سامانه‌های پردازشی و ارتباطی چندرسانه‌ای هوشمند
Intelligent Multimedia Processing and Communication Systems (IMPCS)

Fall 2023, 4 (3), 15-28
DOR:

Received: 13 July 2023
Accepted: 26 Aug 2023

مقاله پژوهشی

An Algorithm for Clustering of Insurance Products and Users in a Collaborative Filtering-Based Insurance Recommender System and evaluating its Performance Based on the Insurance Recommendation

Marzieh Amini Shirkoohi[1], Mohammadreza Yamaghani[2*]

1. MSc Student, Computer Science, Lahijan Branch, Islamic Azad University, Lahijan, Iran.
2. Assistant Professor, Department of Computer Science, Lahijan Branch, Islamic Azad University, Lahijan, Iran.
(*Corresponding Author)* O_yamaghani@iau.ac.ir

Abstract

Introduction: There are many improvements in insurance industries in these decades. So Many people refer to public and private insurance companies to get insurance services. They usually face to some challenges and issues for selecting the best and suitable insurance because of various type of insurance and lack of enough information of insurance service. Choosing the proper insurance service always related to people personal and social features

Method: Prediction of customer's insurance selection according to people personal and social property especially their financial condition play vital role. On one hand Prediction of insurance type can help people who want to utilize insurance service. On the other hand this prediction can facilitate process of insurance for Insurers too. There are multiple important mechanisms and factors like customers clustering, analyze each class feature, detection of popular insurance in each class and using Collaborative filtering technique to offer best insurance that can influence on process of decision and selection the suitable insurance.

Results: The total precision value of the proposed method is 89.98% for joint insurances of similar users. Also, the total value of the F-measure of the proposed method for joint insurances between similar customers is 87.13%.

Discussion: Customer behavior can be predicted by available data of people's personal and social features and type of insurance that they are chosen and rate of their satisfactions. K-means clustering algorithm and recommender systems Techniques like Collaborative filtering are two significant mechanisms to implement prediction of customer's behaviors.

Keywords: Recommender systems, collaborative filtering, k-means clustering, insurance.

سامانه‌های پردازشی و ارتباطی چندرسانه‌ای هوشمند
Intelligent Multimedia Processing and Communication Systems (IMPCS)
انجمن علمی تجارت الکترونیکی ایران
واحد زنجان

ارائه یک الگوریتم برای گروه‌بندی انواع محصولات بیمه و کاربران در سیستم توصیه‌گر بیمه با خوشه‌بندی مبتنی‌بر فیلتر مشارکتی ارزیابی عملکرد آن براساس توصیه بیمه

تاریخ دریافت: ۲۲/ ۰٤/ ۱٤۰۲
تاریخ پذیرش: ۰٤/ ۰٦/ ۱٤۰۲

دورهٔ چهارم، پاییز ۱٤۰۲
شمارهٔ سوم، صص: ۱۳-۲۸

مرضیه امینی شیرکوهی[۱]، محمدرضا یمقانی[۲*]

۱. کارشناسی ارشد، کامپیوتر، واحد لاهیجان، دانشگاه آزاد اسلامی، لاهیجان، ایران. Marzieh.amini@iau.ac.ir
۲. استادیار، کامپیوتر، واحد لاهیجان، دانشگاه آزاد اسلامی، لاهیجان، ایران. (نویسندهٔ مسئول) O_yamaghani@iau.ac.ir

چکیده: با پیشرفت‌هایی که در صنعت بیمه صورت‌گرفته‌است، افراد زیادی برای دریافت خدمات بیمه‌ای به شرکت‌های دولتی و خصوصی بیمه مراجعه‌می‌کنند. پیش‌بینی اینکه مشتریان چه نوع بیمه‌ای مناسب آن‌ها است و با توجه به شرایط فردی و اجتماعی به‌خصوص سطح درآمد کدام بیمه را ممکن است انتخاب‌کنند امری حیاتی است. در صنعت بیمه می‌توان با پیش‌بینی نوع بیمه انتخابی برای هر مشتری کار را برای افرادی که قصد خرید بیمه دارند و همچنین بیمه‌گزاران تا حد زیادی آسان‌کرد. خوشه‌بندی مشتریان، تحلیل ویژگی‌های هر خوشه و دریافت اینکه در هر خوشه کدام بیمه‌ها پرطرفدار هستند سپس استفاده از تکنیک فیلترینگ مشارکتی برای ارائه توصیه خرید بیمه به مشتریان، می‌تواند در روند تصمیم‌گیری و خرید خدمات بیمه‌ای مؤثر باشد. با استفاده از داده‌های موجود در مورد ویژگی‌های فردی و اجتماعی افراد و نوع بیمه‌هایی که انتخاب‌می‌کنند به همراه میزان رضایت آن‌ها از خدمات صنعت بیمه می‌توان به پیش‌بینی رفتار مشتری پرداخت و این امر با استفاده از الگوریتم خوشه‌بندی k-means و استفاده از تکنیک‌های سیستم‌های توصیه‌گر مانند فیلترینگ مشارکتی تا حد زیادی میّسر می‌شود.

واژه‌های کلیدی: سیستم‌های توصیه‌گر، فیلترینگ مشارکتی، خوشه‌بندیk-means، بیمه.

۱. مقدمه

سیستم‌های توصیه‌گر به‌عنوان نمونه‌ای از سیستم‌های پشتیبانی تصمیم شناخته‌می‌شوند که در دهه ۹۰ میلادی به‌عنوان یک شاخه مستقل پا به عرصه پژوهش و پژوهش گ شودند. از این سیستم ها به عنوان یکی از راه‌حل‌های رویارویی با مسئله سربار اطلاعات در تجارت الکترونیک یادمی شود [۱]. از یک سو کسب و کارهای فعال در زمینه تجارت الکترونیک به منظور جذب مشتری بیشتر در بازار پررقابت، نیازمند این سیستم‌ها هستند و از سوی دیگر مشتری نیازمند یاری رسانی در خیل عظیم و رو به‌رشد کالاها و اطلاعات مربوط به آن‌ها است تا بتواند از میان گزینه‌های پیش روی بی شمار، مناسب‌ترین آن ها را برگزیند[۲]. سیستم‌های توصیه‌گر با تحلیل و بررسی داده های متعلق به کاربران، یک‌سری آیتم های خاص را مبنی بر علایق به کاربران پیشنهادمی‌کنند. هدف از آنالیز داده‌های مربوط به کاربران، استخراج الگو های هر کاربر به منظور پیش‌بینی آیتم‌ها می باشد. یکی از مهمترین روش‌ها در سیستم‌های توصیه‌گر، روش فیلترینگ مشارکتی است.[3]

امروزه وب‌سایت‌های تجاری و غیرتجاری زیادی در اینترنت وجود دارند که حجم فزاینده‌ای از محصولات متنوع را ارائه می‌دهند. سیستم‌های توصیه‌کننده(RS[1]) به‌دلیل ظهور تجارت الکترونیک، گستره وسیع و سریع رو به‌رشد انتخاب‌ها برای مشتریان، تنوع ترجیحات بین، عدم آگاهی دقیق از نیازهایشان و فقدان اصطلاحات کلید واژه برای بیان و استفاده از موتورهای جستجو برای پاسخگویی به این خواسته‌ها به بخشی اجتناب ناپذیر از وب جهانی تبدیل‌شده‌اند. الگوریتم های توصیه‌گر سودمندی یک آیتم را برای کاربر هدف پیش‌بینی می‌کنند و با استفاده از رتبه‌بندی‌های گذشته خود برای موارد موجود در سیستم، بهترین مورد را برای ترجیحات کاربر پیشنهاد می‌کنند [۳].

گسترش بسیار زیاد سامانه‌های الکترونیکی انواع بیمه و تنوع محصولات بیمه‌ای موجود در فروشگاه‌ها نیز موجب‌شده‌است و مشتریان فروشگاه‌های الکترونیکی بیمه با مشکل سربار اطلاعاتی روبرو شوند. یافتن کالای مورد نظر به‌صورت برخط که با معیارهای کاربر مطابقت داشته‌باشد نیاز به وقت و جستجوی زیادی دارد و در موراد زیادی هم ممکن است کاربر نتواند بیمه مورد نیاز را مطابق علایق خود پیداکند. از این رو یک راهنمایی و درواقع پیشنهاد مفید مطابق با علایق، وی را از سردرگمی و اتلاف وقت بیهوده بازمی دارد.

با پیشرفت های تکنولوژی و امکان ذخیره‌سازی حجم عظیم اطلاعات، استفاده از داده‌کاوی به عنوان فرایند کشف الگوهای گوناگون در داده‌ها در حال افزایش است و به یک موضوع مهم در حوزه های مختلف از جمله صنعت بیمه تبدیل‌شده‌است.[4]

اخیراً با استفاده از علم داده‌کاوی و به‌کاربستن قوانین موجود در این علم، می‌توان مدل‌هایی روی داده‌ها ایجادکرد که و اطلاعات نهفته در درون داده‌ها را برای ما آشکارسازد. کاربرد داده‌کاوی در علوم مختلف به‌صورت روزافزونی پیشرفت‌می‌کند، اخیراً شاهد گسترش کاربردهای داده‌کاوی در زمینه شبکه اینترنت و وب هستیم [5]. این کاربردها قادرند جنبه‌های مختلفی از وب را پوشش‌دهند و اطلاعات ضمنی موجود در وب‌ها را استخراج‌کنند. این اطلاعات می‌تواند به طریقه استفاده از وب‌ها را برای به منظور رسیدن به مقصود خود آشکار سازند. سیستم توصیه‌گر محصول در تجارت الکترونیکی یکی از واسطهایی است که می‌تواند به منظور افزایش قابلیت کاربرپسندی فروشگاه‌های الکترونیکی استفاده‌شود. همچنین سیستم توصیه‌گر محصول می‌تواند با نظارت بر تاریخچه رفتار خرید مشتریان، اولویت‌ها و نیازهای مشتریان را پیش‌بینی‌کند و محصولاتی که با نیازهای مشتریان ارتباط نزدیکی دارد به‌عنوان توصیه خرید به مشتریان پیشنهادکند. کاربر نیز می‌تواند با اتکا به این توصیه‌ها محصولاتی که با اولویت‌هایش در ارتباط هستند، شناخته و تصمیم به خرید بگیرد. از این رو سیستم توصیه‌گر می‌تواند به شناسایی اقلام مناسب برای نیازها و اولویت‌های خود کمک‌کند و در حل مشکل سربار بیش از حد اطلاعات و رشد فروش در تجارت الکترونیک کمک‌کند [6].

مشتریان فروشگاه‌های الکترونیکی بدون داشتن راهنما در دستیابی به اطلاعات و محصول مورد نظرشان به صورت برخط، دچار سردرگمی خواهند شد. لذا مشتاق هستند تا از یک کارشناس و یا فردی با تجربه راهنمایی بگیرند. در این حال استفاده از سیستم توصیه‌گر محصول، که با توجه به اولویت، محصولات مرتبط را توصیه کند، ضروری است.

سیستم‌های توصیه‌گر محصول مبتنی بر فیلتر مشارکتی، بر اساس تاریخچه خرید و ارتباط کاربران و نظر سایر کاربران، کاربران هدف را پیش بینی کرده و محصولات مرتبط این علایق را به کاربران ارائه‌می دهد. از جمله مشکلات این دسته از سیستم‌های توصیه‌گر می‌توان به عدم اطلاعات کافی در مورد کاربران جدید، عدم توجه به تغییر نیاز کاربران با گذشت زمان، عدم دقت کافی در ارائه توصیه و غیره اشاره کرد. به منظور غلبه بر مشکلات موجود در سیستم‌های توصیه‌گر مبتنی بر محتوا و فیلتر مشارکتی، سیستم‌های توصیه گر ترکیبی ارائه‌شده است که از مزایای روش‌های پیشین استفاده‌کرده و توصیه‌گرهای انعطاف‌پذیری با سطح کاربر پسندی بالا ارائه کرده‌است[7]. در این پژوهش نیز یک سیستم توصیه‌گر ترکیبی ارائه شده‌است که از ترکیب خوشه‌بندی محصولات و روش فیلتر مشارکتی استفاده‌کرده‌است. نوآوری اصلی این پژوهش یافتن محصولات مورد علاقه کاربران، یافتن کاربران مشابه و ارائه محصولات و یا نمونه‌های مشابه به کاربران است. با توجه به این که محصولات مورد علاقه یک کاربر، محصولاتی با بیشترین نمره و رتبه‌بندی است، پس ارائه محصولات به کاربران مشابه می‌تواند مفید و با دقت بالا باشد که سطح کاربر پسندی سیستم توصیه گر را افزایش می‌دهد.

۲. سیستم توصیه‌گر محصول

سیستم توصیه‌گر محصول در تجارت الکترونیکی یکی از واسطهایی است که به‌منظور افزایش قابلیت کاربرپسندی فروشگاه‌های الکترونیکی مورد استفاده قرارمی‌گیرد. همچنین سیستم توصیه‌گر می‌تواند با نظارت بر

تاریخچه رفتار خرید مشتریان، اولویت‌ها و نیازهای مشتریان را پیش-بینی‌کند و محصولاتی که با نیازهای مشتریان ارتباط نزدیکی دارد به-عنوان توصیه خرید به مشتریان پیشنهادکند. کاربر نیز می‌تواند با اتکا به این توصیه‌ها محصولاتی که با اولویت‌های آنان در ارتباط هستند، شناخته و تصمیم به خرید بگیرد. از این رو سیستم توصیه‌گر می‌تواند فرد را در شناسایی اقلام مناسب برای نیازها و اولویت‌های خود کمک‌کند و در حل مشکل سربار بیش از حد اطلاعات و رشد فروش در تجارت الکترونیک کمک‌کند [6].

سامانه‌های توصیه‌گر، الگوریتم‌های نسبتاً ساده‌ای دارند که مناسب-ترین و دقیق‌ترین پیشنهادها را با بررسی و کاوش اطلاعات مرتبط کاربران از بانک اطلاعاتی مربوطه، به کاربر ارائه‌می‌کنند. این سامانه‌ها، با بررسی انتخاب کاربران در گذشته، الگوهایی در داده‌ها پیدا می‌کنند که با توجه به آن الگوی رفتاری، برای هر کاربر توصیه مناسب را نمایش می‌دهند. وجه تمایز سیستم‌های توصیه‌گر و سیستم‌های جستجو این است که اولاً در سیستم توصیه‌گر الزامی برای درخواست صریح از کاربر وجود ندارد و می‌توان در هر لحظه با رعایت ضوابط فنی و غیرفنی آیتم‌ها را به او پیشنهاد داد و ثانیاً یک سیستم توصیه‌گر اصولاً پیشنهادی بودن موارد را به کاربر گوشزدمی‌کند تا کاربر بداند که این موارد نتیجه انتخاب صریح (مثل انتخاب یک زیرشاخه از سایت) و جستجوی صریح (مثل درخواست برای جستجو) نیستند. تاکنون مطالعات متعددی در زمینه سیستم‌های توصیه‌گر و کاربرد آن‌ها توسط محققان داخلی و خارجی انجام‌شده‌است.

تقی‌زاده (۱۳۹۸) در پژوهشی انواع روش‌های فیلترینگ در سیستم‌های پیشنهاددهنده را بررسی کرده‌است. حجم فراوان و رو به رشد اطلاعات بر روی اینترنت، فرایند تصمیم‌گیری و انتخاب اطلاعات، داده یا کالاهای مورد نیاز را، برای بسیاری از کاربران دشوار کرده‌است. سیستم‌های پیشنهاد دهنده، با هدف رفع این چالش به وجود آمده‌اند و تلاش می‌کنند تا از میان حجم عظیم اطلاعات، اطلاعات خاص و مفید را با توجه به علاقه و سلیقه کاربر و تجربیات کاربران گذشته به وی پیشنهاد دهند. برای طراحی سیستم‌های پیشنهاد دهنده اصول بسیاری وجود دارد، اما یکی از اصول مهم در طراحی این سیستم‌ها روش فیلترینگ مورد استفاده در آن است. بدین منظور روش‌های مختلفی تا به امروز ارائه‌شده‌اند. در این پژوهش ضمن دسته بندی این روش‌ها، کاربردها و مزایا و معایب هرکدام ارائه‌شده‌است [8].

خمیسی و همکاران (۱۳۹۸) در پژوهشی الگوریتم فیلترگذاری مشارکتی در سیستم‌های توصیه‌گر را بررسی‌کرده‌اند. سیستم‌های شخصی‌سازی وب متفاوتی جهت پیش بینی اقلام درخواستی آینده کاربر با بهره‌گیری از کاربرد وب‌کاوی پیشنهاد شده‌است. این پژوهش به بررسی الگوریتم فیلتر گذاری مشارکتی در سیستم‌های توصیه‌گر پرداخته‌است [9].

F Pajuelo-Holguera و همکاران در سال ۲۰۲۰ پژوهشی با عنوان پیاده‌سازی سیستم‌های توصیه‌گر برای برنامه‌های فیلترینگ مشارکتی توسعه‌یافته انجام‌داده‌اند. در این پژوهش سیستم توصیه‌گری پیشنهاد شده‌است که روی یک سخت‌افزار با قابلیت برنامه‌ریزی مجدد برای تست روی تراشه برای برنامه‌های فیلترینگ مشارکتی توسعه‌یافته کم انرژی پیاده‌سازی می‌شود[9]. Panda و همکاران در سال ۲۰۲۰ پژوهشی با عنوان یک الگوریتم توصیه‌گر فیلترینگ مشارکتی بر اساس رویکرد نرمال‌سازی پیشنهاد داده‌اند. الگوریتم پیشنهادی از دو فاز ارزیابی و طراحی تشکیل شده‌است. در مرحله اول، الگوریتم میانگین امتیاز کاربر را برای هر مورد پیدامی‌کند و تعداد کاربران خرید کننده را در هر مورد شمارش‌می‌کند. سپس با استفاده از نرم‌افزار min-max تعداد نرمال کاربر در هر مورد و مقیاس میانگین رتبه‌بندی کاربران در یک محدوده معین را مشخص‌می‌کند. در مرحله دوم، الگوریتم پیشنهادی ماتریس رتبه‌بندی آموزش و آزمایش تقسیم‌می‌کند و رتبه‌بندی کاربران را پیش‌بینی‌می-کند. کارایی الگوریتم پیشنهادی در این پژوهش با ده معیار عملکردی به اثبات رسیده‌است [11].

۳. روش‌شناسی پژوهش

داده‌های این پژوهش مربوط به م شتریان شرکت بیمه ایران می با شد که برخی از این مشـــتر یان قبلاً از خد مات بی مه‌ای این شـــر کت استفاده‌کرده اند و برخی تا به‌حال از خدمات بیمه‌ای استفاده نکرده‌اند.

روش گردآوری داده‌ها به‌صورت " پرسشنامه" است. بدین صورت که برای دستیابی به یک‌سری از اطلاعات، به مشتریان حاضر در شرکت بیمه پرسشنامه‌ای ارائه‌شد و از مشتریان درخواست‌شد که به سوالات پرسشنامه صادقانه و با دقت پاسخ‌دهند. همچنین برای شناسایی متغیرهای تاثیرگذار در رفتار پرداختی مشتریان از کمک‌های کارشناس بیمه استفاده‌شد.

با توجه به فرمول آلفای کرونباخ و سوالات پر سشنامه این پژوهش ضریب آلفای کرونباخ ۰/۹۲ محاسبه شده‌است. همچنین پر سشنامه از روایی مطلوبی برخوردار است.

داده‌های پژوهش شامل تعدادی رکورد از مشتریان حقیقی بیمه ایران می باشد که هر رکورد دارای چند فیلد به عنوان خصیصه[۲] های مشتریان می‌باشد. نوع داده‌های مورد استفاده هم از نوع "عددی[۳]" و هم از نوع " اسمی[۴]" می‌باشند. جامعه آماری شامل ۵۰۰ مورد از مشتریان حقیقی بیمه ایران در شهر رشت می‌باشند. داده‌های مورد نظر شامل داده‌های آماری مشتریان و داده‌های بانکی است. داده‌های جمعیت آماری مشتریان شامل جنسیت، سن، وضعیت تأهل، تعداد فرزند، مدرک تحصیلی، شغل، میزان درآمد ماهیانه می‌باشد.

داده‌های بیمه‌ای مشتریان در این پژوهش شامل نوع بیمه درخواستی، مدت زمان استفاده از خدمات بیمه‌ای، میزان رضایت از خدمات بیمه می‌باشد. در نهایت نظر کارشناس بیمه در جهت نوع پرداخت و درصد حداقل پرداخت پرسیده شده‌است. پرسشنامه شامل سه بخش سوال است که بخش اول اطلاعات آماری متقاضی دریافت بیمه، بخش دوم اطلاعات بیمه‌ای دریافت‌کننده بیمه و بخش آخر نظر

کارشناس بیمه به اطلاعات دریافتی در مورد متقاضی است که به شرح زیر می باشد:

- اطلاعات آماری
  ۱. جنسیت
  ۲. سن
  ۳. وضعیت تأهل
  ۴. تعداد فرزند
  ۵. مدرک تحصیلی
  ۶. شغل
  ۷. میزان درآمد ماهیانه
  ۸. محل سکونت
- اطلاعات بیمه‌ای
  ۱. نوع بیمه درخواستی
  ۲. مدت زمان دریافت خدمات بیمه‌ای
  ۳. روش پرداخت حق بیمه
  ۴. میزان رضایتمندی از خدمات بیمه‌ای
- کارشناس بیمه
  ۱. نوع پرداخت
  ۲. درصد حداقل پرداخت

بعد از جمع آوری داده‌ها لازم است که در یک پایگاه جامع به صورت یکپارچه درآیند. بنابراین پایگاه داده یکپارچه در نرم‌افزار Excel تشکیل شد آماده‌سازی داده‌ها و پیش پردازش داده‌ها شامل بخش‌های مقادیر گم‌شده[۵]، کدبندی برچسب‌ها (در یادگیری ماشین، ما معمولاً با مجموعه داده‌هایی سروکار داریم که حاوی چندین برچسب در یک یا چند ستون هستند. این برچسب‌ها می‌توانند به صورت کلمه یا عدد باشند. برای اینکه داده‌ها قابل‌فهم یا قابل خواندن برای تحلیلگر باشند، داده‌ها اغلب با کلمات برچسب‌گذاری می‌شوند. کدبندی برچسب به تبدیل برچسب ها به شکل عددی به منظور تبدیل آن ها به فرم قابل خواندن توسط ماشین اشاره دارد. سپس الگوریتم های یادگیری ماشین می‌توانند به روشی بهتر تصمیم بگیرند که این برچسب ها چگونه باید کار کنند. این یک مرحله پیش پردازش مهم برای مجموعه داده ساختار یافته در یادگیری نظارت شده‌است. کدبندی برچسب داده‌ها را به شکل قابل خواندن توسط ماشین تبدیل‌می‌کند، اما یک عدد منحصر به فرد که از ۰ شروع می‌شود، به هر کلاس داده اختصاص می‌دهد) ، مقیاس‌بندی ویژگی[۶] و تفکیک مجموعه داده[۷] می‌باشد.

۴. روش پیشنهادی

در این پژوهش، یک سیستم توصیه‌گر مبتنی بر فیلترینگ مشارکتی بر پایه خوشه‌بندی کاربران به روش K-means ارائه‌شده‌است. به این صورت که در ابتدا کاربران بر اساس میزان رضایت از خدمات بیمه‌ای و بار دیگر از بین خوشه‌ای که بیشترین رضایت از خدمات بیمه‌ای را شامل‌می‌شود مجدداً بر اساس میزان درآمد ماهانه با الگوریتم k-means خوشه‌بندی می‌شود. دلیل انجام این کار در ابتدا شناسایی کاربرانی است که بیشترین سطح رضایت از خدمات را دارند و از آنجا که درآمد مهمترین ویژگی در بین ویژگی‌های فردی و اجتماعی مشتریان است، سپس روی مشتریانی که بیشترین درصد رضایت را دارند مجدد خوشه‌بندی به روش k-means صورت می‌گیرد. پس از خوشه‌بندی مشتریان بر اساس میزان رضایت و سپس بر اساس درآمد ماهانه هر یک از خوشه‌ها را تحلیل‌می‌کنیم و ویژگی‌های مهم و شاخص هر خوشه را نمایش می‌دهیم، با انجام این کار درمی یابیم که کدام بیمه‌ها بیشترین خرید را در هر خوشه داشته‌اند. بعد از این مرحله، سه بیمه شاخص و مهم را که در همه خوشه‌ها تکرار شده‌اند یافته و در سیستم توصیه‌گر استفاده‌می‌کنیم سپس با استفاده از سیستم فیلترینگ مشارکتی و ضریب همبستگی پیرسون مشتریان مشابه را در پایگاه داده اصلی شناسایی‌می‌کنیم. درپایان انتخاب‌ها بر اساس مشابهت با یکدیگر به کاربر توصیه می‌شوند. روش پیشنهادی به صورت مرحله به مرحله به شرح زیر است:

۱. خوشه‌بندی با روش k-means مشتریان بر اساس میزان رضایت و تشکیل چهار خوشه و در نهایت جداسازی و استفاده از خوشه مشتریان با میزان رضایت عالی(۵).
۲. خوشه‌بندی مجدد آن دسته از مشتریانی که در مرحله قبل جدا سازی شده‌اند و میزان رضایت آن‌ها از دریافت خدمات بیمه‌ای ۵ بوده‌است، با روش k-means این‌بار بر اساس درآمد ماهیانه و به ۴ خوشه تقسیم می‌شوند.
۳. تحلیل نظری خوشه‌ها و شناسایی ویژگی‌های شاخص هر خوشه و ارائه توضیحات.
۴. استفاده از تکنیک فیلترینگ مشارکتی برای هر خوشه بر اساس بیمه‌های پر فروش در هر خوشه و شناسایی مشتریان مشابه و ارائه توصیه .
۵. ارزیابی دقت و صحت و پوشش توصیه‌ها
۶. ارائه لگوریتمی برای پیشنهاد خرید بیمه با توجه به نتایج حاصل.

لازم به ذکر است که در این پژوهش داده‌ها به دو دسته آموزشی و آزمایشی تقسیم‌می شوند. نسبت داده‌های آموزشی ۷۰ در صد کل داده‌ها می‌باشد و ۳۰ درصد باقی‌مانده داده‌ها برای تست و آزمایش درستی پاسخگویی الگوریتم پیشنهادی مورد استفاده قرارمی‌گیرند.

در یک سیستم مبتنی بر فیلتر مشارکتی(CF) فرض‌می‌شود که اگر دو کاربر منافع مشابهی داشته‌با شند، پس آن کاربران علاقه مشابهی نسبت به یکدیگر دارند. به‌طور کلی، یک لیست از کاربران $U = \{U_1. U_2 . \dots . U_{||U||}\}$ و یک لیست از اقلام $I = \{I_1. I_2 . \dots . I_{||I||}\}$ را در نظر بگیرید.

برای هر کاربر U_1 یک لیست m گزینه دارد که برای آن‌ها رتبه‌بندی‌کرده‌است. جدول ۱ نمونه‌ای از ماتریس رتبه‌بندی کاربران را نشان‌می‌دهد. بنابراین، $m \subset I$ باشد (ممکن است که m مجموعه‌ای تهی باشد)، یک کاربر قابل تشخیص $U_A \in U$ ، به نام یک کاربر هدف، که برای

آن کاربر، هدف فیلتر مشارکتی، پیداکردن یک گزینه مورد علاقه در توصیه‌های جستجو خاص است. بنابراین، یک لیست از اقلام n، $n \subset I$، وجودخواهدداشت که کاربر هدف نیز به آن گزینه‌ها بیشتر علاقه‌مند است. لیست توصیه‌شده باید از گزینه‌هایی باشد که توسط کاربر هدف به‌ترتیب کاهش ارزش نمرات پیش‌بینی‌شده و توسط فیلتر مشارکتی مورد بررسی قرار نگرفته‌باشند این رابط از الگوریتم‌های فیلتر مشارکتی به‌عنوان " N‌ـ توصیه برتر" شناخته‌شده‌است [12].

جدول ۱: نمونه‌ای از ماتریس رتبه‌بندی

	Item(I_1)	Item(I_2)	Item(I_3)
User(u_1)	5.0	3.0	2.5
User(u_2)	2.0	2.5	5.0
User(u_3)	2.5	-	-
User(u_4)	5.0	-	3.0
User(u_5)	4.0	3.0	2.0

گام دوم از سیستم‌های مبتنی بر CF، انتخاب یک زیر مجموعه از کاربران با شباهت بالاتر است. پس از آن، گام بعدی محاسبه پیش بینی است. پیش‌بینی عمل استنتاج مقدار ارزیابی کاربر از یک محصول است که تاکنون آن را ارزیابی نکرده‌است، چقدر است. با توجه به جدول ۱، یک نمونه از این محاسبه را می‌توان در شکاف موجود برای کاربر U_3 بر روی گزینه های I_2 و I_3 و کاربر U_4 برای گزینه I_2 پرکرد. بنابراین، پیش‌بینی مستقل از روش مورد استفاده ساخته‌می شود، به این دلیل که توسط یک میانگین وزنی رتبه‌بندی از همسایگان که ضریب قابل‌قبولی از شباهت دارد، تولیدمی‌شود. با توجه به گفته‌های فوق، پیش‌بینی را می‌توان با معادله زیر محاسبه‌کرد [۱۱]:

$$p_{a.i} = \bar{r}_a + \frac{\sum_{u=1}^{h}(r_{u.i} - \bar{r}_u) * w_{a.u}}{\sum_{u=1}^{h} |w_{a.u}|} \quad (۱)$$

که در آن h مقدار بهترین هم سایه‌ها است و آن در اختیار هر یک از سیستم‌هایی است که از فیلترهای مشترک استفاده‌می‌کنند [12].

۵. ارزیابی عملکرد

برای مسائل سیستم‌های توصیه‌گر، یک روش بسیار رایج برای ارزیابی عملکرد، اندازه‌گیری دقت آن است.

جدول ۲: معیارهای ارزیابی دقت توصیه

شاخص	عملکرد شاخص	توضیحات	فرمول
MAE	میانگین خطای مطلق	هرچه مقدار این معیار در ازای مشتریان کمتر باشد، دقت عملکرد روش فیلتر مشارکتی بیشتر خواهدبود.	$\mathrm{MAE} = \frac{1}{N}\sum_{u,i}\lvert p_{u,i} - r_{u,i}\rvert$
Accuracy	دقت	برابر است با تعداد مواردی که درست پیش‌بینی‌کردیم	$Accuracy = \left(1 - \frac{MAE}{5}\right) \times 100$
Recall	پوشش	نسبت موارد مناسب که در میان تمام موارد مربوطه واقعی پیداشده‌است	$Recall = \frac{Correctly\ recommended\ items}{Total\ usful\ recommended\ item}$
Precision	صحت	نسبت موارد یافت‌شده که مربوط به موارد واقعی هستند	$Precisi = \frac{Correctly\ recommended\ items}{Total\ recommended\ items}$
F-mesure	پوشش و صحت	برای کمک به ساده‌سازی دو معیار Recall و Precision در یک معیار واحد تعریف شده‌است.	$F - measure = \frac{2 \times Precision \times Recall}{Precision + Recall}$

۶. پیاده‌سازی مدل پیشنهادی

۱،۶. رکوردهای مجموعه داده

خصیصه‌های مشتریان که همان متغیر مستقل و نظر نهایی کارشناسان که همان متغیر وابسته هستند، فیلدهای مجموع داده را تشکیل‌می‌دهند. این فیل ها شامل ۱۴ خصیصه از مشتریان و نظر کارشناسان است که در جدول ۳ نشان داده‌شده‌است.

جدول ۳: عناوین موجود در مجموعه داده

وضعیت تأهل	سن	جنسیت	نام فیلد
Marital status	Age	Gender	نام اختصاری
گروه شغلی	تحصیلات	تعداد فرزندان	نام فیلد
Job Category	Education	Dependents	نام اختصاری
مدت زمان استفاده از خدمات بیمه	محل زندگی	حقوق ماهیانه	نام فیلد
Period	Location	Mnth_Income	نام اختصاری
نوع پرداخت مورد نظر کارشناس بیمه	میزان رضایت از خدمات بیمه	روش پرداخت مورد نظر مشتری	نام فیلد
Payment type	Level of satisfaction	payment method	نام اختصاری
بیمه بدنه	بیمه ثالث	درصد حداقل پرداخت طبق نظر کارشناس بیمه	نام فیلد
Hull Insurance	Casualty Insurance	Minimum payment percentage	نام اختصاری
بیمه درمان تکمیلی	بیمه آتش‌سوزی	بیمه عمر	نام فیلد
Major Medical Insurance	Fire insurance	life Insurance	نام اختصاری
بیمه زلزله	بیمه کارفرما در قبال کارکنان (ساختمانی)	بیمه کارفرما در قبال کارکنان (غیرساختمانی)	نام فیلد
Earthquake insurance	(construction)	(Non-construction)	نام اختصاری
	بیمه سرقت	بیمه حوادث انفرادی	نام فیلد
	Theft insurance	Individual accident insurance	نام اختصاری

۲،۶. پیش‌پردازش داده‌ها

الگوریتم‌های یادگیری ماشین تنها اعداد را درک‌می‌کنند. در مجموعه داده پژوهش حاضر همان‌گونه که مشاهده‌شد تعدادی متغیر از نوع اسمی وجود دارد، پس لازم است به متغیر عددی تبدیل‌شوند. برای این امر از روش کدبندی استفاده‌می‌کنیم. یعنی برای هر حالت یک خصیصه یک عدد در نظرمی‌گیریم که از ۰ شروع‌می‌شود.

با رسم ماتریس همبستگی می‌توانیم نشان‌دهیم کدام متغیر نسبت به متغیر دیگر همبستگی زیاد یا کم دارد همان‌طور که دیده‌می‌شود وضعیت خرید بیمه ثالث با خرید بیمه بدنه ارتباط نزدیکی دارد و عدد همبستگی آن‌ها ۰/۶۲ نیز است. همچنین می‌توان به نزدیکی تقریبی ویژگی نوع پرداخت و درصد حداقل پرداخت که عدد همبستگی ۰/۵۳ را نشان‌می‌دهد نیز اشاره‌کرد. همچنین گروه شغلی و خرید بیمه کارفرما در قبال کارکنان (ساختمانی) و همی‌نطور گروه شغلی و محل سکونت ضریب همبستگی آن ها ۳۶/. می‌باشد و همبستگی جزئی دارند. در مقابل بیمه آتش سوزی و محل سکونت همبستگی ۰/۲۶- را نشان‌می‌دهد که بیانگر عدم همبستگی این دو ویژگی است.

۷. پیاده‌سازی مدل پیشنهادی

این پژوهش بر روی ۵۰۰ مشتری از مشتریان شرکت بیمه ایران به شماره نمایندگی۳۰۶۹ در شهر رشت انجام‌شده‌است. قابل‌ذکر است که به‌دلیل تنوع بالای بیمه‌های موجود در شرکت بیمه ایران تنها از ۱۰ بیمه برتر و پرفروش استفاده‌شده‌است (جدول ۴) .

جدول ۴: نام بیمه‌های مورد استفاده در پژوهش

ردیف	نام بیمه
۱	بیمه ثالث
۲	بیمه بدنه
۳	بیمه عمر
۴	بیمه آتش‌سوزی
۵	بیمه درمان تکمیلی
۶	بیمه کارفرما در قبال کارکنان (غیر ساختمانی)
۷	بیمه کارفرما در قبال کارکنان (ساختمانی)
۸	بیمه زلزله
۹	بیمه حوادث انفرادی
۱۰	بیمه سرقت

در جدول (۵) بیمه‌های مورد استفاده در دیتاست با شناسه‌های مشخص آمده‌است.

جدول ۵: لیست شناسه‌های مورد استفاده در دیتاست

ردیف	نام بیمه	شناسه بیمه ها
۱	بیمه ثالث	Casualty Insurance
۲	بیمه بدنه	Hull Insurance
۳	بیمه عمر	Life Insurance
۴	بیمه آتش‌سوزی	Fire insurance
۵	بیمه درمان تکمیلی	Major Medical Insurance
۶	بیمه کارفرما در قبال کارکنان (غیر ساختمانی)	(Non-construction)
۷	بیمه کارفرما در قبال کارکنان (ساختمانی)	(construction)
۸	بیمه زلزله	Earthquake insurance
۹	بیمه حوادث انفرادی	Individual accident insurance
۱۰	بیمه سرقت	Theft insurance

خوشه‌بندی مشتریان بر اساس میزان رضایت عالی (۵) به روش k-means و تعداد چهار خوشه در جدول (۶) آمده‌است.

جدول ۶: نتیجه خوشه‌بندی بر اساس میزان رضایت از دریافت خدمات

خوشه ها	خوشه ۱	خوشه ۲	خوشه ۳	خوشه ۴
عنوان خوشه	۵	۴	۳	۲
تعداد	۳۲۱	۱۰۶	۶۲	۱۱

با توجه به اطلاعات جدول و خوشـــه بندی به روش k-means بر اساس میزان رضایت مشتریان به چهار خوشه تقسیم‌شده‌اند. خوشه اول با ۳۲۱ مشـتری که میزان رضـایت ۵ را از خدمات بیمه دارند، خوشـه دوم با ۱۰۶ مشتری که میزان رضایت ۴ را از خدمات بیمه دارند، خوشه شماره ۳ با ۶۲ مشتری که میزان رضایت ۳ را از دریافت خدمات بیمه دارند و خوشـه آخر و چهارم با تعداد ۱۱ مشـتری که میزان رضـایت از خدمات بیمه با امتیاز ۲ را نمایش‌می‌دهد. برای خوشـــه‌بندی مجدد مشتریان بر اساس درآمد ماهیانه، مشتریان خوشه اول با میزان رضایت ۵ و تعداد ۳۲۱ نفر را در نظرمی‌گیریم، مجدد این مشـــتریان را با روش k-means، به چهار خوشـه تقسـیم‌می‌کنیم. نتیجه این خوشـه‌بندی در جدول ۷ آمده‌است.

جدول ۷: خوشه‌بندی مشتریان بر اساس درآمد ماهیانه به روش k-means

خوشه ها	خوشه ۱	خوشه ۲	خوشه ۳	خوشه ۴
عنوان خوشه	بیشتر از ۱۲ میلیون	بین ۱۲-۱۰ میلیون	بین ۱۰-۶ میلیون	بین ۶-۴ میلیون
تعداد	۱۸۳	۷۰	۴۷	۲۱

طبق اطلاعات جدول۷ پس از خوشـــه‌بندی به روش k-means، تعداد ۳۲۱ مشتری در چهار خوشه قرارگرفته‌اند. که از این تعداد ۱۸۳ نفر با حقوق بالای ۱۲ میلیون تومان در ماه در خوشـــه اول، ۷۰ نفر با حقوق ۱۰ الی ۱۲ میلیون تومان در ماه در خوشه دوم، تعداد ۴۷ نفر با حقوق ۶ الی ۱۰ میلیون تومان در ماه در خوشـه سـوم و در آخر تعداد ۲۱ مشـتری با حقوق ماهیانه ۴ الی ۶ میلیون تومان در خوشـه چهارم قراردارند.

۸. تحلیل خوشه‌ها

۱,۸. تحلیل خوشه اول(مشتری با درآمد بالای ۱۲ میلیون تومان)

در خوشه اول ۱۸۳ مشتری با درآمد بالای ۱۲ میلیون تومان است. جدول ۸ بیانگر ویژگی‌های شـاخص مشـتریان با تعداد و درصـد آن‌ها است.

جدول ۸: ویژگی‌های شاخص مشتریان خوشه اول

متغیر	عدد	درصد
مرد	۱۴۳	۷۸,۱٪
۳۰-۴۰	۶۸	۳۷,۲٪
متأهل	۱۴۰	۷۶,۵٪
یک فرزند	۵۷	۳۱,۱٪
لیسانس	۷۹	۴۳,۲٪
صنعت	۷۲	۳۹,۳٪
شهر	۱۵۹	۸۶,۹٪
ماهانه	۸۴	۴۵,۹٪
اقساط	۱۱۹	۶۵٪
۵۰٪ درصد حداقل پرداخت	۸۲	۴۴,۸٪

ویژگی‌های شـاخص در جدول۸ بیانگر این اسـت که این ویژگی‌ها در این خوشه بیشترین تکرار را داشته اند و بخش زیادی از خوشه اول را افراد با این ویژگی ها تشکیل‌می‌دهند.

جدول ۹: بیمه‌هایی که بیشترین خرید را توسط مشتریان خوشه اول داشته‌اند.

اسامی بیمه‌ها	تعداد	درصد
بیمه ثالث	۱۵۹	۸۶,۹٪
بیمه بدنه	۱۵۶	۸۵,۲٪
بیمه عمر	۹۲	۵۰,۳٪

همان‌طور که دیده‌می‌شود، بیمه ثالث، بیمه بدنه و بیمه عمر بیشترین خرید را در خوشه اول با مشتریان با درآمد بالای ۱۲ میلیون تومان دارند.

۲,۸. تحلیل خوشه دوم (مشتری با درآمد ۱۰ الی ۱۲ میلیون)

در خو شه دوم ۷۰ م شتری با درآمد ۱۰ الی ۱۲ میلیون تومان ا ست. جدول ۱۰ بیانگر ویژگی‌های شاخص م شتریان با تعداد و در صد آن‌ها است.

جدول ۱۰: ویژگی‌های شاخص مشتریان خوشه دوم

متغیر	عدد	درصد
مرد	۳۹	۵۵,۷٪
۳۰-۴۰	۳۲	۴۷,۷٪
یک فرزند	۲۵	۳۷,۷٪
لیسانس	۳۴	۴۸,۶٪
خدمات	۲۲	۳۱,۴٪
شهر	۴۹	۷۰٪
بالای یک سال	۵۲	۷۴,۳٪
ماهانه	۳۱	۴۴,۳٪
اقساط	۶۵	۹۲,۹٪
۵۰٪ درصد حداقل پرداخت	۳۷	۵۲,۹٪

ویژگی‌های شـــاخص در جدول (۱۰) بیانگر این هســـتند که این ویژگی‌ها در این خوشـــه بیشـــترین تکرار را داشـــته‌اند و بخش زیادی از خوشـــه دوم را افراد با این ویژگی‌ها تشـــکیل‌می‌دهند. جدول ۱۱ بیانگر بیمه‌های شاخص خوشه دوم با تعداد و درصد آن‌هاست.

جدول ۱۱: بیمه‌هایی که بیشترین خرید را توسط مشتریان خوشه دوم داشته‌اند.

اسامی بیمه‌ها	تعداد	درصد
بیمه ثالث	۴۵	۶۴,۳٪
بیمه عمر	۳۵	۵۰٪
بیمه درمان تکمیلی	۴۰	۵۷,۱٪

همان‌طور که دیده‌می‌شـود، بیمه ثالث، بیمه درمان تکمیلی و بیمه عمر بیشترین خرید را در خوشه دوم با مشتریان دارای درآمد ۱۰ الی ۱۲ میلیون تومان دارند.

۳,۸. تحلیل خوشه سوم (مشتری با درآمد ۶ الی ۱۰ میلیون)

در خوشـــه ســـوم ۴۷ مشـــتری با درآمد ۶ الی ۱۰ میلیون تومان اســـت. جدول ۱۲ بیانگر ویژگی‌های شـــاخص مشـــتریان با تعداد و درصـــد آن‌هاست.

جدول ۱۲: ویژگی‌های شاخص مشتریان خوشه سوم

متغیر	تعداد	درصد
زن	۲۴	۵۰٪
۲۰-۳۰	۱۶	۳۳,۳٪
متأهل	۲۷	۵۶,۳٪
بدون فرزند	۱۸	۳۷,۵٪
لیسانس	۲۹	۵۶,۳٪
خدمات	۲۶	۵۴,۲٪
شهر	۴۲	۸۷,۵٪
بالای یک سال	۲۹	۶۰,۴٪
ماهانه	۲۲	۴۵,۸٪
اقساط	۳۵	۷۴,۴٪
۵۰٪ درصد حداقل پرداخت	۱۷	۳۶,۲٪

ویژگی‌های شاخص در جدول ۱۳ نشان‌می‌دهد که این ویژگی‌ها در این خوشه بیشترین تکرار را داشته‌اند و بخش زیادی از خوشه سوم را افراد با این ویژگی‌ها تشکیل می‌دهند.

جدول ۱۳: بیمه‌هایی که بیشترین خرید را توسط مشتریان خوشه سوم داشته‌اند.

اسامی بیمه‌ها	تعداد	درصد
بیمه ثالث	۲۹	۶۰,۴٪
بیمه درمان تکمیلی	۲۴	۵۰٪

همان‌طور که دیده‌می‌شـــود، بیمه ثالث و بیمه درمان تکمیلی بیشترین خرید را در خوشه سوم با مشتریان دارای درآمد ۶ الی ۱۰ میلیون تومان دارند.

۴,۸. تحلیل خوشه چهارم(مشتری با درآمد ۴ الی ۶ میلیون)

در خوشـــه چهارم ۲۱ مشـــتری با درآمد ۴ الی ۶ میلیون تومان اســـت. جدول ۱۴بیانگر ویژگی های شاخص مشتریان با تعداد و در صد آن‌ها است.

جدول ۱۴: ویژگی‌های شاخص مشتریان خوشه چهارم

ویژگی	تعداد	درصد
زن	۱۷	۸۱٪
۲۰-۳۰	۹	۴۲,۹٪
متأهل	۱۲	۵۷,۱٪
بدون فرزند	۱۰	۴۷,۶٪
دیپلم	۱۰	۴۷,۶٪
خدمات	۱۱	۵۲,۴٪
شهر	۱۳	۶۱,۹٪
یک سال	۸	۳۸,۱٪
ماهانه	۱۱	۵۲,۴٪
اقساط	۲۱	۱۰۰٪
۵۰٪ درصد حداقل پرداخت	۱۲	۵۷,۱٪

ویژگی‌های شاخص در جدول ۱۴ بیانگر آنند که این ویژگی‌ها در این خوشه بیشترین تکرار را داشته‌اند و بخش زیادی از خوشه چهارم را افرادی با این ویژگی‌ها تشکیل‌می‌دهند.

۵,۸. الگوریتم فیلترینگ مشارکتی

پس از خوشه‌بندی مشتریان ابتدا از نظر میزان رضایت از خدمات بیمه‌ای و سپس درآمد ماهیانه و تحلیل ویژگی‌های هر خوشه، به کمک الگوریتم فیلترینگ مشارکتی مشتریان مشابه در انتخاب بیمه را تشخیص‌می‌دهیم. برای این کار بیمه‌های ثالث، عمر و درمان تکمیلی که بیشترین خرید را در خوشه‌ها داشته‌اند به عنوان بیمه‌های پرفروش انتخاب‌می‌کنیم. برای ادامه کار با فیلترینگ مشارکتی به داده‌های اسامی بیمه‌ها و شناسه‌های آن‌ها و داده‌های شناسه مشتری مشتریان و میزان رتبه‌ای که به بیمه‌ها داده‌اند نیاز است. شکل ۱ و ۲ این اطلاعات را نمایش می‌دهد.

	insuranceId	title
0	1	Casualty Insurance
1	2	Hull Insurance
2	3	life Insurance
3	4	Fire insurance
4	5	Major Medical Insurance
5	6	(Non-construction)
6	7	(construction)
7	8	Earthquake insurance
8	9	Individual accident insurance
9	10	Theft insurance

شکل ۱: اسامی بیمه‌ها و شناسه‌های مربوط

	userId	insuranceId	rating
0	1	1	5
1	1	2	5
2	1	3	5
3	1	7	5
4	2	1	5
...	...	...	...
1256	497	2	5
1257	498	3	5
1258	498	6	4
1259	499	5	5
1260	500	6	5

1261 rows × 3 columns

شکل۲:اسامی مشتریان و نوع بیمه‌های انتخابی و رتبه اعطا شده به هر بیمه

حال با استفاده از قطعه کدی که در شکل ۲ نمایش‌داده‌شده است، ویژگی‌های مشتری ورودی را به سیستم می‌دهیم که به‌ترتیب بیمه‌های ثالث با امتیاز ۵، بیمه عمر با امتیاز ۴ و بیمه درمان تکمیلی با امتیاز ۳ را انتخاب‌کرده‌است.

```
userInput = [
            {'title': 'Casualty Insurance', 'rating':5},
            {'title': 'life Insurance', 'rating':5},
            {'title': 'Major Medical Insurance', 'rating':5},
         ]
inputInsurance = pd.DataFrame(userInput)
inputInsurance
```

شکل ۳: بیمه‌های منتخب مشتری اصلی و امتیاز مشتری به هر بیمه

خروجی این قطعه کد به علاوه دریافت اطلاعات نام بیمه، شناسه بیمه و رتبه‌ای است که مشتری به آن داده است (شکل ۳).

	title	rating
0	Casualty Insurance	5
1	life Insurance	5
2	Major Medical Insurance	5

شکل ۴: شناسه بیمه، نام بیمه و رتبه مشتری به هر بیمه

حالا مجموعه مشتریانی که این سه بیمه را انتخاب‌کرده‌اند و امتیاز های متفاوتی به آن ها داده‌اند، می یابیم، زیرا هدف اصلی انتخاب‌های مشابه در زمینه بیمه است. در شکل ۴ و ۵ خروجی را می‌بینیم.

```
[(12,
     userId  insuranceId  rating
 30      12            1       4
 32      12            3       5
 33      12            5       4),
(15,
     userId  insuranceId  rating
 38      15            1       4
 40      15            3       4
 41      15            5       5),
(18,
     userId  insuranceId  rating
 48      18            1       3
 50      18            3       5
 52      18            5       5)]
```

شکل ۵: مجموعه مشتریانی که سه بیمه مشابه را انتخاب‌کرده‌اند

در مرحله بعد، همه مشتری‌ها با مشتری مشخص‌شده خود مقایسه شد و مشابه‌ترین مشتری پیدا شد. از طریق ضریب همبستگی پیرسون متوجه‌می‌شویم که هر مشتری چقدر به ورودی شباهت دارد. برای اندازه‌گیری قدرت ارتباط خطی بین دو متغیر استفاده‌می شود. شکل ۶ شبیه‌ترین مشتریان به مشتری ورودی را با محاسبه میزان شباهت به روش ضریب همبستگی پیرسون نمایش‌می‌دهد.

	similarityIndex	userId
51	1.0	449
60	1.0	32
76	1.0	96
70	1.0	67
68	1.0	62
...	...	...
94	-1.0	164
95	-1.0	175
96	-1.0	177
62	-1.0	43
87	-1.0	133

100 rows × 2 columns

شکل ۶: محاسبه شباهت مشتریان به مشتری ورودی با ضریب همبستگی پیرسون

با توجه به اطلاعات شـکل ۶ برای مثال مشـتری شـماره ۴۴۹ با میزان شـباهت ۱ بیشـترین شـباهت را به مشـتری ورودی و مشـتری شماره ۱۳۳ بی شباهت ترین مشتری به مشتری ورودی از نظر انتخاب بیمه و رتبه دادن به بیمه‌ها است.

برای تشـخیص اینکه مشـتری ورودی چه امتیازی ممکن بود به بیمه‌های دیگری که مشـتریان مشـابه انتخاب‌کرده‌اند بدهد، وزن رتبه‌دهی توسط مشتری ورودی را با فرمول زیر محاسبه‌می‌کنیم.

$$\text{weighted Rating} = \text{similarity Index} * \text{rating} \quad (۲)$$

فرمول توضیح‌می‌دهد که وزن رتبه احتمالی مشـتری اصـلی برابر است با میزان شباهت مشتری مشابه با مشتری ورودی در رتبه‌ای که مشتری مشابه به بیمه داده‌است. خروجی کار در شکل ۷ آمده است.

	similarityIndex	userId	insuranceId	rating	weightedRating
0	1.0	449	1	5	5.0
1	1.0	449	2	4	4.0
2	1.0	449	3	4	4.0
3	1.0	449	5	3	3.0
4	1.0	32	1	5	5.0

شکل ۷: محاسبه میزان رتبه احتمالی(وزن) مشتری ورودی به بیمه-های مورد انتخاب مشتریان مشابه

در مرحله بعدی کار برای تشخیص و پیش‌بینی بهتر دیگر بیمه‌های منتخب مشـتری اصـلی جمع میزان شـباهت‌های کل بیمه‌ها و جمع میزان وزن‌های رتبه‌دهی احتمالی توسـط مشـتری اصـلی و سـپس میانگین بیشـترین وزن رتبه‌دهی احتمالی توسـط مشـتری اصـلی را محا سبه و بی شترین آن‌ها را انتخاب‌می‌کنیم. خروجی در شکل ۷ و ۸ آمده‌است.

insuranceId	sum_similarityIndex	sum_weightedRating
1	10.0	48.0
2	8.0	35.0
3	2.0	7.0
4	4.0	17.0
5	9.0	32.0
6	2.0	10.0
8	1.0	5.0
10	2.0	8.0

شکل ۸: محاسبه جمع شباهت‌ها و جمع وزن‌ها

insuranceId	weighted average recommendation score	insuranceId
6	5.000	6
8	5.000	8
1	4.800	1
2	4.375	2
4	4.250	4

شکل ۹: محاسبه میانگین وزن‌ها برای پیداکردن بهترین توصیه

با توجه به خروجی مرحله قبل و شکل‌های بالا مشاهده‌می‌شود که بیمه‌ها با شناسه ۶ و ۸ در بالاترین خانه‌های جدول با امتیاز ۵ قرار دارند. پس دو بیمه‌ای که می‌توان به مشتری اصلی توصیه‌کرد بیمه ۶ و ۸ یعنی بیمه زلزله و بیمه کارفرما در قبال کارکنان (غیرساختمانی) است. شکل ۱۰ توصیه‌ها را با توجه به این توضیحات نشان‌می‌دهد.

	insuranceId	title
5	6	(Non-construction)
7	8	Earthquake insurance

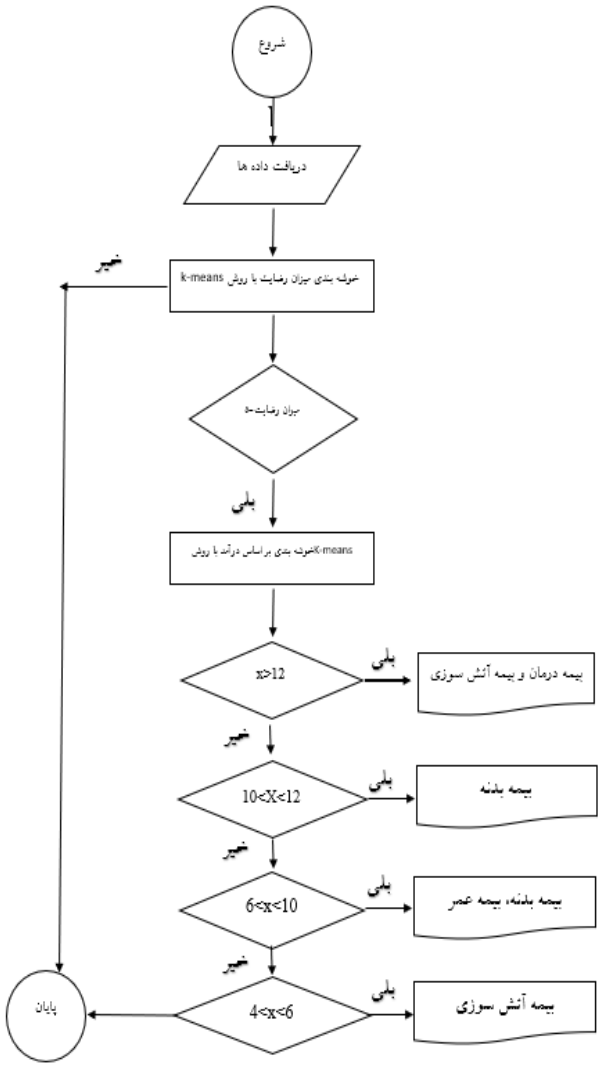
شکل ۱۰: ارائه توصیه بیمه به مشتری ورودی

در نهایت با به نتیجه رسیدن سیستم توصیه‌گر در این پژوهش برای کاربر ورودی هر خوشه نیز می‌توان بیمه‌هایی که کاربران مشابه دیگر انتخاب‌کرده‌اند را پیشنهادداد. نتیجه کار در جدول ۱۵ آمده‌است. شکل ۱۲ نیز فرایند الگوریتم پیشنهادی را به‌خوبی نشان‌می‌دهد. توضیحات به شرح زیر است :

۱. خوشـه‌بندی با روش k-means مشـتریان بر اسـاس میزان رضـایت و تشـکیل چهار خوشـه و در نهایت جداسـازی و استفاده از خوشه مشتریان با میزان رضایت عالی۵.
۲. خوشه بندی مجدد آن دسته از مشتریانی که در مرحله قبل جدا سازی شده‌اند و میزان ر ضایت آن‌ها از دریافت خدمات بیمه‌ای ۵ بوده‌اسـت، با روش k-means این‌بار بر اسـاس درآمد ماهیانه به ۴ خوشه تقسیم‌می‌شوند.
۳. اگر حقوق کاربر جدید بالای ۱۲ میلیون تومان باشد، بیمه درمان و بیمه آتش‌سوزی را به کاربر پیشنهاد می‌دهیم.
۴. اگر حقوق کاربر جدید بین ۱۰ الی ۱۲ میلیون تومان باشد، بیمه بدنه را پیشنهاد می‌دهیم.
۵. اگر حقوق کاربر جدید بین ۶ الی ۱۰ میلیون تومان باشد، بیمه بدنه و بیمه عمر را پیشنهاد می‌دهیم.
۶. اگر حقوق کاربر جدید بین ۴ الی ۶ میلیون تومان باشد، بیمه آتش‌سوزی را به کاربر پیشنهاد می‌دهیم.

جدول ۱۵: ارائه پیشنهاد خرید بیمه به هر خوشه

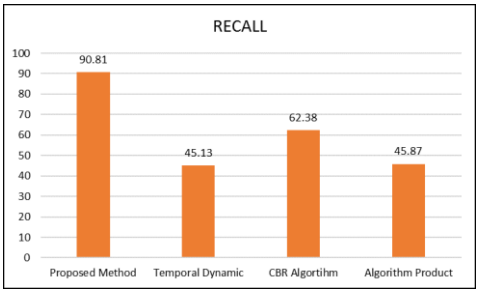
ردیف	ویژگی خوشه	بیمه های پر طرفدار	بیمه پیشنهادی
۱	درآمد بالای ۱۲ میلیون تومان	بیمه ثالث، بیمه بدنه، بیمه عمر	بیمه درمان و بیمه آتش‌سوزی
۲	درآمد ۱۰ الی ۱۲ میلیون تومان	بیمه ثالث، بیمه عمر، بیمه درمان تکمیلی	بیمه بدنه
۳	درآمد ۶ الی ۱۰ میلیون تومان	بیمه ثالث، بیمه درمان تکمیلی	بیمه بدنه، بیمه عمر
۴	درآمد ۴ الی ۶ میلیون تومان	بیمه عمر، بیمه درمان تکمیلی	بیمه آتش‌سوزی



شکل ۱۲: الگوریتم

۹،۸. ارزیابی عملکرد روش پیشنهادی بر اساس توصیه بیمه

در ارزیابی عملکرد در روش پیشنهادی بر اساس توصیه بیمه، توصیه صحیح به‌صورت نسبت بیمه‌هایی است که به صورت مشترک بین مشتریان مشابه بررسی‌شده و بالاترین امتیاز را گرفته‌است. بیمه‌ها مشترک بین دو مشتری مشابه به
منزله توصیه‌ای است که از سوی یک مشتری به مشتری مشابه صورت گرفته و به صورت قطعی مورد قبول واقع‌شده و خریداری شده‌است. البته دیگر بیمه‌هایی که توسط یک مشتری به مشتری مشابه دیگر ارائه‌می‌شود ممکن است مورد قبول واقع شده و خریداری شود ولی به دلیل عدم دسترسی به نتایج واقعی از این موارد صرف‌نظر می‌شود. جدول ۱۶ معیار Recall برای ۴ مشتری است که بیشترین شباهت را دارند و نشان‌دهنده نسبت توصیه‌های دقیق به کل توصیه‌های ارائه‌شده به مشتریان مشابه است. دقت کل روش پیشنهادی از میانگین دقت کسب‌شده برای هر مشتری به‌دست آمده‌است.



جدول۱۶: معیار Recall برای همه مشتریان

#	449	181	219	455	Total Recall
449	100	83.88	93.88	92.67	
181	89.31	100	89.37	90.5	
219	84.19	81.29	100	82.44	
455	82.72	90.6	92.18	100	
avg	89.05	88.94	93.85	91.40	90.81
#	449	181	219	455	Total Recall
449	100	82.63	92.85	91.25	
181	88.22	100	88.36	89.6	
219	83.26	80.26	100	81.23	
455	81.69	89.3	91.20	100	
avg	88.29	88.04	93.10	90.52	89.98
#	449	181	219	455	
449	#	83.25	93.36	91.95	Total
181	88.76	#	88.86	90.04	Recall
219	83.72	80.77	#	81.83	
455	82.20	89.94	91.68	#	
avg	84.89	84.65	91.3	87.70	87.13

طبق جدول ۱۶ ، مقدار کل Recall روش پیشنهادی به ازای بیمه‌های مشترک بین کابران مشابه ۹۰/۸۱٪ می‌باشد. مقدار کل Precision روش پیشنهادی به‌ازای بیمه‌های مشترک بین کاربران مشابه ۸۹/۹۸٪ می‌باشد. همچنین مقدار کل معیار F-measure روش پیشنهادی به‌ازای بیمه‌های مشترک مشتریان مشابه ۸۷/۱۳٪ است.

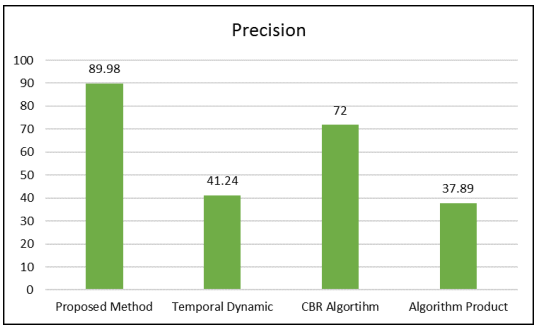
۹.مقایسه روش پیشنهادی با الگوریتم‌ها و روش‌های دیگر

به منظور اعتبار سنجی روش پیشنهادی، در این بخش به مقایسه آن با دیگر الگوریتم ها خواهیم‌پرداخت. با توجه به اینکه روش پیشنهادی از روش فیلتر مشارکتی توصیه بیمه بر اساس خوشه‌بندی است آن را از نظر معیارهای دقت توصیه با دیگر الگوریتم‌ها مقایسه خواهیم‌کرد. در شکل ۴-۲۶ مقایسه روش پیشنهادی با دیگر الگوریتم ها در این حوزه از نظر دقت توصیه‌هایی که به‌درستی ارائه‌شده‌اند، آمده‌است. از آنجا که در این پژوهش توصیه صحیح به‌صورت نسبت بیمه‌هایی است که بین کابران مشابه بوده و بالاترین امتیاز را گرفته‌است، پس می‌توان این مقدار را به عنوان معیار مورد استفاده Recall معرفی‌کرد. بیمه‌های مشترک بین دو کاربر مشابه به‌منزله توصیه مفیدی است که از سوی

یک م شتری به م شتری م شابه صورت‌گرفته و به صورت قطعی مورد قبول واقع‌شده و خریداری شده‌است. نمودار ۱ مقایسه روش پیشنهادی با روش‌های متفاوت دیگری[13,14,15] که در نمودار ذکر شده‌ا ست را از نظر معیار Recall نشان‌می‌دهد.

نمودار۱: مقایسه روش پیشنهادی با روش‌های دیگر از نظر معیار Recall

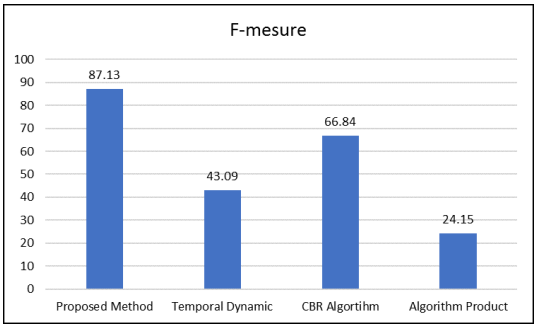
همان‌طور که در نمودار بالا نشان داده شده‌است. روش پیشنهادی در مقایسه با سایر روش‌های ذکرشده از نظر معیار Recall نتیجه بهتری را نشـــان‌می‌دهد. نتیجه بهتر در معیار Recall نشـــان‌دهنده دقت بالای سیستم تو صیه‌گر پیشنهادی در ارائه تو صیه به کاربران م شابه ا ست. همچنین نمودار ۲ مقایسه روش پیشنهادی با روش‌های دیگر را از نظر



معیار Precision نشان‌می‌دهد.

نمودار2: مقایسه روش پیشنهادی با روش‌های دیگر از نظر معیار Precision

همچنین با توجه به نمودار۲ روش پیشـــنهادی در مقایســـه از نظر معیار Precision نتیجه بهتری را نشـــان‌می‌دهد. علاوه بر این نمودار ۳ مقایسه روش پیشنهادی با تفاوت در روش نتیجه‌گیری را از نظر معیار F-measure نشان‌می‌دهد.



نمودار3: مقایسه روش پیشنهادی با الگوریتم‌های دیگر از نظر معیار F-mesure

۱۰. نتیجه‌گیری

هدف از پژوهش حاضر ارائه الگوریتمی برای گروه‌بندی انواع محصولات بیمه و کاربران در سیستم توصیه‌گر بیمه با خوشه‌بندی مبتنی بر فیلتر مشـــارکتی ارزیابی عملکرد آن بر اســـاس توصـــیه بیمه می‌باشـــد. فاز شنا سایی م شتریان برای ارتقا و مدیریت تقا ضای م شتری و همچنین ک سب دانش سازمانی درخ صوص م شتریان حال حا ضر بیمه ایران با انجام داده‌کاوی در داده‌های مشتریان این بانک با اجرای خوشه‌بندی به روش K-means و فیلترینگ م شارکتی صورت‌گرفته و اهمیت متغیرها به عنوان الگو در داده‌ها تلقی می شود. بیمه‌ها و مؤ سساتی که خدمات بیمه ارائه‌می‌دهند، با ا ستفاده از این الگوها و سرعت بالای پا سخگویی الگوریتم می‌توانند با سرعت بالاتر و دقت بیشتر در رابطه با شایستگی مشـــتریان خود برای ارائه خدمات بیمه تصـــمیم‌گیری کنند. بنابراین پاسخ سوال مثبت می‌باشد.

فیلترینگ مشارکتی از تکنیک‌های سیستم‌های توصیه‌گر است که در دنیای اقتصاد به‌ویژه در بخش بازاریابی و ارائه خدمات و محصولات از آن استفاده‌می‌شود. با ساتفاده از این تکنیک توانستیم میزان شباهت مشتریان و خدمات بیمه‌ای که خریداری‌می‌کنند را بررسی‌کنیم که منجر به بهبود شناخت مشتریان و نوع خدمات بیمه‌ای که استفاده می کنند، شد. با دقت بالایی که این تکنیک در اختیار می‌گذارد، مشتریان مشابه به‌خوبی شناسایی‌شده و باعث ارائه خدمات بیمه‌ای بهتر از سوی شرکت بیمه‌ای شده و در نتیجه کاهش ضرر در صنعت بیمه و افزایش سود می‌شود.

ویژگی‌های میزان رضایت، درآمد ماهانه مهمترین عناصر در بررسی شباهت م شتریان جهت دریافت خدمات بیمه می با شد. مدل حا صل شـــده می‌تواند ابزاری در بهبود ارائه خدمات بیمه به‌خصـــوص در فاز شناخت مشتریان باشد.

در این پژوهش تنها فیلترینگ مشارکتی مورد بررسی و پیاده‌سازی قرارگرفت. با وجود این می‌توان از الگوریتم‌های دیگر ســـیســـتم‌های تو صیه‌گر به شنا سایی م شتریان م شابه پرداخت. همچنین ا ستفاده از روش‌های ترکیبی هوش مصـنوعی برای سـنجش شـباهت مشـتریان و مقایسه آن‌ها پیشنهاد می‌شود.

References

[1] J. Neidhardt, T. Kuflik and W. Wörndl, “Special section on recommender systems in tourism,” Information Technology & Tourism, vol. 19, no. 1, pp. 83-85, 2018

[2] T.N. Nguyen and F. Ricci, “A chat-based group recommender system for tourism,” Information Technology & Tourism, vol. 18, no. 1, pp. 5-28, 2018.

[3] Bazargani, Mehdi, & Homayunpour, Zainab. (2019). Presenting a new method to discover the nearest neighbor in recommender systems based on collaborative filtering. Intelligent Multimedia Communication and Processing Systems, 1(1), 55-64.

[4] Manteghipour, Mahnaz, & Rahimkhani, Parisa. (1401). Designing a hybrid model for classification of imbalanced data in the field of Casualty Insurance. Intelligent multimedia processing and communication systems, 3(2), 1-9.

[5] M. Jalili, S. Ahmadian, M. Izadi, P. Moradi and M. Salehi, “Evaluating collaborative filtering recommender algorithms: a survey,” IEEE access, vol. 6, pp. 74003-74024, 2018.

[6] W.S. Lin, N. Cassaigne and T.C. Huan, “A framework of online shopping support for information recommendations,” Expert Systems with Applications, vol. 37, no. 10, pp. 6874-6884, 2010.
[7] L. Zhang, J. Zhu and Q. Liu, “A meta-analysis of mobile commerce adoption and the moderating effect of culture”, Computers in human behavior, vol. 28, no. 5, pp. 1902-1911, 2012.
[8] A. Salah, N. Rogovschi and M. Nadif, “A dynamic collaborative filtering system via a weighted clustering approach”, Neurocomputing, vol. 175, pp. 206-215, 2016.
[9] M. Taghizadeh, “A review of filtering methods in recommender systems”, 6th National Conference on Computer Science and Engineering and Information Technology, 859097, 2018. [Persian].
[10] M. Khamisi, S. Samiri, “Review and comparison of collaborative filtering algorithm in recommender systems”, International Conference on Science, Engineering, Technology and Technological Businesses, 903155, 2018. [Persian].
[11] F. Pajuelo-Holguera, J.A. Gómez-Pulido, F. Ortega and J.M. Granado-Criado, “Recommender system implementations for embedded collaborative filtering applications”, Microprocessors and Microsystems, vol. 73, p. 102997, 2020.
[12] S.K. Panda, S. K. Bhoi and M. Singh, “A collaborative filtering recommendation algorithm based on normalization approach”, Journal of Ambient Intelligence and Humanized Computing, vol. 11, no. 11, pp. 4643-4665, 2020.
[13] Yao, S., Halpern, Y., Thain, N., Wang, X., Lee, K., Prost, F., ... & Beutel, A. (2021). Measuring recommender system effects with simulated users. *arXiv preprint arXiv:2101.04526*.
[14] Teixeira, B., Martinho, D., Novais, P., Corchado, J., & Marreiros, G. (2022, August). Diabetic-Friendly Multi-agent Recommendation System for Restaurants Based on Social Media Sentiment Analysis and Multi-criteria Decision Making. In *EPIA Conference on Artificial Intelligence* (pp. 361-373). Cham: Springer International Publishing.
[15] Cui, Y. (2021). Intelligent recommendation system based on mathematical modeling in personalized data mining. *Mathematical Problems in Engineering*, *2021*, 1-11.

پی‌نوشت

[1] Recommendation System
[2] Feature
[3] Numeric
[4] Nominal
[5] Missing Values
[6] Feature Scaling
[7] Splitting the Dataset