نشـــریه بین المللـــی مهنـدسی صنـایع و مدیـــریت تولیـد زمستـــان ۱۳۸۹- صفحه ۲۰۲-۱۹۲

شمـــاره ۴، جلـــد ۲۱، زمستـــان ۱۳۸۹

نشـــریه بین المللـــی مهنـــدسی صنایـــع و مدیـــریت تولیـــد

نشریه بین المللی مهندسی صنایع و مدیریت تولید دانشگاه علم و صنعت ایران

IJIE

ISSN: 2008-4870

http://IJIEPM.iust.ac.ir/

# توسعه یک سیستم تشخیص مشتری تلفیقی مبتنی بر درخت رگرسیونی هرس شده و شبکه عصبی بهبودیافته

**علیرضا سروش و اردشیر بحرینی‌نژاد***

**کلمات کلیدی**

انتخاب ویژگی،
پیش‌بینی،
درخت رگرسیونی هرس شده،
سیستم تشخیص مشتری،
شبکه عصبی بهبودیافته

**چکیده:**

در دنیای رقابتی امروزی، شیوه‌های جذب مشتری یکی از با اهمیت‌ترین حوزه‌های کاربردی داده‌کاوی بوده و پرواضح است که یکی از مهمترین ابعاد آن پیش‌بینی رفتار خرید مشتری است. زیرا، پیش‌بینی خوب می‌تواند به توسعه استراتژیهای بازاریابی دقیقتر و صرف کاراتر منابع کمک نماید. ایجاد یک سیستم تشخیص مشتری(CRS) به دلیل وجود تعداد زیادی ویژگی در دسترس طراح کاری بسیار مشکل است. بعلاوه، نیاز شدیدی به ایجاد یک CRS وجود دارد که همزمان پیچیدگی کم و قابلیت پیش‌بینی خوبی را داشته باشد. از اینرو، مقصود این مقاله، توسعه یک CRS تلفیقی (HCRS) است که از نظر محاسباتی کارا و اثربخش است. نوآوری مدل HCRS، هم طراحی و هم پیاده‌سازی سیستم مذکور با ایجاد یک درخت رگرسیونی هرس شده (PRT) و طراحی یک شبکه عصبی پیشخوراند بهبودیافته (IFFNN) جهت افزایش سرعت، دقت و کاهش پیچیدگی را شامل می‌شود. از آنجاییکه، شناسایی مشتریان یکی از دغدغه‌های صنعت بیمه است، از داده‌های یک شرکت بیمه هلندی استفاده شده است. نتایج نشان داد که HCRS تنها ۷٪ از ویژگی‌ها را در حالت بهینه انتخاب می‌کند که به میزان قابل‌توجهی هزینه محاسبات را کاهش می‌دهد. به‌علاوه، نتایج نشان داد که PRT نسبت به روش منحنی مشخصه عملیاتی دریافت‌کننده کاراتر بوده و IFFNN نسبت به FFNN و PRT پیش‌بینی‌های دقیقتری را ارائه می‌کند.

## ۱. مقدمه

مساله انتخاب ویژگیها، یک مساله مستقل در نظریه تشخیص الگو بوده و تاکنون حل نشده است[۱]. فرایند انتخاب ویژگی‌ها بعنوان مساله‌ای از بهینه‌سازی ترکیبی کلی در یادگیری ماشین شناخته می‌شود که تعداد ویژگیها را کاهش داده و داده‌های غیرمرتبط و زائد را حذف می‌کند. انتخاب ویژگیها در حوزه‌های مختلفی که هزاران ویژگی در دسترس است، بسیار مورد توجه قرار گرفته است. هدف اصلی انتخاب ویژگی، شناسایی زیرمجموعه‌ای از ویژگیها است که تاثیر بیشتری بر روی یک متغیر پاسخ معلوم دارند[۱]. کشف زیرمجموعه بهینه‌ای از ویژگی‌ها معمولاً مشکل بوده و نشان داده شده است که بسیاری از مسائل مرتبط NP-hard شناخته می‌شوند[۲]. پیاده‌سازی مناسب انتخاب ویژگی‌ها نه تنها اطلاعات مهمی را برای پیشگویی و طبقه‌بندی فراهم می‌کند، بلکه تلاشهای موردنیاز برای تحلیل داده‌های چندبعدی را کاهش می‌دهد.

از طرف دیگر، انتخاب ویژگیها موفقیت‌های بسیاری را در کاربردهای دنیای واقعی داشته است، زیرا غالباً می‌تواند به میزان قابل توجهی ابعاد را برای به کارگیری الگوریتهای داده‌کاوی جهت کار بر روی داده‌ها با ابعاد بزرگ کاهش دهد. از اینرو، در سالهای اخیر، مدیریت ارتباط با مشتری[۲] (CRM) یکی از زمینه‌های تحقیقاتی بوده است که انتخاب ویژگی‌ها به کار گرفته شده است[۳]. امروزه، توجه به

**تاریخ وصول: ۸۹/۱/۲۳**

**تاریخ تصویب: ۸۹/۷/۱۰**

**علیرضا سروش**، دانشجوی دکتری مهندسی صنایع، فنی و مهندسی، تربیت مدرس a.soroush@modares.ac.ir

***نویسنده مسئول مقاله: دکتر اردشیر بحرینی‌نژاد**، استادیار مهندسی صنایع، فنی و مهندسی، تربیت مدرس. bahreininejad@modares.ac.ir

[2]- *Customer Relationship Management*

CRM بواسطه افزایش میزان رقابت میان شرکتها حیاتی‌تر شده است. تغییرات سریع در احتیاجات مشتریان مجزا از یکدیگر است. CRM، وسیله‌ی عمده‌ای است که کسب و کارها می‌توانند با این چالشها مواجه شوند و قادر است به آنها کمک نماید که تقاضاهای مختلف مشتریان را شناسایی نموده و از این طریق مزیت رقابتی بدست آورد[۴]. به این دلیل، مطالعه تاثیرات انتخاب ویژگی برای آماده‌سازی یک سیستم CRM دارای اهمیت است[۵].

به صورت سنتی، انتخاب بهینه مشتریان هدف یکی از مهمترین عوامل برای یک سیستم CRM در نظر گرفته شده است. از اینرو، مدلهای بسیاری برای شناسایی بسیاری از مشتریان ممکن که یک محصول مشخص را خواهند خرید یا کسانی که روابط بیشتری را با شرکت ادامه خواهند داد، ارائه شده است. از اینرو، شرکت تلاش می‌کند که مدلهای پیشگویی را به صورت دقیقی توسعه دهد تا بتواند شناسایی کند که کدام مشتریان با احتمال بیشتری خرید می‌کنند. این مدلها بعنوان یک سیستم تشخیص مشتری توصیف می‌شوند[۳]. تعداد ویژگیهای در دسترس طراح یک سیستم تشخیص مشتری معمولاً بسیار زیاد است. این تعداد می‌تواند دهها یا حتی هزاران ویژگی باشد. بیش از یک دلیل برای کاهش تعداد ویژگی‌ها به یک حداقل کافی وجود دارد. پیچیدگی محاسباتی یک دلیل قابل مشاهده است. دلیل دیگر آن است که اگرچه دو ویژگی ممکن است به صورت جداگانه اطلاعات تشخیصی خوبی را ارائه کنند، اما اگر آنها با یکدیگر در یک بردار ورودی ترکیب شوند، به دلیل همبستگی بالا عایدی کمی حاصل می‌شود و پیچیدگی بدون هیچ عایدی افزایش می‌یابد. دلیل اصلی دیگر مربوط به خصیصه‌های تعمیمی موردنیاز تشخیص دهنده برای پیشگویی مشتریان آتی می‌شود. نسبت بالاتر تعداد الگوهای آموزشی به تعداد پارامترهای تشخیص دهنده مشتری به تعمیم بهتر ویژگی‌های تشخیص دهنده منجر می‌شود.

تعداد ویژگی‌های زیاد مستقیماً موجب تعداد پارامترهای زیاد تشخیص دهنده می‌شود. از اینرو، اگر تعداد الگوهای آموزشی محدود باشد، بهتر است برای طراحی تشخیص دهنده با قابلیت تعمیم خوب تعداد ویژگیها نیز کم باشد. یک مرحله مهم در طراحی یک سیستم تشخیص مشتری مرحله ارزیابی عملکرد است که در آن احتمال خطای پیشگویی سیستم طراحی شده برآورد می‌شود. باید تاکید شود که این مرحله بسیار حیاتی است. اگر ویژگی‌هایی با قدرت کم تشخیص مشتری انتخاب شوند، متعاقباً سیستم عملکرد ضعیفی خواهد داشت. بنابراین، سناریوهای مختلفی را می‌توان اتخاذ نمود. یک روش بررسی ویژگی‌ها به صورت جداگانه و حذف آنهایی است که قابلیت تمایز کمتری دارند. راه دیگر، بررسی آنها به صورت ترکیبی است. گاهی اوقات کاربرد یک تبدیل خطی یا غیرخطی برای یک بردار ویژگی ممکن است به یک بردار جدید با ویژگی‌های تشخیصی بهتر منجر شود.

در سالهای اخیر، چندین الگوریتم برای انتخاب ویژگی‌ها به صورت مناسب بر روی CRM به کار گرفته شده‌اند و تعدادی مطالعه مقایسه‌ای نیز انجام شده است. بعلاوه، از آنجاییکه هر ویژگی مورد استفاده بعنوان بخشی از یک رویه پیشگویی می‌تواند هزینه و زمان اجرای یک سیستم تشخیص مشتری را افزایش دهد، انگیزه قوی در جامعه تشخیص مشتری برای طراحی و پیاده‌سازی سیستم‌هایی با مجموعه ویژگی‌های کم وجود دارد. همزمان یک نیاز متضاد برای لحاظ مجموعه کافی از ویژگی‌ها جهت دستیابی به نرخهای تشخیص بالا تحت شرایط مشکل وجود دارد. این موضوع منجر به توسعه تکنیکهای متنوعی در جامعه تشخیص مشتری برای کشف یک زیرمجموعه بهینه از ویژگی‌ها از میان مجموعه‌ای بزرگتر از ویژگی‌های ممکن شده است. بطوریکه، در مقاله [۴] از یک رویه آماری ساده یعنی انتخاب ویژگی رو به جلو مبتنی بر امتیاز مربع خی دو استفاده شده است که با مجموعه خالی از متغیرها شروع نموده و به تدریج متغیرهایی اضافه می‌شود که بیشترین بهبود عملکرد را دارند. در مقاله [۳]، انتخاب ویژگی با اجرای یک الگوریتم استقرایی بر روی مجموعه داده‌ها انجام می‌شود. همچنین، رویکرد درخت تصمیم برای انتخاب ویژگی پیشنهاد شده است. در مقاله[۶]، یک رویه منحنی مشخصه عملیاتی دریافت کننده[1] (ROC) برای انتخاب ویژگی‌ها پیشنهاد شده است یا بطور خاصتر آنها بیان شده است که منطقه زیر منحنی را می‌توان برای انتخاب ویژگی استفاده نمود. در مقاله [۷]، رویکرد جدیدی برای هدفگذاری مشتریان در بازاریابی پایگاه داده ارائه شده است. یک الگوریتم ژنتیک (GA) استاندارد برای جستجو در میان ترکیبات ممکن از ویژگی‌ها استفاده می‌شود. ویژگی‌های ورودی انتخابی توسط GA جهت آموزش شبکه‌های عصبی استفاده می‌شود. همچنین، در مقاله[۸]، یک روش ایجاد تجمعی مبتنی بر GA بر مبنای مکانیزم انتخاب زیرمجموعه‌ای از ویژگی‌های پوشه‌ای[2] پیشنهاد شده است که مبتنی بر اطلاعات گذشته خرید، مقداری را که هر مشتری خرج خواهد کرد پیش‌بینی نماید. از نقاط ضعف روش GA زمان‌بر بودن آن است. در مقاله[۹]، یک سیستم استدلال مبتنی بر مورد بعلاوه تکنیک کاهش دو بعدی برای رضایت مشتریان پیشنهاد شده است که رفتار خرید مشتریان را برای یک محصول خاص با استفاده از مشخصه‌های آماری آنها پیشگویی می‌کند. در مقاله[۱۰]، یک مدل رگرسیونی خطی چندگانه برای جلوگیری از تطبیق بیش از حد بعنوان یک رویه انتخاب ویژگی به کار گرفته شده است و وفاداری رفتاری مشتری با استفاده از پایگاه داده تراکنشها پیشگویی شده است. نقطه ضعف روشهای مورد استفاده در مقالات [9,10]، عدم لحاظ روابط غیرخطی در رفتار مشتریان است. در مقاله [۱۱]، الگوریتم افراز تودرتو[3] و روش ذوب شبیه‌سازی شده[1] جهت انتخاب

[1]- *Receiver Operating Characteristic*
[2]- *Wrapper*
[3]- *Nested Partition*

ویژگی‌ها برای تشخیص مشتری ترکیب شده است. این روش بسیار کند بوده و لزوماً به راه‌حل بهینه نرسیده و ممکن است در یک حداقل موضعی متوقف شود. در مقاله [۱۲]، نظریه مجموعه زبر[۲] برای انتخاب ویژگی‌ها در CRM به کار برده شده است و از داده‌های گذشته خرید یک سیستم بازی تصویری برای پیشگویی رفتار خرید مشتری استفاده شده است. این تکنیک قابلیت سر و کار داشتن با داده‌های زیاد را ندارد. همچنین، در مقاله [۱۳]، یک مدل مرجع سلسله مراتبی برای ماشین بردار پشتیبان[۳] (SVM) مبتنی بر طبقه‌بندی در کاربردهای CRM دنیای واقعی پیشنهاد شده است که در آن حذف ویژگی بازگشتی بعنوان یک رویه حذف رو به عقب برای رتبه‌بندی ویژگی‌ها مبتنی بر SVM پیشنهاد شده است. معایب این روش شامل محدودیت در انتخاب کرنل، سرعت و اندازه کم است.

هدف اصلی این مقاله ایجاد یک سیستم تشخیص مشتری[۴] (CRS) بهینه برای بهبود کارایی برنامه‌های CRM است. با مرور ادبیات، ما کشف کردیم که مقالات در دسترس در خصوص CRS معمولاً رویکردی با یک روش را به کار می‌گیرند، بطوریکه تنها یک تکنیک همچون یک درخت تصمیم، شبکه عصبی یا سایر روشها در حوزه هوش محاسباتی به کار برده می‌شوند و از طریق به کارگیری چنین رویکردی، معمولاً دستیابی به هر دو نیاز پیچیدگی کم و عملکرد پیش‌بینی خوب مشکل است.

بر این اساس، مقصود این مقاله توسعه یک سیستم تشخیص مشتری تلفیقی[۵] (HCRS) است که از نظر محاسباتی کارا و اثربخش بوده و هر دو نیاز را تامین می‌کند. رویکرد انتخابی بدین صورت است: ابتدا، یک درخت رگرسیون هرس شده[۶] (PRT) برای انتخاب ویژگی (مبتنی بر حداقل‌سازی هزینه میانگین مربعات خطا) طراحی می‌شود. سپس، یک شبکه عصبی پیشخوراند بهبودیافته[۷] (IFFNN) برای شناسایی مشتریان آتی ایجاد می‌شود. نوآوری مدل HCRS شامل طراحی و پیاده‌سازی سیستم تلفیقی مذکور برای تشخیص مشتریان براساس یک PRT است که با انتخاب زیرمجموعه‌ای بهینه از ویژگی‌ها پیچیدگی را کاهش و بالتبع سرعت محاسبات پیش‌بینی را افزایش می‌دهد و ایجاد یک IFFNN که عملاً نتایج پیش‌بینی بهتری را فراهم می‌کند، می‌شود. بدین ترتیب، نقص به کارگیری تنها یک تکنیک جهت شناسایی مشتریان پوشش داده می‌شود. نتایج پیش‌بینی با و بودن انتخاب ویژگی تحلیل می‌شود و تکنیک ROC بمنظور ارزیابی روش انتخاب ویژگی و دو تکنیک دیگر بنامهای درخت رگرسیونی و شبکه عصبی پیشخوراند بمنظور سنجش نتایج پیش‌بینی به کار گرفته شده و مقایسه می‌شوند. در ادامه مدل HCRS طراحی شده بر روی داده‌های یک شرکت بیمه هلندی پیاده‌سازی شده و قابلیت مدل تحلیل می‌شود. بقیه مقاله به صورت زیر سازماندهی می‌شود. در بخش دوم، مورد کاوی تحقیق توصیف می‌شود. در بخش سوم، انتخاب ویژگی از طریق PRT طراحی شده انجام می‌شود. در بخش چهارم، پیش‌بینی با توسعه شبکه عصبی پیشخوراند بهبودیافته صورت گرفته و نتایج مقایسه می‌شوند. نهایتاً، در بخش پنجم، نتیجه‌گیری‌ها ارائه می‌شوند.

## ۲. توصیف داده‌ها: موردکاوی یک شرکت بیمه

یکی از پرکاربردترین زمینه‌هایی که می‌توان برای شناسایی مشتریان از آن استفاده نمود، صنعت بیمه است. بویژه اینکه، شناسایی ویژگی‌های مشتریان یک محصول شرکت بیمه از طریق رفتاری که در قبال سایر محصولات بیمه‌ای شرکت از خود نشان داده‌اند، می‌تواند جالب توجه باشد. از اینرو، در این تحقیق، مجموعه داده‌ها از طریق یک شرکت بیمه هلندی بنام کویل چلنج تهیه شده و مربوط به یک کسب و کار در دنیای واقعی می‌شود. این شرکت بیمه قصد دارد مشتریان بالقوه برای یک محصول معین را شناسایی نماید. داده‌ها مربوط به ۹۸۲۲ مشتری این شرکت بیمه می‌شود که تعدادی از آنها اقدام به خرید بیمه‌نامه محصولی بنام خانه متحرک (نوعی خودرو) نموده‌اند. این داده‌ها امکان ارزیابی قابلیت مدل HCRS در پیش‌بینی مشتریان احتمالی را فراهم می‌سازد.

در این مقاله، دو مجموعه داده جداگانه به کار برده می‌شود: یک مجموعه آموزشی با ۵۸۲۲ مشتری و یک مجموعه ارزیابی با ۴۰۰۰ مشتری. هر رکورد متشکل از ۸۶ ویژگی شامل داده‌های آمارگیری اجتماعی (۴۳ ویژگی) و مالکیت محصول (۴۲ ویژگی) می‌شود. ویژگی آخر یعنی «تعداد بیمه‌نامه خانه متحرک»، متغیر هدف است. داده‌های آموزشی جهت آماده‌سازی مدل HCRS استفاده می‌شود و نرخ هدف مورد انتظار براساس مجموعه ارزیابی برآورد می‌شود. از ۵۸۲۲ مشتری احتمالی در مجموعه داده آموزشی، ۳۴۸ نفر بیمه نامه خانه متحرک را خریده‌اند که از آن نرخ هدف ٪۵،۹۷=۳۴۸/۵۸۲۲ حاصل می‌شود.

از دیدگاه مدیریت، این نرخ هدف در صورتی بدست خواهد آمد که متقاضیان به صورت تصادفی بعنوان مصرف‌کنندگان در پایگاه داده شرکت ثبت شوند. مجموعه داده ارزیابی جهت اعتبارسنجی مدل HCRS استفاده می‌شود. همچنین، از ۴۰۰۰ مشتری احتمالی در مجموعه داده ارزیابی، ۲۳۸ نفر بیمه‌نامه خانه متحرک را خریده‌اند که نرخ هدف ٪۵،۹۵=۲۳۸/۴۰۰۰ حاصل می‌شود. همانطور که مشاهده می‌شود نرخ هدف تقریباً در هر دو مجموعه برابر است. مدل HCRS برای شناسایی ٪۲۰ اول از مشتریانی در مجموعه داده ارزیابی طراحی می‌شود که انتظار می‌رود محتملترین افراد برای خرید بیمه‌نامه خانه متحرک باشند. دقت پیش‌بینی مدل HCRS از

[1]- *Simulated Annealing*
[2]- *Rough Set*
[3]- *Support Vector Machine*
[4]- *Customer Recognition System*
[5]- *Hybrid Customer Recognition System*
[6]- *Pruned Regression Tree*
[7]- *Improved Feedforward Neural Network*

طریق محاسبه نرخ هدف مشاهده شده در میان مشتریان منتخب بررسی می‌شود. ذکر این نکته ضروری است که تنها اطلاعات در مجموعه داده آموزشی در توسعه مدل HCRS استفاده می‌شود و مجموعه داده ارزیابی منحصراً برای پیش‌بینی استفاده می‌شود.

## ۳. انتخاب ویژگی با طراحی یک درخت رگرسیونی هرس شده

انتخاب ویژگی، فرایند انتخاب از ویژگی‌های (یا متغیرها) اصلی، آن ویژگی‌هایی که برای پیشگویی یا تقسیم‌بندی دارای اهمیت هستند. یک معیار ویژگی، J، بر روی زیرمجموعه‌ای از ویژگی‌ها تعریف می‌شود و ما ترکیبی از ویژگی‌ها را جستجو می‌کنیم که برای آن J حداقل می‌شود. دستیابی به زیرمجموعه‌ای خوب از ویژگی‌ها تحت یک یا چند سنجه نیاز به جستجوی فضای زیرمجموعه‌های ویژگی دارد. انتخاب متغیرهای مرتبط در بهبود عملکرد یک مدل پیش‌بینی بسیار مهم است. هدف شناسایی حداقل زیرمجموعه‌ای از متغیرها است که بالاترین دقت را ارائه می‌کنند. این مساله بنام انتخاب ویژگی از نظر مفهومی مشابه با مساله انتخاب متغیر در یک مدل رگرسیونی کلاسیک است. از آنجاییکه اندازه فایل داده‌های مشتری در حال افزایش است، علاقمندی به مساله انتخاب متغیر بشدت رشد داشته است. در انتخاب ویژگی، رویه‌هایی برای رتبه‌بندی ویژگی‌های داده‌های اصلی براساس رابطه آنها تعریف می‌شوند: تنها حفظ مرتبط‌ترین ویژگی‌ها امکان اجرای کاهش ابعاد با حداقل هزینه اطلاعات داده‌ای را می‌دهد. اگرچه، روشها و تکنیک‌های انتخاب ویژگی متنوعی معرفی شده است، کلیه آنها چند هدف مشترک را دنبال می‌کنند:

- حداکثرسازی دقت همزمان با حداقل‌سازی تعداد ویژگی‌ها
- بهبود دقت با حذف ویژگی‌های غیرمرتبط
- کاهش پیچیدگی داده‌ها و هزینه محاسبات
- بهبود تغییراتی که یک راه‌حل هم قابل درک و هم عملی باشد

از آنجاییکه، درخت تصمیم یک روش غیرخطی است و مدل رفتاری مشتری نیز طبیعتی غیرخطی دارد، بعلاوه، از سرعت محاسباتی بالایی برخوردار است؛ از اینرو، درخت تصمیم می‌تواند رویکردی مناسب جهت شناسایی ویژگی‌های منتخب برای مساله تشخیص مشتری باشد.

### ۳-۱. درخت رگرسیونی هرس شده

درختهای تصمیم یکی از ساده‌ترین و موفقترین الگوریتمهای یادگیری در داده‌کاوی و یادگیری ماشین هستند. درختهای تصمیم به این دلیل مشهور هستند که بسادگی قابل تفسیر و از نظر محاسباتی کم هزینه هستند. در صنعت و محیط کسب و کار، تکنیکها برای ارزیابی اعتبار، کشف کلاهبرداری و مدیریت ارتباط با مشتری استفاده شده‌اند[۱۴]. معمولاً هدف پیدا نمودن درخت تصمیم بهینه با حداقل‌سازی خطای تعمیم است. ویژگی‌های ورودی و مقادیر خروجی می‌توانند گسسته یا پیوسته باشند. درخت تصمیم دو نام دیگر دارد: یک درخت تصمیم با محدوده‌ای از برچسب‌های دسته‌ای گسسته (قیاسی یا طبقه‌ای)، درخت طبقه‌بندی نامیده می‌شود، درحالیکه، یک درخت تصمیم با محدوده‌ای از مقادیر خروجی پیوسته (عددی)، یک درخت رگرسیونی نامیده می‌شود. بطور کلی، این درختهای تصمیم، درخت طبقه‌بندی و رگرسیونی[1] (CART) نامیده می‌شوند [۱۵].

اهمیت یک ویژگی بر مبنای مجموع بهبودها در کلیه گره‌ها است که ویژگی نقش دو نیم کننده (وزن‌دهی شده توسط بخشی از داده‌های آموزشی در هر شکاف گره) را دارد. جانشینها نیز در محاسبات اهمیت لحاظ می‌شوند، بدین معنی که حتی به متغیری که هرگز یک گره را دو نیم نمی‌کند، ممکن است یک امتیاز اهمیت بزرگ اختصاص داده شود. این موضوع امکان رتبه‌بندی اهمیت متغیر برای مشخص‌سازی پوشانه‌های متغیر و همبستگی غیرخطی میان ویژگی‌ها را می‌دهد. امتیازهای اهمیت را می‌توان به صورت اختیاری به دو نیم کننده‌ها محدود نمود و صرفاً مقایسه دو نیم کننده‌ها و رتبه‌بندی‌های اهمیت کل تشخیصی مفید است [۱۴].

از آنجاییکه درختهای تصمیم با پیچیدگی کمتر جامعتر هستند، معمولاً تصمیم‌گیرندگان آنها را ترجیح می‌دهند. بعلاوه، پیچیدگی درخت به دلیل یادگیری جزئیات خاص تاثیر نامطلوبی بر روی عملکرد دقت آن دارد و منجر به عملکرد تعمیم ضعیفی خواهد شد. متداولترین رویکرد مورد استفاده، ابتدا رشد یک درخت تا یک اندازه بزرگ و سپس هرس کردن گره‌ها براساس یک معیار هرس است. معیار توقف مورد استفاده و روش هرس به کار برده شده پیچیدگی را کنترل می‌کند [۱۶].

عملکرد را می‌توان به چندین روش مقایسه نمود. استفاده از یک معیار ارزیابی درخور مساله واقعی تحت بررسی مهم است. هزینه میانگین مربعات خطا باید مورد محاسبه قرار گیرد، زیرا آنها در انتخاب روش تاثیر می‌گذارند. فرض برابری هزینه‌های میانگین مربعات خطا در موارد بسیار کمی مناسب است. معمولاً حتی اگر آنها دقیقاً معلوم نیستند، تا اندازه‌ای می‌توان در خصوص هزینه‌ها صحبت نمود.

در این مقاله، PRT از دو مرحله مفهومی تشکیل شده است: رشد و هرس. تکنیک PRT دو نیم کننده‌هایی را جستجو می‌کند که مربعات خطای (انحراف حداقل مربعات) پیشگویی را حداقل می‌کنند. پیشگویی در هر گره پایانی بر مبنای میانگین وزنی برای گره تعیین می‌شود. نشان داده شده است که درخت تصمیم می‌تواند یک روش انتخاب ویژگی کارا بوده و می‌تواند پیش پردازشگر داده‌ها پیش از استفاده از شبکه‌های عصبی باشد. آن را می‌توان جهت کاهش تعداد ورودی‌ها برای دستیابی به نتایجی دقیقتر که به

[1]- *Classification and Regression Tree*

صورت معنی‌دار زمان آموزش را کوتاهتر می‌سازد، استفاده نمود. ما نیز قصد داریم از خروجی‌های PRT طراحی شده بعنوان ورودی‌های شبکه عصبی جهت کسب نتایجی بسیار بهتر استفاده نماییم. منافع استفاده از درخت تصمیم بعنوان یک پیش پردازشگر برای شبکه عصبی شامل موارد زیر می‌شود[۳]:

- زمان آموزش بسیار سریع
- به تبدیل یا آماده‌سازی داده‌ها نیاز ندارد (درخت تصمیم به سادگی می‌تواند داده‌های خام را استفاده نماید).
- بگارگیری خودکار پیشگویی کننده‌های قیاسی (اسمی).
- قابلیت پیشتیبانی تعداد پیشگویی کننده‌های بسیار زیاد (حداکثر ۸۰۰۰ عدد)
- قابلیت پشتیبانی فایلهای داده آموزشی بسیار بزرگ.

### ۲-۳. الگوریتم هرس[1]

کارهای اولیه در حوزه درختهای تصمیم امکان هرس را نمی‌دادند. در عوض، درختها رشد می‌کردند تا آنها با شرط توقف برخورد کرده و درخت منتج بعنوان درخت نهایی در نظر گرفته می‌شد. هرس، فرایند کاهش یک درخت از طریق تبدیل برخی گره‌های شاخه‌ای به گره‌های پایانی و حذف گره‌های پایانی زیر شاخه اصلی است. مکانیزم هرس اکیداً بر داده‌های آموزشی مبتنی می‌شود و با یک سنجه هزینه پیچیدگی آغاز می‌کند[۱۷].

الگوریتم هرس بطور کلی برای درختهایی به کار گرفته می‌شود که لزوماً طبقه‌بندی نبوده بلکه درختهای رگرسیونی هستند. فرض کنید که $R(t)$ اعداد حقیقی همبسته با هر گره $t$ از یک درخت معلوم $T$ باشد. فرض کنید مقدار $R(t)$ توسط رابطه زیر تعیین شود[۱۵]:

$$R(t) = e(t)p(t) \qquad (۱)$$

که $e(t)$ میانگین مربعات خطا است که معلوم می‌کند که یک مورد در گره می‌افتد و اگر ما فرض کنیم که $N(t)$ تعداد نمونه‌های $L=\{(x_i,y_i),i=1,...,n\}$ را نشان دهد، آنگاه $p(t)$ را می‌توان تعریف نمود:

$$p(t) = \frac{N(t)}{n} \qquad (۲)$$

از اینرو، اگر t بعنوان یک گره پایانی درنظر گرفته شود، $R(t)$، سهم آن گره به خطای کل است. فرض کنید که $T_t$ زیردرختی با ریشه t باشد. اگر $R_\alpha(T_t) < R_\alpha(t)$، آنگاه سهم به هزینه پیچیدگی زیر درخت کمتر از آن برای گره t است. این موضوع برای $\alpha$ کوچک اتفاق می‌افتد. همچنانکه $\alpha$ افزایش می‌یابد، تساوی حاصل می‌شود، زمانیکه[۱۷]:

$$\alpha = \frac{R(t) - R(T_t)}{N_d(t) - 1} \qquad (۳)$$

که $N_d(t)$ تعداد گره‌های پایانی در $T_t$ است، یعنی $N_d(t) = |\tilde{T}_t|$ و حذف درخت در t ارجح می‌شود. بنابراین، ما نهایتاً تعریف می‌کنیم:

$$g(t) = \frac{R(t) - R(T_t)}{N_d(t) - 1} \qquad (۴)$$

بعنوان یک سنجه از شدت پیوند از گره t. اولین مرحله الگوریتم، گره با کوچکترین مقدار $g(t)$ را جستجو می‌کند. آن گره تبدیل به یک گره پایانی می‌شود و مقدار $g(t)$ برای کلیه اجدادش محاسبه می‌شود. این فرایند تکرار می‌شود و ادامه می‌یابد تا به گره ریشه برسیم. بدین ترتیب، الگوریتم هرس توالی از درختها را تولید می‌کند.

### ۳-۳. وارسی اعتبار[2]

وارسی اعتبار (CV)، روشی برای برآورد میزان خطا است که دارای ایده‌ای ساده می‌باشد. مجموعه داده‌ها به اندازه نمونه به دو قسمت تفکیک می‌شوند. پارامترهای مدل با استفاده از یک مجموعه (با حداقل‌سازی چند معیار بهینه‌سازی) برآورد شده و معیار خوبی برازش[3] بر روی مجموعه دوم ارزیابی می‌شود. نسخه معمول وارسی اعتبار یک روش صرفنظر از یکی[4] است که در آن مجموعه دوم متشکل از تنها یک نمونه است. آنگاه، برآورد وارسی اعتبار معیار خوبی برازش،CV، متوسط کلیه مجموعه‌های آموزشی ممکن به اندازه $n-1$ است. خطای وارسی اعتبار،CV، بعنوان وسیله‌ای برای تعیین یک مدل مناسب، برای هر عضو از خانواده مدلهای کاندید $\{M_k, k=1,...,K\}$ و مدل $M_{\hat{k}}$ انتخاب شده محاسبه می‌شود، که:

$$\hat{k} = \arg\min CV(k) \qquad (۵)$$

وارسی اعتبار هنگام انتخاب یک مدل صحیح گرایش به برازش بیش از حد دارد، بطوریکه، برای مجموعه داده یک مدل بسیار پیچیده

[1] *Pruning Algorithm*

[2] *- Cross-Validation*
[3] *- Leave-one-out*
[4] *- Goodness-of-fit*

انتخاب می‌کند. شواهدی وجود دارد که وارسی اعتبار چندتایی، زمانیکه $d > 1$ نمونه از مجموعه آموزشی حذف می‌شود، انتخاب مدل بهتر از وارسی اعتبار صرفنظر از یکی انجام می‌شود. برای $n$ بزرگ، میزان محاسبات زیاد بوده و به طراحی $n$ طبقه‌بندی کننده نیاز دارد. با اینکه، با وجود هزینه یک افزایش در واریانس برآوردکننده، به صورت تخمینی نااریب است[۱۷].

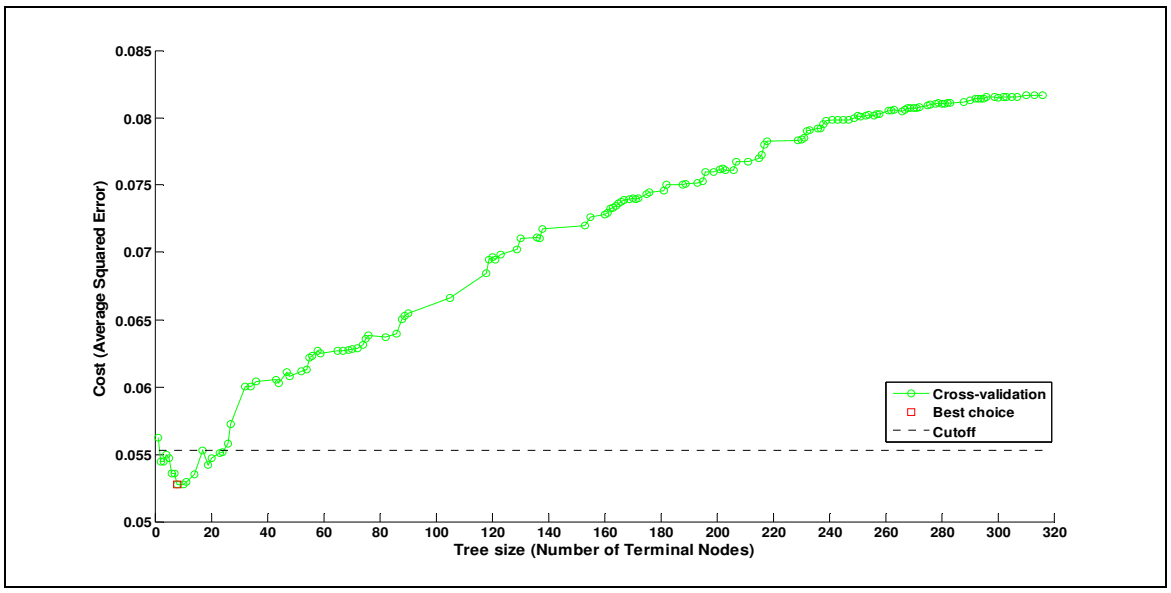
### ۴-۳. مشخصه‌های PRT

در قسمتهای قبل نحوه عملکرد PRT بعنوان یک ابزار جهت انتخاب ویژگی‌ها توصیف شد. در این بخش، به بیان مشخصه‌های آن جهت انتخاب ویژگی‌ها بر روی موردکاوی می‌پردازیم. مشخصات PRT به کار گرفته شامل موارد زیر می‌شود:

- نوع درخت، درخت رگرسیونی است، زیرا نوع متغیر هدف عددی است (۰ یا ۱).
- از وارسی اعتبار ۵۰ تکه‌ای جهت برآورد خطای واقعی برای درختها در اندازه‌های مختلف استفاده می‌شود. بدین معنی که، تابع، نمونه را به ۵۰ زیرنمونه که به صورت تصادفی انتخاب شده و تقریباً دارای اندازه برابر هستند، تفکیک می‌نماید. برای هر یک از زیرنمونه‌ها، یک درخت را به داده‌های باقیمانده برازش می‌دهد و از آن برای پیش‌بینی زیرنمونه استفاده می‌نماید. سپس، اطلاعات کلیه زیرنمونه‌ها را برای محاسبه هزینه برای کل نمونه با یکدیگر ترکیب می‌کند. همچنین برداری شامل خطای استاندارد هر مقدار هزینه، برداری شامل تعداد گره‌های پایانی برای هر زیردرخت و اسکالری شامل بهترین سطح برآورد شده هرس محاسبه می‌شود. بهترین سطح، کوچکترین درختی را تولید می‌کند که در محدوده یک خطای استاندارد از زیردرخت حداقل هزینه است. متغیرهای منتخب آنهایی خواهند بود که بهترین عملکرد را دارند. بدین معنی که، درخت با حداقل هزینه انتخاب می‌شود.
- جهت هرس نمودن درخت، معیار میانگین مربعات خطا به کار برده می‌شود.
- هزینه درخت، مجموع کلیه گره‌های پایانی شامل احتمال برآورد شده هر گره در هزینه گره است. از آنجاییکه درخت یک درخت رگرسیونی است، هزینه یک گره میانگین مربعات خطا در کلیه مشاهدات در آن گره است.
- خطا برای هر گره، واریانس مشاهدات تخصیص داده شده به آن گره است.
- احتمال یک گره از طریق نسبتی از مشاهدات از داده‌های اصلی که شرایط گره را برآورده می‌کنند، محاسبه می‌شود.
- اندازه یک گره به صورت تعدادی از مشاهدات از داده‌های مورد استفاده جهت ایجاد درختی که شرایط را برای گره برآورده کند، تعریف می‌شود.

### ۵-۳. نتایج PRT

مبتنی بر مشخصه‌های ذکر شده، درخت رگرسیونی برای ۵۸۲۲ مشتری ترسیم می‌شود که شامل ۸۵ ویژگی می‌شود. با نگاه به اندازه درخت می‌توان متوجه شد که پیچیدگی درخت به میزان قابل توجهی زیاد است و بنابراین مجموعه قواعد تولید شده نمی‌توانند خیلی شهودی باشند. جهت دستیابی به توصیفی فشرده‌تر داده‌ها، ما برخی از شاخه‌های درخت را هرس می‌کنیم، بطوریکه به صورت مختصر خواهیم دید. بنابراین، درختی که کمترین هزینه با تعداد ویژگی کمتر و تعداد گره‌های پایانی کمتر را دارد، انتخاب می‌شود. شکل۱، نقطه بهینه برای هرس درخت رگرسیونی با کمترین هزینه را نمایش می‌دهد.




**شکل۱. نقطه بهینه برای هرس درخت رگرسیونی با کمترین هزینه**

همانطور که در تصویر مشاهده می‌شود، خط ترسیم شده، هزینه برآورد شده برای هر اندازه درخت را نشان می‌دهد، خط‌چین یک خطای استاندارد بالاتر از حداقل خطا را مشخص می‌کند و مربع کوچک بر روی خط حداقل هزینه درخت زیر خط‌چین را مشخص می‌کند. درخت کامل شامل ۳۱۶ گره پایانی است که هزینه بر مبنای معیار میانگین مربعات خطا برای زیردرختهای مختلف محاسبه شده است.

بطور کلی، درخت ۶۳۱ گره دارد که هزینه درخت به تدریج که اندازه درخت رشد می‌کند و به مقدار ۰،۰۵۲۸ با ۸ گره پایانی می‌رسد، کاهش می‌یابد. در ادامه، این هزینه همراه با یک روند صعودی نوسان می‌کند. بنابراین، بهترین انتخاب، بهترین سطح برآورد شده هرس را نشان می‌دهد که شامل هشت گره پایانی می‌شود. در این حالت، می‌توان متوجه شد که پیچیدگی‌های درخت از طریق کاستن تعداد گره‌های پایانی کاهش می‌یابد. زمانیکه درخت رگرسیونی هرس شده در مجموعه آموزشی ترسیم می‌شود، ویژگی‌هایی که در ترسیم آن بنظر می‌رسند، انتخاب می‌شوند. بطوریکه، ویژگی‌هایی که در مسیرهای منتهی به هر گره پایانی در درخت هرس شده بنظر می‌رسند، زیرمجموعه منتخب می‌شوند. جدول۱، قواعد درخت، اندازه درخت و مقادیر p(t)، e(t)، R(t) در هر گره بعد از مرحله هرس بهینه را نشان می‌دهد.

**جدول ۱. قواعد درخت بعد از مرحله هرس بهینه.**

| گره (t) | قاعده | اندازه | احتمال p(t) | خطا e(t) | ریسک R(t) |
|---|---|---|---|---|---|
| ۱ | اگر بیمه‌نامه خودرو کوچکتر از ۵،۵ است برو به گره ۲ درغیر اینصورت گره ۳. | ۵۸۲۲ | ۱ | ۰،۰۵۶۲ | ۰،۰۵۶۲ |
| ۲ | تطابق= ۰،۰۲۴۹ | ۳۴۵۹ | ۰،۵۹۴۱ | ۰،۰۲۴۲ | ۰،۰۱۴۴ |
| ۳ | اگر نوع عمده مشتری کوچکتر از ۲،۵ است، برو به گره ۴ درغیر اینصورت گره ۵. | ۲۳۶۳ | ۰،۴۰۵۸ | ۰،۰۹۸۶ | ۰،۰۴۰۰ |
| ۴ | اگر بیمه‌نامه آتش‌سوزی کوچکتر از ۳،۵ است، برو به گره ۶ درغیر اینصورت گره ۷. | ۴۶۴ | ۰،۰۷۹۶ | ۰،۱۶۱۵ | ۰،۰۱۲۹ |
| ۵ | اگر بیمه‌نامه قایق کوچکتر از ۰،۵ است، برو به گره ۸ درغیر اینصورت گره ۹. | ۱۸۹۹ | ۰،۳۲۶۱ | ۰،۰۸۰۶ | ۰،۰۲۶۳ |
| ۶ | تطابق= ۰،۱۲۸۶ | ۲۴۱ | ۰،۰۴۱۳ | ۰،۱۱۲۱ | ۰،۰۰۴۶ |
| ۷ | اگر کارگر بی‌تجربه کوچکتر از ۳،۵ است، برو به گره ۱۰ درغیر اینصورت گره ۱۱. | ۲۲۳ | ۰،۰۳۸۳ | ۰،۲۰۲۷ | ۰،۰۰۷۸ |
| ۸ | اگر بیمه‌نامه آتش‌سوزی کوچکتر از ۲،۵ است، برو به گره ۱۲ درغیر اینصورت گره ۱۳. | ۱۸۸۲ | ۰،۳۲۳۲ | ۰،۰۷۷۳ | ۰،۰۲۵۰ |
| ۹ | اگر بیمه شخص ثالث شخصی کوچکتر از ۱ است، برو به گره ۱۴ درغیر اینصورت گره ۱۵. | ۱۷ | ۰،۰۰۲۹ | ۰،۲۴۹۱ | ۰،۰۰۰۷ |
| ۱۰ | تطابق= ۰،۲۶۱۷ | ۲۱۴ | ۰،۰۳۶۷ | ۰،۱۹۳۲ | ۰،۰۰۷۱ |
| ۱۱ | تطابق= ۰،۷۷۷۸ | ۹ | ۰،۰۰۱۵ | ۰،۱۷۲۸ | ۰،۰۰۰۳ |
| ۱۲ | تطابق= ۰،۰۵۱۰ | ۹۶۱ | ۰،۱۶۵۰ | ۰،۰۴۸۴ | ۰،۰۰۸۰ |
| ۱۳ | تطابق= ۰،۱۱۹۴ | ۹۲۱ | ۰،۱۵۸۱ | ۰،۱۰۵۲ | ۰،۰۱۶۶ |
| ۱۴ | تطابق= ۰،۸۱۸۲ | ۱۱ | ۰،۰۰۱۸ | ۰،۱۴۸۸ | ۰،۰۰۰۳ |
| ۱۵ | تطابق= ۰ | ۶ | ۰،۰۰۱۰ | ۰ | ۰ |

همانطور که می‌توان در جدول ۱ مشاهده نمود، پس از مرحله هرس بهینه ۱۵ گره باقی می‌ماند و شماره گره‌های پایانی ۲، ۶، ۱۰، ۱۱، ۱۲، ۱۳، ۱۴ و ۱۵ هستند و هر یک نشان می‌دهند که چه تعداد مشتری در ستون اندازه براساس چه قاعده‌ای و چه احتمالی در نظر گرفته می‌شوند. خطا برای هر گره در ستون پنجم، واریانس مشاهدات تخصیص داده شده به آن گره است و ریسک برای هر گره در ستون ششم، خطای گره وزن‌دهی شده توسط احتمال گره است. نهایتاً، شش ویژگی بنامهای بیمه‌نامه خودرو، نوع عمده مشتری، بیمه‌نامه آتش‌سوزی، بیمه‌نامه قایق، کارگر بی‌تجربه و بیمه شخص ثالث شخصی بعنوان تاثیرگذارترین ویژگی‌ها بر روی هدف یعنی خرید بیمه‌نامه خانه متحرک انتخاب می‌شوند. این رویه انتخاب ویژگی منجر به کاهش قابل توجهی معادل با ۹۳٪ در داده‌های ورودی می‌شود. بعلاوه، جهت ارزیابی اثربخشی شش ویژگی منتخب توسط PRT، انتخاب زیرمجموعه‌ای از ویژگی‌ها را با تکنیک قوی دیگری بنام منحنی ROC ارائه شده در مقاله [۶] انجام داده‌ایم. نتیجه نشان داد که ۱۲ ویژگی اول منتخب توسط این تکنیک شامل بیمه‌نامه خودرو، تعداد بیمه‌نامه خودرو، درآمد متوسط، بیمه‌نامه آتش‌سوزی، توان خرید، تحصیلات سطح پایین، بیمه شخص ثالث شخصی، صاحبخانه، تعداد بیمه شخص ثالث شخصی، اجاره نشین، درآمد کمتر از ۳۰ هزار و تحصیلات سطح بالا می‌شوند. نکته جالب توجه این است که مقایسه دو تکنیک نشان می‌دهد که تنها سه ویژگی منتخب مشابه هستند. در مرحله بعد پیاده‌سازی مدل HCRS طراحی شده، خریداران بیمه‌نامه خانه متحرک بر مبنای ویژگی‌های منتخب توسط هر دو تکنیک انتخاب ویژگی پیش‌بینی می‌شوند و نتایج در حالتهای مختلف مقایسه می‌شوند.

## ۴. پیش‌بینی با توسعه FFNN بهبودیافته

انتخاب ویژگی، فرایند شناسایی مواردی است که بعنوان یک فرایند حیاتی برای پیش‌بینی بهترین هستند. مرحله بسیار مهمی است، زیرا به پاکسازی داده‌ها و کاهش داده‌ها با لحاظ ویژگی‌های مهم، حذف موارد زائد و حاوی اطلاعات کمتر کمک می‌کند. از آنجاییکه، ویژگی‌ها از یک مجموعه داده استخراج شده‌اند، نیاز به اعتبارسنجی دارند. بنابراین، پس از استفاده از PRT برای انتخاب ویژگی، یک FFNN بهبودیافته برای پیش‌بینی بر مبنای زیرمجموعه‌ای بهینه از ویژگی‌ها طراحی شده است.

معروفیت شبکه‌های عصبی مصنوعی را نه تنها براساس جاذبه‌های نظری آنها بلکه با قابلیتهای وسیع آنها بعنوان ابزار حل کاربردی می‌توان توصیف نمود.

تحقیقات در زمینه شبکه‌های عصبی مصنوعی[1] (ANNs) در بسیاری از حوزه‌ها نتایج امیدوارکننده‌ای را ارائه نموده است و آنها بعنوان ابزار حل مساله در تنظیمات علمی و مهندسی متنوع مورد استفاده قرار گرفته‌اند. موفقیت این مدلهای ANN به قابلیت‌هایش بعنوان یک برآوردگر کلی مرتبط می‌شود[۱۸]. در یک شبکه عصبی پیشخوراند[2] (FFNN)، اتصالات بین واحدها تشکیل سیکل نمی‌دهند. معمولاً FFNNها به سرعت برای یک ورودی تولید پاسخ می‌کنند.

FFNNها، زیرمجموعه‌ای از دسته مدلهای رگرسیون غیرخطی و متمایزسازی هستند[۱۹]. مزیت FFNNها قدرت آنها بعنوان یک برآوردگر احتمالی بیزین است. آنها این مشخصه را نشان داده‌اند که خروجی‌های آنها می‌تواند بعنوان برآوردهایی از توزیع احتمال پسین[3] ملاحظه شود[۱۸].

در این تحقیق، بمنظور بهبود کارایی FFNN، تابع عملکرد مجموع سه عامل را اندازه می‌گیرد. به طوری که، علاوه بر میانگین مربعات خطا، دو عامل میانگین مربعات وزنها و بایاسها و میانگین مربعات خروجی را لحاظ می‌کند که ما این شبکه عصبی طراحی شده را شبکه عصبی پیشخوراند بهبودیافته (IFFNN) می‌نامیم. جهت سنجش خطای پیش‌بینی، معیار میانگین مربعات خطا (MSE) استفاده می‌شود.

پس از آموزش، IFFNN با کمترین MSE انتخاب می‌شود و قابلیت پیش‌بینی مشتری مورد بررسی قرار می‌گیرد. بطوریکه، ۲۰٪ اول از کل مشتریان را انتخاب می‌کنیم که احتمال بیشتری برای خرید بیمه‌نامه خانه متحرک دارند. نهایتاً، تعداد خریداران پیش‌بینی شده در میان ۲۰٪ مشتریان منتخب مشخص می‌شود. بدین ترتیب، می‌توانیم کیفیت مدل HCRS طراحی شده را تعیین نماییم. در زیربخش بعدی، به طراحی معماری IFFNN اقدام خواهیم نمود.

### ۴-۱. معماری IFFNN

در بسیاری از حالتهای عملی، یک طراح با ویژگی‌هایی مواجه می‌شود که مقادیر در محدوده پویای متفاوتی قرار می‌گیرند. بطوریکه، ویژگی‌ها با مقادیر بزرگ می‌توانند تاثیر بیشتری نسبت به مقادیر کوچک در تابع هزینه داشته باشند که لزوماً اهمیت نسبی آنها در طراحی پیشگویی‌کننده منعکس نمی‌کند. انجام مراحل پیش‌پردازش روی ورودی‌ها و اهداف شبکه می‌تواند منجر به فرایند آموزش کاراتری شود.

محدودسازی ورودی‌ها و اهداف در یک محدوده خاص با مقیاس‌گذاری آنها غالباً بهتر است. این مشکل با نرمالسازی ویژگی‌ها غلبه می‌شود، آنچنانکه مقادیر در محدوده‌ای مشابه قرار گیرد. در این تحقیق، ورودی‌ها و هدف را نرمالسازی نموده‌ایم، آنچنانکه آنها میانگین صفر و انحراف معیار واحد دارند. بعلاوه، الگوریتم لونبرگ-مارکوآرت (LM) بمنظور آموزش IFFNN استفاده شده است.

بطور کلی، در خصوص مسائل تخمین تابع، برای شبکه‌هایی که حداکثر صدها وزن را شامل می‌شوند، الگوریتم LM سریعترین همگرایی را خواهد داشت. این مزیت بویژه زمانی قابل توجه است که آموزش بسیار دقیق مورد نیاز است. در موارد بسیاری، LM قادر است میانگین مربعات خطای کمتری را نسبت به هر الگوریتم دیگری بدست بیاورد [20, 21].

یک IFFNN دولایه متشکل از یک لایه مخفی توسعه داده شده است. تنها یک نرون در لایه خروجی جهت انجام پیش‌بینی خریداران بیمه‌نامه خانه متحرک وجود دارد. با توجه به طبیعت مساله، چندین تابع انتقال را می‌توان برای لایه‌های مختلف شبکه به کار برد.

استفاده از توابع انتقال لجیستیک سیگموید در لایه مخفی خروجی نرونهای آن را در محدوده [0, 1] تثبیت می‌کند. از اینرو، بخاطر طبیعت مثبت ورودی‌ها و خروجی، یک تابع انتقال لجیستیک سیگموید در لایه مخفی و لایه خروجی به کار گرفته شده است. شبکه برای ۱۰۰۰ دوره آموزشی آموزش داده شده است. جهت پیدا نمودن بهترین تعداد نرون در لایه مخفی، از یک تا ۵۰ نرون آزمون شده است و بمنظور افزایش احتمال دستیابی به یک حداقل مطلق برای هر تعداد نرون ۱۰۰ تکرار انجام شده است.

### ۴-۲. نتایج

همانطور که بیان شد، پیش‌بینی خریداران بیمه‌نامه خانه متحرک با توسعه یک IFFNN که از مجموعه داده‌های شش ویژگی منتخب بعنوان ورودی استفاده می‌کند، اجرا شده است. بیان این نکته لازم است که مجموعه داده‌ای که انتخاب ویژگی بر پایه آن انجام شده بود، باید بطور کامل از مجموعه داده ارزیابی مستقل باشد؛ در غیر اینصورت، ریسک تطبیق بیش از حد وجود خواهد داشت. بعلاوه، جهت اطمینان از آموزش خوب شبکه (بدون تطبیق بیش از حد)،

[1]- *Artificial Neural Networks*
[2]- *Feedforward Neural Network*
[3]- *Posterior Probability Distribution*

۷۵٪ مجموعه داده‌های آموزشی جهت آموزش شبکه (برای تعیین وزنها و بایاسها) و ۲۵٪ باقیمانده برای اعتبارسنجی استفاده می‌شود. سپس، مجموعه داده ارزیابی جهت آزمون شبکه عصبی توسعه داده شده استفاده می‌شود.

نتیجه نشان داد که تعداد نرون بهینه برای اجرای IFFNN بر روی مجموعه داده شش ویژگی کاهش یافته توسط مدل PRT برابر ۴۴ نرون است.

جهت ارزیابی اثربخشی انتخاب ویژگی با استفاده از PRT، IFFNN دیگری توسعه داده شده است که از داده‌ها بدون انتخاب ویژگی برای پیش‌بینی استفاده می‌کند. بعلاوه، جهت نمایش برتری IFFNN بر FFNN (بدون انتخاب ویژگی) و درخت رگرسیونی بدون هرس، عملکرد آنها با بررسی MSE و دقت تکنیکها بر روی مجموعه داده ارزیابی مقایسه می‌شوند. نتایج کلی بدست آمده با استفاده از مدل HCRS و آنهایی که با استفاده از IFFNN، FFNN و درخت رگرسیونی (که از داده‌ها با ۸۵ ویژگی استفاده می‌کنند) در جدول ۲ نشان داده می‌شوند.

**جدول ۲. نتایج کلی بدست آمده با و بدون انتخاب ویژگی**

| تکنیک پیش‌بینی | مجموعه داده آموزشی | | خریداران میان ۲۰٪ مشتری منتخب مجموعه داده ارزیابی | |
|---|---|---|---|---|
| | MSE | تعداد خریداران | MSE | تعداد خریداران |
| HCRS (IFFNN با انتخاب شش ویژگی) | ۰.۰۵۱۸ | ۱۹۲ | ۰.۰۵۳۲ | ۱۲۵ |
| IFFNN بدون انتخاب ویژگی (۸۵ ویژگی) | ۰.۰۵۰۹ | ۱۹۵ | ۰.۰۵۳۳ | ۱۲۵ |
| FFNN بدون انتخاب ویژگی (۸۵ ویژگی) | ۰.۰۵۲۳ | ۱۷۲ | ۰.۰۵۴۴ | ۱۱۹ |
| درخت رگرسیونی بدون هرس (۸۵ ویژگی) | ۰.۰۴۴۲ | ۲۶۹ | ۰.۰۶۱۸ | ۹۲ |

جدول ۲ نتایج بهینه بدست آمده برای اجرای چهار حالت طراحی شده با تغییر تعداد ویژگی‌ها و تکنیکهای پیش‌بینی مبتنی بر MSE و تعداد خریداران بیمه‌نامه خانه متحرک نشان می‌دهد. همانطور که در جدول مشاهده می‌شود، به کارگیری صرفاً یک درخت رگرسیونی، یک FFNN و یک IFFNN با ۸۵ ویژگی بر مبنای معیار MSE، به ترتیب مقادیر ۰.۰۶۱۸، ۰.۰۵۴۴ و۰.۰۵۳۳ بر روی مجموعه داده ارزیابی تولید می‌کند، درحالیکه با استفاده از داده‌های ویژگی بسیار کم (شش ویژگی) بدست آمده، از طریق توسعه IFFNN به مقدار ۰.۰۵۳۲ کاهش می‌یابد. همچنین، لحاظ معیار عملکرد دیگر (یعنی تعداد خریداران بیمه‌نامه خانه متحرک) ثابت می‌کند که استفاده از مدل HCRS به صورت معنی‌داری نتایج پیش‌بینی (به استثنای IFFNN بدون انتخاب ویژگی که نتیجه برابر است) را در میان ۲۰٪ مشتری منتخب در مجموعه ارزیابی بهبود می‌دهد، یعنی ۱۲۵ خریدار در مقابل به ترتیب ۹۲، ۱۱۹ و ۱۲۵ خریدار.

بنابراین، داشتن ۶ ویژگی بجای ۸۵ ویژگی پیچیدگی محاسبات را کاهش می‌دهد، درحالیکه نتیجه پیش‌بینی مشابه است. باید اضافه شود که متوسط زمان هر بار اجرای IFFNN برابر ۰،۵ ثانیه در مقابل ۵ ثانیه برای FFNN است که حاکی از سرعت محاسباتی بسیار بالاتر IFFNN است.

همچنین، جهت ارزیابی دقت نتایج انتخاب ویژگی و پیش‌بینی بدست آمده با طراحی مدل HCRS، تکنیک ROC جهت انتخاب ویژگی و تکنیک‌های FFNN و PRT برای پیش‌بینی خریداران بیمه‌نامه خانه متحرک به کار برده شده و نتایج مقایسه شدند. برای انجام این کار، ۱۲ ویژگی منتخب از تکنیک انتخاب ویژگی ROC بعنوان ورودی برای IFFNN استفاده می‌شوند (ما آن را ROC تلفیقی(HROC) می‌نامیم) و دو تکنیک FFNN و PRT که از مجموعه داده‌های شش ویژگی منتخب PRT استفاده می‌کنند (ما آنها را مدلهای شبکه عصبی پیشخوراند تلفیقی (HFFNN) و PRT دوبل (DPRT) می‌نامیم) توسعه داده شده‌اند. جدول ۳ پیش‌بینی‌های بدست آمده با توسعه مدلهای HCRS،HROC، HFFNN و DPRT را نمایش می‌دهد.

**جدول ۳. پیش‌بینی‌های بدست آمده برای خریداران بیمه‌نامه خانه متحرک با توسعه چهار مدل**

| نوع مدل | تکنیک انتخاب ویژگی | تکنیک پیش‌بینی | خریداران میان ۲۰٪ مشتری منتخب مجموعه داده آموزشی MSE | تعداد خریداران | درصد از کل خریداران | درصد از مشتریان تشخیص داده شده | مجموعه داده ارزیابی MSE | تعداد خریداران | درصد از کل خریداران | درصد از مشتریان تشخیص داده شده |
|---|---|---|---|---|---|---|---|---|---|---|
| HCRS | PRT | IFFNN | ۰.۰۵۱۸ | ۱۹۲ | ٪۵۵.۲ | ٪۱۶.۵ | ۰.۰۵۳۲ | ۱۲۵ | ٪۵۲.۵ | ٪۱۵.۶ |
| HROC | ROC | IFFNN | ۰.۰۵۲۵ | ۱۹۸ | ٪۵۶.۹ | ٪۱۷.۰ | ۰.۰۵۳۳ | ۱۲۲ | ٪۵۱.۳ | ٪۱۵.۳ |
| HFFNN | PRT | FFNN | ۰.۰۵۲۲ | ۱۸۶ | ٪۵۳.۵ | ٪۱۶.۰ | ۰.۰۵۴۲ | ۱۱۹ | ٪۵۰.۰ | ٪۱۴.۹ |
| DPRT | PRT | PRT | ۰.۰۵۱۳ | ۱۷۸ | ٪۵۱.۲ | ٪۱۵.۳ | ۰.۰۵۴۹ | ۱۰۰ | ٪۴۲.۰ | ٪۱۲.۵ |

با مقایسه نتایج پیش‌بینی در جدول ۳ مشاهده می‌شود که پیش‌بینی‌های انجام شده با مدل HCRS برتری قابل توجهی در کلیه معیارهای عملکرد نسبت به مدلهای HROC، HFFNN و DPRT روی مجموعه داده ارزیابی نشان می‌دهد. آنچنانکه، به کارگیری PRT و FFNN بعنوان تکنیکهای پیش‌بینی به ترتیب میانگین مربعات خطایی برابر ۰،۰۵۴۹ و ۰،۰۵۴۲ تولید می‌کند و به کارگیری ROC بعنوان تکنیک انتخاب ویژگی پیش از اجرای IFFNN میانگین مربعات خطایی برابر ۰،۰۵۳۳ را ارائه می‌کند، درحالیکه این مقدار با استفاده از داده‌های بدست آمده از مرحله انتخاب ویژگی و با توسعه IFFNN به ۰،۰۵۳۲ کاهش می‌یابد. همچنین، لحاظ سه معیار عملکرد دیگر یعنی تعداد خریداران، درصد از کل خریداران و درصد از مشتریان تشخیص داده شده اثبات می‌کند که استفاده از مدل HCRS به میزان قابل توجهی نتایج پیش‌بینی را بهبود می‌دهد. این مقادیر به ترتیب ۱۲۵، ۵۲،۵٪ و ۱۵،۶٪ برای مدل HCRS در مقابل ۱۲۲، ۵۱،۳٪، ۱۵،۳٪ برای مدل HROC، ۱۰۰، ۴۲،۰٪ و ۱۲،۵٪ برای مدل DPRT و ۱۱۹، ۵۰،۰٪ و ۱۴،۹٪ برای مدل HFFNN میان ۲۰٪ مشتری منتخب روی مجموعه داده ارزیابی هستند. همچنین، این موضوع قابل توجه است که مدل HROC با وجود به کارگیری ویژگیهایی دو برابر مدل HCRS به جواب بدتری منجر می‌شود. بعلاوه، همانطور که در جداول ۲ و ۳ دیده می‌شود، MSE با استفاده از درخت رگرسیونی بدون هرس برابر ۰،۰۶۱۸ است؛ در عین حال، با استفاده از مدل DPRT خطای پیش‌بینی به مقدار ۰،۰۵۴۹ کاهش می‌یابد. بطور خاص، این موضوع برای سایر معیارهای عملکرد نیز صدق می‌کند که نشاندهنده آن است که مدل انتخاب ویژگی پیشنهادی با استفاده از PRT تاثیراتش را بر روی درخت رگرسیونی نیز دارا است.

بطور کلی، نتایج در جدول ۳ نشان می دهد که مدل HCRS (تلفیقی از PRT و IFFNN) نمایش قوی و برتری را از داده‌ها ارائه می‌کند، همچنانکه آن قادر بود تعداد ویژگی‌ها را به میزان ۹۳٪ کاهش دهد و کاهش قابل توجهی را در خطای تشخیص مشتریان جدید نشان دهد.

بعلاوه، مقایسه بهترین تعداد مشتری پیش‌بینی شده (۱۲۱ خریدار با روش بیز ساده بر روی سایت به آدرس: (http://www.liacs.nl/~putten/library/cc2000/report2.html بر روی داده‌های مورد استفاده برتری عملکرد مدل HCRS را نمایش می‌دهد. سایر مدلهای ارائه شده همچون SVM توسط محققین بر روی این سایت نیز نتایج ضعیفتری را ارائه می‌کنند.

## ۵. نتیجه‌گیری

توسعه‌های اخیر در انتخاب ویژگی مساله بهبود عملکرد پیشگویی کننده‌ها را از نقطه نظر عملی مورد توجه قرار داده است. در این مقاله، ما تلاش نمودیم که یک سیستم تشخیص مشتری تلفیقی (HCRS) را برای پیش‌بینی خریداران یک محصول مشخص بنام بیمه‌نامه خانه متحرک توسعه دهیم. مدل جدیدی برای پیش‌بینی رفتار مشتری با طراحی یک شبکه عصبی پیشخوراند بهبودیافته (IFFNN) مبتنی بر شش ویژگی منتخب بهینه با استفاده از درخت رگرسیونی هرس شده پیشنهاد شده است. بطوریکه، ما کمترین نرخ میانگین مربعات خطا را برای تعیین بهترین درخت رگرسیونی هرس شده و بهترین معماری IFFNN لحاظ نمودیم. این مدل HCRS به یک بهینه‌سازی ترکیبی نامحدود منجر شد که در آن نرخ MSE معیار جستجو است. معیار حداقل میانگین مربعات خطا جهت تعیین نقطه بهینه برای هرس درخت رگرسیونی به کار برده شده بود که نهایتاً منجر به انتخاب شش ویژگی شد. به کارگیری این رویه انتخاب ویژگی پیشنهادی بهبود قابل توجهی را در کاهش ویژگی‌ها نشان داد. همچنین، این رویه سرعت محاسباتی بسیار زیادی در انتخاب ویژگی دارد. سپس، پیش‌بینی خریداران با توسعه IFFNN براساس ویژگی‌های منتخب انجام شده است. جهت بهبود کارایی FFNN، تابع عملکرد مجموع وزنی سه عامل میانگین مربعات خطا، میانگین مربعات وزنها و بایاسها و میانگین مربعات خروجی را مورد سنجش قرار داد. بطوریکه، IFFNN طراحی شده سرعت محاسباتی بسیار بالاتری داشته و نتیجه بهتری را ارائه کرد. جهت ارزیابی اثربخشی مدل HCRS، پیش‌بینی با استفاده از صرفاً یک IFFNN، یک FFNN و یک درخت رگرسیونی همگی بدون انتخاب ویژگی اجرا شده است. نتایج، برتری و اثربخشی مدل HCRS مبتنی بر معیارهای عملکرد را اثبات کرد. علاوه بر این، جهت ارزیابی قوی بودن خروجی تکنیک انتخاب ویژگی PRT، تکنیک ROC و جهت سنجش دقت نتایج پیش‌بینی بدست آمده با استفاده از IFFNN، دو تکنیک شامل FFNN و PRT که از مجموعه داده‌های شش ویژگی منتخب PRT استفاده می‌کنند، توسعه داده شده‌اند. مقایسه نتایج پیش‌بینی برتری معنی‌دار تلفیق PRT و IFFNN (مدل HCRS) بر هر سه مدل دیگر در کلیه معیارهای عملکرد نشان داد. بعلاوه، مقایسه بین PRT و درخت رگرسیون کلی که تنها از داده‌های بدون هرس استفاده می‌کند، تاثیر رویه انتخاب ویژگی پیشنهادی بر روی عملکرد تکنیک درخت رگرسیونی نشان داد. نتایج کلی عملکردهای پیش‌بینی نشان می‌دهد که رویه انتخاب ویژگی پیشنهادی با به کارگیری درخت رگرسیونی هرس شده به میزان قابل توجهی نتایج پیش‌بینی بر روی درخت رگرسیونی را نیز بهبود می‌دهد.

## ۶. تشکر و قدردانی

نویسندگان مقاله تمایل دارند از پشتیبانی مالی مرکز تحقیقات مخابرات ایران (ITRC) در پیشبرد اهداف این تحقیق تشکر و قدردانی نمایند.

## مراجع


[1] Kohavi, R., John, G.H., "*Wrappers for Feature Subset Selection", Artificial Intelligence,* (1–2), 1997, pp. 273–324.

[2] Blum, A.L., Rivest, R.L., "*Training a 3-Node Neural Networks is NP-Complete*", *Neural Networks*, 5, 1992, pp.117-127.

[3] Ng, K.S., Liu, H., "*Customer Retention Via Data mining*", Artificial Intelligence Review*,* 14(6), 2000, pp.569-590.

[4] Anderson, E.T., "*Sharing the Wealth: When Should Firms Treat Austomers as Partners?", Management Science,* 48(8), 2002, 955–71.

[5] Kim, Y.S., "*Toward a Successful CRM: Variable Selection, Sampling, and Ensemble", Decision Support Systems,* 41, 2006, pp. 542– 553.

[6] Yan, L., Wolniewicz R., Dodier, R., "*Predicting Customer Behaviour in Telecommunications*", *IEE Intelligent Systems,* 19, 2004, pp. 50–58.

[7] Kim, Y.S., Street, W.N., "*An Intelligent System for Customer Targeting: a Data Mining Approach", Decision Support Systems,* 37, 2004, pp. 215– 228.

[8] Yu, E., Cho, S., "*Constructing Response Model Using Ensemble Based on Feature Subset Selection"*, Expert Systems with Applications, 30, 2006, pp. 352–360.

[9] Ahn, H., Kim, K., Han, I., "*A Case-Based Reasoning System with the Two-Dimensional Reduction Technique for Customer Classification"*, Expert Systems with Applications, 32, 2007, pp. 1011–1019.

[10] Buckinx, W., Verstraeten, G., Poel, D. V., "*Predicting Customer Loyalty using the Internal Transactional Database*", Expert Systems with Applications, 32, 2007, pp. 125–134.

[11] Yan, L., Changrui, Y., "A *New Hybrid Algorithm for Feature Selection and its Application to Customer Recognition"*, *LNCS 4616*, 2007, pp. 102-111.

[12] Tseng, T.L., Huang, C.C., "*Rough Set-Based Approach to Feature Selection in Customer Relationship Management"*, *Omega,* 35, 2007, pp. 365-383.

[13] Lessmann, S., Voß, S., "*A Reference Model for Customer-Centric Data Mining with Support Vector Machines", European Journal of Operational Research*, 199, 2009, pp. 520–530.

[14] Wu, X., Kumar, V., Quinlan, J.R., Ghosh, J., Yang, Q., Motoda, H., McLachlan, G. J., Ng, A., Liu, B., Yu, Ph. S., Zhou, Zh.H., Steinbach, M., Hand, D. J., Steinberg, D., "*Top 10 Algorithms in Data Mining",* Knowledge Information System 14, 2008, pp. 1–37.

[15] Breiman, L., Friedman, J., Olshen, R., Stone, C., *"Classification and Regression Trees*". Boca Raton, FL: CRC Press, 1984.

[16] Theodoridis, S., Koutroumbas, K., *"Pattern Recognition*", Academic Press, Third Edition, 2006.

[17] Webb, A.R., "*Statistical Pattern Recognition"*, John Wiley & Sons, Ltd., Second Edition, 2002.

[18] Wan, E.E., *"Neural Network Classification: A Bayesian Isterpretation*", IEEE Transactions on Neural Networks, 1(4), 1990, 303-305.

[19] Bishop, C.M., *"Neural Networks for Pattern Recognition*", Oxford: Oxford University Press, 1995.

[20] Demuth, H., Beale, M., Hagan, M., *"Neural Network Toolbox™ User's Guide",* Version 6.0.3, The MathWorks, Inc., pp. 1-901, 2009.

[21] Hagan, M., Demuth, H., Beale. M., "*Neural Network Design"*, first Edition, USA, PWS Publishing Company, 1996.