

پیشینی وفاداری و رویگردانی بیمه‌شدگان خویش‌فرمای سازمان تأمین اجتماعی^۱

اکبر نجفی^۲

تاریخ دریافت مقاله: ۱۳۹۸/۲/۱۷

تاریخ پذیرش مقاله: ۱۳۹۸/۵/۲۰

چکیده

هدف: وفاداری و رویگردانی مشتریان یکی از چالش‌های مهم شرکت‌ها و مؤسساتی است که خدمات و کالا را به مشتریان خود عرضه می‌کنند. سازمان تأمین اجتماعی هم در موضوع وفاداری بیمه‌شدگان خویش‌فرما جهت ادامه بیمه‌پردازی، در سال‌های اخیر با چالش فوق‌مواجهه بوده است. هدف این مقاله شناسایی میزان رویگردانی بیمه‌شدگان و همچنین یافتن مدلی بر مبنای الگوریتم‌های هوش مصنوعی است که میزان رویگردانی بیمه‌شدگان در سال‌های آتی را به‌درستی تخمین زده، میزان ازدست‌رفتن منابع درآمدی مربوط به آنها را محاسبه نماید.

روش: در این پژوهش اطلاعات و شاخص‌های مهم حدود ۲۱۴۰۷ نفر در قالب ۲۷ ویژگی از بانک‌های اطلاعاتی سازمان تأمین اجتماعی استخراج گردید. سپس با استفاده از الگوریتم NSGA-II، تعداد ۷ ویژگی مهم که حداقل خطای طبقه‌بندی را داشتند، انتخاب شد.

یافته‌ها: در ابتدا عملیات طبقه‌بندی داده‌ها با شبکه عصبی چندلایه بر روی داده‌های موجود و با استفاده از ۲۷ ویژگی انجام شد و دقت طبقه‌بندی ۹۷٫۶ درصد به دست آمد. پس از انتخاب بهترین طبقه‌بندی که شبکه عصبی بود، عملیات طبقه‌بندی با شبکه عصبی بر روی همان داده‌ها و در قالب ۷ ویژگی انجام شد و دقت ۹۷٫۵۵ درصد به دست آمد. در مرحله انتخاب بهترین طبقه‌بندی برای انجام پیش‌بینی‌های مورد نیاز، از ۳ الگوریتم شبکه عصبی چندلایه، الگوریتم ماشین بردار پشتیبان (SVM) و الگوریتم KNN استفاده شد که نهایتاً شبکه عصبی چندلایه، بهترین دقت طبقه‌بندی را با مقدار ۹۶٫۸ درصد به دست آورد. سپس برای شبکه عصبی مذکور با استفاده از ۷ ویژگی مربوط به داده‌های سال‌های ۱۳۶۷ تا ۱۳۹۵ آموزش انجام شد. از شبکه عصبی آموزش‌دیده برای پیش‌بینی وفاداری و رویگردانی مشتریان سال‌های ۱۳۹۶ و ۱۳۹۷ که به تعداد ۸۳۶۴ رکورد می‌باشد، استفاده شد.

نتیجه‌گیری: با در نظر گرفتن همه احتمالات، حدود ۲۷٫۶۵ درصد از بیمه‌شدگان خویش‌فرمای سال‌های ۱۳۹۶ و ۱۳۹۷ در کلاس رویگردان دسته‌بندی می‌شوند.

واژگان کلیدی: شبکه عصبی، وفاداری، رویگردانی، بیمه‌شدگان، خویش‌فرما، تأمین اجتماعی.

۱- این مقاله از پایان نامه استخراج گردیده و با حمایت مالی موسسه عالی پژوهش تأمین اجتماعی انجام شده است.

۲- کارشناسی ارشد مهندسی فناوری اطلاعات، گرایش تجارت الکترونیک، دانشگاه آزاد اسلامی، شیراز، ایران. Najafi209@gmail.com

مقدمه

یکی از ارزشمندترین دستاوردهای بشر برای رویارویی با خطرات و تهدیدات و همچنین تأمین شرایط اقتصادی، اجتماعی و روانی، بیمه^۱ است. بیمه افزون بر جبران خسارت‌های اقتصادی ناشی از حوادث و ارتقای سطح زندگی افراد، باعث آرامش خاطر اعضای جامعه می‌شود و این آرامش خاطر خود سبب پویایی حیات اجتماعی، رشد و شکوفایی استعدادها و افزایش کارایی و بهره‌وری در جامعه می‌گردد. (رستمی و شرافت، ۱۳۹۴: ۹۵-۸۴)

صنعت بیمه در کشور به علت رویارویی با مسائلی چون خصوصی‌سازی، اصلاحات در قوانین موجود، آزادسازی تعرفه‌ها و پیشرفت فناوری اطلاعات در حال تغییر و تحول بنیادین است. در سال‌های اخیر، صنعت بیمه از صنعتی محصول‌محور به خدمتی مشتری‌محور تبدیل شده است. در واقع، در سال‌های اخیر شرکت‌های بیمه، به شرکت‌هایی مشتری‌گرا تبدیل شده‌اند که مشتری در مرکز توجه این شرکت‌هاست. در ادامه و در جدول ۱ تفاوت محصول‌گرایی و مشتری‌گرایی نشان داده شده است. (مظلومی و دیگران، ۱۳۹۳: ۵۹-۱)

جدول ۱: تفاوت محصول‌گرایی و مشتری‌گرایی [۲]

ویژگی	محصول‌گرایی	مشتری‌گرایی
تمرکز	محصولات	مشتریان با ارزش
سازمان	یک فرد مسئول محصولات	یک فرد مسئول مشتریان
استراتژی	یافتن مشتری برای محصول	یافتن محصول برای مشتری
پیشنهاد	محصولات یکسان	محصولات یکپارچه و شخصی‌شده با تولید انبوه
معیار موفقیت	سودآوری از محصول، رضایت مشتری	سودآوری از مشتری، وفاداری مشتری

با توجه به نکات مطرح‌شده در این جدول، می‌توان گفت در عصر حاضر، سازمان‌هایی که به نیازمندی‌های مشتریان بی‌اعتنا باشند، از صحنه رقابت حذف می‌شوند. در طرف مقابل، مشتریان راضی از خدمت‌رسانی، منبع بقا و سودآوری شرکت‌های بیمه هستند. به عبارت دیگر، وفاداری مشتریان به شرکت‌های بیمه معیار موفقیت در عصر حاضر است.

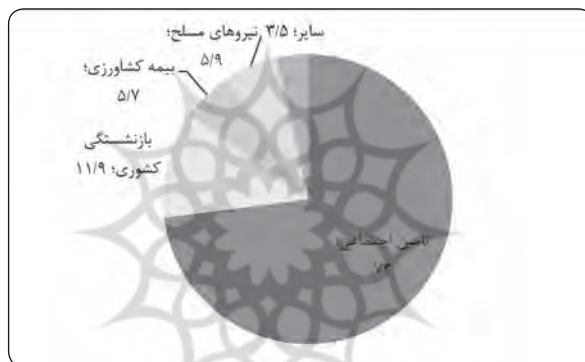
بیان مسئله

بیان مسئله در این پژوهش، شناسایی میزان رویگردانی بیمه‌شدگان و همچنین یافتن مدلی بر مبنای الگوریتم‌های هوش مصنوعی که میزان رویگردانی بیمه‌شدگان موضوع پژوهش در سال‌های آتی را به‌درستی تخمین زده و میزان از دست رفتن منابع درآمدی مربوط به آنها را محاسبه نماید. در ابتدا

مناسب است به تعدادی از چالش‌های موجود بر سر درآمد و هزینه‌ها در سازمان تأمین اجتماعی اشاره کنیم:

چالش‌های سازمان تأمین اجتماعی

در حال حاضر حدود ۱۸ صندوق بیمه‌ای و بازنشستگی در کشور وجود دارد. در این میان ۱۴ صندوق اختصاصی مانند صندوق بازنشستگی وزارت نفت، صدا و سیما، فولاد و بیمه مرکزی وجود دارد. از نظر وضعیت پوشش جمعیت و جمعیتی که در اقتصاد رسمی فعالیت می‌کنند، ۷۳ درصد این جمعیت تحت پوشش سازمان تأمین اجتماعی، ۱۱/۹ درصد تحت پوشش صندوق بیمه بازنشستگی کشوری، ۵/۷ درصد تحت پوشش صندوق بیمه اجتماعی کشاورزان، روستاییان و عشایر و ۵/۹ درصد نیز تحت پوشش بیمه نیروهای مسلح هستند. در شکل ۱ نمودار تحلیلی بیمه‌شدگان در کشور نشان داده شده است.



شکل ۱: نمودار تحلیلی بیمه‌شدگان در کشور

همان‌طور که در نمودار تحلیلی بیمه‌شدگان کشوری مشخص است، سازمان تأمین اجتماعی از بزرگ‌ترین سازمان بیمه اجتماعی کشور است که با توجه به تعداد زیاد بیمه‌شدگان، این سازمان با چالش‌های متعددی روبرو شده است. آشنایی با چالش‌های سازمان تأمین اجتماعی می‌تواند به محققان در بررسی رویگردانی بیمه‌شدگان یاری رساند. در ادامه به بررسی چالش‌های اساسی سازمان تأمین اجتماعی از دیدگاه معاون پیشین حقوقی و امور مجلس سازمان تأمین اجتماعی، محسن ایزدخواه پرداخته می‌شود.

پیشی گرفتن مصارف بر منابع

یکی از چالش‌های بزرگ و اساسی سازمان تأمین اجتماعی در چند سال اخیر، پیشی گرفتن مصارف بر منابع این سازمان است؛ بطوری‌که سازمان تأمین اجتماعی در بودجه سال ۱۳۹۵ خود اولاً برای تأمین این کسری مجوز دریافت وام از سیستم بانکی را به میزان بیست‌هزار میلیارد ریال از هیئت‌امنای سازمان اخذ کرده است؛ ثانیاً، با توجه به این‌که دولت بابت پرداخت دیون خود به این سازمان در

بودجه سالانه کل کشور رقمی نزدیک به ۱۶ هزار میلیارد ریال قرار داده، با نگاهی خوشبینانه و بدون توجه به سازوکارهای قانونی دولت و مجلس، سازمان در بودجه سال ۱۳۹۵ خود رقمی بالغ بر هزارمیلیارد ریال قرار داده است که تحقق آن امکان‌پذیر نشد. به‌عبارت دیگر، به صورت آشکار و پنهان سازمان تأمین اجتماعی در سال ۱۳۹۵ بین ۲۵ تا ۲۷ درصد کسری بودجه داشته و این به آن معناست که در سال‌های اخیر و پیش رو، سازمان تأمین اجتماعی برای انجام تعهدات قانونی خود با مشکل جدی مواجه خواهد بود. این امر باعث بروز تاخیر در پرداخت افزایش حقوق مستمری‌بگیران، تأخیر در پرداخت چند ماهه اسناد مربوط به حوزه درمان و دریافت وام‌های کلان از بانک‌های مختلف کشور برای تأمین هزینه‌های کمرشکن طرح تحول سلامت و پرداخت به بیمارستان‌ها و مراکز دولتی، به‌تأخیرانداختن پرداخت حقوق مستمری‌بگیران و انجام این تعهدات به آخرین روزهای هر ماه، وادار کردن شرکت سرمایه‌گذاری شستا و شرکت‌های زیرمجموعه برای اخذ وام و پرداخت آن به سازمان تأمین اجتماعی تحت عنوان سود شرکت‌ها از جمله اقدامات این سازمان در به‌تأخیرانداختن این تعهدات به دلیل کسری بودجه است.

هزینه‌های طرح تحول سلامت

طرح تحول سلامت با تمام مزایا و رضایت‌مندی‌هایی که در سطح کشور ایجاد کرده، به علت ضعف بنیان‌های کارشناسی در وزارت تعاون، کار و رفاه اجتماعی و همچنین صندوق‌های درگیر مثل سازمان تأمین اجتماعی و سازمان بیمه سلامت، منابع این سازمان‌ها را هدف قرار داده است. مطابق ماده ۲۹ قانون تأمین اجتماعی، ۹ درصد از مآخذ حق بیمه، به درمان اختصاص داده می‌شود. همچنین در اجرای تبصره ۲ ماده ۷ قانون بیمه همگانی، سازمان تأمین اجتماعی موظف است دونهم سهم درمان ناشی از دریافت حق بیمه را برای دوران مستمری‌گیری اختصاص دهد. با توجه به این نکات، سهم خالص درمان ناشی از دریافت حق بیمه در سال ۱۳۹۵ بالغ بر ۱۴ هزار و پانصد میلیارد تومان بوده است که با احتساب سایر درآمدهای حوزه درمان در صورت تحقق بالغ بر ۱۶ هزار میلیارد تومان است. این در حالی است که در بودجه سال ۱۳۹۵، سازمان تأمین اجتماعی بیش از ۱۸ هزار و نهصد میلیون تومان به درمان اختصاص داده شده که بر خلاف مقررات حاکم بر سازمان تأمین اجتماعی است و اگر این روند ادامه پیدا کند، در سال‌های آتی، حوزه درمان سایر منابع بلندمدت این سازمان را از بین خواهد برد.

تقابل با چالش‌های سازمان تأمین اجتماعی

همان‌طور که پیش‌تر به آن اشاره شده است، سازمان تأمین اجتماعی با دو چالش اساسی در حوزه درآمد و هزینه‌ها روبه‌روست. هر کدام از چالش‌های مطرح شده می‌توانند باعث ناکارآمدی سازمان مذکور شده که در نتیجه این عدم کارآمدی، مشتریان بیشتری از این سازمان رویگردان خواهند شد. در این راستا، این تحقیق بر آن است پیشنهاد زیر را که یکی از چند پیشنهاد مطرح است، برای مقابله با چالش حوزه درآمد و هزینه در سازمان تأمین اجتماعی مطرح کند.

● پیشنهاد: جذب و نگه‌داشت بیمه‌شدگان غیراجباری که حق بیمه را بدون کمک نهادهای حاکمیتی پرداخت می‌نمایند که از جمله مهم‌ترین آن‌ها می‌توان به بیمه‌شدگان مشاغل آزاد و اختیاری اشاره کرد. در راستای مقابله با چالش‌های موجود، می‌توان پیشنهادات مختلفی را در جهت بهبود وضعیت سازمان تامین اجتماعی ارائه نمود. در تعیین سیاست‌های سازمان تامین اجتماعی باید به این نکته توجه نمود که قسمت عمده نیروی کار کشور را نیروی جوان فعال تشکیل می‌دهد. از آنجاییکه این نیروها به عنوان منابع غنی تولید درآمد از محل پرداخت حق بیمه می‌باشند، سیاست‌گذاری در جهت جذب این نیروها امری بسیار مهم و ضروری می‌باشد. با توجه به نکات مطرح شده، جذب و نگهداشت بیمه‌شدگان غیراجباری و همچنین تصمیم‌گیری راهبردی در جهت ممانعت از رویگردانی بیمه‌شدگان موضوع مورد بحث در این پژوهش می‌باشد. به منظور دستیابی به این هدف مهم، می‌توان از تکنیک‌های داده‌کاوی به منظور کشف دانش مورد نیاز استفاده نمود.

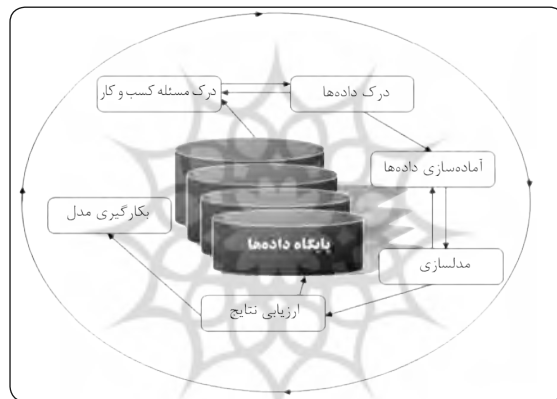
با توجه به پیشرفت‌های گسترده در حوزه فناوری اطلاعات، حجم داده‌های ذخیره شده در پایگاه داده شرکت‌های بیمه به سرعت در حال افزایش است. در واقع، پایگاه داده شرکت‌های بیمه دارای حجم بسیار گسترده‌ای از داده‌های بیمه‌شدگان هستند که این داده‌ها اطلاعات تجاری بالقوه و بازرشی دارند. کشف دانش^۱ از پایگاه داده شرکت‌های بیمه می‌تواند به این شرکت‌ها برای بهبود خدمات و تصمیم‌گیری‌های مدیریتی کمک بسیار برساند. یکی از روش‌های شناخته‌شده برای کشف دانش از داده‌های ذخیره‌شده در پایگاه داده شرکت‌های بیمه، روش داده‌کاوی^۲ است.

داده‌کاوی ابزار بسیار ارزشمندی است که در سال‌های اخیر به طور گسترده‌ای برای کشف دانش، جست‌وجوی روابط و الگوها بین حجم عظیمی از داده‌ها مورد استفاده قرار گرفته است. داده‌کاوی از ترکیب چند علم مختلف تشکیل شده است. آمار، یادگیری ماشین^۳، روش‌های بهینه‌سازی، روش‌های تشخیص و شناخت الگو، شبکه‌های عصبی^۴، مدل‌های ریاضی و هوش مصنوعی^۵ از جمله فنونی هستند که داده‌کاوی از آن‌ها بهره می‌برد. ترکستانی و دیگران (۱۳۹۵)

با توجه به این‌که موضوع کشف دانش از سیستم اطلاعات شرکت‌های بیمه مورد بحث است، در این پژوهش از استاندارد جهانی فرآیند داده‌کاوی در صنعت بهره می‌بریم. در شکل ۲، استاندارد جهانی فرآیند داده‌کاوی در صنعت نشان داده شده است. ترکستانی و (۱۳۹۵). همان‌طور که در شکل مشخص است، استاندارد جهانی فرآیند داده‌کاوی در صنعت از شش مرحله اصلی تشکیل شده است:

- 1 . Knowledge Discovery
- 2 . Data Mining
- 3 . Machine Learning
- 4 . Neural Networks
- 5 . Artificial Intelligence

مرحله اول مربوط به درک مسئله کسب و کار است. در این مرحله تلاش می‌شود مسئله مورد نظر در فضای کسب و کار شرکت بیمه، تعیین و مشخص شود. مرحله دوم مربوط به درک داده‌ها است. در این مرحله داده‌های مورد نیاز از پایگاه داده شرکت بیمه استخراج می‌شود. داده‌های استخراج شده یک مجموعه داده با تعداد مشخصی رکورد را تشکیل می‌دهند. مرحله سوم مربوط به آماده‌سازی داده‌ها است. در این مرحله عملیات‌های مختلفی همچون پاکسازی مجموعه داده، کمی‌سازی مجموعه داده و گسسته‌سازی مجموعه داده انجام می‌شود. مرحله چهارم مربوط به مدل‌سازی است. در این مرحله با استفاده از تکنیک‌های داده‌کاوی به کشف قواعد وابستگی در مجموعه داده پرداخته خواهد شد. مرحله پنجم، ارزیابی نتایج است. در این مرحله، قواعد استخراج شده بر اساس شاخص‌های مختلف فیلتر شده و برخی از مهم‌ترین قواعد جداسازی می‌شوند. در نهایت، در مرحله ششم بر اساس قواعد جداسازی شده، تصمیمات مدیریتی و سازمانی اعمال خواهد شد.



شکل ۲: استاندارد جهانی فرآیند داده‌کاوی در صنعت [۳].

پیشینه تحقیق

در این بخش، برخی از مهم‌ترین و جدیدترین کارهای انجام شده در زمینه کاربرد داده‌کاوی در صنعت بیمه بحث و بررسی خواهد شد. تحقیق‌های مورد بررسی در دو گروه تحقیقات داخلی و تحقیقات خارجی هستند.

یئو^۱ و همکارانش (۲۰۰۱) یک مدل کارآمد بر روی تغییرات حق بیمه برای بیمه اتومبیل مشتریان ارائه کردند. تحقیق ارائه شده مبتنی بر تکنیک‌های داده‌کاوی بوده است. این تحقیق دو فاز اصلی دارد. در فاز اول، مشتریان با استفاده از الگوریتم خوشه‌بندی K-Means، خوشه‌بندی می‌شوند. به منظور خوشه‌بندی مشتریان، از معیارهای جمعیت‌شناسی همچون سن راننده و جنسیت راننده استفاده

1. Yeo

شده است. در فاز دوم، هر خوشه با استفاده از شبکه‌های عصبی، مدل‌سازی شده و تغییرات حق بیمه پیش‌بینی شده است. نتایج پیاده‌سازی این تحقیق نشان داد که ترکیب الگوریتم خوشه‌بندی K-Means و شبکه‌های عصبی می‌تواند به ۷۸ درصد دقت پیش‌بینی دست یابد.

وو^۱ و همکارانش (۲۰۰۵) از تکنیک‌های داده‌کاوی برای کشف دانش در مجموعه داده بزرگ صنعت بیمه استفاده کرده‌اند. این تحقیق از شش مرحله آماده‌سازی داده، پیش‌پردازش داده، داده‌کاوی، تفسیر، کاربرد و ارزیابی تشکیل شده است. از این مطالعه می‌توان برای استخراج قوانین تصمیم‌گیری برای شرکت‌های بیمه استفاده کرد. قوانین تصمیم‌گیری می‌توانند برای تحقیق بر روی مشتریان بالقوه یک محصول بیمه‌ای استفاده شوند. این تحقیق پس از طراحی و پیاده‌سازی، بر روی مجموعه داده چند شرکت بیمه بزرگ اجرا شد. محققان از بازخورد تأثیر تحقیق انجام‌شده بر روی شرکت‌های بیمه برای ارزیابی تحقیق ارائه شده، استفاده کردند. نتایج به‌دست‌آمده نشان داد که این تحقیق می‌تواند به ۸۰ درصد دقت تشخیص دست یابد.

دلن^۲ و همکارانش (۲۰۰۹) با استفاده از تکنیک‌های داده‌کاوی، پوشش سیستم سلامت در امریکا را بررسی کردند. تناقض‌های موجود در پوشش سیستم سلامت در امریکا موضوعی بسیار مهم و کلیدی برای سیاست‌گذاران است. متأسفانه افراد بسیار زیادی در امریکا تحت پوشش سیستم سلامت قرار ندارند و از این‌رو محققان تلاش نموده‌اند تا فاکتورهای مهم در بروز این پدیده را شناسایی نمایند. به عبارت دیگر، در این تحقیق با استفاده از تکنیک‌های داده‌کاوی تلاش شده است تا فاکتورهای مؤثر در بروز این پدیده شناسایی شود. به منظور تشخیص فاکتورهای مؤثر، از تکنیک شبکه عصبی مصنوعی و تکنیک درخت تصمیم استفاده شده است. نتایج شبیه‌سازی نشان داد که تحقیق ارائه‌شده مبتنی بر شبکه عصبی کارایی بالاتری دارد و توانایی تشخیص صحیح ۷۸ درصد از مجموعه داده تست را دارد. همچنین، تحقیق پیشنهادی تشخیص داد که سطح درآمد، موقعیت شغلی، تحصیلات و وضعیت تأهل مهم‌ترین فاکتورهای بروز این پدیده هستند. ایراد اصلی این تحقیق آن است که با افزایش تعداد فاکتورهای ورودی به شبکه عصبی هوشمند، زمان اجرای الگوریتم به صورت نمایی افزایش می‌یابد.

توکلی و همکارانش (۱۳۸۹) مقاله‌ای با عنوان «به‌کارگیری فرآیند داده‌کاوی برای پیش‌بینی الگوهای رویگردانی مشتری در بیمه» ارائه کردند. در این مقاله عنوان شده است که با اعمال فرآیند داده‌کاوی در مقام یکی از فناوری‌های مدیریت دانش می‌توان طی کاوشی در پایگاه‌های داده حجیم، الگوهایی را برای پیش‌بینی رفتار رویگردانی مشتریان استخراج کرد. الگوهای استخراج‌شده می‌توانند به مدیران

1 . Wu
2 . Delen

برای اخذ تصمیمات لازم یاری برساند. در این مقاله، با بهره‌گیری از تکنیک‌های مختلف داده‌کاوی به کاوش در پایگاه‌های داده یکی از شرکت‌های سهامی عام بیمه‌ای در رشته بیمه آتش‌سوزی پرداخته شده است. نتایج به‌دست‌آمده نشان داد ویژگی‌های کانال جذب مشتری، سابقه خرید و کاربری و مکان بیمه‌شده، عوامل تأثیرگذار بر پیش‌بینی رویگردانی هستند.

از آنجا که در دهه‌های اخیر در حوزه مدیریت بازاریابی، سیاست‌ها از رویکرد محصول محور به مشتری محور تغییر یافته است، در این تحقیق به جای تمرکز بر ویژگی‌های محصول و سنجش و تحلیل آن، بر روی مشتری و ویژگی‌های مربوط به مشتری تمرکز شده است. در واقع، با سنجش و تحلیل ویژگی‌های مربوط به خدمت‌گیرنده (مشتری)، این تحقیق انجام گرفته است. خط‌مشی مشتری‌مداری و تمرکز بر ویژگی‌های مشتری در این تحقیق بدین صورت انتخاب شده است که در بسیاری از کسب‌وکارهای فعال در بازار رقابتی، بقای جریان و تداوم سود، وابسته به مشتری است. همان‌طور که قبلاً اشاره شد، پیش‌بینی رفتار مشتری برای به‌اجرا در آوردن هر یک از راهبردهای مدیریت ارتباط با مشتری از جمله رویگردانی و وفاداری منوط به جمع‌آوری اطلاعات مشتری و تجزیه و تحلیل الگوهای رفتار اوست. در این تحقیق، داده‌های در نظر گرفته شده برای یک برش ۱۸ ماهه است. به نظر می‌رسد انتخاب بازه زمانی ۱۸ ماهه محدود و چندان مناسب نیست. انتخاب داده‌ها مربوط به برش زمانی کوتاه، احتمالاً به دلایل ذیل انجام شده است:

- داده‌های مربوط به دوره طولانی‌تر وجود نداشته یا در اختیار پژوهشگر نبوده است. در واقع، دسترسی پژوهشگر به داده‌ها محدود بوده است.
- به جهت سبک‌تر و روان‌تر بودن پردازش‌ها و دستیابی به سرعت بالاتر اجرای الگوریتم، محدوده زمانی ۱۸ ماهه از اطلاعات انتخاب شده است.

انتخاب داده مربوط به دوره زمانی کوتاه مدت ۱۸ ماهه از جمله نقاط ضعف این تحقیق به حساب می‌آید، زیرا هرگونه تغییر در خط‌مشی فروش شرکت در طول زمان، الگوی به‌دست‌آمده در این مقاله را نامعتبر می‌کند. در این صورت نیاز به اجرای مجدد و مکرر فرآیند داده‌کاوی برای داده‌های بعدی خواهد بود. از جمله مزیت‌های اصلی این تحقیق، انتخاب مناسب ویژگی‌ها در مجموعه داده مورد نظر می‌باشد. در واقع، در این مقاله تعداد ۳۶ ویژگی مختلف شناسایی شده است که ۳ ویژگی کلیدی از بین آن‌ها که بیشترین تأثیر را داشته انتخاب شده‌اند.

بروفر و همکارانش (۱۳۹۵) مقاله‌ای با عنوان «شناسایی الگوی رفتاری مشتریان در بیمه عمر و تشکیل سرمایه با استفاده از داده‌کاوی» ارائه نمودند. هدف اصلی در این مقاله ارائه یک مدل مناسب و کارآمد جهت بخش‌بندی مشتریان بر اساس برخی از مهم‌ترین ویژگی‌های مالی و جمعیت‌شناسی در قالب عوامل موثر بر شاخص‌های ارزش دوره عمر مشتری (آ.ا.ف.ام) می‌باشد. ارزش دوره عمر

مشتری، ارزشی است که مشتری در طول عمر خود برای سازمان ایجاد می‌نماید. این مفهوم، علاوه بر ارزش فعلی مشتری به ارزش بالقوه و آتی مشتری نیز اشاره دارد و هدف اصلی از محاسبه آن، ایجاد یک برداشت وزنی از مشتریان به منظور تخصیص بهینه منابع به آن‌هاست. پس از تعیین مقادیر شاخص‌های مدل آ.راف.ام شامل تازگی مبادله، تعداد دفعات مبادله و ارزش پولی مبادله از ۱۸۰۰۰۰ مشتری و وزن‌دهی آن‌ها با استفاده از فرآیند تحلیل سلسله مراتبی و همچنین تعیین تعداد خوشه بهینه بر اساس شاخص سیلوئت و تعیین نرخ تاثیر شاخص‌های آ.راف.ام، خوشه‌بندی مشتریان با استفاده از الگوریتم کارآمد خوشه‌بندی K-Means انجام شد.

روش

انگیزه اصلی در این پژوهش، کشف دانش از سیستم اطلاعات سازمان تأمین اجتماعی مبتنی بر رویکرد داده‌کاوی می‌باشد. در واقع، انگیزه اصلی در این پژوهش، پیش‌بینی وفاداری و رویگردانی بیمه‌شدگان خویش‌فرمای سازمان تأمین اجتماعی مبتنی بر رویکرد داده‌کاوی است. روش پیشنهادی که به منظور پیش‌بینی وفاداری و رویگردانی بیمه‌شدگان خویش‌فرمای سازمان تأمین اجتماعی طراحی می‌شود، از پنج مرحله اصلی تشکیل خواهد شد. در شکل ۳ مراحل اصلی روش پیشنهادی نشان داده شده است.



شکل ۳: مراحل اصلی روش پیشنهادی

همان‌طور که در شکل مشخص است، مرحله اول مربوط به جمع‌آوری مجموعه داده می‌باشد. در این مرحله، تعداد مشخصی رکورد از بیمه‌شدگان سازمان تأمین اجتماعی در قالب یک مجموعه داده جمع‌آوری

می‌شوند. مرحله دوم مربوط به شناسایی ویژگی‌ها در مجموعه داده جمع‌آوری شده، می‌باشد. در این مرحله می‌بایست ویژگی‌های مورد نیاز را یافته و همچنین مقادیر مجاز برای هر کدام از ویژگی‌ها نیز تعیین و مشخص شوند. مرحله سوم مربوط به استانداردسازی مجموعه داده می‌باشد. در این مرحله، عملیات مختلفی همچون کمی‌سازی، گسسته‌سازی و پاکسازی بر روی مجموعه داده مورد نظر اعمال و اجرا می‌گردند. در مرحله چهارم، تکنیک خوشه‌بندی اعمال می‌گردد. در این مرحله، مجموعه داده استاندارد بدست آمده توسط تکنیک خوشه‌بندی، به خوشه‌های مختلفی تقسیم‌بندی می‌شود. در واقع، داده‌هایی که دارای مقادیر نزدیک به یکدیگر هستند، در یک خوشه قرار گرفته و از سایر داده‌ها جدا و تفکیک می‌شوند. مرحله پنجم مربوط به استخراج قواعد است. در این مرحله با استفاده از الگوریتم‌های استخراج قواعد وابستگی، تمامی قواعد مورد نیاز در هر خوشه استخراج خواهد شد.

با توجه به مراحل مطرح‌شده در شکل ۳، در این بخش تصمیم داریم الگوهای را از مجموعه داده موجود که شامل ۲۷ ویژگی و ۲۱۴۰۶ نمونه است، استخراج نموده، سپس با استفاده از الگوهای بدست‌آمده، پارامترها و ویژگی‌های مؤثر در وفاداری و رویگردانی مشتریان را بررسی و تحلیل کنیم.

پایاده‌سازی الگوریتم NSGA-II انتخاب ویژگی‌ها

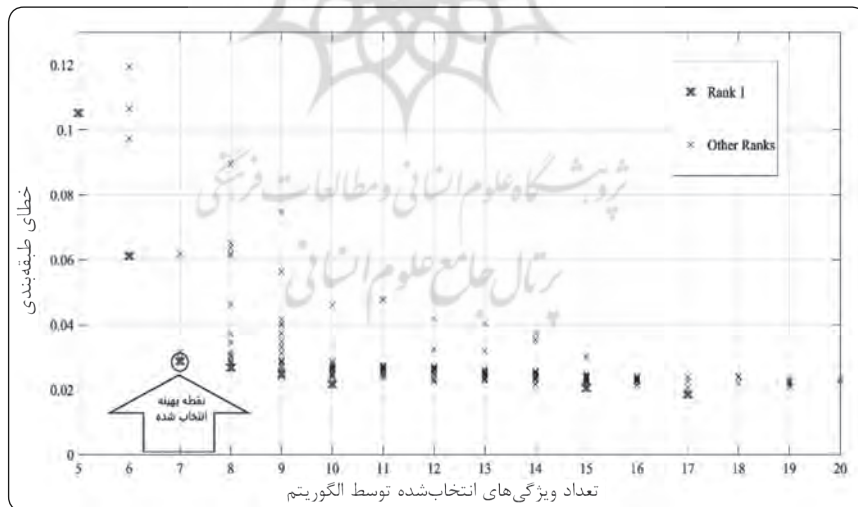
همان‌طور که بیان شد، الگوریتم ژنتیک استفاده شده در این تحقیق دو هدفه تعریف شده است که یکی از اهداف تعداد ویژگی‌های انتخاب‌شده به عنوان یک زیرمجموعه از ویژگی‌های اصلی و دیگری درصد صحت طبقه‌بندی شبکه عصبی به ازای این زیرمجموعه ویژگی انتخاب‌شده است. در طول الگوریتم زیرمجموعه‌هایی از مجموعه ویژگی‌های ورودی به صورت تصادفی انتخاب می‌گردد، تعداد این حدس‌های تصادفی برابر با اندازه جمعیت در یک تکرار الگوریتم ژنتیک است. به ازای هر یک از این زیرمجموعه‌های تصادفی دو معیار تعداد ویژگی و صحت طبقه‌بندی شبکه عصبی به ازای این زیرمجموعه به دست می‌آید. پس از آن الگوریتم ژنتیک با دو معیار غلبه و فاصله ازدحامی این پاسخ‌ها را رتبه‌بندی می‌کند. در این حالت، ماتریسی از پاسخ‌ها موجود است که یک بعد آن به طول اندازه جمعیت است و بعد دیگر آن به طول تعداد اهداف که در این تحقیق برابر با ۲ می‌باشد. به تعداد تکرار مشخص شده الگوریتم ژنتیک جواب‌های نامغلوب را انتخاب و جواب‌های مغلوب را حذف می‌کند تا به مجموعه جواب‌های بهینه دست یابد.

الگوریتم ژنتیک پایاده‌سازی شده در این تحقیق با سه تنظیم مختلف اجرا شده و نتایج این سه اجرای مختلف در کنار هم قرار گرفته است و دوباره با دو معیار غلبه و فاصله ازدحامی مرتب شده است. در جدول ۲ تنظیمات مربوط به این سه اجرا آورده شده است، همین‌طور در شکل ۲-۳ پاسخ‌های خروجی الگوریتم انتخاب ویژگی نمایش داده شده است که پاسخ‌های مرتبه اول با ضربدرهای قرمز رنگ و بقیه پاسخ‌ها با ضربدرهای مشکی نمایش داده شده است.

جدول ۲: تنظیمات مربوط به الگوریتم NSGA-II

درصد جهش	درصد تقاطع	اندازه جمعیت	تعداد تکرار
۰/۱	۰/۹	۲۰	۴۰ اجرای شماره ۱
۰/۴	۰/۶	۲۰۰	۱۰۰ اجرای شماره ۲
۰/۲	۰/۸	۱۰۰	۲۰۰ اجرای شماره ۳

همان‌طور که بیان شد، ضربدرهای پررنگ پاسخ‌های مرتبه اول هستند. این پاسخ‌ها جواب‌های بهینه الگوریتم NSGA-II می‌باشند. در این شکل خطای طبقه‌بندی بر حسب تعداد ویژگی انتخابی رسم شده است. باید از بین پاسخ‌های مرتبه اول پاسخی که طبق مسئله ما بهینه‌ترین حالت است را انتخاب کرد. در شکل ۴، پاسخی که از نظر نویسنده، به عنوان بهینه‌ترین حالت در بین پاسخ‌های مرتبه اول می‌باشد، انتخاب شده است. این پاسخ یک زیرمجموعه ۷ عضوی از ویژگی‌های ورودی می‌باشد که خطای طبقه‌بند در حدود ۳/۴۷ درصد دارد. علت انتخاب این نقطه به عنوان حالت بهینه، این است که پاسخ مرتبه اول قبلی آن که یک مجموعه ۶ عضوی است، خطایی حدود ۶ درصد دارد که این خطا از نظر مقداری زیاد است. نقطه بهینه انتخاب‌شده نسبت به نقطه بعدی که ۸ عضو دارد تقریباً خطای یکسان دارد؛ در نتیجه نقطه انتخاب‌شده، نقطه بهینه است. مشخصات دقیق نقطه انتخاب‌شده در ادامه بیان شده است.



شکل ۴: پاسخ بهینه انتخاب شده از بین پاسخ‌های الگوریتم NSGA-II

ویژگی‌های انتخاب‌شده شکل ۴، در جدول ۳ نمایش داده شده است.

جدول ۳: ویژگی‌های استخراج شده از بانکهای اطلاعاتی و ویژگی‌های انتخاب شده توسط الگوریتم NSGA-II

ویژگی‌های موجود	ویژگی‌های انتخاب شده توسط الگوریتم ژنتیک
۱ تاریخ درخواست	
۲ نوع بیمه	*
۳ سال تولد بیمه شده	*
۴ نرخ حق بیمه	
۵ دستمزد روزانه قرارداد	
۶ وضعیت درمان در قرارداد	
۷ تاریخ قرارداد	*
۸ محل سکونت (شهر)	
۹ وضعیت سلامت	
۱۰ وضعیت اخذ SSN	
۱۱ جنسیت	*
۱۲ ملیت	
۱۳ نحوه شناسایی بیمه شده	
۱۴ جمعیت شهر محل تولد	
۱۵ کاربر ثبت درخواست قرارداد	
۱۶ تعداد روز سابقه قبل از اخذ قرارداد	
۱۷ تعداد روز سابقه ناشی از قرارداد	*
۱۸ وضعیت تاهل	
۱۹ تعداد اولاد	
۲۰ وضعیت کفالت والدین (تعداد تحت کفالت)	
۲۱ تعداد سال بعد از ازدواج	
۲۲ سن بزرگ‌ترین فرزند	*
۲۳ سال شروع قرارداد	*
۲۴ سن مشتری	
۲۵ فاصله ثبت درخواست تا انعقاد قرارداد (روز)	
۲۶ دستمزد شخص	
۲۷ انعقاد قرارداد نسبت به زمان تاهل (قبل/بعد)	

پیاده‌سازی الگوریتم‌های طبقه‌بندی ویژگی‌ها

داده‌های ورودی با در نظر گرفتن ۷ ویژگی به‌دست‌آمده در قسمت قبل به عنوان ورودی به سه طبقه‌بند شبکه عصبی چندلایه پرسپترون، ماشین بردار پشتیبان و K نزدیک‌ترین همسایگی داده شده و نتایج طبقه‌بندی هر کدام به دست آمد. در ضمن داده‌های ورودی به طبقه‌بندها پیش‌تر نرمال شده است. ترتیب داده‌های آموزش و آزمایش به نحوی است که به ترتیب ۸۰ درصد داده‌ها برای آموزش و ۲۰ درصد آن‌ها برای آزمایش مورد استفاده قرار گرفته است. نتایج هر یک از طبقه‌بندها برای هر یک از مجموعه‌های داده‌های آموزش و آزمایش در ادامه آمده است.

شبکه عصبی چندلایه پرسپترون

شبکه عصبی استفاده‌شده در این تحقیق از نوع پرسپترون چندلایه است که تعداد نورون‌های آن در لایه پنهان با یک بررسی اولیه ۷ نورون در نظر گرفته شده است. (مطابق شکل ۵) در این بررسی اولیه میزان صحت طبقه‌بندی میانگین داده‌های آزمایش به ازای تعداد نورون‌های مختلفی که در لایه پنهان می‌باشد محاسبه شده است و تعداد نورونی که بیشترین صحت طبقه‌بندی را ایجاد می‌کند در نظر گرفته شده است. در شکل ۶ نمودار میانگین صحت طبقه‌بندی بر حسب تعداد نورون‌های لایه پنهان رسم شده است.



شکل ۵: شبکه عصبی MLP استفاده شده جهت طبقه‌بندی



شکل ۶: میانگین صحت طبقه‌بندی شبکه عصبی MLP بر حسب تعداد نورون‌های لایه پنهان

مطابق شکل ۶ تعداد ۷ نورون در لایه مخفی بیشترین صحت عملکرد را می‌تواند ایجاد کند. در جدول ۴ و شکل ۵ بقیه مشخصات این شبکه بیان شده است.

جدول ۴: مشخصات شبکه عصبی MLP

تعداد نورون در لایه پنهان	۷
تابع کارایی	'mse'
الگوریتم آموزش	'trainlm'

رویدان	وفادار	رویدان
۵۷	۱۷۵۰	وفادار
۲۴۳۱	۴۳	رویدان

شکل ۷: ماتریس درهم‌ریختگی مربوط به داده‌های آزمایش در طبقه‌بندی به وسیله شبکه عصبی MLP

در ادامه نتایج مربوط به طبقه‌بندی داده‌های با استفاده از شبکه عصبی آورده شده است. در شکل ۷ ماتریس درهم‌ریختگی^۱ نتایج مربوط به داده‌های آزمایش ارائه شده است؛ همچنین در جدول ۵ نتایج مربوط به این طبقه‌بندی ارائه شده است.

جدول ۵: نتایج طبقه‌بندی به وسیله شبکه عصبی MLP

	Accu- Train (%)	Accura- Test (%)	(%) Sensitivity	(%) Specificity	(%) Precision
وفاداری	۹۶,۰۹	۹۶,۸۴	۹۷,۶	۹۸,۲۶	۹۶,۸۴
رویدگدانی	۹۷,۵	۹۸,۲۶	۹۷,۷	۹۶,۸۴	۹۸,۲۶

همچنین شبکه عصبی با تعداد لایه‌های مخفی ۲ و ۴ و ۶ و ۸ و ۱۰ و هر بار با نورون‌های متفاوت از ۱ تا ۱۰، اجرا گردید که بهترین نتایج حاصل از آن در جدول ۶ منعکس شده است. با توجه به شکل ۵ که بیانگر نتایج شبکه عصبی تک لایه است و نتایج حاصله از اجرای شبکه عصبی دو تا ده لایه، که در جدول ۶ نمایش داده شده، نتیجه گرفته می‌شود که بهترین گزینه برای داده‌های ما، انتخاب شبکه عصبی تک لایه و با ۷ نورون است. در شبکه عصبی تک لایه به دقت ۹۷,۵۵ درصد برای داده‌های آزمایش دست یافتیم.

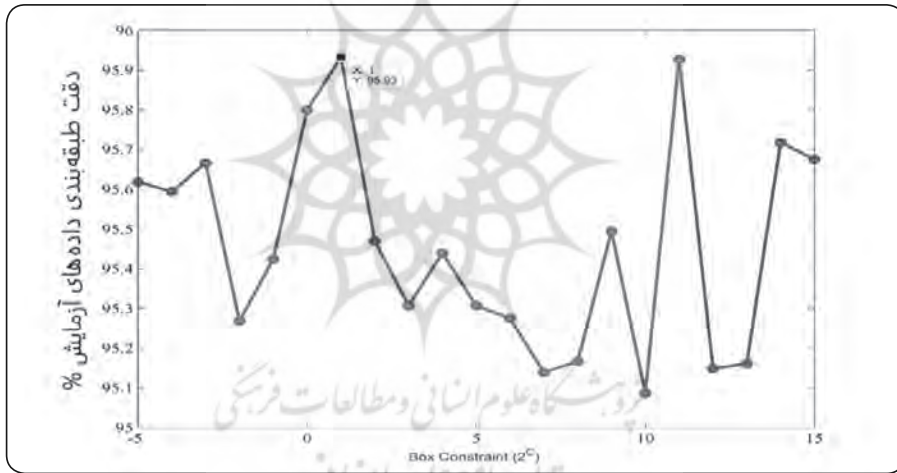
1 Confusion Matrix

جدول ۶: نتایج درصد دقت طبقه‌بندی به وسیله شبکه عصبی MLP با لایه‌های مخفی متغیر

	تعداد لایه پنهان : ۲ تعداد نورون: ۸	تعداد لایه پنهان : ۴ تعداد نورون: ۳	تعداد لایه پنهان : ۶ تعداد نورون: ۴	تعداد لایه پنهان : ۸ تعداد نورون: ۴	تعداد لایه پنهان : ۱۰ تعداد نورون: ۷
درصد دقت طبقه‌بندی داده‌های آزمایش	۹۷,۴۴	۹۷,۴۹	۹۷,۴۷	۹۷,۲۹	۹۷,۱۱

ماشین بردار پشتیبان

در این قسمت داده‌های ورودی با یک طبقه‌بند دیگر به نام «ماشین بردار» پشتیبان طبقه‌بندی می‌شود. جهت پیاده‌سازی Grid Search ابتدا کرنل خطی را در نظر گرفته و به ازای ضرایب Box constraint در بازه $[2^{-5}, 2^{15}]$ مقادیر صحت طبقه‌بندی SVM را بدست می‌آوریم. در شکل ۸ مقادیر صحت طبقه‌بندی SVM به ازای کرنل خطی با ضریب Box constraint در بازه بیان شده رسم شده است. بیشترین مقدار صحت طبقه‌بندی زمانی رخ می‌دهد که ضریب مربوطه در کرنل خطی برابر با ۲ باشد.

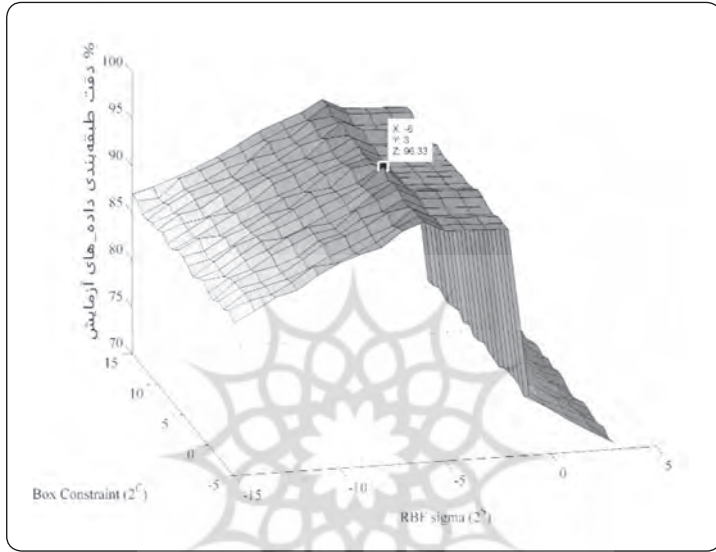


شکل ۸: بررسی مقادیر صحت طبقه‌بندی SVM با کرنل خطی در مقادیر مختلف Box constraint

حال به سراغ کرنل گوسی رفته و به ازای ضرایب مربوط به آن، صحت طبقه‌بندی SVM را محاسبه می‌کنیم. همان‌طور که می‌دانید کرنل گوسی دارای دو ضریب box constraint و RBF sigma می‌باشد، که باید با استفاده از روش Grid Search و تغییر این دو ضریب به ازای تمام حالات ممکن صحت طبقه‌بند بدست آید. بنابراین به ازای ضرایب Box constraint در بازه $[2^{-5}, 2^{15}]$ و ضریب RBF sigma در بازه $2^{-15}, 2^3$ مقادیر صحت طبقه‌بندی SVM را بدست می‌آوریم. در شکل ۸ مقادیر صحت طبقه‌بندی SVM به ازای کرنل گوسی با ضریب Box constraint در بازه بیان شده و ضریب RBF sigma در بازه بیان شده رسم شده است. بیشترین مقدار صحت طبقه‌بندی زمانی رخ می‌دهد که

ضرایب مربوطه در کرنل گوسی به ترتیب برابر با ۸ و ۱۵۶۲۵ و ۰٪ باشند.

بررسی دو کرنل با ضرایب مربوطه آن نشان می‌دهد که کرنل گوسی با ضرایب مربوطه به آن عملکرد مناسب‌تری نسبت به دیگر حالت موجود دارد و می‌تواند میانگین صحت عملکرد در حدود ۹۶/۳۳ درصد را ایجاد کند. در شکل ۱۰ و جدول ۷ نتایج مربوط به طبقه‌بندی از طریق ماشین بردار پشتیبان بیان شده است.



شکل ۹: بررسی مقادیر صحت طبقه‌بندی SVM با کرنل گوسی

	وفادار	رویگردان
وفادار	۱۷۲۳	۸۷
رویگردان	۶۳	۲۴۰۶

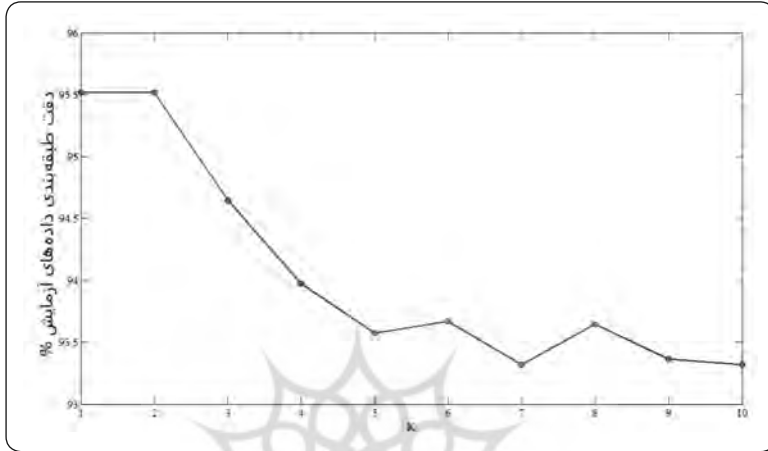
شکل ۱۰: ماتریس درهم‌ریختگی داده‌های آزمایش در ماشین بردار پشتیبان SVM

جدول ۷: نتایج طبقه‌بندی به وسیله ماشین بردار پشتیبان SVM

	Accuracy Train (%)	Accuracy Test (%)	(%) Sensitivity	(%) Specificity	(%) Precision
وفاداری	۹۵,۸۶	۹۵,۱۹	۹۵,۱۹	۹۷,۴۵	۹۶,۴۷
رویگردانی	۹۷,۳	۹۷,۴۵	۹۷,۴۵	۹۵,۱۹	۹۶,۵۱

K نزدیک ترین همسایگی

از روش های دیگری که در این تحقیق به عنوان روش طبقه بندی استفاده شد، روش KNN می باشد. در شکل ۱۱ صحت عملکرد متوسط به ازای تعداد همسایگی های مختلف رسم شده است، با توجه به این نمودار بیشترین صحت عملکرد زمانی رخ می دهد که ۱ یا ۲ همسایگی داشته باشیم.



شکل ۱۱: بررسی مقادیر صحت طبقه بندی KNN به ازای مقادیر مختلف تعداد همسایگی k

در ادامه نتایج مربوط به پیاده سازی روش KNN به ازای یک همسایگی آورده شده است، در شکل ۱۲ ماتریس درهم ریختگی مربوط به داده های آزمایش و همچنین در جدول ۸ نتایج این پیاده سازی آورده شده است.

	وفادار	روبگردان
وفادار	۱۷۱۳	۹۷
روبگردان	۸۹	۲۳۸۲

شکل ۱۲: ماتریس درهم ریختگی داده های آزمایش در طبقه بندی K نزدیک ترین همسایگی

جدول ۸: نتایج طبقه بندی به وسیله K نزدیک ترین همسایگی (KNN)

	Accura- Train (%) cy	Accura- Test (%) cy	(%) Sensitivity	(%) Specificity	(%) Precision
وفاداری	۱۰۰	۹۴,۶۴	۹۴,۶۴	۹۶,۳۹	۹۵,۰۶
روبگردانی	۱۰۰	۹۶,۳۹	۹۶,۳۹	۹۴,۶۴	۹۶,۰۸

مقایسه طبقه‌بندها

جهت مقایسه این سه طبقه‌بند و انتخاب بهترین آن‌ها، از میانگین درصد صحت طبقه‌بندی برای داده‌های آزمایش بین دو کلاس استفاده شده است. در جدول ۹ نتایج این مقایسه آورده شده است. طبق اطلاعات این جدول مقایسه‌ای، شبکه عصبی با درصد صحت طبقه‌بندی میانگین ۹۷/۵۵ درصد بیشترین توانایی را در طبقه‌بندی داده‌ها دارد.

جدول ۹: نتایج مقایسه طبقه‌بندها

	NN	SVM	KNN
Accuracy (%)	۹۷,۵۵	۹۶,۳۲	۹۵,۵۲

بررسی نتایج و جمع بندی

هدف اصلی در این پژوهش استخراج دانش از مجموعه داده بیمه‌شدگان خویش‌فرمای سازمان تأمین اجتماعی به منظور پیش‌بینی وفاداری و رویگردانی بیمه‌شدگان بود. برای دستیابی به هدف اصلی این پژوهش روش‌های انتخاب و طبقه‌بندی ویژگی‌ها مورد استفاده قرار گرفت. در قسمت انتخاب ویژگی‌ها سعی بر آن شد از بین ۲۷ ویژگی اولیه که از مشتریان به دست آمده بود، ویژگی‌های مؤثرتر با استفاده از الگوریتم ژنتیک چندهدفه (NSGA-II) به دست آورده شود. نتیجه به دست آمده در این قسمت به این صورت بود که از بین ویژگی‌های اولیه تعداد ۷ ویژگی انتخاب شد که این ویژگی‌ها عبارتند از: ۱. نوع بیمه افراد، ۲. سال تولد، ۳. تاریخ درخواست بیمه، ۴. جنسیت، ۵. سابقه قرارداد بیمه، ۶. تاریخ تولد بزرگ‌ترین فرزند، ۷. سال قرارداد بیمه.

بررسی صحت توزیع کلاس‌ها براساس ویژگی‌های انتخاب‌شده

پس از انجام و پیاده‌سازی عملیات طبقه‌بندی در بخش قبل و انتخاب شبکه عصبی به عنوان مناسب‌ترین طبقه‌بندی که بهترین پاسخگویی را با شرایط این مسئله داشته است، لازم است این موضوع بررسی شود که آیا می‌توان از ۷ ویژگی انتخاب‌شده برای نمونه‌برداری‌های بعدی و پیش‌بینی میزان وفاداری و رویگردانی مشتریان استناد کرد یا خیر؟

با توجه به توصیف ارائه شده از مقوله انتخاب ویژگی و همین‌طور حفظ قابلیت طبقه‌بندی ویژگی‌های انتخاب شده در مقایسه با همه ویژگی‌ها، در ابتدا همه داده‌های مسئله که به تعداد ۲۱۴۰۶ رکورد و در قالب ۲۷ ویژگی دسته‌بندی شده‌اند، بوسیله شبکه عصبی طبقه‌بندی می‌نماییم و دقت طبقه‌بندی آن را بدست می‌آوریم. از طرفی نتیجه حاصله از شبکه عصبی که بر روی همان داده‌ها و در قالب ۷ ویژگی انتخابی انجام شده را با نتیجه قبلی مقایسه می‌نماییم.

از تحلیل ماتریس درهم‌ریختگی شکل ۱۳ موارد ذیل قابل برداشت می‌باشد:

- از تعداد ۹۱۳۹ مشتری رویگردان: شبکه عصبی توانسته تعداد ۸۷۹۹ نفر را به درستی تشخیص داده و تعداد ۱۸۴ نفر را به اشتباه، وفادار تشخیص داده است.
 - از تعداد ۱۲۲۶۷ مشتری وفادار: شبکه عصبی توانسته تعداد ۸۳ و ۱۲۰ نفر را به درستی تشخیص داده و تعداد ۳۴۰ نفر به اشتباه، رویگردان تشخیص داده است.
 - مجموعاً با دقت ۹۷,۶ درصد، در تشخیص کلاسهای موجود، موفق بوده و ۲,۴ درصد از جمعیت مشتریان را نتوانسته در کلاس واقعی خود، پیش‌بینی نماید.
- در ادامه، نتایج حاصل از طبقه‌بندی همه داده‌ها در قالب همه ویژگی‌ها و همچنین ۷ ویژگی انتخابی در جدول ۱۰ آمده است:

جدول ۱۰: مقایسه طبقه‌بندی داده‌ها در دو قالب (همه ویژگی‌ها با ویژگی‌های انتخابی)

درصد صحت طبقه‌بندی داده‌های آزمایش	
۹۷,۶	همه ویژگی‌ها (۲۷ ویژگی)
۹۶,۸	ویژگی‌های انتخاب شده (۷ ویژگی)

کلاس واقعی			
	رویگردان	وفادار	
رویگردان	۸۷۹۹ ٪۴۱,۱	۱۸۴ ٪۰,۹	
وفادار	۳۴۰ ٪۱,۶	۱۲۰۸۳ ٪۵۶,۴۶	
			٪۹۷,۶ ٪۲,۴

جدول ۱۱: ماتریس درهم‌ریختگی طبقه‌بندی شبکه عصبی براساس همه ۲۷ ویژگی موجود

با توجه به نزدیکی اعداد حاصل نتیجه طبقه‌بندی برای داده‌های آزمایش مطابق جدول ۱۰، این نتیجه حاصل می‌شود که انتخاب ویژگی‌ها به درستی انجام شده و جهت پیش‌بینی وفاداری و رویگردانی مشتریان، می‌توان به جای اندازه‌گیری و تحلیل ۲۷ ویژگی، از ۷ ویژگی انتخاب شده استفاده نمود. همچنین جهت کاربردهای بعدی، اندازه‌گیری و جمع‌آوری اطلاعات در قالب این ۷ ویژگی جهت کلاس‌بندی وفادار یا رویگردان مشتریان کفایت می‌نماید.

بررسی صحت طبقه‌بندی و تفکیک دو گروه مشتریان وفادار و رویگردان

در قسمت طبقه‌بندی ویژگی‌ها سه طبقه‌بند مختلف از جمله شبکه عصبی چندلایه پرسپترون،

ماشین بردار پشتیبان و K نزدیک‌ترین همسایگی جهت طبقه‌بندی مورد استفاده قرار گرفت که طبق نتایج بدست آمده در بخش قبل شبکه عصبی چندلایه پرسپترون با صحت طبقه‌بندی میانگین ۹۷/۵۵ درصد بیشترین توانایی را در طبقه‌بندی داده‌ها داشت.

بنابراین با مراجعه به جدول ۵ مشاهده می‌شود که برای داده‌های آزمایش، الگوریتم طبقه‌بندی می‌تواند مشتریان وفادار را با صحت ۹۶/۸۴ درصد و مشتریان رویگردان را با صحت ۹۸/۲۶ درصد تشخیص دهد. به طور متوسط می‌توان بیان کرد که الگوریتم طبقه‌بندی مورد نظر می‌تواند با صحت ۹۷/۵۵، مجموعه داده بیمه‌شدگان خویش‌فرمای سازمان تأمین اجتماعی را تفکیک نماید و یا با همین صحت می‌تواند پیش‌بینی وفاداری و رویگردانی بیمه‌شدگان در آینده را انجام دهد.

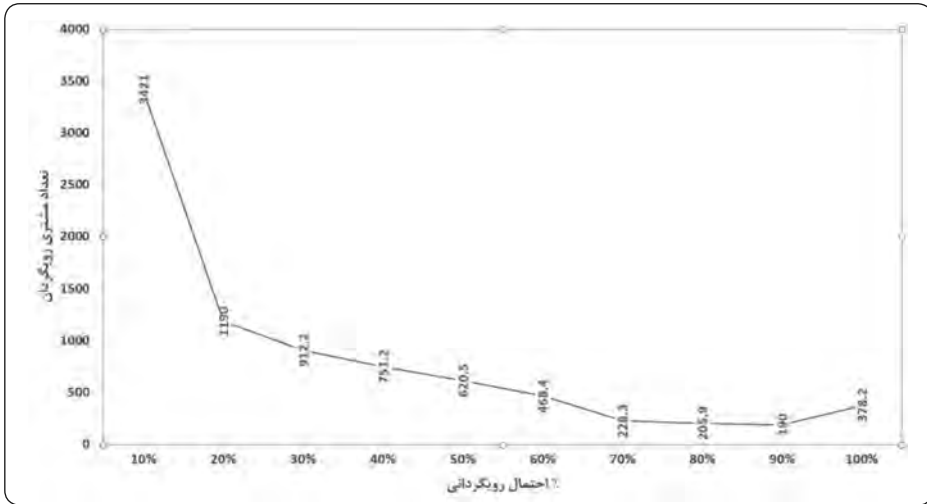
با توجه به این نکته که داده‌های آموزش و آزمایش کاملاً از یکدیگر جدا بوده است، می‌توان بیان کرد که اگر داده جدیدی تولید شود، سیستم پیش‌بینی می‌تواند وضعیت وفاداری و یا رویگردانی داده جدید را با همین میانگین صحت ۹۷/۵۵ درصد مشخص کند. برای پیش‌بینی سال‌های آینده هم می‌توان از این شبکه عصبی آموزش دیده استفاده شود. با داشتن یک شمای کلی از مشتریان آینده می‌توان پیش‌بینی مناسبی از آینده کسب و کار ایجاد کرد.

بررسی نتایج مربوط به آینده پژوهی

در این قسمت به بررسی میزان توانمندی مدل ایجاد شده توسط شبکه عصبی مصنوعی جهت بررسی داده‌های آینده پرداخته می‌شود.

پیش‌بینی وضعیت داده‌های سالهای ۱۳۹۶ و ۱۳۹۷

برای این کار، داده‌های مربوط به سال ۱۳۹۶ و ۱۳۹۷ از دیتابیس استخراج شد. البته باید در نظر گرفت که جهت پیش‌بینی میزان رویگردانی و وفاداری مشتریان باید یک دید کلی نسبت به مشتریان خویش‌فرمای آینده وجود داشته باشد. خوشبختانه این دید کلی با داده‌های موجود، برای مشتریان خویش‌فرمای سال‌های ۱۳۹۶ و ۱۳۹۷ وجود دارد. برای پیش‌بینی وضعیت مشتریان خویش‌فرمای جدید که در سال‌های ۱۳۹۶ و ۱۳۹۷ اقدام به ایجاد قرارداد و بیمه پردازی نموده‌اند، اطلاعات رکوردهای مربوط به ۸۳۶۴ نفر، انتخاب شد. سپس با استفاده از شبکه عصبی که با داده‌های قبل از سال ۱۳۹۶ آموزش دیده است، اقدام به کلاس‌بندی این دسته از مشتریان جدید نمودیم. شکل ۱۴ و جدول ۱۱ پیش‌بینی وضعیت وفاداری و رویگردانی مشتریان خویش‌فرمای سال‌های ۱۳۹۶ و ۱۳۹۷ است.



شکل ۱۴: نمودار تعداد مشتریان رویگردان سالیهای ۹۶ و ۹۷ براساس احتمالات متفاوت

در رابطه ۱، p مقدار احتمال وقوع رویگردانی افراد است و $\text{Number of person in } p$ نمایانگر تعداد نفرات رویگردان در احتمال p می‌باشد.

$$\text{تعداد مشتریان رویگردان} = \sum_{p=1}^{100} (p * \text{Number of person in } p) \quad \text{رابطه ۱:}$$

رابطه ۲:

$$\text{درصد مشتریان رویگردان} = \frac{\text{تعداد مشتریان رویگردان}}{\text{کل مشتریان}} \times 100$$

تحلیل نمودار شکل ۱۴ با توجه به روابط ۱ و ۲، به این شرح می‌باشد که از مجموع ۸۳۶۴ نفر مشتری فعال سال‌های ۱۳۹۶ و ۱۳۹۷:

- تعداد ۳۷۸ نفر به احتمال ۹۱ تا ۱۰۰ درصد در کلاس رویگردان تشخیص داده شده است.
- تعداد ۱۹۰ نفر به احتمال ۸۱ تا ۹۰ درصد در کلاس رویگردان تشخیص داده شده است.
- تعداد ۲۰۶ نفر به احتمال ۷۱ تا ۸۰ درصد در کلاس رویگردان تشخیص داده شده است.
- تعداد ۲۲۸ نفر به احتمال ۶۱ تا ۷۰ درصد در کلاس رویگردان تشخیص داده شده است.
- تعداد ۴۶۸ نفر به احتمال ۵۱ تا ۶۰ درصد در کلاس رویگردان تشخیص داده شده است.

- تعداد ۶۲۰ نفر به احتمال ۴۱ تا ۵۰ درصد در کلاس رویگردان تشخیص داده شده است.
- تعداد ۷۵۱ نفر به احتمال ۳۱ تا ۴۰ درصد در کلاس رویگردان تشخیص داده شده است.
- تعداد ۹۱۲ نفر به احتمال ۲۱ تا ۳۰ درصد در کلاس رویگردان تشخیص داده شده است.
- تعداد ۱۱۹۰ نفر به احتمال ۱۱ تا ۲۰ درصد در کلاس رویگردان تشخیص داده شده است.
- تعداد ۳۴۲۱ نفر به احتمال ۰ تا ۱۰ درصد در کلاس رویگردان تشخیص داده شده است.

جدول ۱۱: وضعیت رویگردانی مشتریان سال‌های ۱۳۹۶ و ۱۳۹۷

	رویگردان قطعی بیش از ۸۱٪	در معرض رویگردانی از ۶۱٪ تا ۸۰٪	احتمال رویگردانی از ۴۱٪ تا ۶۰٪	احتمال رویگردانی از ۲۱٪ تا ۴۰٪
تعداد مشتری رویگردان	۵۱۳	۱۰۴۱	۱۱۳۹	۱۷۹۵

با توجه به نمودار ۱۴، چنانچه مجموع احتمال رویگردانی تک تک مشتریان را با شبکه عصبی آموزش دیده، محاسبه نماییم، این نتیجه حاصل می‌گردد که حدود ۲۷,۶۵ درصد از مجموع بیمه‌شدگان خویش فرمای سال‌های ۱۳۹۶ و ۱۳۹۷ در کلاس رویگردان دسته‌بندی خواهند شد.

پژوهشگاه علوم انسانی و مطالعات فرهنگی
پرتال جامع علوم انسانی

منابع

- بروفرا، ا، رضائیان، ع، شکوهیار، س. (۱۳۹۵)، «شناسایی الگوی رفتاری مشتریان در بیمه عمر و تشکیل سرمایه با استفاده از داده کاوی»، پژوهش‌های مدیریت در ایران، ۶۵ تا ۹۴.
- ترکستانی، محمد صالح و ده پناه، آرمان و تقوی فرد، محمدتقی و شفیعی، شهرام. (۱۳۹۱)، «ارائه چارچوبی برای اصلاح نرخ حق بیمه در رشته بدنه اتومبیل با استفاده از مدل شبکه‌های عصبی (مطالعه موردی: شرکت بیمه آسیا)»، مدیریت فناوری اطلاعات، ۷۱۱ تا ۷۳۲.
- توکلی، ا، مرتضوی، س، کاهانی، م و حسینی، زهرا. (۱۳۸۹)، «به‌کارگیری فرآیند داده کاوی برای پیش‌بینی الگوهای رویگردانی مشتری در بیمه»، چشم‌انداز مدیریت بازرگانی، ۴۱ تا ۵۵.
- رستمی، شرافت و مرادی، م. (۱۳۹۴)، «بررسی تاثیر عوامل رفتاری در انتخاب بیمه توسط مشتری (مطالعه موردی: بیمه عمر کارآفرین شهرستان سنج)»، فصلنامه مطالعات مدیریت و حسابداری، ۸۴ تا ۹۵.
- مظلومی، ن، حقیقت، ف، مفاخری، ف، ترحمی، ف. (۱۳۹۳)، «مشتری مداری در صنعت بیمه»، گزارش موردی شماره ۲۴، پژوهشکده بیمه، ۱۱ تا ۵۹.
- Ai Cheo Y. and Kate A.S. and Robert J.W. and Malcolm B. (2001), Modelling the Effect of Premium Changes on, School of Business Systems. Monash University, Clayton, Victoria 3800. Australia.
- Dursun D., Christie F., Charles M., Deepa R. (2009), Analysis of healthcare coverage: A data mining approach, Expert Systems with Applications 36, p 995–1003.
- Ziming W. and Weiwei L. Zilongzh. and Angzhan W. (2017), An Ensemble Random Forest Algorithm for Insurance Big Data Analysis. IEEE International Conference on Computational Science and Engineering (CSE) and IEEE International Conference on Embedded and Ubiquitous Computing (EUC), P5.



پروفیسر گاہ علوم انسانی و مطالعات فرہنگی
پرتال جامع علوم انسانی

سایر مقالات

پژوهشگاه علوم انسانی و مطالعات فرهنگی
پرتال جامع علوم انسانی