیت دانش آموزان استفاده نمود. همچنین خطای بین مقادیر پیشگویی شده و مقادیر واقعی برای داده‌های آموزشی و ارزیابی در شبکه عصبی طراحی شده، به سه صورت RMSE، MPE و RSE محاسبه شده است و ملاحظه می شود خطای بسیار کمتری نسبت به روشهای رگرسیونی وجود دارد. به طوریکه خطا برای داده‌های آموزشی بیشتر از ارزیابی بوده است. مقادیر پایین سایر پارامترهای خطا و همچنین مقادیر بسیار خوب ضرایب همبستگی نیز مؤید این ادعا می باشد.

منابع :

۱. استفاده از قوانین انجمنی داده کاوی ایران. *تهران، ششمین کنفرانس داده کاوی*

۲. اعلایی، غلامحسین (۱۳۷۱).ارزشیابی ملاک‌های هدایت تحصیلی نظام جدید متوسطه. *پایان نامه کارشناسی ارشد، دانشگاه تربیت مدرس، تهران.*

۳. آموزگار ،حمید رضا (۱۳۸۹) . مقایسه ی دقت تکنیکهای کلاسه بندی در داده کاوی با استفاده از دیتاست پزشکی. *پایان نامه کارشناسی ، دانشگاه پیام نور، مشهد .*

۴. بهادر ، حمید، علیایی . صغری، باقری . افسانه(۱۳۹۴) .ارایه مدلی برای شناسایی عوامل موثر در هدایت تحصیلی دانش آموزان با استفاده ازدرخت تصمیم و شبکه عصبی . *تهران، دومین کنگره سراسری فناوریهای نوین ایران با هدف دستیابی به توسعه پایدار.*

۵. حاتملو، عبدالرضا، هاشمی‌نژاد، سیدجواد (۱۳۸۷). تحلیل رفتار آموزشی دانشجویان با استفاده از تکنیک‌های داده‌کاوی. *سومین کنفرانس داده‌کاوی ایران، تهران .*

۶. خیرآبادی، اعظم، مینایی بیدگل، بهروز (۱۳۸۹). بررسی و خوشه‌بندی نتایج ارزشیابی اساتید دانشگاه با استفاده از روش-های داده‌کاوی. *چهارمین کنفرانس داده‌کاوی ایران، تهران .*

۷. شکورنیاز، جمال شهرابی، ونوس (۱۳۸۶). مروری بر مفاهیم، وظایف و فرآیند داده کاوی سازمانی.*اولین کنفرانس داده کاوی*

ایران، دانشگاه صنعتی امیرکبیر، تهران.

۸. صافی، احمد (۱۳۹۲). اصول و فنون راهنمایی در دوره‌های تحصیلی. *انتشارات تهران، تهران.*

9. Abel, Dr V Rama Bai, G Krishna Chaitanya(2017). Academic Guidance in the Field Selection by Ranking Method Two and Up. *International Journal of Innovative Research in Computer and Communication Engineering (An ISO 3297: 2007 Certified Organization) Vol. 2, Issue 1.*
10. Abhijit Raorane(2011). DATA MINING TECHNIQUES: A SOURCE FOR CONSUMER BEHAVIOR ANALYSIS. *International Journal on Computer Science and Engineering ,IJCSE 1.*
11. Abner Tanna, Casey(2016). Academic orientation in field selection by linear regression. *International Journal of Computer Trends and Technology (IJCTT) – Volume 4 Issue 7.*
12. Baker, R. S. J. D. (2010). Data mining for education. *International encyclopedia of education*, *7*(3), 112-118.
13. Baker, R.. S. & Yacef, K. (2009). The state of educational data mining in 2009: A review and future visions. *JEDM-Journal of Educational Data Mining, 1(1), 3-17.*
14. D.A. Adeniyi, Z. Wei, Y. Yongquan(2016). Automated web usage data mining and recommendation system using K-Nearest Neighbor (KNN) classification method. *Applied Computing and Informatics 12, 90–108*
15. Damon Wook, Hani Yahaya(2018). Educational guidance in choosing a job with a neural network approach. *Second International Conference on Computer and Electrical Engineering.*
16. Damon, Romero Morales, Ventura Soto(2018). Academic Guidance in Field Selection by Fuzzy Neural Systems.*IEEE JOURNAL OF LATIN-AMERICAN LEARNING TECHNOLOGIES.*
17. Dr. Mamta Madan(2015). A Review on: Data Mining for Telecom Customer Churn Management. *International Journal of Advanced Research in Computer Science and Software Engineering, Volume 5, Issue 9*
18. Han, J., Pei, J., & Kamber, M. (2011). Data mining: concepts and techniques. *Elsevier.*
19. K. Mizue, and O. Toshio(2011) .Neural network navigation support-knowledge-navigation in hyperspace: The sub-symbolic approach. *Journal of Educational Multimedia and Hypermedia, Vol. 10, No. 1, pp. 85–103.*