



پژوهش‌های نوین در تصمیم‌گیری

دوره ۶، شماره ۲، تابستان ۱۴۰۰، صص ۴۷-۲۰

نوع مقاله: پژوهشی

طراحی سیستم رتبه‌بندی محصولات خریداری‌شده براساس نظرات مصرف‌کنندگان و متخصصین با استفاده از رویکرد تحلیل احساسات مبتنی بر ویژگی

امید افشاری‌زاده جعفری^۱، مرتضی ملکی مین باش رزگاه^{۲*}، عظیم زارعی^۳،
محسن شفیعی نیک آبادی^۴

۱. دانشجوی دکتری، دانشکده اقتصاد، مدیریت و علوم اداری، دانشگاه سمنان، سمنان، ایران

۲. دانشیار، گروه مدیریت بازرگانی، دانشکده اقتصاد، مدیریت و علوم اداری، دانشگاه سمنان، سمنان، ایران

۴. دانشیار، گروه مدیریت صنعتی، دانشکده اقتصاد، مدیریت و علوم اداری، دانشگاه سمنان، سمنان، ایران

تاریخ ارسال: ۱۳۹۹/۱۱/۲۱

تاریخ پذیرش: ۱۳۹۹/۰۷/۱۵

چکیده

با گسترش سایتهای فروش اینترنتی تمایل به خرید برخط با توجه به مزایای زیاد آن روزبه‌روز در حال افزایش است. اکثر مشتریان قبل از تصمیم‌گیری و انتخاب محصول، نظرات خریداران قبلی محصول را بررسی کرده و براساس آن محصول خود را انتخاب می‌کنند. وجود برندهای مختلف در بازار و از سویی بررسی حجم بالایی از نظرات جهت تصمیم‌گیری در خصوص خرید محصول چالش بزرگی است که خریداران با آن مواجه هستند. وجود ابزارهای خودکار متن‌کاوی و آنالیز احساسات برای بررسی نظرات کاربران، می‌تواند در این زمینه بسیار مفید باشد و به عنوان راهکاری در رتبه‌بندی محصولات براساس نظر مشتریان مورد استفاده قرار گیرد. در این پژوهش با بررسی بیش از ۴۵۰۰ مورد از نظرات مشتریان و ارزیابی متخصصان در خصوص ۷۰ کالا، سه نوع ویژگی شامل: ویژگی‌های حسی موجود در نظر مشتریان و متخصصین، ویژگی‌های ساخت محصول و ویژگی‌های نظرسنجی استخراج گردیده و با بکارگیری تحلیل احساسات مبتنی بر ویژگی، سیستم رتبه‌بندی محصولات ایجاد گردیده است.

کلیدواژه‌ها: رتبه‌بندی محصولات، تحلیل احساسات مبتنی بر ویژگی، متن‌کاوی، دیجی‌کالا



۱- مقدمه

با گسترش سایتهای فروش اینترنتی، به علت سهولت و قابل اعتماد بودن این نوع معاملات، مشتریان زیادی به خرید از این سایتهای رو آورده‌اند [۱]. وجود برندهای مختلف در بازار، تصمیم‌گیری و انتخاب را برای مصرف‌کننده بسیار سخت کرده است. خریداران براساس نظرات موجود در سایتهای فروش و رتبه‌بندی محصولات که در این سایتهای انجام می‌گیرد، اقدام به تصمیم‌گیری می‌کنند [۴]. نتایج یک پژوهش نشان می‌دهد، حدود ۶۳ درصد از کسانی که به صورت برخط خرید می‌کنند، به مرور نظرات سایر مصرف‌کنندگان کالا می‌پردازند، ۶۴ درصد از این افراد حداقل ۱۰ دقیقه و ۳۳ درصد بیش از ۳۰ دقیقه به بررسی نظرات می‌پردازند [۳].

هر سایت معمولاً دارای حجم زیادی نظرات متنی است که استخراج و خلاصه‌سازی چنین اطلاعاتی کار ساده‌ای نیست، بنابراین سیستم‌های تحلیل احساسات خودکار مورد نیاز است [۱۲]. کاوش خودکار نظرات برای ایجاد سیستم رتبه‌بندی محصول، ابزاری ارزشمند است که باعث می‌شود مشتریان بالقوه تصمیم‌گیری آگاهانه‌تری به جای اتکا به تعداد کمی از این بررسی‌ها و نظرات داشته باشند [۲۴].

در مطالعاتی که تاکنون در زمینه نظرات کاربران در خصوص رتبه‌بندی محصولات انجام گرفته است، معمولاً رابطه رتبه‌بندی محصول با میزان فروش و یا کمک به تولیدکننده در شناخت ویژگی‌های محصول مورد توجه بوده است. در هیچ‌کدام از مطالعات قبلی، به نقد و بررسی متخصصین سایتهای برای تولید مفاهیم رتبه‌بندی محصولات توجه نشده است و ترکیب ویژگی‌های حسی موجود در نظرات و ویژگی‌های ساخت محصول و ویژگی‌های نظرسنجی موجود در تارنمای خرده‌فروشی‌های برخط مد نظر قرار نگرفته است. این پژوهش جهت پاسخگویی به این سؤال که عناصر کلیدی موجود در دیدگاه‌های مصرف‌کنندگان و متخصصان در رابطه با رتبه‌بندی محصولات خریداری‌شده از تارنمای خرده‌فروشی برخط دیجی‌کالا چیست و چگونه می‌توان این مفاهیم را در قالب سیستم رتبه‌بندی محصولات بکار برد؟ رویکردی را ارائه می‌دهد که براساس آن با در نظر گرفتن نظرات کاربران و متخصصین و با ترکیب داده‌های حسی، داده‌های نظرسنجی و مشخصات ساخت محصول و استفاده از روش‌های متن‌کاوی و تحلیل احساسات ویژگی‌محور رتبه‌بندی کلی محصولات



انجام می‌گیرد.

۲- پیشینه نظری پژوهش

در مبحث عقیده‌کاوی، تعیین ویژگی‌ها در عقاید مشتریان و تشخیص و رده‌بندی اطلاعات ذهنی و تعیین گرایش احساسی‌شان به فرآیند تصمیم‌گیری سهولت می‌بخشد [۷]. امروزه حجم کثیری از داده‌های متنی برای پردازش موجود است و هر ساله بر میزان این داده‌ها افزوده می‌شود [۱۱]. افراد نظرات خود را درباره محصولات، اخبار و موضوعات مختلف بیان می‌کنند. به عنوان مثال وقتی فردی قصد خرید کالایی را دارد، شروع به جستجوی نظرات دیگران نموده و با توجه به این که اکثر داده‌ها به صورت متون غیرساخت یافته است، کشف اطلاعات از این متون بسیار سخت می‌شود [۵]. آنالیز احساسات یا عقیده‌کاوی شاخه‌ای از متن‌کاوی و به معنی یافتن نگرش نویسنده متن درباره یک موضوع خاص است که می‌تواند در این زمینه راهگشا باشد. رویکردهای تحلیل احساسات از سه جنبه قابل بررسی است که عبارتند از روش‌ها؛ یادگیری ماشین، مبتنی بر قانون و مبتنی بر لغت، سطح متن؛ سند، جمله و واژه، سطح رتبه‌بندی؛ ویژگی و عمومی. یادگیری ماشین: مجموعه داده‌ها از قبل آموزش داده می‌شوند و سپس با استفاده از الگوریتم‌های استاندارد یادگیری ماشین، قطبیت متن مشخص می‌شود [۳]. مبتنی بر قانون: اطلاعات از مجموعه داده‌ها استخراج شده و سعی می‌شود با توجه به قطبیت واژه‌ها، این اطلاعات ارزیابی شوند. قوانین متنوعی چون، کلمات نفی، اصطلاحات، شکلک‌ها و غیره در این راستا وجود دارد [۱۵]. مبتنی بر لغت‌نامه: با استفاده از جهت‌گیری معنایی، مثلاً ارزیابی یک نظر، ذهنی یا عینی بودن آن مشخص شده و قطبیت آن از لحاظ مثبت یا منفی بودن آشکار می‌گردد [۱۸]. براساس ساختار مجموعه داده، تحلیل احساسات در سطح ویژگی، سند و جمله تقسیم‌بندی می‌شود. در سطح سند، هدف بررسی متن و مشخص نمودن قطبیت آن به عنوان یک مجموعه واحد است. در سطح جمله هدف مشخص نمودن قطبیت یک جمله مجزا از متن است که در واقع از نظر بنیادی با نوع اول تفاوتی ندارد، زیرا جملات در حقیقت سندهای کوچک‌شده در حد یک جمله هستند. رتبه‌بندی یک محصول در دو سطح مبتنی بر ویژگی و کلی قابل انجام است. البته بیشتر سایت‌های فروش یا سایت‌هایی که به جمع‌آوری نظرات در مورد فیلم‌ها می‌پردازند، میزان قطبیت یک متن یا نظر را در سطح کلی (کلیت محصول یا فیلم) ارزیابی می‌کنند [۱۵].



۳- پیشینه تجربی پژوهش

مطالعات فراوانی در رشته‌هایی چون سیستم‌های اطلاعات، علوم کامپیوتر و بازاریابی در زمینه رتبه‌بندی نظرات کاربران انجام شده است [۸]. این مطالعات از رتبه‌بندی کاربران و اظهارنظرهای آنها جهت تشریح ویژگی‌های متعدد بررسی‌های برخط و رفتار مشتری استفاده نموده‌اند. پردازش زبان طبیعی، روش‌های یادگیری با نظارت^۲ و بدون نظارت^۳ و قوانین وابستگی^۴ در تحلیل نظرات کاربرد دارند. این روش‌ها از زیرمجموعه‌های داده‌کاوی می‌باشند که علم استخراج الگوهای جالب، کوچک اما با اهمیت، ضمنی، ناشناخته و بالقوه مفید و یا دانش از مقدار بسیار عظیمی از داده‌ها می‌باشد [۲۱].

در پژوهشی یک الگوریتم یادگیری با نظرات ارائه گردید که می‌تواند به آنالیز عقیده افرادی که محصولات را در پست‌ها، وبلاگ‌ها و توییتهای مورد تحلیل قرار داده‌اند، بپردازد. در این الگوریتم ویژگی مهم محصول استخراج شده و امکان پیش‌بینی جهت‌گیری هر کدام از این مشخصه‌ها در آینده وجود دارد [۱۰]. در تحقیقی که در سال ۲۰۱۰ انجام گرفت، مکانیزم رتبه‌بندی مبتنی بر ویژگی محصول ارائه شد. ابتدا ویژگی‌های محصول را براساس دسته‌بندی محصول و تحلیل بسامد و استفاده از محصول مشخص نموده و جهت‌گیری احساسی این نظرات مشخص گردید. در ادامه گراف جهت‌دار و وزن‌داری مبتنی بر ویژگی‌ها براساس آمار کلیه نظرات کاربران تشکیل شد. در این روش از رهیافت واژگان کلیدی برای تشخیص نظرات حاوی ویژگی‌ها استفاده شده است. در نهایت ارزیابی پروژه با الگوریتم PageRank و داده‌های موجود در وب سایت آمازون انجام گرفت [۲۵].

در مقاله‌ای دیگر، روش رتبه‌بندی محصول براساس ویژگی ارائه شد که با استفاده از تقسیم کاربران به دو دسته کاربران آشنا و ناآشنا و اختصاص اوزان مختلف به ایشان براساس میزان اعتبار کاربران (کاربران بالای لیست نسبت به پائین از درجه اعتبار بالاتری برخوردارند) به محاسبه رتبه براساس وزنی که کاربر به محصول داده است پرداختند [۱۱]. در پژوهشی دیگر، با کاوش میلیون‌ها نظر، سیستم رتبه‌بندی را پیشنهاد دادند که با توجه به بررسی اعتبار و تاریخ ارسال نظرات عمل می‌کند. این سیستم با فیلتر نمودن جملات و نظراتی که نامربوط بوده و بیشتر جنبه خدماتی کالا را مد نظر قرار می‌دهند به مصرف‌کننده به تصمیم‌گیری و خرید با کیفیت کمک می‌کنند [۲۴].

در پژوهشی که در سال ۲۰۱۴ انجام گرفت، پژوهشگران به این نتیجه رسیدند که رتبه‌بندی



محصول، سنجشی از کیفیت محصول است به طوری که درک بالا از کیفیت محصول با تمایل به خرید آن در ارتباط کامل است. رتبه‌ای که مصرف‌کننده به محصول می‌دهد بعنوان یک منبع معتبر اطلاعاتی برای سایر مصرف‌کنندگان تلقی می‌شود و مصرف‌کننده بالقوه در حین خرید به متوسط این رتبه توجه می‌کند [۱۲]. در مقاله‌ای، روشی برای رتبه‌بندی محصولات براساس تحلیل احساسات و تئوری فازی ارائه گردید که وزن هر ویژگی محصول براساس تعداد تکرار واژه‌های حسی موجود در واژه‌نامه Hownet مشخص می‌شود [۱۳]. در سال ۲۰۱۹ اصغر و همکاران، چارچوبی پیشنهاد داده‌اند که شامل چهار ماژول اصلی استخراج ویژگی-احساس، گروه‌بندی ویژگی، دسته‌بندی ویژگی-احساس و خلاصه‌سازی ویژگی است. آن‌ها تعدادی الگوی جدید ارائه داده‌اند که در آن اسم، عبارات اسمی و فعل‌ها، بیانگر ویژگی‌های کاندید هستند. پس از استخراج ویژگی‌ها، گروه‌بندی ویژگی‌ها را براساس اندازه‌گیری شباهت معنایی پیشنهاد داده‌اند. به منظور انتساب امتیاز به نظرات داده‌شده درباره ویژگی‌های استخراج‌شده، یک تکنیک ترکیبی برای امتیازدهی احساسات پیشنهاد داده شد که مفاهیم مبتنی بر پیکره و مبتنی بر لغت را ترکیب و در نهایت نقاط مثبت و منفی هر ویژگی را بیان می‌کند [۲]. در تحقیقی که در سال ۲۰۲۰ توسط زارعی و همکاران با عنوان "ارائه چارچوب هوشمندی بازار اجتماعی مبتنی بر وب. ۲/ با استفاده از تکنیک متن‌کاوی در وبسایت‌های رسانه‌های اجتماعی" انجام گرفت، با مرور ادبیات و مصاحبه با خبرگان، ۴ بعد اصلی هوشمندی بازار اجتماعی. ۲/ مشخص گردید. سپس از تکنیک خوشه‌بندی برای استخراج مهم‌ترین شاخص‌ها و زیر شاخص‌های مؤثر در هر یک از ابعاد، استفاده گردید. با استفاده از تحلیل احساسات کلمات مشخص گردید که برند سامسونگ نسبت به امرسان در اکثر شاخص‌ها نظرات مثبت مشتریان را به خود اختصاص داده است [۲۳]. فیرمانتو و همکاران (۲۰۱۹) با این فرض که اسم‌ها و عبارات اسمی در قالب ویژگی‌ها معمولاً نزدیک به نظرات استفاده می‌شوند، طی فرآیندی، جملات پیچیده و ترکیبی را تکه تکه کرده و به جملات ساده تبدیل نمودند تا با پیدا نمودن یک واژه در برگزیده نظر بتوانند نزدیکترین ویژگی به آن را پیدا نمایند [۵]. در سال ۲۰۱۶ توسط یانگ و همکاران^۲ یکی از مهم‌ترین پژوهش‌های حوزه رتبه‌بندی محصول انجام گرفت، در این تحقیق با عنوان "یکپارچه‌سازی اطلاعات ارزشمند و ناهمگن برای طراحی سیستم رتبه‌بندی چندین محصول"، از انواع مختلفی از نظرات برخط مشتریان شامل: رتبه عددی،^۴ توصیف‌های متنی و لغات مقایسه‌ای^۱ استفاده شده است. با توجه به این‌که اطلاعات ناهمگون درباره محصولات باعث سردرگمی خریداران می‌شود، محققین در این

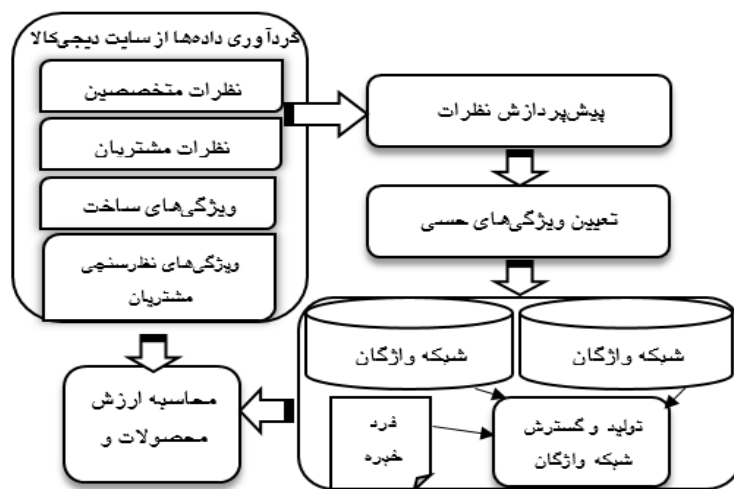


مقاله سعی نمودند که اطلاعات ترکیب شده و سیستم رتبه‌بندی دقیقی ارائه گردد. برای رسیدن به این هدف، اطلاعات یکپارچه به دو دسته تقسیم شدند: اطلاعات توصیفی و اطلاعات تطبیقی. اطلاعات توصیفی نظرات مستقیم مصرف‌کنندگان در مورد یک کالای خاص هستند که به صورت نظرات حسی و رتبه عددی بیان می‌شوند. اطلاعات تطبیقی از جملات تطبیقی می‌آید که به طور ضمنی در نظرسنجی‌ها وجود دارد و در آرای تطبیقی^۱ برخط برای مقایسه بیش از یک محصول به کار می‌روند. اطلاعات توصیفی و اطلاعات تطبیقی به صورت یک ساختار گرافی یکپارچه شده و نمره هر محصول و رتبه‌بندی تمام محصولات به دست می‌آید. نظرات مربوط به ۳ نوع محصول در این تحقیق جمع‌آوری گردید (تلفن همراه، لپ‌تاپ و دوربین دیجیتال). نتایج نشان داد که روش به کار رفته در این تحقیق عملکرد بهتری در مقایسه با سایر روش‌های رتبه‌بندی محصول دارد. رتبه‌بندی مبتنی بر این روش به مصرف‌کنندگان برای مقایسه چندین محصول و اتخاذ تصمیم خرید مناسب، بدون نیاز به اتلاف وقت زیاد کمک می‌کند [۲۲]. یکی دیگر از پژوهش‌های مهم در مورد رتبه‌بندی محصولات توسط گو و همکاران^۲ در سال ۲۰۱۸ انجام گرفت. در این تحقیق یک روش جدید رتبه‌بندی ارائه شده است که براساس آن ارزش‌های عینی و ذهنی هر کدام از نظرات ترکیب شده‌اند و وزن هر کدام از ویژگی‌ها براساس روش تحلیل تفکیک کننده خطی^۳ به دست آمد. در ادامه مقایسه جفتی بین محصولات انجام شده و براساس الگوریتم توسعه یافته PageRank، امتیاز نهایی هر ویژگی مشخص گردید [۹]. مطالعات ذکر شده دارای نواقصی می‌باشند، از جمله این که امتیاز کلی یک محصول به خوبی احساسات مصرف‌کننده را در مورد هر کدام از ویژگی‌ها محصول بیان نمی‌کند. به عبارتی دو محصول که امتیاز کلی یکسانی را از مصرف‌کننده دریافت نموده‌اند، قطعاً براساس تک تک ویژگی‌هایشان کاملاً با هم متفاوتند. از طرفی در هیچ کدام از مطالعات انجام گرفته دیدگاه متخصصین سایت‌ها در تحلیل احساسات مورد توجه قرار نگرفته است. در این پژوهش سعی شده است تا این موارد مرتفع گردد و در سیستم رتبه‌بندی پیشنهادی، ترکیبی از تحلیل نظرات متنی کاربران، امتیاز عددی موجود در سایت (مربوط به هر محصول) و مشخصات ساخت محصول که توسط شرکت سازنده ارائه می‌شود، مورد استفاده قرار گیرد.



۴- روش شناسی پژوهش

روش پیشنهادی مبتنی بر تحلیل احساسات است که در آن با استخراج نظرات حسی کاربران، محصولات را رتبه بندی کرده و در نهایت محصول مناسب به مشتری پیشنهاد می شود. در شکل (۱) ساختار کلی روش پیشنهادی نشان داده شده و در ادامه به تشریح هر مرحله خواهیم پرداخت.



شکل ۱. ساختار کلی پژوهش

۴-۱- گردآوری داده‌ها

در این مرحله داده‌های مربوط به پژوهش شامل، متن نظرات مشتریان و متخصصان، داده‌های عددی نظرسنجی مشتریان و داده‌های مربوط به مشخصات ساخت محصول جمع‌آوری گردیده است.

۴-۲- پیش پردازش

در مرحله پیش‌پردازش، متن ورودی به ساختاری قابل پردازش برای مراحل بعد تبدیل می‌شود. مهم‌ترین فرآیندهای پیش‌پردازش متون فارسی عبارتند از: یکدست‌سازی متن، تعیین مرز جمله‌ها و واژه‌ها، حذف واژه‌های غیرمهم، ریشه‌یابی و شناسایی نقش کلمات.



۳-۴- تعیین ویژگی‌های حسی

در ابتدا بایستی ویژگی‌های حسی مهم در متن نظرات مشتریان و متخصصین شناسایی شود. تشخیص مهم‌ترین ویژگی‌های مجموعه داده‌های متنی باعث بهبود تجزیه و تحلیل و در نتیجه تولید معنا از این داده‌های خام می‌شود. یکی از ساده‌ترین روش‌های تعیین ارزش کلمات در متون، استفاده از اطلاعات آماری است. اما نقطه ضعف آن عدم استفاده از ساختار معنایی کلمات است. در این روش با ترکیب روش آماری TF-IDF با شبکه وازگان معنایی یک روش TF-IDF مفهومی معرفی می‌شود. معیار TF-IDF از رابطه (۱) جهت محاسبه ارزش کلمات در یک متن استفاده می‌کند.

$$tfidf_{x,y} = tf_{x,y} \times \log\left(\frac{N}{df_x}\right) \quad \text{رابطه (۱)}$$

که در آن $tf_{x,y}$ تعداد تکرارهای کلمه x در سند y ، N تعداد کل اسناد و df_x تعداد اسنادی است که در آن‌ها کلمه x تکرار شده است.

استفاده از معیار TF-IDF به تنهایی نمی‌تواند دقت لازم را داشته باشد. به همین دلیل در روش پیشنهادی برای استخراج ویژگی‌ها به روش زیر عمل می‌شود:

- پیش‌پردازش نظرات و حذف کلمات بی‌ارزش
- توکن‌بندی جملات و شناسایی نقش کلمات
- جدا کردن اسامی و محاسبه TF-IDF به صورت مفهومی
- انتخاب یک مقدار آستانه برای معیار TF-IDF
- شناسایی کلماتی که معیار TF-IDF آن‌ها بیشتر از مقدار آستانه و تعیین آن‌ها به عنوان ویژگی‌های انتخابی

در این الگوریتم منظور از TF-IDF مفهومی این است که برای هر کلمه قبل از محاسبه مقدار TF-IDF، هم خانواده‌های آن نیز شناسایی شده و به عنوان یک کلمه در محاسبات در نظر گرفته می‌شوند. با این کار، ارزش اختصاص یافته به کلمات علاوه بر اطلاعات آماری، براساس محتوا نیز خواهد بود. در نتیجه دقت آن افزایش می‌یابد. اما نکته دیگر اینکه برای کاهش پردازش و افزایش دقت، این فرآیند تنها بر روی اسامی انجام می‌گیرد، زیرا تمامی ویژگی‌ها مانند، دوربین، باتری، پردازنده و ... جزء اسامی هستند.



۴-۴- شبکه واژگان و تولید مجموعه واژگان حسی

شبکه واژگان شامل مجموعه‌ای از واژگان است که ارتباط بین این کلمات در آن مشخص شده است و برای ابهام زدایی و تعیین شباهت معنایی در کاربردهای مختلف پردازش زبان طبیعی (و بازیابی اطلاعات) مانند ترجمه ماشینی، استخراج اطلاعات و خلاصه‌سازی استفاده می‌شود. تلاش‌هایی برای ایجاد شبکه‌واژگان زبان فارسی انجام شده است که می‌توان به مواردی مانند فارس‌نت [۲۵] شبکه واژگان دانشگاه تهران [۱۹] و فردوس‌نت [۱] اشاره کرد. برای ایجاد مجموعه واژگان دو روش کلی وجود دارد؛ یکی توسعه یا ترجمه عبارات حسی از روی مجموعه واژگان حسی موجود و دیگری گسترش فهرست کلمات حسی اولیه با استفاده از شبکه واژگان، پیکره‌های از متون نظرات و سایر منابع زبانی. در روش پیشنهادی از راهکار دوم استفاده می‌شود یعنی در ابتدا مجموعه‌ای از واژه‌ها به صورت دستی انتخاب شده، سپس این مجموعه از واژه‌ها با استفاده از فارس‌نت و فردوس‌نت گسترش می‌یابند. برای تولید حس‌نگار واژه‌های حسی به سه بخش: صفت (مانند: خوب، بد، زشت، ضعیف)، اسم (قوت، ضعف) و قید (به‌خوبی، به‌سختی) تقسیم‌بندی می‌شوند. با توجه به ساختار نظرات کاربران از فعل‌ها در روش پیشنهادی استفاده نمی‌شود. زیرا اگر فعلی بار حسی هم داشته باشد، با توجه به اینکه ریشه کلمات استخراج می‌شود، می‌تواند بار معنایی را پوشش دهد.

در این تحقیق از شبکه واژگان فردوس‌نت استفاده می‌شود که منطبق با وردنت است. همچنین از منابع زبانی و پایگاه‌های دانشی چون شبکه واژگان دانشگاه پرینستون (وردنت)، مترجم متن گوگل، پایگاه دانش ویکی‌پدیا و هستان‌شناسی یاگو [۱۶]، پیکره‌های متنی فارسی، شبکه واژگان دانشگاه تهران و فارس‌نت نیز استفاده شده است.

هدف این است که واژگان حسی تولید شده، تا حد امکان مختص تحلیل نظرات ثبت شده در مورد گوشی‌های تلفن همراه باشد. این امر باعث می‌شود مجموعه واژگان گسترده نبوده و در عین حال تمام کلمات موردنظر را دربر گیرد. ابتدا توسط فرد خبره فهرستی از عبارات حسی که متعلق به موضوع مورد بحث (گوشی‌های هوشمند تلفن همراه) هستند، شناسایی می‌شوند. بعد از شناسایی کلمات اولیه، با استفاده از شبکه واژگان فردوس‌نت این مجموعه، گسترش داده می‌شود. سعی می‌کنیم با استفاده از ساختار مفهومی کلمات ارزش آن‌ها را در جمله محاسبه کنیم و در نهایت برای هر کلمه در هر نظر یک عدد بین ۴- الی ۴+ اختصاص داده شود. اگر واژه در بین واژگان اولیه وجود نداشته باشد، با توجه به فارس‌نت هم‌خانواده‌های این کلمه استخراج شده و



ارزشی برابر با کلمه هم‌خانواده به آن اختصاص می‌یابد. میزان و شدت حس، با توجه به برخی کلمات جمله از قبیل: هم خانواده، متضاد، منفی‌کننده‌ها و معکوس‌کننده‌های حسی تنظیم می‌شود. در جدول (۱) نمونه‌ای از واژگان اولیه و ارزش نسبی اختصاص داده شده به آن‌ها مشخص شده است.

جدول ۱ نمونه‌ای از واژگان حسی و ارزش آن‌ها

ارزش حسی	کلمه	ارزش حسی	کلمه
+۱	نسبتاً خوب	-۴	اقتضاح
+۲	قوی	+۴	شگفت‌انگیز
-۲	ضعیف	+۴	فوق‌العاده
+۱	خوش‌دست	+۳	عالی
+۲	زیبا	۰	قابل قبول
-۱	نچندان خوب	-۱	بد
-۲	سنگین	-۳	ضایع

با فرض اینکه مجموعه $NR_{ij} = \{nr_{ij}^1, nr_{ij}^2, \dots, nr_{ij}^k\}$ نظرات حسی مشتریان و متخصصین در مورد ویژگی‌ها باشد که در آن nr_{ij}^k صفت مرتبط با نظر j ام برای ویژگی k ام محصول نام است.

در جدول (۲) فرض شده است که محصول P_i دارای K ویژگی است و تعداد q_i نظر برای آن ثبت شده است. S_{ij}^k مشخص‌کننده مقدار عددی اختصاص یافته توسط فرد j برای ویژگی k ام محصول i است. سطر آخر از جدول (۲) مشخص‌کننده میانگین ارزش ویژگی‌ها برای همه نظرات است.

جدول ۲ ماتریس ارزش عددی ویژگی‌های محصول P_i

ویژگی نظر	f_1	f_2	...	f_k	...	f_K
۱						
۲						
:						
j				S_{ij}^k		
:						
q_i						
میانگین	S_i^1	S_i^2	...	S_i^k	...	S_i^K

نظر متخصصین برای هر محصول یک نظر واحد است؛ یعنی خروجی نظر متخصصین یک ماتریس تک سطری است. نظرات مشتریان و متخصصین با هم ترکیب شده و میانگین آن محاسبه می‌شود. در جدول (۳)، \bar{S}_i^k مشخص کننده میانگین ارزش نظرات کاربران و متخصصین است که از رابطه (۲) بدست می‌آید.

$$\bar{S}_i^k = \frac{S_i^k + M_i^k}{2} \quad \text{رابطه (۲)}$$

جدول ۳ ترکیب ارزش عددی نظر مشتریان و متخصصین

ویژگی ارزش حسی نظرات	f_1	f_2	...	f_k	...	f_K
مشتریان	S_i^1	S_i^2	...	S_i^k	...	S_i^K
متخصصین	M_i^1	M_i^2	...	M_i^k	...	M_i^K
ارزش حسی کل	\bar{S}_i^1	\bar{S}_i^2	...	\bar{S}_i^k	...	\bar{S}_i^K

اما برای محاسبه ارزش تمامی محصولات از جدول (۴) استفاده می‌شود که مقدار S_i در این جدول از رابطه (۳) محاسبه می‌شود.



$$S_i = \frac{\sum_{k=1}^K \bar{S}_i^k}{K} \quad \text{رابطه (۳)}$$

جدول ۴ ماتریس ارزش گذاری برای همه محصولات

ویژگی \ محصول	f ₁	f ₂	...	f _k	...	f _K	
P ₁							S ₁
P ₂							S ₂
⋮							⋮
P _i				\bar{S}_i^k			S _i
⋮							⋮
P _N							S _N

۴-۵- محاسبه ارزش محصولات و رتبه‌بندی آنها

در این تحقیق استخراج اطلاعات به سه دسته (۱) اطلاعات حسی متنی (۲) اطلاعات نظرسنجی مشتریان (۳) مشخصات ساخت محصول، تقسیم‌بندی می‌شود. در بخش ۴-۴ به طور مفصل روش محاسبه ارزش عددی ویژگی‌های موجود در نظرات مشتریان و متخصصین با استفاده از تحلیل احساسات مبتنی بر ویژگی تشریح گردید. در ادامه به بررسی نحوه استخراج اطلاعات ساخت محصول و نظرسنجی مشتریان می‌پردازیم.

۴-۵-۱- اطلاعات ساخت محصول

هر گوشی دارای مشخصات سخت‌افزاری است که در این تحقیق به عنوان ویژگی‌های ساخت محصول از آن استفاده می‌شود. برای هر محصول ماتریسی مطابق جدول (۵) ایجاد می‌شود که در آن TS_i^k مشخص کننده مقدار مرتبط با ویژگی ساخت k ام برای محصول i ام است. ویژگی‌های ساخت محصول با نماد fS نمایش داده شده است.

جدول ۵ بردار ارزش‌گذاری داده‌های ساخت محصول P_i

ویژگی‌ها	fS ₁	fS ₂	...	fS _k	...	fS _K
ارزش ویژگی‌های ساخت	TS_i^1	TS_i^2	...	TS_i^k	...	TS_i^K



از آنجایی که ویژگی‌های ساخت محصول دارای واحدهای اندازه‌گیری متفاوتی می‌باشند، از نرمال‌سازی داده‌ها استفاده می‌شود که اهمیت آن‌ها به واحد اندازه‌گیری‌شان بستگی نداشته باشد. برای این منظور در روش پیشنهادی، ویژگی‌های ساخت محصول و ارزش حسی ویژگی‌های موجود در متن نظرات مشتریان و متخصصین بین ۰ تا ۵ نرمال سازی می‌شوند. برای این منظور از رابطه (۴) استفاده گردیده است.

$$x_{normalized} = \frac{x - x_{min}}{x_{max} - x_{min}} \times 5 \quad \text{رابطه (۴)}$$

که در آن x ، x_{min} و x_{max} به ترتیب مقدار اولیه، حداقل مقدار و حداکثر مقدار مربوط به یک ویژگی هستند. همچنین $x_{normalized}$ مقدار نرمال شده آن ویژگی است.

۲-۵-۴- اطلاعات نظرسنجی مشتریان

در تارنمای خرده‌فروشی دیجی‌کالا در صفحه هر محصول، قسمتی با عنوان امتیاز کاربران به محصول وجود دارد که براساس میانگین نظرات کاربران، ویژگی‌های: کیفیت ساخت، ارزش خرید به نسبت قیمت، نوآوری، امکانات و قابلیت‌ها، سهولت استفاده و طراحی و ظاهر، از صفر تا پنج مقداردهی شده‌اند. برای استفاده از این ویژگی‌ها، ماتریسی مطابق جدول (۶) در نظر گرفته شده است که در آن PS_i^k مقدار میانگین ویژگی k ام برای محصول i ام است. چون این ویژگی‌ها بین صفر تا پنج هستند، نیاز به نرمال‌سازی همانند دو دسته ویژگی‌های قبلی ندارند و ارزش آن‌ها به همین شکل وارد جدول می‌شود. ویژگی‌های نظرسنجی مشتریان با نماد fn نمایش داده شده است.

جدول ۶ بردار ارزش گذاری ویژگی‌های نظرسنجی محصول P_i

ویژگی‌ها	fn_1	fn_2	...	fn_k	...	fn_K
ارزش ویژگی‌های نظرسنجی	PS_i^1	PS_i^2	...	PS_i^k	...	PS_i^K

با توجه به جداول ۴، ۵ و ۶ و برای محاسبه امتیاز نهایی براساس ویژگی‌های ذکر شده، از جدول ۷ استفاده می‌شود. اعداد موجود در این جدول بین صفر و پنج نرمالیزه شده و آماده مرحله رتبه‌بندی محصولات می‌باشد.



جدول ۷ بردار ارزش گذاری نهایی ویژگی‌ها

ویژگی‌ها	حسی متنی				ساخت محصول				نظرسنجی			
	f ₁	f ₂	...	f _k	fs ₁	fs ₂	...	fs _k	fn ₁	fn ₂	...	fn _k
ارزش ویژگی‌ها	S ₁	S ₂	...	S _k	TS _i ¹	TS _i ²	...	TS _i ^k	PS _i ¹	PS _i ²	...	PS _i ^k

۳-۵-۴- رتبه‌بندی محصولات

بعد از محاسبه ارزش حسی نظرات مشتریان و متخصصین، اطلاعات نظرسنجی مشتریان و اطلاعات ساخت محصول و با ترکیب آن‌ها براساس رابطه (۵)، ارزش کلی محصولات محاسبه شده و رتبه‌بندی براساس آن انجام می‌شود. برای تعیین میزان تأثیر هر بخش از ضرایب مشخص استفاده گردیده است [۲۲].

$$W_Score(P_i) = \alpha S_i + \beta TS_i + (1 - (\alpha + \beta)) PS_i \quad \text{رابطه (۵)}$$

که در آن $0 \leq \alpha + \beta \leq 1$ و $W_Score(P_i)$ مشخص کننده ارزش کلی تعیین شده برای محصول i ام است.

۴-۵-۴- معیار ارزیابی رتبه‌بندی محصولات

یکی از چالش‌های اصلی در روش پیشنهادی، نحوه ارزیابی آن است. اکثر کارهای انجام شده در این زمینه مانند [۲۶] و [۲۲] از میزان فروش محصول به عنوان معیار ارزیابی استفاده کرده‌اند. یعنی محصولی که فروش بیشتری داشته باید در رتبه‌بندی، در رتبه‌های بالا قرار بگیرد. در سایت دیجی‌کالا محصولات براساس میزان فروش رتبه‌بندی شده‌اند. با توجه به رتبه فروش محصول در سایت و رتبه بدست آمده در سیستم پیشنهادی، معیاری تحت عنوان ضریب همبستگی توسط رابطه (۶) معرفی می‌شود [۲۲].

$$\rho(\overrightarrow{R_{x_i}} - \overrightarrow{R_{y_i}}) = 1 - \frac{6 \sum_{i=1}^n (R_{x_i} - R_{y_i})^2}{n(n^2 - 1)} \quad \text{رابطه (۶)}$$

که در آن $\overrightarrow{R_{x_i}}$ رتبه کسب شده توسط سیستم و $\overrightarrow{R_{y_i}}$ رتبه مربوط به فروش آن محصول است. با توجه به رابطه (۶)، هرچه اختلاف بین رتبه کسب شده توسط سیستم رتبه‌بندی ارائه شده و رتبه فروش محصول کمتر باشد ضریب همبستگی بیشتر خواهد بود.



۵- یافته‌های پژوهش

۵-۱- مجموعه داده

برای جمع‌آوری نظرات کاربران، چهار برند پرطرفدار گوشی همراه شامل: سامسونگ، اپل، هواوی و شیائومی انتخاب شده‌اند و براساس قیمت به سه رده قیمت مناسب، میان‌رده و پرچم‌دار تقسیم شده‌اند. مجموع نظرات استخراج شده برای این ۷۰ گوشی، ۴۳۵۰ نظر است که به صورت میانگین برای هر گوشی ۶۲ نظر ذخیره شده است. با توجه به این‌که در تارنمای خرده‌فروشی دیجی‌کالا می‌توان نظرات مشتریان را براساس مفیدترین نظرات مرتب‌سازی نمود، در این تحقیق نظرات انتخاب شده از بین مفیدترین نظرات بوده و به طور میانگین هر نظر شامل ۸۰ کلمه می‌باشد.

۵-۲- نرمال‌سازی و تشخیص کلمات عامیانه

زمانی که متن مورد پردازش شامل جملات عامیانه باشد، معمولاً پیچیدگی پردازش بیشتر می‌شود. معمولاً نظر متخصص با نگارش رسمی نوشته می‌شود اما نظرات کاربران در اکثر موارد به صورت عامیانه است. مثلاً کلمه "هم‌رده"، "هم‌رده" یا هم رده " هر سه ممکن است در متن وجود داشته باشند. در ساختار پیشنهادی عبارات و فعل‌های عامیانه حذف شده و به جای آن عبارت رسمی جایگزین می‌شود. در شکل (۲) متن مربوط به نظر یک کاربر در مورد گوشی iPhone7 و در شکل (۳) متن نرمال‌شده آن نشان داده شده است.

دوستان خود من الان s6 دارم و با آیفون ۷ رفیقم هم خیلی کار کرده‌ام. در کل میتونم بگم واسه کسی که s6 داره خرید این گوشی کمترین جذابیتی نداره و در واقع می‌توان گفت ۷ همون s6 بروز شده است.

شکل ۲ متن خام نظر کاربر در مورد گوشی iPhone7

دوستان خود من الان s6 دارم و با آیفون ۷ رفیقم هم خیلی کار کرده‌ام. در کل می‌توانم بگویم برای کسی که s6 دارد خرید این گوشی کم‌ترین جذابیتی ندارد و در واقع می‌توان گفت ۷ همان s6 بروز شده است.

شکل ۳ متن نرمال شده مربوط به شکل ۲



۳-۵- جداسازی جملات و توکن‌ها

چون هر ویژگی در یک جمله مشخص می‌شود، بنابراین جداسازی جملات و توکن‌ها بسیار مهم است. در شکل (۴) بخشی از نظر مربوط به شکل (۳) نشان داده شده و جملات تفکیک گردیده و برای هر جمله، توکن‌ها شناسایی شده‌اند.

دوستان|خود|من|الان|دارم|۶|دارم.
و|با|آیفون|۷|رفیق|هم|خیلی|کار|کرده|ام|۷|.
در|اکل|می|توانم|بگویم|۷|.

شکل ۴ شناسایی جملات و توکن‌ها

۴-۵- برچسب‌زنی نقش کلمات

برای برچسب‌زنی نقش کلمات، از ابزار برچسب‌زنی دانشگاه فردوسی مشهد استفاده شده است. خروجی این بخش در شکل (۵) نشان داده شده است. در روش پیشنهادی از الگوریتم ویتربای [۱۴] برای برچسب‌زنی نقش کلمات استفاده می‌شود. این روش با استفاده از مدل مخفی مارکوف برای متون فارسی نیز کاربرد مناسبی دارد.

دوستان <N> خود <PR> من <PR> الان <PR> دارم <V> و <CONJ> با <PREP>
آیفون <PRENUMvN> رفیق <N> هم <ADV> خیلی <ADV> کار کرده‌ام <V>

شکل ۵ تعیین نقش کلمات در جملات

۵-۵- استخراج ویژگی‌ها

پس از اعمال پردازش بیان شده روی مجموعه داده، ۹ ویژگی برای محاسبه ارزش حسی کلمات مرتبط با گوشی‌های همراه انتخاب شدند. در مجموع ویژگی‌های استخراجی برای هر محصول در جدول (۸) به تفکیک نشان داده شده است.

جدول ۸ ویژگی‌های مورد استفاده در رتبه‌بندی گوشی‌های همراه

دوربین، باتری، طراحی، ضخامت، صفحه نمایش، پردازنده، صدا، اسپیکر و قیمت	ویژگی‌های حسی متنی
کیفیت ساخت، ارزش خرید نسبت به قیمت، نوآوری، امکانات و قابلیت‌ها، سهولت استفاده و طراحی و ظاهر	ویژگی‌های نظر سنجی مشتریان



ویژگی‌های ساخت محصول	، حداکثر RAM وزن، فرکانس پردازنده، حافظه داخلی، مقدار ظرفیت حافظه، رزولوشن دوربین اصلی، رزولوشن دوربین جلو، ظرفیت باتری و اندازه بر حسب اینچ
----------------------	----------------------------------------------------------------------------------------------------------------------------------------------

۶-۵- تعیین ارزش ویژگی‌ها توسط شبکه واژگان حسی

برای تعیین ارزش ویژگی‌های حسی استخراج شده از متن نظر مشتریان و متخصصان و براساس توضیحات بخش ۴-۴، مقادیر عددی به ازای هر نظر، استخراج می‌شود. برای این ویژگی‌ها تقریباً صفات مشخصی وجود دارند. بنابراین ابتدا فهرستی از عبارات حسی اولیه با کمک افراد خبره تهیه شده و میزان بار حسی آن‌ها تعیین می‌شود. برای بسط و تکمیل این مجموعه کلمات اولیه از شبکه واژگان فردوس نت استفاده شده است. مقادیر صفات اولیه در نظر گرفته شده شامل واژه‌های: عالی، قدرتمند، زیبا، خوش‌دست، فوق‌العاده، شگفت‌انگیز، خوب، سنگین، سبک، بی‌رقیب، محشر، ساده، شارژ سریع، ضعیف، راضی، ناراحتی، باکیفیت، بی‌کیفیت است.

به طور خلاصه مراحل محاسبه ارزش هر ویژگی حسی متنی به شکل زیر است:

۱. پیش‌پردازش نظرات
 ۲. تفکیک نظرات مربوط به هر گوشی در یک فایل متنی
 ۳. استخراج و تفکیک جملات
 ۴. در هر جمله به دنبال ویژگی‌های حسی تعیین شده می‌گردیم. (این ویژگی‌ها شامل دوربین، باتری، پردازنده، صفحه نمایش، ظاهر و قیمت.)
 ۵. اگر یک ویژگی یافت شد سپس در جمله دنبال یک صفت می‌گردیم که بیان کننده حس کاربر نسبت به آن ویژگی باشد.
 ۶. اگر صفت استخراج شده جزء صفتهای اولیه تعیین شده باشد، مقدار عددی آن استخراج می‌شود. در غیر اینصورت در شبکه واژگان دنبال هم معنی یا متضاد آن می‌گردیم و عدد متناسب با آن را استخراج کرده و به ویژگی اختصاص می‌دهیم.
 ۷. موارد ۴ الی ۶ برای تک تک جملات تکرار می‌شود.
 ۸. موارد ۳ الی ۷ برای تک تک نظرات تکرار می‌شود.
 ۹. در نهایت میانگین مقادیر استخراج شده برای هر ویژگی محاسبه می‌شود.
- در نهایت یک فایل اکسل از تمام محصولات ایجاد می‌شود؛ به طوری که هر سطر آن یک محصول و هر ستون یک ویژگی محصول (ویژگی‌های حسی متنی، ویژگی‌های نظرسنجی و



ویژگی‌های ساخت محصول) است. فایل ایجاد شده شامل ۷۰ مدل مختلف گوشی است که در مجموع برای آن‌ها ۲۴ ویژگی استخراج شده است. در پایان این مرحله، فایل خام داده‌ها برای رتبه‌بندی محصولات آماده شده است.

۷-۵- رتبه‌بندی محصولات

رتبه‌بندی نهایی بر اساس رابطه (۵) انجام می‌گیرد که در آن هر سه گروه ویژگی استفاده شده‌اند. برای ارزیابی دقیق‌تر روش پیشنهادی چند سناریوی مختلف مطرح می‌شود و بر اساس تنوع آن‌ها، سیستم پیشنهادی ارزیابی می‌گردد.

- سناریو (۱): در رابطه (۵) مقدار α برابر صفر در نظر گرفته شده و ویژگی‌های نظرسنجی و ساخت محصول در رتبه‌بندی مورد استفاده قرار می‌گیرند.
- سناریو (۲): در رابطه (۵) مقدار α برابر ۱ و β برابر صفر در نظر گرفته می‌شود و در نتیجه تنها ویژگی‌های حسی جهت محاسبه W_Score استفاده می‌شوند.
- سناریو (۳): هر سه گروه ویژگی مورد استفاده قرار می‌گیرند. تأثیر دو گروه ویژگی‌های نظرسنجی و مشخصات ساخت محصول یکسان و برابر $(1-\alpha)/2$ در نظر گرفته می‌شوند و مقدار (متغیر خواهد بود).
- سناریو (۴): در سناریوهای قبل تأثیر ویژگی‌ها به صورت انفرادی، برابر در نظر گرفته شد. به عنوان مثال تأثیر دوربین، باتری و طراحی ظاهر یکسان است که برای سلیق مختلف نمی‌تواند مناسب باشد. برای محاسبه میزان اهمیت یک ویژگی برای کاربران، بهترین گزینه محاسبه فراوانی تکرار آن ویژگی در نظرات است. به عنوان مثال اگر تمام ویژگی‌های استخراجی برابر ۱۰۰ و ویژگی دوربین و هم‌خانواده‌های آن ۱۵ بار تکرار شده باشند، ضریب این ویژگی برابر $0/15$ خواهد بود.

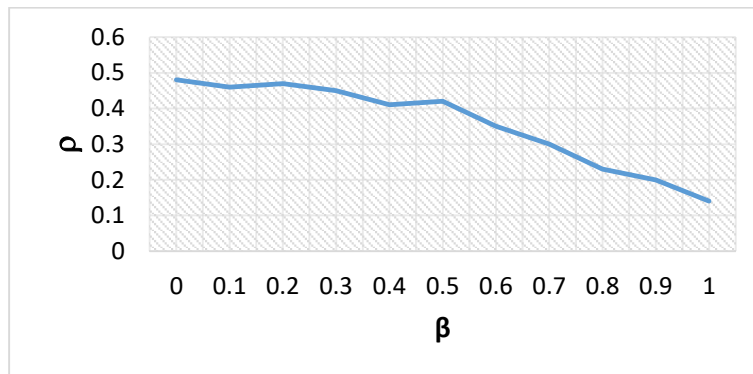
۸-۵- ارزیابی نتایج

• ارزیابی سناریو (۱)

اگر مقدار α برابر صفر و β بین ۰ تا ۱ باشد؛ همان‌طور که در شکل (۶) نشان داده شده است، با افزایش مقدار β ضریب همبستگی کاهش می‌یابد به طوری که بیشترین مقدار ضریب همبستگی به ازای $\beta=0$ و کمترین آن برای $\beta=1$ است. در واقع وجود مشخصات سخت افزاری بالا باعث افزایش قیمت آن شده لذا خرید آن در توان همه افراد نخواهد بود. بنابراین



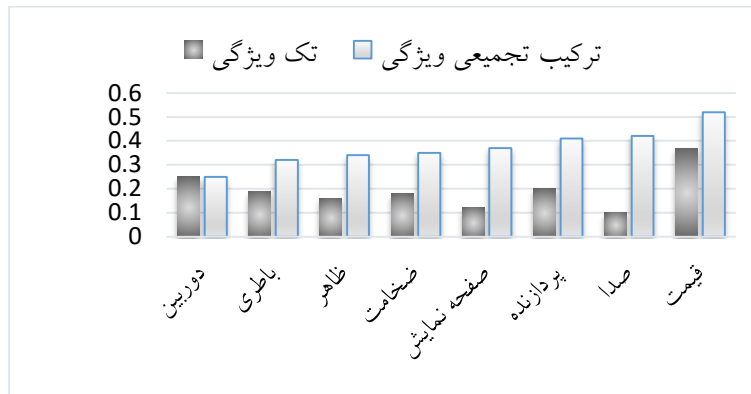
ویژگی های سخت افزاری معیار مناسبی جهت رتبه بندی محصولات نیستند.



شکل ۶ تغییرات ضریب همبستگی به ازای مقادیر β

• ارزیابی سناریو (۲)

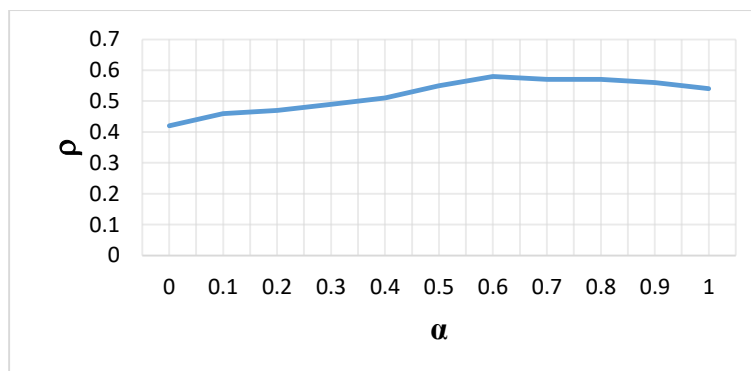
در این بخش تنها از ویژگی های حسی استفاده می شود. در آزمایش اول رتبه بندی تنها با تکیه بر یکی از ویژگی ها انجام می شود و در آزمایش دوم استفاده از ترکیب ویژگی ها به صورت تجمیعی انجام می گیرد. یعنی در مرحله اول تنها ویژگی دوربین، در مرحله دوم دوربین و باطری و در نهایت در حالت هشتم از همه ویژگی ها استفاده شده و نتایج مقایسه می شوند. در شکل (۷) در کنار ویژگی های تکی، نتایج مربوط به تجمیع ویژگی ها نشان داده شده است. در این شکل از چپ به راست ویژگی ها افزوده می شوند. نتایج نشان می دهد که استفاده از همه ویژگی ها باعث افزایش ضریب همبستگی شده است و مطابق انتظار در آخرین مرحله که ویژگی قیمت اضافه گردیده، ضریب همبستگی بیشترین مقدار تغییر را دارد.



شکل ۷ نتایج مربوط به تأثیر ترکیب تجمیعی ویژگی‌ها بر ضریب همبستگی

• ارزیابی سناریو (۳)

در این سناریو از هر سه گروه ویژگی استفاده شده و ضریب α متغیر بوده و ضریب ویژگی‌های عددی و نظرسنجی برابر $(1-\alpha)/2$ خواهد بود. در شکل (۸) نتایج نشان داده شده است. با افزایش α و به تناسب افزایش تأثیر ویژگی‌های حسی مقدار ضریب همبستگی بهبود می‌یابد و بیشترین مقدار آن برای $\alpha=0/6$ است. بهترین حالت برای زمانی است که ضریب تأثیر ویژگی‌های حسی $0/6$ و ویژگی‌های عددی و نظرسنجی هر کدام برابر $0/2$ باشد.

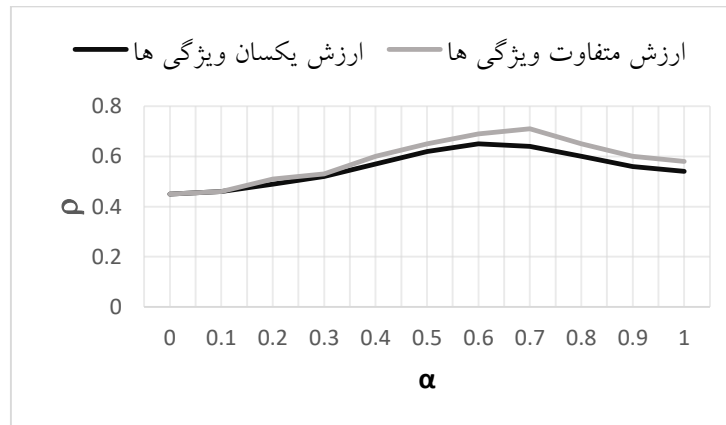


شکل ۸ تأثیر تغییر α در حالتی که هر سه گروه ویژگی استفاده می‌شوند



• ارزیابی سناریو (۴)

در این حالت، اهمیت ویژگی‌های مختلف برای کاربران متفاوت است. استفاده از ضرایب متفاوت برای ویژگی‌های حسی باعث افزایش ضریب همبستگی می‌شود. بعد از انجام همه آزمایشات مشخص شد که استفاده از ویژگی‌های حسی و ویژگی‌های نظرسنجی با مقدار $\alpha=0.7$ و استفاده از ویژگی‌های حسی با ضرایب متفاوت، با توجه به معیار ارزیابی، بهترین نتیجه را می‌دهد که در شکل (۹) مشخص است.



شکل ۹ نتایج تأثیر ویژگی‌های حسی با ضرایب متفاوت

با توجه به سیستم رتبه بندی پیشنهادی و انواع سناریو ذکر شده، نهایتاً از سناریوی ۴ برای رتبه بندی ۷۰ گوشی همراه جدول (۹) استفاده شد.

جدول ۹ لیست گوشی‌های مورد ارزیابی

لیست گوشی‌ها	برند
iPhone7_128, iPhone11_A2217, iPhone11_A2223, iPhone11_pro_A2220, iPhoneX_256, iPhone_XS_512, iPhone_8plus, iPhone_6s, iPhone_6s_plus, iPhone8, iPhone_SE, iPhone_SE_2020, iPhone_X64	اپل
Galaxy_A50s_SM, Galaxy_A51_SM, Galaxy_A71_SM, Galaxy_Note10_Lite_SM, Galaxy_A70_SM, Galaxy_A80_SM, Galaxy_J2_Core, GalaxyA01_SM, Galaxy_A20s, Galaxy_A30s, Galaxy_A31, Galaxy_A10s, Galaxy_A11, Galaxy_A21s, Galaxy_M11, Galaxy_S10, Galaxy_Note10_Pluss, Galaxy_Note10, Galaxy_S20, Galaxy_A9, Galaxy_Note8, Galaxy_J7	سامسونگ
Huawi_Y6_Prime, Huawe_Y9_Prime, Huawe_Y5, Huawe_P30Lite,	هواوی



لیست گوشی‌ها	برند
Honor_Lite_HRY-LX1T204.2, Honor_Lite_HRY-Lx, Huawe_p30Lite_MARLX1A, Huawe_P30Lite_MARLX1M, Huawe_Y8, Huawe_Y9s, Huawe_Y8s, Huawe_Y7prime, Huawe_Y6s, Huawe_Y7p, Huawe_Nova_3i, Huawe_Nova_5T	
Xiaomi_Mi A3_M1906F9SH, Xiaomi_Mi_Note10, Xiaomi_Mi9T_M1903F10G, Xiaomi_Mi9T_Pro_M1903F11G, Xiaomi_Redmi_Note7, Xiaomi_Redmi_Note8, Xiaomi_Redmi_Note8_Pro, Xiaomi_Redmi_Note8T, Xiaomi_Redmi_Note9S, Xiommi_Mi8_SE	شیائومی

در جدول (۱۰) نتیجه رتبه‌بندی برای ۱۰ گوشی اول به ترتیب نشان داده شده است. براساس رتبه‌بندی بدست آمده، رتبه هر گوشی ارتباط مستقیمی با قیمت گوشی‌ها ندارد و می‌توان نتیجه‌گرفت که کاربران معیارهای متفاوتی را برای برتری گوشی‌ها در نظر می‌گیرند.

جدول ۱۰ رتبه‌بندی ۱۰ گوشی برتر براساس روش پیشنهادی

گوشی	رتبه
Xiaomi_Redmi_Note8_Pro	۱
Galaxy_A50s_SM	۲
Huawe_Y9_Prime	۳
Galaxy_A71_SM	۴
iPhone11_A2223	۵
Xiaomi_Redmi_Note9S	۶
Xiaomi_Mi9T_Pro_M1903F11G	۷
iPhone11_A2217	۸
Galaxy_Note10	۹
Xiaomi_Mi_Note10	۱۰

۶- نتیجه‌گیری و پیشنهادات آتی

با گسترش خرید از خرده‌فروشی‌های برخط، مشتریان سعی می‌کنند قبل از خرید، بررسی‌های برخط متخصصین و سایر مشتریان را که از محصول استفاده کرده‌اند، مورد توجه قرار دهند. در مطالعات مشابهی که در بخش ۳ به تفصیل تشریح گردید، در بحث رتبه‌بندی محصولات براساس نظرات کاربران، ویژگی‌های ذهنی و عینی مورد توجه قرار گرفته است و در برخی از پژوهش‌ها، ادغام اطلاعات متنی، رتبه‌بندی عددی موجود در



تارنمای فروش و جملات تطبیقی برای رتبه بندی محصولات انجام گرفته است [۲۲]. اما در هیچ کدام از پژوهش‌های انجام گرفته به نظرات تخصصی خبرگان که در قالب نقد و بررسی محصول است و ویژگی‌های سخت افزاری محصولات برای رتبه بندی توجهی نشده است. در این پژوهش با ترکیب سه نوع از داده‌ها شامل: نظرات متنی مشتریان و متخصصین، داده‌های مربوط به نظرسنجی مشتریان و داده‌های مربوط به ساخت محصول، سیستم رتبه بندی جدیدی ارائه گردید. از طرفی با به کارگیری نظراتی که در تارنمای دیجی کالا به عنوان مفیدترین نظرات مرتب شده‌اند، بر غنای نتایج حاصل از رتبه بندی افزوده گردید. پس از پیش برداشتن نظرات متنی مشتریان و متخصصین، در ادامه و با استفاده از رویکرد TFIDF معنایی، ۹ مورد از مهم‌ترین ویژگی‌های حسی استخراج گردید. سپس مجموعه اولیه از کلمات حسی مرتبط با ویژگی‌های استخراج شده، توسط فرد خبره بین ۴- الی ۴+ ارزش گذاری شدند. در صورتی که کلمات حسی موجود در متن نظرات در بین این کلمات نباشد، از شبکه واژگان حسی فردوسنت و فارسنت استفاده گردیده و یک ارزش متناسب با نوع رابطه برای آن‌ها مشخص می‌شود. در نهایت با نرمال سازی مقادیر ویژگی‌های سه گانه در فاصله صفر تا پنج و ترکیب این ویژگی‌ها در یک جدول و استفاده از سناریوهای مختلف، روش رتبه بندی مناسب مورد استفاده قرار گرفت و در نهایت محصولات رتبه بندی شدند.

با توجه به تلاش‌های فراوانی که در انجام دقیق این تحقیق صورت پذیرفته است، مشخصاً محدودیت‌هایی نیز وجود داشته است. از جمله این محدودیت‌ها، گستردگی جامعه آماری می‌باشد. در این پژوهش جامعه آماری صرفاً تارنمای دیجی کالا و نظرات مربوط به گوشی‌های همراه بود، با توجه به این که نوع پژوهش داده‌محور می‌باشد، گسترش جامعه آماری به سایر سایت‌های خرده‌فروشی برخط داخلی و حتی خارجی و از سویی استفاده از نظراتی که مصرف‌کنندگان در خصوص سایر محصولات ثبت می‌کنند، در دقت و اعتبار مدل پیشنهادی رتبه بندی محصولات می‌افزاید. از جمله محدودیت‌های دیگری که محقق در بحث جمع آوری داده‌ها با آن مواجه گردید، ورود دستی نظرات حسی متنی، داده‌های نظرسنجی و سخت‌افزاری گوشی‌های همراه به فایل‌های متنی و اکسل بود. علت ورود دستی داده‌ها، افزایش دقت بیشتر سیستم پیشنهادی می‌باشد اما مشخصاً باعث کاهش استفاده از نظرات بیشتر و جزو محدودیت‌های تحقیق به‌شمار می‌رود. در بحث روش تحقیق و رویکردهای آن، در این پژوهش از تحلیل احساسات مبتنی بر ویژگی استفاده گردید، این روش نیازمند به تفکیک جملات از لحاظ نحوی دارد و در ادامه بایستی از فرهنگ واژه‌هایی که در این زمینه



توسعه داده شده است، استفاده نمود. زبان فارسی محدودیت‌های خاص خود را در این خصوص دارد. در بررسی‌های انجام گرفته توسط محقق، نواقص فراوانی در پایگاه‌هایی چون فردوس نت و غیره به عنوان واژه‌نامه‌های مورد استفاده مشاهده گردید که بر دقت تحلیل احساسات متون فارسی تأثیرگذار است و این پژوهش نیز از این قاعده مستثنی نیست.

به پژوهشگران آتی پیشنهاد می‌گردد در جمع‌آوری داده‌های اولیه پژوهش، قلمرو مکانی تحقیق را که صرفاً تارنمای خرده‌فروشی دیجی‌کالا بوده است، گسترش داده و از سایر تارنماها و خرده‌فروشی‌های داخلی استفاده نمایند. از طرفی با گسترش قلمرو محصولات به یک حوزه خاص، به طور مثال کالاهای دیجیتالی، می‌توان نتایج جدید و قابل توجه‌تری کسب نمود. در بحث داده‌های اولیه می‌توان از افرادی که اخیراً در شبکه‌های اجتماعی فعالیت گشایش بسته را انجام می‌دهند، به‌عنوان متخصصین این امر استفاده نموده و نظرات ایشان را به عنوان ورودی فرد خبره در تحلیل‌های آتی وارد نمود. در خصوص استفاده از مشخصات ساخت محصول در رتبه‌بندی، در این پژوهش شش متغیر استفاده شده است که در سایت‌های خرده‌فروشی برخط به مراتب متغیرهای سخت افزاری بیشتری موجود است و می‌تواند در پژوهش‌های آتی مورد استفاده قرار گیرد. نکته دیگری که در بحث داده‌های اولیه وجود دارد این است که اخیراً تارنمای دیجی‌کالا پایگاه داده بسیار بزرگی از داده‌های متنی (نظرات مشتریان) در مورد کالاهای مختلف را تهیه نموده و در اختیار پژوهشگران قرار می‌دهد. جهت پژوهش‌های آتی محققین می‌توانند از این داده‌ها استفاده نمایند.

روش استفاده شده در این پژوهش تحلیل حساسیت مبتنی بر ویژگی بوده است، در بحث استخراج ویژگی‌های موجود در نظرات از روش TF-IDF معنایی استفاده گردید، روش‌های بسیار زیادی مبتنی بر واژگان یا الگوریتم‌های باناظر و بدون ناظر در حوزه یادگیری ماشین وجود دارد که می‌توان از آن‌ها بهره برد و با مقایسه روش‌های مذکور بهترین روش را انتخاب نمود. در خصوص استخراج احساس مربوط به هر ویژگی از روش مبتنی بر لغت‌نامه استفاده گردید که روش‌های متنوع دیگری وجود دارد که محققین قادر به استفاده از آن‌ها می‌باشند.



۷. پی‌نوشت

1. Sentiment analysis
2. Aspect based sentiment analysis
3. Natural language processing (NLP)
4. Supervised learning methods
5. Unsupervised learning methods
6. Association rules
7. Asghar et al
8. Firmanto et al
9. Yang et al
10. Numeric rating
11. Text descriptions
12. Comparative words
13. Comparative votes
14. Guo et al
15. Linear Discriminant Analysis(LDA)
16. Natural Language Processing NLP
17. Information Retrieval
18. Yago
19. Viterbi
20. On boxing

۸- منابع

- [1] Asgarian E, Kahani M, Sharifi S. Hesnegar: Persian sentiment wordnet. Signal and Data Processing. 2018 Jun 10;15(1):71-86.
- [2] Asghar MZ, Khan A, Zahra SR, Ahmad S, Kundi FM. Aspect-based opinion mining framework using heuristic patterns. Cluster Computing. 2019 May;22(3):7181-99.
- [3] Collomb A, Costea C, Joyeux D, Hasan O, Brunie L. A study and comparison of sentiment analysis methods for reputation evaluation. Rapport de recherche RR-LIRIS-2014-002. 2014 Mar.
- [4] Feldman R. Techniques and applications for sentiment analysis. Communications of the ACM. 2013 Apr 1;56(4):82-9.
- [5] Firmanto A, Sarno R. Aspect-Based Sentiment Analysis Using Grammatical Rules, Word Similarity and SentiCircle. International Journal of Intelligent Engineering and Systems. 2019 Oct;12(5):190-201.
- [6] Flanagan AJ, Metzger MJ, Pure R, Markov A, Hartsell E. Mitigating risk in ecommerce transactions: perceptions of information credibility and the role of user-generated ratings in product quality and purchase intention. Electronic Commerce Research. 2014 Mar;14(1):1-23.



- [7] Ghobakhloo M, Rajabzadeh Ghatari A, Toloie Eshlaghy A, Alborzi M. Provide an Opinion Analysis-Based Recommender System for Personalized Personal Banking Services. *Journal of Modern Research in Decision Making*. 2020 Jan Vol. (5) 1, 25-49.
- [8] Ghose A, Ipeiritis PG. Estimating the helpfulness and economic impact of product reviews: Mining text and reviewer characteristics. *IEEE transactions on knowledge and data engineering*. 2010 Sep 30;23(10):1498-512.
- [9] Guo C, Du Z, Kou X. Products ranking through aspect-based sentiment analysis of online heterogeneous reviews. *Journal of Systems Science and Systems Engineering*. 2018 Oct;27(5):542-58.
- [10] Jeyapriya A, Selvi CK. Extracting aspects and mining opinions in product reviews using supervised learning algorithm. In 2015 2nd international conference on electronics and communication systems (ICECS) 2015 Feb 26 (pp. 548-552). IEEE.
- [11] Kong R, Wang Y, Xin W, Yang T, Hu J, Chen Z. Customer reviews for individual product feature-based ranking. In 2011 first international conference on instrumentation, measurement, computer, communication and control 2011 Oct 21 (pp. 449-453). IEEE.
- [12] Liu B. Sentiment analysis and opinion mining. *Synthesis lectures on human language technologies*. 2012 May 22;5(1):1-67.
- [13] Liu Y, Bi JW, Fan ZP. Ranking products through online reviews: A method based on sentiment analysis technique and intuitionistic fuzzy set theory. *Information Fusion*. 2017 Jul 1;36:149-61.
- [14] Liu Y. Investigation of Viterbi Algorithm Performance on Part-of-Speech Tagger of Natural Language Processing. In 2017 International Conference on Computer Systems, Electronics and Control (ICCSEC) 2017 Dec 25 (pp. 1430-1433). IEEE.
- [15] Maas A, Daly RE, Pham PT, Huang D, Ng AY, Potts C. Learning word vectors



- for sentiment analysis. In Proceedings of the 49th annual meeting of the association for computational linguistics: Human language technologies 2011 Jun (pp. 142-150).
- [16] Mahdisoltani F, Biega J, Suchanek F. Yago3: A knowledge base from multilingual wikipedias. In 7th biennial conference on innovative data systems research 2014. CIDR Conference.
- [17] Maleki Minbashrazgah, M, Afsharizadeh Jafari, O. Adaptive Clustering of Food Industry Companies' Vision, Mission, and Values using Text Mining. *Journal of Business Management Perspective*, 2018 Vol. (16) 32, 163-179.
- [18] Mäntylä MV, Graziotin D, Kuutila M. The evolution of sentiment analysis—A review of research topics, venues, and top cited papers. *Computer Science Review*. 2018 Feb 1;27:16-32.
- [19] Montazery M, Faili H. Automatic Persian wordnet construction. In *Coling 2010: Posters 2010 Aug* (pp. 846-850).
- [20] Shamsfard M, Hesabi A, Fadaei H, Mansoory N, Famian A, Bagherbeigi S, Fekri E, Monshizadeh M, Assi SM. Semiautomatic development of farsnet; the persian wordnet. In *Proceedings of 5th global WordNet conference, Mumbai, India 2010 Jan 31* (Vol. 29).
- [21] Sohrabi B, Raeesi Vanani I, Zareh Mirkabad F. Designing a Recommender System for Optimizing and Managing Bank Facilities through the Utilization of Clustering and Classification Algorithms. *Journal of Modern Research in Decision Making*. 2016 Jan Vol. (1) 2, 53-76.
- [22] Yang X, Yang G, Wu J. Integrating rich and heterogeneous information to design a ranking system for multiple products. *Decision Support Systems*. 2016 Apr 1;84:117-33.
- [23] Zarei A, Feiz D, Taheri G. Providing Social Market Intelligence Framework based on web 2.0 Using Text-Mining Technique on Social Media Websites



(Case Study: Competitive Analysis between Samsung and Emersun Brands).
Journal of Management Research in Iran. 2020 Vol. (24) 4, 98-125.

- [24] Zhang K, Cheng Y, Liao WK, Choudhary A. Mining millions of reviews: a technique to rank products based on importance of reviews. In Proceedings of the 13th international conference on electronic commerce 2011 Aug 3 (pp. 1-8).
- [25] Zhang K, Narayanan R, Choudhary AN. Voice of the Customers: Mining Online Customer Reviews for Product Feature-based Ranking. WOSN. 2010 Jun 22;10:11-.
- [26] Zhang Z, Guo C, Goes P. Product comparison networks for competitive analysis of online word-of-mouth. ACM Transactions on Management Information Systems (TMIS). 2013 Jan 1;3(4):1-22.