

مقایسه تأثیر تحلیل احساسات و رتبه‌بندی کاربران بر عملکرد سیستم‌های پیشنهاددهنده

فاطمه عباسی^۱، آمنه خدیور^{۲*}

۱- استادیار، مدیریت فناوری اطلاعات، گروه مدیریت فناوری اطلاعات، دانشکده فناوری اطلاعات، موسسه آموزش عالی مهر البرز، تهران، ایران.

۲- دانشیار، مدیریت فناوری اطلاعات، گروه مدیریت، دانشکده علوم اجتماعی و اقتصاد، دانشگاه الزهراء، تهران، ایران.

پذیرش: ۱۴۰۰/۴/۹

دریافت: ۱۳۹۹/۱۱/۹

چکیده

در سال‌های گذشته، ظهور شبکه‌های اجتماعی منجر به افزایش توجه به سمت سیستم‌های پیشنهاددهنده مبتنی بر نظرها شده است. هدف از توسعه چنین سیستم‌هایی استفاده از اطلاعات ارزشمند نظرهای متنی کاربران در فرایند الگوسازی و ارائه پیشنهاد است. در محیط شبکه‌های اجتماعی به‌طور معمول سیستم‌های پیشنهاددهنده مبتنی بر پالایه نمودن مشارکتی برای ارائه توصیه به کاربران استفاده می‌شود. اساس کار این رویکرد، تجربه و نظر سایر افراد برای خرید اقلام و محصولات است. در این پژوهش سیستمی برای ارائه توصیه به کاربران برای خرید کتاب با ترکیب فیلترینگ مشارکتی و تحلیل احساسات ارائه شد. برای تحلیل احساسات از الگوهای ترکیبی برای استخراج عقاید نظرهای کاربران استفاده شد. در رویکرد ترکیبی از رأی‌گیری مبتنی بر وزن جهت الگوسازی استفاده گردیده است. الگو پیاده‌سازی شده بر نظرهای ۷۲۱۰ کاربر و خریدار کتاب تارنمای آمازون که از راه خزنده وب از تارنمای آمازون استخراج شده‌اند، ارزیابی شده است. برای ارائه توصیه به کاربران پس از تشکیل پروفایل اقلام، شباهت میان اقلام استخراج می‌شود و در انتها اقلام مشابه با محصولات که هر کاربر به



آن نمره خوبی داده است، به‌عنوان محصول پیشنهادی ارائه می‌شود. نتایج نشان می‌دهند تحلیل احساسات نظر کاربران بر پیشنهاد کالاهای موردعلاقه کاربر و عملکرد سیستم‌های پیشنهاددهنده تأثیر مثبتی دارد.

واژه‌های کلیدی: سیستم پیشنهاددهنده، ترجیحات، تحلیل احساسات، فیلترینگ مشارکتی.

۱- مقدمه

افزایش رقابت در حوزه کسب‌وکار منجر به اهمیت روزافزون برقراری، حفظ و گسترش روابط با مشتریان شده است. از سوی دیگر استفاده از فناوری اطلاعات در کسب‌وکارها، روش‌های جدیدی برای برقراری ارتباط میان مشتریان فراهم کرده است [۱]. حجم عظیمی از اطلاعات تارنماها در قالب‌های متنی و داده‌های نیمه‌ساختار نیافته می‌باشند که تجزیه و تحلیل این متون از اهمیت ویژه‌ای برخوردار است [۲]. به منظور تحلیل این داده‌ها نیاز به روش‌هایی است که بتوانند امکان دستیابی به داده‌ها و استخراج اطلاعات از آنها را فراهم کنند. برای استخراج دانش از واقعیت‌ها و عقاید از روش‌های متن‌کاوی و تحلیل احساسات استفاده می‌شود [۳]. تحلیل احساسات تشخیص، استخراج و طبقه‌بندی عقاید، احساسات و نگرش‌ها در موضوعات مختلف می‌باشد [۴-۵]. تحلیل احساسات در حوزه‌های مختلف از جمله انتخاب و خرید کتاب برای کتابخانه‌های دانشگاهی بسیار مؤثر است. شواهد نشان می‌دهند میزان مراجعه و استفاده از کتاب‌های کتابخانه‌های دانشگاه‌ها نسبت به هزینه انجام‌شده برای خرید کتاب به‌ویژه کتاب‌های لاتین بسیار پایین است. نتایج پژوهش تفرشی و دروگرکلخوران (۱۳۸۸) در ارتباط با میزان آشنایی و بهره‌گیری دانشجویان از کتابخانه‌های دانشگاهی نشان می‌دهند که سبک آموزش استادان، استفاده از رایانه به‌عنوان ابزار جستجو، عدم کفایت و به‌روزنبودن منابع از علل مراجعه‌نکردن دانشجویان به کتابخانه‌های دانشگاهی است [۶]. فرایند خرید کتاب برای اغلب کتابخانه‌های دانشگاهی به این صورت است که براساس نیازهای گروه‌های علمی خرید انجام می‌شود. با توجه به پایین‌بودن میزان مراجعه دانشجویان برای استفاده از منابع کتابخانه‌ای



نسبت به هزینه انجام‌شده، اصلاح فرایند و رویه خرید کتاب برای کتابخانه‌های دانشگاه‌های کشور ضروری به نظر می‌رسد. تحلیل احساسات یکی از ابزارهایی است که می‌تواند به اصلاح روند کمک شایانی کند. این مهم یکی از مشوق‌های اصلی پژوهش حاضر و آغاز فرایند پژوهش و مطالعه بود [۷].

امروزه بسیاری از سایت‌های تجارت الکترونیک به ارائه پیشنهاد محصولات به کاربران می‌پردازند [۸]. سیستم‌های پیشنهاددهنده اقلام مورد علاقه کاربران را براساس استخراج سلیقه‌ها و ترجیحات آنها پیشنهاد می‌دهند. هدف سیستم‌های پیشنهاددهنده در واقع رتبه‌بندی موضوع‌های سیستم به لحاظ نزدیک‌بودن به علایق کاربران است تا زمان ارائه پیشنهاد، آیتم‌هایی با رتبه بالاتر را به کاربر پیشنهاد دهند [۹].

پژوهش‌های انجام‌شده در داخل و خارج از کشور به طراحی سیستم پیشنهاد پرداخته‌اند [۱۲-۱۶]. با وجود این به سیستم‌های پیشنهاددهنده‌ای که مبتنی بر تحلیل احساسات خریداران و استفاده از روش‌های متن کاوی و استخراج عقاید واقعی کاربران باشد، کمتر توجه شده است و اغلب پژوهش‌های این حوزه بر مبنای داده‌های ساختاریافته می‌باشند. باتوجه به آنکه صحت خروجی تحلیل بر داده‌های ساختاریافته از اعتبار بالایی برخوردار است، لازم است بیشتر به آن توجه شود. باتوجه به خلأ شناسایی‌شده در این حوزه، پژوهش پیش‌رو به طراحی سیستم پیشنهاددهنده خرید مبتنی بر تحلیل احساسات خریداران می‌پردازد. در این مقاله سیستم پیشنهاددهنده‌ای براساس فیلترینگ مشارکتی و با ترکیب تحلیل احساسات ارائه شد. بخش دوم مروری بر پیشینه پژوهش‌های ارائه شده در این حوزه دارد. در بخش سوم روش‌شناسی و فرایند کلی طی‌شده در این پژوهش ارائه شده است. در بخش چهارم نیز یافته‌ها و دستاوردهای پژوهش و درنهایت در بخش آخر خلاصه و نتیجه‌گیری بیان شده است.

۲- پیشینه پژوهش

باتوجه به رشد روزافزون اطلاعات، داشتن یک سیستم هوشمند که قادر باشد علاقه‌های کاربران را یاد بگیرد و براساس این علایق به‌طور خودکار علایق غیرمرتبط را فیلتر کند یا اطلاعات مرتبط را در زمان کم به کاربر پیشنهاد دهد، ضروری است [۱۰]. هدف سیستم



پیشنهاددهنده در واقع رتبه‌بندی آیتم‌های سیستم به لحاظ نزدیک‌بودن به سلیقه‌های کاربران است تا در زمان پیشنهاد آیتم‌هایی با رتبه بالاتر را به کاربر پیشنهاد دهند.

تحلیل احساسات حوزه مطالعاتی است که عقاید، احساسات، ارزیابی‌ها و نگرش مردم را نسبت به موجودیت‌هایی چون محصولات، خدمات، سازمان‌ها، افراد، مسائل، رویدادها، موضوعات و ویژگی‌های آنها تجزیه و تحلیل می‌کند [۲].

سینگ^۱ و همکاران رویکرد جدیدی برای بهبود نتایج رویکرد مبتنی بر محتوا با تحلیل احساسات ترکیب ارائه کرده‌اند. در این پژوهش اقلامی که از جهت محتوایی یکسان هستند و نیز در طبقه‌بندی نظرها به‌عنوان مثبت برچسب‌گذاری شده‌اند، به کاربران پیشنهاد شده است [۱۱]. کومار^۲ و همکاران با ترکیب فیلترینگ مشارکتی و تحلیل احساسات، عملکرد سیستم پیشنهاددهنده فیلم را بهبود داده‌اند [۱۲]. کوریکوس^۳ و همکاران تحلیل احساسات را بر نظرهای کاربران در زمینه منابع آموزشی انجام داده‌اند که خروجی‌های این پژوهش به‌عنوان نرخ درک‌شده کاربر خاص در سیستم پیشنهاددهنده استفاده شد [۱۳]. گورنی^۴ و همکاران (۲۰۱۳) روش پیشنهاددهنده کاربر را با تابع عینیت اندازه احساسات ارائه کرده‌اند که در آن نه تنها علاقه کاربر بلکه احساساتش نیز در نظر گرفته شد [۱۴]. چن^۵ و همکاران (۲۰۱۵) بر چگونگی استفاده از نظرهای کاربران برای توسعه سیستم‌های پیشنهاددهنده مبتنی بر محتوا و مشارکتی مرور کلی داشته‌اند [۱۵]. الحمدي و زنگ (۲۰۱۵) پارچویی را برای استفاده از منبع داده‌ای جدید برای ارائه پیشنهادهای شخصی‌سازی شده با کاوش در پست‌های متنی دوستان در میکرو بلاگ‌ها ارائه می‌دهند [۱۶]. جایاشری و کولکارنی^۶ (۲۰۱۷) سیستم هوشمندی را برای پیشنهاد گزینه‌های مختلف جهت تسهیل تصمیم‌گیری ارائه کرده‌اند [۱۷].

همان‌طور که در بخش ادبیات اشاره شد، در داخل کشور بیشتر پژوهش‌های انجام‌شده در حوزه تحلیل احساسات است و ترکیب سیستم‌های پیشنهاددهنده و تحلیل احساسات مورد توجه قرار نگرفته است [۱۸-۱۹]، [۷].

1. Singh
2. Kumar
3. Koukourikos
4. Gurini
5. Chen
6. Jayashree and Kulkarni

درباره پژوهش‌های خارج از کشور نیز دو دسته مقاله به چشم می‌خورد، برخی تنها به بررسی تحلیل احساسات پرداخته‌اند و دسته دیگر به ترکیب سیستم‌های پیشنهاددهنده و تحلیل احساسات پرداخته‌اند. در بخش تحلیل احساسات، مقاله‌های بیشتری به دو رویکرد یادگیری ماشین و رویکرد مبتنی بر واژه‌نامه پرداخته‌اند. در رویکرد نظارت‌شده نیز بیشتر مقاله‌ها به بررسی یادگیری نظارت‌شده و مقایسه نتایج دو تا سه الگوریتم پرداخته‌اند که براساس نتیجه الگوریتم با دقت بالاتر به الگوسازی تحلیل احساسات پرداخته‌اند. همان‌طور که از بررسی مقاله‌ها مشخص است، رویکرد ترکیبی برای تحلیل احساسات مورد توجه نبوده است [۱۱-۱۷]. با توجه به آنکه رویکردهای ترکیبی در تحلیل احساسات می‌تواند منجر به نتایجی با دقت بالاتر شود، لازم است به این روش بیشتر توجه شود. بررسی پژوهش‌ها نشان می‌دهد ترکیب سیستم پیشنهاددهنده و رویکرد ترکیبی برای تحلیل احساسات بررسی نشده است که می‌تواند موضوع بررسی‌های بیشتر در این زمینه باشد.

۳- روش پژوهش

جامعه هدف این پژوهش، داده‌های یک شرکت فروش آنلاین است. در این مقاله از روش‌های ترکیبی برای استخراج نظرها و احساسات نظرهای کاربران و از فیلترینگ مشارکتی برای توسعه سیستم پیشنهاددهنده استفاده شد که در شکل ۱ مراحل و رویکردهای هر مرحله نشان داده شده است.



شکل ۱. چارچوب پژوهش

۴- جمع‌آوری داده‌ها

برای جمع‌آوری داده‌ها از یک نرم‌افزار خزننده وب استفاده شده است. با استفاده از نرم‌افزار مذکور ۷۲۱۰ کامنت و سایر اطلاعات مرتبط با کتاب‌هایی با موضوع هوشمندی کسب‌وکار استخراج شده‌اند.

۵- آماده‌سازی داده‌ها

در متن کاوی، پیش‌پردازش داده‌ها برای استخراج دانش از داده‌های متنی بدون ساختار استفاده می‌شود [۲۰]. متن‌های آنلاین اغلب شامل بخش‌های نامطلوب و فاقد ارزش^۱ می‌باشند. برخی از کلمه‌ها نیز در متن بر جهت‌گیری کلی متن تأثیری ندارند که نگهداشتن این بخش‌ها و کلمات باعث ایجاد مشکلاتی در پردازش و دقت تحلیل می‌شود و از این جهت پیش‌پردازش داده‌ها برای افزایش کیفیت و دقت تحلیل بسیار ضروری است [۲۱]. در این پژوهش از نشانه‌گذاری کلمه‌ها^۲، جایگزینی کلمات بی‌معنا^۳، برچسب‌گذاری اجزای واژگانی کلام^۴، حذف ایست‌واژه‌ها^۵ و ریشه‌یابی کلمه^۶ برای آماده‌سازی داده‌ها استفاده شد.

نشانه‌گذاری کلمه‌ها: نشانه‌گذاری تقسیم‌بندی متن به واحدهای کوچک‌تری چون کلمات، اصطلاحات و نمادها است که در اصطلاح نشانه‌نامیده می‌شوند [۲۲]. در این پژوهش تمام نظرها به واحدهای کوچک‌تری تقسیم و درنهایت کلمات به‌دست آمده از متن‌ها به‌صورت متغیرهای مجزا در نظر گرفته شدند.

جایگزینی کلمه‌های بی‌معنی: کلمه‌های بی‌معنی اشاره به اصطلاحات کوتاه و اختصارات دارند [۲۳]. در این پژوهش برای جایگزینی این کلمه‌های بی‌معنی فهرستی از اختصارها تهیه شد تا در صورت وجود این کلمه‌ها در متن، کلمه اصلی جایگزین این کلمه‌ها شوند.

1. Noise and uninformative parts
2. Tokenization
3. Noisy data
4. Part of Speech tagging (POS tagging)
5. Stop-word
6. Stemming



حذف ایست واژه‌ها: ایست واژه‌ها کلماتی هستند که با وجود تکرار فراوان در متن از جهت معنایی اهمیت کمی دارند. این کلمات حاوی اطلاعات نیستند مانند ضمائر، حروف اضافه و حروف ربط [۲۴].

برچسب‌گذاری اجزای واژگانی کلام: برچسب‌گذاری اجزای واژگانی کلام به هر کلمه‌ای براساس نقش دستوری در متن برچسبی چون فعل، صفت یا قید را تخصیص می‌دهد. برچسب‌گذاری اجزای واژگانی کلام برای استخراج ویژگی‌ها از نظر دقت بسیار کارآمد است [۲۵].

ریشه‌یابی کلمه‌ها: در این مرحله تمام کلمه‌ها به فرمت ریشه اصلی خود در می‌آیند. الگوریتم‌های ریشه‌یابی پسوند کلمه براساس قوانین گرامری حذف می‌کند. در این پژوهش از کتابخانه ریشه‌یاب گلوله برفی^۱ که محبوب‌ترین و استانداردترین رویکرد است، استفاده شد [۲۶].

۶- الگوسازی

برای الگوسازی از روش ترکیبی مبتنی بر چند الگوریتم طبقه‌بندی‌کننده استفاده شد که این الگوریتم‌ها برای دسته‌بندی کردن هر نظر با یکدیگر همکاری می‌کنند. در این پژوهش هفت دسته‌بندی‌کننده بیز ساده^۲، طبقه‌بندی‌کننده بردار پشتیبان^۳، بیزین ساده چندجمله‌ای^۴، پرسپترون چندلایه^۵، بیز ساده برنولی^۶، رگرسیون لجستیک^۷ و طبقه‌بندی‌کننده بردار پشتیبان^۸ Nu از کتابخانه جعبه ابزار زبان طبیعی^۹ پایتون جهت ایجاد الگو استفاده شده‌اند. این الگوریتم‌ها جزء دسته‌بندی‌کننده‌هایی هستند که در روش یادگیری نظارت شده برای پیش‌بینی استفاده می‌شوند. رویکرد یادگیری ماشین نظارت‌شده، نوعی از یادگیری است که در آن ورودی و خروجی

1. the Snowball stemmer library
2. Naïve Bayes
3. Linear Support Vector Classification (Linear SVC)
4. multinomial naive Bayes (MNB)
5. Multilayer perceptron (MLP)
6. Bernoulli Naive Bayes (BernoulliNB)
7. Logistic Regression
8. Nu - Vector Classification (Nu - SVC)
9. Natural Language Toolkit (NLTK)



مشخص است و ناظری وجود دارد که اطلاعاتی را در اختیار یادگیرنده قرار می‌دهد و به این ترتیب سیستم سعی می‌کند تا تابعی را از ورودی به خروجی فراگیرد [۲۷]. در نهایت نتیجه نهایی هر نظر با استفاده از روش وزن‌دهی در رویکرد رأی‌گیری مشخص می‌شود. رویکرد ترکیبی مبتنی بر ترکیب مجموعه‌ای از الگوهای دسته‌بندی است که ترکیب آنها صحت عملکرد بیشتری را نسبت به استفاده منفرد و مستقل از آنها در اختیار کاربر خواهد گذاشت [۲۸]. در روش رأی‌گیری مبتنی بر وزن، وزن رأی دسته‌بندی‌کننده‌های مختلف متفاوت است و وزن بالاتر به دسته‌بندی‌کننده‌ای که عملکرد بهتری داشته باشد، تخصیص پیدا می‌کند [۲۹]. در این پژوهش براساس دقت هفت دسته‌بندی‌کننده، وزن‌های متفاوتی تخصیص پیدا می‌کند و الگو مبتنی بر رأی‌گیری براساس وزن‌های مختلف ارزیابی می‌شوند.

۷- ارزیابی الگو برای تحلیل احساسات

پس از پیاده‌سازی الگو با استفاد از رویکرد ترکیبی لازم است دقت، بازخوانی و معیار ارزیابی F تخمین زده شود. دقت و بازخوانی معیارهای کاربردی در حوزه ارزیابی اطلاعات هستند که میزان تناسب اسناد ارزیابی شده به وسیله سیستم را با نیاز کاربر تعیین می‌کنند [۳۰]. معیارهای ارزیابی دسته منفی:

$$\text{رابطه (۱) بازخوانی} = \frac{TN}{FP+TN}$$

$$\text{رابطه (۲) دقت} = \frac{TN}{TN+FN}$$

$$\text{رابطه (۳) معیار F} = \frac{2*Recall*Precision}{Recall+Precision}$$

معیارهای ارزیابی دسته مثبت:

$$\text{رابطه (۴) بازخوانی} = \frac{TP}{FN+TP}$$

$$\text{رابطه (۵) دقت} = \frac{TP}{TP+FP}$$

$$\text{رابطه (۶) معیار F} = \frac{2*Recall*Precision}{Recall+Precision}$$

برای کاهش خطا در الگوسازی برای جداسازی داده‌های آموزش و آزمون از روش اعتبارسنجی چندلایه‌ای استفاده شد. در این روش کل مجموع داده‌ها به K قسمت مساوی



تقسیم می‌شود. از این K زیرمجموعه، هر بار یکی برای آزمون و K-1 زیرمجموعه باقیمانده برای آموزش به کار گرفته می‌شوند. به این ترتیب فرایند مذکور K بار تکرار می‌شود [۳۱-۳۲]. معمول‌ترین مقداری که برای K در متون علمی در نظر گرفته می‌شود برابر با ۱۰ است [۳۴-۳۳].

۸- تحلیل احساسات

پس از ارزیابی الگو، تحلیل احساسات روی ۷۲۱۰ نظر گردآوری شده از تارنمای آمازون انجام می‌شود و نظرهای هر کاربر نسبت به هر کتاب، یعنی نظر مثبت یا منفی کاربران به کتاب‌ها مشخص می‌شود.

۹- شباهت میان اقلام

با استفاده از نمره‌های به‌دست‌آمده از تحلیل احساسات (عددی بین -۱ تا +۱) و نمره‌هایی که کاربر به محصولات تخصیص داده (عددی انتخابی از بین اعداد ۱، ۲، ۳، ۴، ۵)، ماتریس کاربر - قلم تشکیل می‌شود. یکی از بخش‌های حیاتی در سیستم‌های توصیه‌کننده مبتنی بر فیلترینگ مشارکتی محاسبه شباهت میان اقلام و انتخاب شبیه‌ترین اقلام و کالاها می‌باشد. ایده اولیه در محاسبه شباهت میان دو قلم i و j، جداسازی کاربرانی است که این دو قلم را رتبه‌بندی کرده‌اند و پس از آن شباهت دو قلم i, j با استفاده از روش‌های مختلفی محاسبه می‌شود. در سیستم‌های پیشنهاددهنده مبتنی بر فیلترینگ مشارکتی، روش‌های مختلفی برای محاسبه میان هریک از اقلام وجود دارد که یکی از محبوب‌ترین این روش‌ها شباهت مبتنی بر همبستگی^۱ است. در این رویکرد شباهت میان دو قلم i و j با استفاده از روش همبستگی پیرسون^۲ و فرمول ۱۰ تخمین زده می‌شود.

$$\text{Sim}(i, j) = \frac{\sum_{u \in U} (R_{u,i} - \bar{R}_i)(R_{u,j} - \bar{R}_j)}{\sqrt{\sum_{u \in U} (R_{u,i} - \bar{R}_i)^2} \sqrt{\sum_{u \in U} (R_{u,j} - \bar{R}_j)^2}} \quad (\text{رابطه ۷})$$

1. correlation-based similarity
2. Pearson



که در آن $R_{u,i}$ نمره کاربر u به قلم i ، \bar{R}_i میانگین نمره‌های قلم i ، u کاربری که به قلم i و j امتیاز داده است و متعلق به مجموعه کاربران U می‌باشد. در این پژوهش با استفاده از روش همبستگی پیرسون براساس امتیاز کاربر - قلم شباهت میان اقلام محاسبه شده، سپس نتایج این مرحله (شباهت مبتنی بر فیلترینگ مشارکتی) در پیش‌بینی ترجیحات کاربر استفاده می‌شوند [۳۳]

۱۰- ارائه پیشنهاد

مهم‌ترین بخش در سیستم‌های پیشنهادی مبتنی بر فیلترینگ مشارکتی، ارائه پیشنهاد است. پس از شناسایی اقلام مشابه، در این مرحله با توجه به شباهت اقلام، کالا و محصولات جدید به کاربر هدف پیشنهاد می‌شود. با محاسبه مجموع وزنی امتیازهای داده شده به وسیله کاربر، ترجیحات کاربر هدف پیش‌بینی می‌شوند.

۱۱- ارزیابی سیستم پیشنهاددهنده

یکی از مراحل مهم برای اطمینان از اینکه یک سیستم پیشنهاددهنده به خوبی قابل تعمیم است، ارزیابی عملکرد آن است. معیارهای شناخته شده‌ای برای بررسی دقت یا عملکرد سیستم‌های پیشنهاددهنده وجود دارد؛ معیارهایی مانند صحت، پوشش و معیار F (به منظور ایجاد توازن میان این دو معیار) که می‌توانند به خوبی عملکرد پیشنهاددهنده را نشان دهند [۳۴].

صحت: نشان‌دهنده این است که چند درصد از مجموعه پیشنهادها درست است و درستی پیشنهادها را می‌سنجد. منظور از اقلام مطلوب^۱ مجموعه‌ای از کالاهایی است که باید به کاربر پیشنهاد داده شوند و اقلام پیشنهادی^۲ مجموعه کالاهایی است که به وسیله سیستم به کاربر پیشنهاد شده است [۳۵].

-
1. Favorite Items
 2. Recommended Items



$$\frac{\text{Recommended Items} \cap \text{Favorite Items}}{\text{Recommended Items}} = \text{صحت} \quad \text{رابطه ۸}$$

پوشش: نشان‌دهنده این است که چند درصد از کالاهای موردعلاقه کاربر به او پیشنهاد شده است.

$$\frac{\text{Recommended Items} \cap \text{Favorite Items}}{\text{Favorite Items}} = \text{پوشش} \quad \text{رابطه ۹}$$

معیار - F: میانگین هارمونیک دقت و پوشش است [۳۶].

$$\frac{\text{دقت} * \text{پوشش} * 2}{\text{دقت} + \text{پوشش}} = F - \text{معیار} \quad \text{رابطه ۱۰}$$

۱۲- یافته‌ها

پس از تبدیل داده‌های متنی که ساختارنیافته هستند به داده‌های ساختاریافته با استفاده از روش برچسب‌گذاری اجزای واژگانی کلام، کلمه‌ها و واژگان برای استفاده در الگوسازی استخراج شدند. باتوجه‌به هدف پژوهش (که الگوسازی برای تحلیل احساسات است) در این مرحله کلمه‌های با نقش فعل، قید و صفت جهت الگوسازی انتخاب شدند. برای ارزیابی هریک از هفت الگوریتم از معیارهای دقت، بازخوانی، معیار F، نرخ دقت و نرخ خطا استفاده شد که در جدول ۱ دقت هر یک از الگوریتم‌ها ارائه شده است. نتایج جدول ۱ از ده بار تکرار الگو هریک از دسته‌بندی‌کننده‌ها حاصل شد. نتایج نشان می‌دهند دسته‌بندی‌کننده بی‌زین ساده چندجمله‌ای با نرخ دقت ۰/۷۹۱۱، دقت بالاتری نسبت به الگوریتم‌ها برای الگوسازی دارد.

جدول ۱ ارزیابی الگوریتم‌های دسته‌بندی کننده

الگوریتم	دقت	بازخوانی	معیار F	نرخ دقت	نرخ خطا
بیز ساده	۰/۶۹	۰/۶۷۵	۰/۶۷۸	۰/۷۰۴	۰/۲۹۵
طبقه‌بندی کننده بردار پشتیبان	۰/۷۴۴	۰/۷۴۳	۰/۷۴۳	۰/۷۵۵۵	۰/۲۴۳
بیزین ساده چندجمله‌ای	۰/۷۸۳	۰/۷۷۴	۰/۷۷۷	۰/۷۹۱۱	۰/۲۰۷
پرسپترون چندلایه	۰/۷۵۳	۰/۷۴۸	۰/۷۴۷	۰/۷۵۹۹	۰/۲۳۹
بیز ساده برنولی	۰/۶۶۸	۰/۶۴۷	۰/۶۵۰	۰/۶۸۳۶	۰/۳۱۵
رگرسیون لجستیک	۰/۷۸۳	۰/۷۷۴	۰/۷۷۷	۰/۷۹۰۷	۰/۲۰۸
طبقه‌بندی کننده بردار پشتیبان Nu	۰/۷۶۱	۰/۷۵۲	۰/۷۵۳	۰/۷۶۸	۰/۲۳۱

پس از آنکه نتایج هریک از الگوریتم‌ها مشخص شد، از روش رأی‌گیری مبتنی بر وزن‌دهی برای الگو ترکیبی استفاده شد. در جدول ۲ نتایج ارزیابی رویکرد ترکیبی ارائه شده است. همان‌طور که نتایج نشان می‌دهند، دقت و عملکرد رویکرد ترکیبی نسبت به هریک از دسته‌بندی کننده‌ها بهتر است. دلیل این امر پیش‌بینی صحیح‌تر در نواحی است که دسته‌بندی کننده‌ها نتوانستند عملکرد مناسبی در آن بخش‌ها داشته باشند. نتایج نشان می‌دهند رویکرد ترکیبی خطای دسته‌بندی کننده‌ها را جبران می‌کند و الگویی با دقت بالاتر را توسعه می‌دهد.

جدول ۲. ارزیابی رویکرد ترکیبی رأی‌گیری مبتنی بر وزن‌دهی

رویکرد ترکیبی رأی‌گیری مبتنی بر وزن‌دهی	دقت	بازخوانی	معیار F	نرخ دقت	نرخ خطا
	۰/۷۹۵۸	۰/۷۸۰۴	۰/۷۸۵۴	۰/۸۰۰۳	۰/۱۹۸۹

در گام بعد الگو پیشنهادی بر ۷۲۱۰ کامنت مربوط به ۲۰۰ عنوان کتاب مرتبط با موضوع هوشمندی کسب‌وکار منتج از تارنمای آمازون پیاده‌سازی شده و نظرهای مثبت و منفی هر نظردهنده نسبت به هر کتاب مشخص می‌شود.

پس از تحلیل نظرهای کاربران ماتریس کاربر-قلم تشکیل می‌شود. در این مرحله شباهت میان اقلام براساس سه نوع داده، یعنی امتیازهای کاربر به کتاب‌ها (عددی انتخابی از بین اعداد ۱، ۲، ۳، ۴، ۵)، عدد به‌دست‌آمده از تحلیل احساسات نظرهای کاربران (عددی بین -۱ و +۱) و



ترکیب امتیاز کاربر و امتیاز تحلیل احساسات ارزیابی می‌شود. پس از تخمین شباهت میان اقلام، اقلام شبیه به اقلامی که پیش‌تر کاربر در مورد آنها نظر داده است، به کاربر پیشنهاد می‌شود. برای ترکیب امتیاز کاربر و امتیاز به‌دست‌آمده از تحلیل احساسات، امتیازهای تحلیل احساسات نرمال شد تا در بازه ۰ تا ۵ قرار گیرد. برای ترکیب این دو امتیاز از روش میانگین‌گیری استفاده شد. در این گام، امتیاز پیشنهادها براساس مجموع وزنی تخمین زده می‌شود، برای مثال اگر کاربر به کتابی امتیاز ۵ داده باشد و شباهت میان این کتاب و کتاب دیگری که کاربر نظردهی نکرده است، $0/9$ باشد، کتاب جدید با امتیاز $4/5$ به کاربر پیشنهاد می‌شود. در این روش کتاب‌هایی به کاربر پیشنهاد می‌شود که هم امتیاز بالا و هم امتیاز بالایی به‌وسیله کاربر کسب کرده باشند. در جدول ۳ نمونه‌ای از پیشنهادها به کاربران ارائه شد.

جدول ۳. نمونه‌ای از پیشنهادها به کاربران

نام کاربر	کتاب امتیاز داده‌شده به‌وسیله کاربر	کتاب‌های پیشنهادی	امتیاز کتاب پیشنهادی
Wanderley Oliveira Mendes	Predictive Analytics: The Power to Predict Who Will Click, Buy, Lie, or Die	Data Smart: Using Data Science to Transform Information into Insight	+4/745678
		Emotional Intelligence ۲.۰	+2/543639
		Emotional Intelligence: Why It Can Matter More Than IQ	-1/017456

برای ارزیابی سیستم، کتاب‌هایی که امتیاز بالای ۴ داشته باشند، برای ارزیابی انتخاب می‌شوند و نمره‌های پیشنهادی با امتیاز واقعی کاربر مقایسه می‌شود تا میزان دقت سنجیده شود. در این مرحله دقت نتایج سیستم پیشنهاددهنده با مقادیر مختلف امتیازها با استفاده از معیارهای صحت، پوشش و معیار F با یکدیگر مقایسه می‌شوند (جدول ۴).



جدول ۴. مقایسه شاخص‌های سیستم پیشنهاددهنده براساس سه نوع داده

معیار- F	پوشش	صحت	معیار داده‌های استفاده شده
۰/۷۴۸۹	۰/۸۳۶۷۲	۰/۶۷۷۸۲	امتیازهای کاربر
۰/۷۱۶۲	۰/۹۳۱۲۴	۰/۵۸۱۹۰	امتیاز به دست آمده از تحلیل احساسات
۰/۶۹۲۳	۰/۹۲۱۴۸	۰/۵۵۴۴	ترکیب امتیاز کاربر و تحلیل احساسات

نتایج جدول ۴ نشان می‌دهد که سیستم‌های پیشنهاددهنده مبتنی بر تحلیل احساسات درصد بیشتری از کالاهای مورد علاقه کاربر را پیشنهاد می‌دهد و تأثیر مثبتی بر عملکرد سیستم‌های پیشنهاددهنده دارد.

۱۳- بحث و نتیجه‌گیری

در این پژوهش به ارائه سیستم پیشنهاددهنده خرید کالا با استفاده از تحلیل احساسات خریداران پرداخته شد. نوآوری ساختاری و مفهومی و همچنین عمده تفاوت پژوهش حاضر با سایر پژوهش‌ها می‌تواند به آن اشاره کرد. استفاده از متن کاوی و تحلیل احساسات با هدف دستیابی به نظرها و نگرش‌های واقعی مشتریان به جای ابزارهای مرسوم چون پرسشنامه و مصاحبه است. همچنین از نتایج حاصل از تحلیل احساسات در طراحی سیستم پیشنهاددهنده استفاده شده است که وجه تمایز مهم این پژوهش از پژوهش‌های پیشین می‌باشد. به طور خلاصه مراحل انجام پژوهش به صورت زیر است:

در این پژوهش ارائه پیشنهادها براساس سه نوع داده انجام شده است. این داده‌ها به ترتیب نمره‌های کاربر به هر کتاب که به صورت کمی و دارای مقداری از پیش تعیین شده از اعداد ۱، ۲، ۳، ۴، ۵ می‌باشد. دسته دوم نمره‌های تخمینی براساس تحلیل احساسات بر نظراتی متن کاربران نسبت به هر کتاب است. دسته سوم داده‌ها، میانگین امتیاز کاربر و نمره تحلیل احساسات می‌باشند. نتایج نشان می‌دهند که تحلیل داده‌های متن و ساختار نیافته بر درصد پیشنهادهای مورد علاقه کاربران و عملکرد سیستم پیشنهاددهنده تأثیر مثبتی دارد. با استفاده از نتایج پژوهش می‌توان استدلال کرد که به کارگیری نظرات کاربران بر ارائه پیشنهاداتی هم‌سو با



علاقه کاربر آن تأثیر مثبتی دارد. در واقع نظرهای کاربران بیان‌کننده احساسات و عقاید کاربران در مورد محصولات هستند و می‌توانند مبنای دقیق‌تری برای ارائه پیشنهادهای دقیق‌تر به کاربران باشند.

در اغلب پژوهش‌هایی که تاکنون انجام شده‌اند، یک دسته‌بندی برای ارائه الگو تحلیل احساسات مدنظر قرار گرفته است [۱۱؛ ۱۲؛ ۱۵]. حال آنکه در پژوهش حاضر از رویکردهای ترکیبی که حاصل ترکیب هفت دسته‌بندی‌کننده بیز ساده، طبقه‌بندی‌کننده بردار پشتیبان، بیزین ساده چندجمله‌ای، پرسپترون چندلایه، بیز ساده برنولی رگرسیون لجستیک و طبقه‌بندی‌کننده بردار پشتیبان Nu استفاده شد. در نهایت نتیجه نهایی هر نظر با استفاده از دو روش رأی اکثریت و وزن‌دهی در رویکرد رأی‌گیری مشخص شده است که در پژوهش‌های پیشین سابقه نداشته است. همچنین الگو ارائه‌شده در این پژوهش می‌تواند در سایر حوزه‌ها نیز استفاده شود و با کمترین تغییرات به تحلیل احساسات نظرهای کاربران پردازد.

انجام این پژوهش با محدودیت‌هایی مواجه بوده است که در ادامه به برخی از این محدودیت‌ها اشاره می‌شود. در پژوهش حاضر اقلام اطلاعاتی کاربر مانند اطلاعات جمعیت‌شناختی و خرید کاربران در دسترس نبود. همچنین سیستم حاضر تنها می‌تواند به ارائه پیشنهاد برای مشتریانی پردازد که سابقه استفاده از محصول را داشته‌اند، به عبارتی مشتری فعلی محصول هستند. در صورتی که اطلاعات مربوط به اقلام و کاربران افزایش پیدا کند، نمونه توسعه یافته می‌تواند به کاربران تازه‌وارد نیز به ارائه پیشنهاد پردازد. در ادامه با توجه به تجربه پژوهشگران در انجام فرایند و همین‌طور بررسی محدودیت‌های پژوهش، پیشنهادهایی ارائه می‌شود. در اینجا پیشنهاد می‌شود روش پیشنهادی با مجموعه داده‌های صنایع دیگر جهت تعمیم‌پذیری اجرا شود. همچنین پیشنهاد می‌شود فاکتور زمان برای دقیق‌تر شدن پیشنهادها در طراحی سیستم‌های آینده استفاده شود. از این‌رو به پژوهشگران توصیه می‌شود سیستم پیشنهاددهنده گروهی مبتنی بر تحلیل احساسات با استفاده از یادگیری عمیق را ارائه کنند.

۱۴- منابع

- [1] Yu Y., Duan W., Cao Q "The impact of social and conventional media on firm equity value: A sentiment analysis approach", *Decision Support Systems*, 2013, 55(4): 916-926.
- [2] Liu B. "Sentiment analysis and opinion mining", *synthesis lectures on human language technologies*, Williston: Morgan & Claypool Publishers, 2012, 5(1): 1-67.
- [3] Radovanović M., Ivanović M. "Text mining: Approaches and Applications", *Novi Sad J. Math*, 2008, 38(3): 227-234.
- [4] Yan G., He W., Shen J., Tang C. "A bilingual approach for conducting Chinese and English social media sentiment analysis", *Computer Networks*, 2014, 75: 491-503.
- [5] Williams L., Bannister C., Arribas-Ayllon M., Preece A., Spasic I. "The role of idioms in sentiment analysis", *Expert Systems with Applications*, 2015, 42(21): 7375-7385.
- [۶] تفرشی ش.، دروگر کلخوران س. «بررسی میزان آشنایی و بهره‌گیری دانشجویان دانشگاه آزاد اسلامی واحد کرج از کتابخانه‌ها»، فصلنامه دانش‌شناسی، ۱۳۸۸، صص ۳۳-۴۴.
- [۷] عباسی ف.، سهرابی ب.، مانیان ا.، خدیور آ. «ارائه مدلی برای پیش‌بینی موضوعات مرتبط با هوشمندی کسب‌وکار»، مدیریت اطلاعات، ۱۳۹۷.
- [8] Schafer J., Konstan J., Riedl J. "Recommender systems in e-commerce", *Proceedings of the 1st ACM conference on Electronic commerce*, 1999, pp.158-166.
- [9] Aggarwal C. *Recommender Systems: The Textbook*, NY: Springer, 2016.
- [۱۰] مهران‌فر آ.، رستگارپور م. سیستم‌های پیشنهاددهنده در شبکه‌های اجتماعی، اهواز: انتشارات علوم و فنون پزشکی اهواز، ۱۳۹۰.
- [11] Singh V., Mukherjee M., Mehta G. "Combining a content filtering heuristic and sentiment analysis for movie recommendations", *Springer-Verlag Berlin Heidelberg*, 2011, pp. 659-664.
- [12] Kumar Singh V., Mukherjee M., Kumar Mehta G. "Combining collaborative filtering and sentiment classification for improved movie recommendations", *International Workshop on Multi-disciplinary Trends in Artificial Intelligence*, 2011, pp. 38-50.
- [13] Koukourikos A., Stoitsis G., Karampiperis P. "Sentiment analysis: A tool for rating attribution to content in recommender systems", *2nd Workshop on Recommender Systems for Technology Enhanced Learning*, 2012, pp. 61-70.



- [14] Gurini D., Gasparetti F., Micarelli R., Tre R. *A sentiment-based approach to twitter user recommendation*, *RecSys*, 2013.
- [15] Chen L., Chen G., Wang F. " Recommender systems based on user reviews: The state of the art", *User Model User-Adap Inter* , 2015, 25(2): 99-154.
- [16] Alahmadi D., Zeng X.-J. "ISTS: Implicit social trust and sentiment based approach to recommender systems", *Expert Systems With Applications*, 2015, 42(22): 8840-8849.
- [17] Jayashree R., Kulkarni D. "Recommendation system with sentiment analysis as feedback component", *Proceedings of Sixth International Conference on Soft Computing for Problem Solving*, 2017, pp. 359-367.
- [۱۸] طیرانی نجاران ن.، جلالی م. «ارائه رویکرد ترکیبی نوین جهت متن کاوی تحلیل احساسات در توییتر با استفاده از درخت تصمیم CART»، *دوفصلنامه مدیریت مهندسی و رایانس نرم*، ۱۳۹۷، ۴(۱): ۵۹-۷۶.
- [۱۹] محمدی ع.، یزدیان دهکردی م.، نعمت‌بخش م. «تشخیص ویژگی‌های ضمنی با استفاده از قواعد نحوی زبان فارسی و خوش‌بندی صفات»، *مجله مهندسی برق دانشگاه تبریز*، ۱۳۹۹، ۵(۳): ۱۳۹۵-۱۴۰۴.
- [20] Gurusamy V., Kannan S. *Preprocessing techniques for text mining*, *RTRICS.Podi*, 2014.
- [21] Haddi E., Liu X., Shi Y. "The role of text pre-processing in Sentiment Analysis", *Procedia Computer Science*, 2013, pp. 26-32.
- [22] Katariya N. P., Chaudhari M. S. " Text preprocessing for GText mining used side information", *International Journal of Computer Science and Mobile Applications*, 2015, pp. 1-5.
- [23] Stavrianou A., Andritsos P., Nicoloyannis N. "Overview and semantic issues of text mining", *ACM SIGMOD Record*, 2007, 36(3): 23-34.
- [24] Ramasubramanian C., Ramya R. "Effective pre-processing activities in text mining using improved porter's stemming algorithm", *International Journal of Advanced Research in Computer and Communication Engineering*, 2013, 2(12): 4536-4538.
- [25] Zubair Asghar M., Khan A., Ahmad S., Kundi F. A "Review of Feature Extraction in Sentiment Analysis", *Journal of Basic and Applied Scientific Research*, 2015, pp. 181-186.
- [26] Krouska A., Troussas C., Virvou M. *The effect of preprocessing techniques on Twitter*, Chalkidiki .Greece: *IEEE*, 2016.

- [27] Medhat W., Hassan A., Korashy H. "Sentiment analysis algorithms and applications: A survey", *Ain Shams Engineering Journal*, 2014, 5(4), pp.1093-1113.
- [۲۸] الهی ش.، خدیور آ.، حسن‌زاده ع. طراحی یک سیستم خبره تصمیم‌یار برای کمک به فرآیند ایجاد استراتژی مدیریت دانش، نشریه مدیریت فناوری اطلاعات، ۱۳۹۰، صص. ۴۳-۶۲.
- [29] Zhang Y., Zhang H., Cai J., Yang B. (2014) "A weighted voting classifier based on differential evolution", *Abstract and Applied Analysis*. 2014. بازبایی از <http://dx.doi.org/10.1155/2014/376950>.
- [30] Raghavan V., Bollmann P., Jung G.S. "A critical investigation of recall and precision as measures of retrieval system Performance", *ACM Transactions on Information Systems*, 1989, 7(3): 206-229.
- [۳۱] نبوتی ا.، عزیزی ا.، عباسی ا.، وکیلی ارکی ح.، زارعی ج.، رضوی ا. (۱۳۹۲) «کاربرد داده‌کاوی در پیش‌بینی مرگ بیماران سوختگی: مقایسه عملکرد چند الگوریتم»، مدیریت اطلاعات سلامت، ۱۳۹۲، صص. ۷۷۹-۷۸۹.
- [۳۲] صنیعی‌آباده م.، محمودی س. داده‌کاوی کاربردی، تهران: نیاز دانش، ۱۳۹۴.
- [33] Kohavi R. (1995) "A study of cross validation for accuracy estimation and model selection", *Appears in the International Joint Conference on Artificial Intelligence (IJCAI)*, 1995, 14(2).
- [34] Bradford JP, Brodley CE. "The effect of instance-space partition on significance", *Machine Learning*, 2001, 42(3): 269-286.
- [35] Sarwar B., Karypis G., Konstan J., Riedl J. *Item-based collaborative filtering recommendation algorithms*, 10th Proceedings of the 10th international conference on World Wide Web, 2001, pp. 285-295.
- [36] Hassan M., Hamada A. "Neural networks approach for improving the accuracy of multi-criteria recommender systems", *Appl. Sci*, 2017, 7(9).