

# کمی سازی گرایش احساسی نظرات متنی فارسی مشتریان بر روی ویژگی های کالا در وب

سمانه کریمیان<sup>۱</sup>، محمدجواد کارگر<sup>۲</sup>

<sup>۱</sup> گروه مهندسی کامپیوتر، دانشکده فنی مهندسی، دانشگاه علم و فرهنگ، تهران، ایران  
k.samaneh841@yahoo.com

<sup>۲</sup> گروه مهندسی کامپیوتر، دانشکده فنی مهندسی، دانشگاه علم و فرهنگ، تهران، ایران  
kargar@usc.ac.ir

## چکیده

امروزه، اینترنت به مشتریان امکان می دهد تا نظرات و عقاید خود را در مورد سازمان ها و محصولات بیان نموده و سایر افراد را در مورد تجربه ی خود مطلع کنند، بدین ترتیب تأثیر نظرات مشتریان بر موفقیت یک کسب و کار غیر قابل چشم پوشی است. هدف این پژوهش، معرفی یک روش مبتنی بر واژه نامه برای استخراج و کمی سازی گرایش احساسی مشتریان در مورد ویژگی های یک کالا با بررسی و تحلیل نظرات آنها به زبان فارسی می باشد. پژوهش حاضر از جنبه هدف، کاربردی و از منظر روش تحقیق از نوع پیمایشی مقطعی می باشد. در این تحقیق، نظرات مشتریان پیرامون ۱۱ ویژگی ۳ مدل گوشی تلفن همراه در فروشگاه آنلاین دیجی کالا در تابستان سال ۱۳۹۳ جمع آوری گردید و در هشت گام شامل: انتخاب کالای مورد نظر و جمع آوری نظرات، پیش پردازش داده های متنی، استخراج ویژگی های کالا، ایجاد واژه نامه ها، استخراج الگوهای نظرات، تعیین قطبیت یا گرایش احساسی بخش های نظرات، پس پردازش و در انتها جمع بندی و کمی سازی گرایش احساسی نظرات مورد تحلیل قرار گرفت. نتایج اعتبارسنجی نشان می دهد روش معرفی شده از دقت بالایی در کمی سازی گرایشهای احساسی مشتریان برخوردار می باشد.

## کلمات کلیدی

متن کاوی، عقیده کاوی، تحلیل احساسات، کمی سازی نظرات در وب

### ۱- مقدمه

بسیاری از مشتریان به بررسی ویژگی های کالا/خدمات در نظرات خود می-پردازند که توجه به آن می تواند بر بهبود کیفیت کالا/خدمات تأثیر بسزایی داشته باشد.

در مطالعات متعددی سعی شده است تا سیستم هایی ارائه شود که به تحلیل خودکار یا نیمه خودکار این نظرات پرداخته و نتایج قابل قبول و قابل کاربردی را نمایش دهند. این سیستم ها، که به سیستم های عقیده کاوی یا تحلیل احساسات شهرت دارند، با استفاده از تکنیک های داده کاوی و متن-کاوی بر روی این متون انتقادی، نظیر خلاصه سازی، طبقه بندی، خوشه بندی و ...، فروشندگان را برای تصمیم گیری بهتر در مورد ارائه کالاها و خدمات یاری می کند [۱].

با گسترش سریع تجارت و کسب و کار الکترونیکی و فروش کالاها/خدمات تحت وب، تعداد خریداران این کالاها/خدمات نیز رو به افزایش است. به منظور افزایش رضایت مشتری و نیز تجربه ی خرید وی، بیشتر فروشندگان و سازمان ها این امکان را برای مشتریان خود فراهم آورده اند تا در مورد کالاها/خدمات آن ها نظرات و عقاید خود را بیان نمایند. تمایل افراد برای بیان نظرات خود و به اشتراک گذاری آن با دیگران، در رفتار سایرین نیز اثرات بالقوه ای خواهد داشت. عدم توجه به این تأثیرات می تواند منجر به شکست و در نهایت نابودی یک کسب و کار در دنیای مجازی شود. افزون بر این،

[۱۸]. تمرکز روش معرفی شده در این مقاله بر روی اسناد انتقادی به زبان چینی در مورد رستوران‌ها است و استخراج ویژگی‌ها بر مبنای قید صورت می‌گیرد. برای تشخیص قیده‌ها هم از برچسب‌های ادات سخن استفاده می‌شود.

## ۱-۲- تعیین گرایش احساسی در سطح کلمات و عبارات

در این رویکرد تلاش بر این است تا گرایش احساسی کلمات و عبارات منحصر به فرد تعیین شود. مبنای کار در این رویکرد این است که در جملات یک متن انتقادی، مشتری ممکن است ویژگی‌های مختلفی از کالا را با گرایش‌های احساسی متفاوتی بیان نماید. به همین جهت لازم است تا ابتدا گرایش احساسی کلمات و عبارات تعیین گردد.

پاسکیو و اتزیونی، از روش برچسب‌سزنی ساده شده (که یک روش رده‌بندی بدون ناظر است) برای یافتن گرایش معنایی واژگان در یک زمینه خاص، جهت کارایی بهتر استفاده می‌کنند [۱۶]. در تکنیک بیان شده توسط لیو، که بر روی الگوی دوم بیان نظرات کاربرد دارد، گرایش احساسی هر الگوی استخراجی توسط نقاط قوت و ضعف مشخص می‌گردد [۱].

استفاده از واژه نامه برای تعیین گرایش احساسی در سطح کلمات نیز در مطالعات متعددی به کار رفته است [۱۳، ۱۲، ۶-۱۴]. در این روش مترادف‌ها و متضادهای یک صفت در شبکه‌ی واژه‌ها، بسته به اینکه صفت دارای چه گرایش احساسی باشد، به لیست مثبت‌ها و منفی‌ها اضافه می‌شود. از این لیست، در مراحل بعدی برای تعیین گرایش احساسی در سطح جمله و عبارت استفاده می‌شود.

## ۱-۳- تعیین گرایش احساسی در سطح جملات متن انتقادی

پس از تعیین گرایش احساسی کلمات و عبارات، در این رویکرد، سعی می‌شود تا جملات هر متن انتقادی به صورت مجزا رده‌بندی گردند. اساس این رویکرد این است که در یک متن انتقادی لزوماً تمام جملات در مورد یک کالا منفی یا مثبت نیست، بلکه ترکیبی از این دو می‌باشد.

در روش هو و لیو، برای خلاصه‌سازی در سطح ویژگی کالاها، جملاتی که در برگیرنده‌ی یک ویژگی خاص هستند تجمیع شده و تعداد نظرات مثبت و منفی در مورد آن شمارش می‌گردد [۱۳].

در [۲۰]، علاوه بر ماژول تعیین احساسات در سطح کلمه، ماژولی برای تعیین احساسات در سطح جمله نیز وجود دارد. برای این مرحله نیز از دو مدل میانگین هارمونیک و میانگین هندسی استفاده شده است.

در [۱۹]، که به خلاصه‌سازی نظرات مشتریان در وب می‌پردازد، بخش استخراج ویژگی کالاها از روشی مشابه [۱۳] می‌کند که شامل فرآیندهای زیر است: تحلیل برچسب‌های ادات سخن، محاسبه‌ی فراوانی عبارت-معکوس فراوانی جمله (tf-isf)، اگر صفتی در نزدیکی این اسامی نماینده قرار داشته باشد، با آن نماینده به عنوان یک ویژگی عقیده‌ای مکرر رفتار می‌شود.

## ۱-۴- تعیین گرایش احساسی در سطح متن نظر

به طور کلی، مهم‌ترین مزیت رده‌بندی احساسی در سطح اسناد این است که باعث فراهم آوردن عقیده غالب بر سند می‌شود. با این وجود، کاستی اصلی

بر خلاف پژوهش‌های گسترده‌ای که در چند سال اخیر بر روی عقیده کاوی بر روی متون زبان انگلیسی صورت گرفته است، روش و سیستمی ساختیافته برای عقیده کاوی نظرات بیان شده به زبان فارسی ارائه نشده است. از این رو در این پژوهش، سیستمی برای انجام عقیده کاوی مبتنی بر ویژگی‌ها و مشخصات فنی یک گروه کالایی در زبان فارسی، توسعه داده می‌شود. با تحلیل این نظرات، از دید درون‌کشوری و درون سازمانی، سازمان‌ها می‌توانند برنامه‌های بازاریابی، روش‌های بهبود ارائه خدمات و کیفیت محصولات، راه‌هایی برای توسعه و ورود به بازارهای جهانی را در اهداف و برنامه‌های بلند مدت کسب و کار خود در نظر بگیرند.

برای بیان عقاید و نظرات در اینترنت، سه قالب کلی وجود دارد: نقاط قوت و ضعف، نقاط قوت و ضعف همراه با نظر کوتاه و قالب آزاد [۱]. برای تعیین گرایش احساسی در قالب اول و دوم، تنها کافی است تا ویژگی‌های کالا/خدمت مشخص گردد، زیرا گرایش احساسی در بخش نقاط قوت و ضعف به وضوح تعیین شده است. اما در قالب سوم، هم ویژگی‌ها و هم گرایش احساسی هر یک باید تشخیص داده شود. بنابراین اولین مسئله در عقیده کاوی متون با قالب آزاد، تعیین ویژگی‌های کالا/خدمت در نظرات مشتریان است. دومین مسئله در عقیده کاوی، تشخیص اطلاعات ذهنی مشتریان و سپس رده‌بندی عقاید بر اساس نظرات یافت شده برای ویژگی کالاها است. در این حالت رویکردهای عقیده کاوی به سه دسته کلی تقسیم می‌شوند:

۱. تعیین گرایش احساسی ویژگی‌ها یا صفات مطرح شده برای کالا/خدمات.

۲. تعیین گرایش احساسی جملات عقیده‌ای در هر نظر.

۳. تعیین گرایش احساسی هر نظر به صورت کلی.

علاوه بر این، روش‌های رده‌بندی احساسی متون انتقادی به دو دسته نظارت شده و بدون ناظر نیز تقسیم می‌شوند. در روش‌های نظارت شده، بر چسب کلاس هر متن یا جمله انتقادی مشخص بوده و میزان صحت نتایج بر اساس روابط رده‌بندی مشخص می‌گردد. در این روش‌ها، معمولاً از تکنیک‌های یادگیری ماشین در آموزش و ایجاد مدلی برای رده‌بندی اسناد استفاده می‌شود. در روش‌های بدون ناظر، معمولاً از واژه‌نامه‌ها استفاده می‌گردد و گرایش احساسی با شمارش مثبت‌ها و منفی‌ها به دست می‌آید.

## ۱-۱- استخراج ویژگی‌ها و عبارات کلیدی

همان‌طور که بیان شد، اولین مسئله در عقیده کاوی یافتن ویژگی‌های کالا در نظرات مشتریان است که در این بخش به این موضوع پرداخته می‌شود. ترنی، برای استخراج ویژگی‌ها و عبارات کلیدی به یافتن عباراتی پرداخته است که حاوی صفت یا قید هستند، زیرا این دو بیانگر یک عقیده شخصی در جمله می‌باشند [۳].

لئونگ و همکاران نیز، برای استخراج ویژگی‌ها از یک متن نظر، از برچسب ادات سخن استفاده کرده‌اند [۱۲]. زی و همکاران، برای استخراج ویژگی و بهبود نتایج از روش ترکیبی برچسب ادات سخن به همراه ارتباط کلمات استفاده و رده‌بندی بر اساس آن‌ها صورت می‌گیرد [۵].

در [۱۷] برای استخراج ویژگی‌ها از یک سری نشانه‌های خاص نظیر حروف بزرگ، برچسب‌های ادات سخن، نوع کلمه و مانند آن بهره برده شده است. لیو و همکاران، توصیه‌گری مبتنی بر اولویت‌ها و عقاید مشتریان ارائه نموده‌اند

## شکل (۱): خلاصه روش انجام تحقیق

### ۲-۱- انتخاب کالای مورد نظر و جمع آوری نظرات

هدف نخست در این گام انتخاب یک کالا جهت بررسی و تحلیل نظرات آن است که در این پژوهش نظرات بیان شده در مورد تلفن همراه مورد بررسی قرار می‌گیرد و سایت دیجی کالا ۴ برای جمع آوری نظرات انتخاب شده است. همان‌طور که در بخش (۲) بیان شد، نظرات ممکن است در سه قالب کلی بیان شوند، از آنجا که شیوه‌ی مطرح شده در این پژوهش به صورت بدون ناظر عمل می‌کند، قابلیت به کارگیری بر روی هر سه قالب بیان نظرات را دارد. نظرات می‌تواند با استفاده از یک خزنده‌ی وب ۵ و یا به صورت دستی جمع‌آوری شود. هر یک از متون انتقادی به صورت یک فایل متنی مجزا ذخیره می‌گردد. علت این امر، سادگی محاسبه‌ی تعداد مشتریان برای به کارگیری در مراحل بعدی و نیز کاهش زمان پردازش در متون کوتاه است.

### ۲-۲- پیش پردازش داده‌های متنی

در این گام، که به نام آماده‌سازی متون نیز شناخته می‌شود، عملیات ابتدایی متن کاوی انجام می‌گیرد. یکی از مشکلات کار با متون فارسی شیوه‌های نوشتاری متفاوت برای نوشتن یک کلمه‌ی خاص است. در زبان فارسی نوشتار عامیانه علاوه بر نوشتار کتابی و اداری برای نوشتن متن نظرات رایج است و بسیاری از کاربران فارسی زبان، نوشتار عامیانه را نسبت به کتابی ترجیح می‌دهند. به عبارت دیگر در هنگام بیان نظرات توسط مشتریان به زبان فارسی قواعد این زبان رعایت نمی‌گردد و همین امر سبب دشواری تحلیل متون فارسی می‌گردد.

در این پژوهش برای انجام این کار، افزونه‌ی زبان فارسی که برای پلتفرم گیت نوشته شده است [۲۳]، به کار رفته است. این گام با استفاده از این افزونه به وظایف زیر شکسته می‌شود:

- تفکیک جملات به کلمات و نشانه‌ها: در این مرحله هر نشانه در متن شناسایی می‌شود. همان‌طور که قبلاً نیز بیان شد، هر نشانه می‌تواند یک واژه، عدد و علائم نقطه‌گذاری و غیره باشد.
- تفکیک متون به جملات: هر یک از متون به جملات تفکیک می‌شوند تا در مراحل بعدی قابل استفاده باشند. در واقع با استفاده از جملات محدودی اثر یک بخش عقیده‌ای کوچکتر می‌شود و از گذاشتن اثر جانبی بر سایر بخش‌های متن نظر جلوگیری می‌کند.
- برچسب‌گذاری ادات سخن یا تعیین مشخصه‌ی دستوری نشانه‌ها: نوع برچسب و نقشی که می‌تواند به یک نشانه در جمله اختصاص داده شود، در این مرحله تعیین می‌گردد. این نقش‌ها برای تعیین و شناسایی الگوهای زبانی مورد استفاده قرار می‌گیرد.

### ۲-۳- استخراج ویژگی‌های کالا

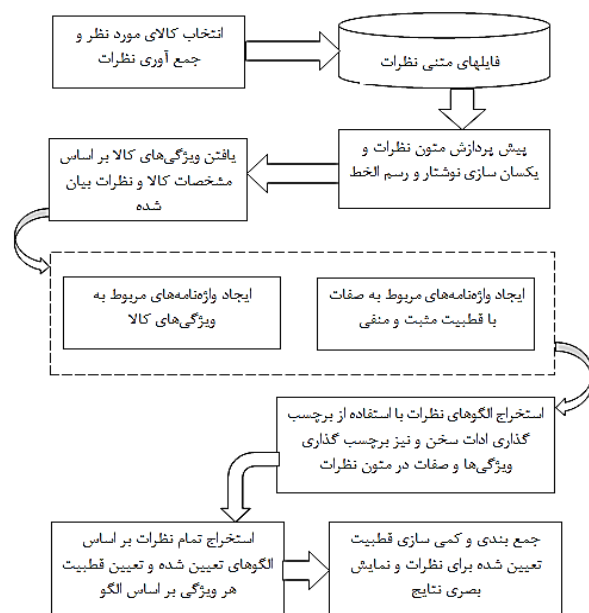
هدف از این مرحله مشخص نمودن ویژگی‌هایی است که در متون نظرات به دنبال آنها هستیم. از آنجا که نظرات مود بررسی در مورد تلفن‌های همراه می‌باشد، ویژگی‌های مطرح برای یک تلفن همراه در نظر گرفته می‌شود. این ویژگی‌ها و در واقع مشخصات فنی کالا است که در سایت دیجی کالا برای هر دستگاه نشان داده شده است. با در نظر گرفتن این حالت، ویژگی‌های

رده‌بندی در سطح اسناد به این شرح است که جزئیات تمايلات و عدم تمايلات افراد را مشخص نمی‌کند.

ترتی با محاسبه میانگین گرایش‌های معنایی به دست آمده برای لغات، مثبت یا منفی بودن هر یک از متون انتقادی را تعیین نموده است [3]. روش‌های یادگیری ماشین نیز نظیر بیز ساده [۴، ۵، ۶]، ماکزیمم آنتروپی [۴، ۵]، ماشین‌های بردار پشتیبان [۴، ۵، ۷، ۸]، SMO<sup>۲</sup>، که حالت بهینه شده ماشین بردار پشتیبان است، [۱۸] و شبکه‌های عصبی مصنوعی [۸] برای رده‌بندی احساسی متون انتقادی به کار برده شده است. دیو وهمکاران، برای تعیین گرایش احساسی هر متن یا رده‌بندی آن به یکی از دو کلاس مثبت و منفی از یک تابع امتیاز استفاده می‌کنند [۱۵]. در [۱۲] نیز، تعیین گرایش احساسی در سطح متن انتقادی بر اساس رتبه‌بندی‌های پنج‌گانه و یک تابع امتیاز صورت می‌گیرد، برخلاف روش‌های دیگر که فقط مثبت یا منفی بودن یک متن را در نظر می‌گیرند.

### ۲- روش و ابزار تحقیق

آنچه در این بخش بدان پرداخته می‌شود ارائه‌ی یک روش مبتنی بر قانون و واژه نامه و بدون ناظر برای استخراج بخش‌های عقیده‌ای در نظرات ثبت شده به زبان فارسی توسط مشتریان در خصوص یک کالای خریداری شده است. روش ارائه شده در این پژوهش یک رویکرد نیمه خودکار است، چرا که ایجاد واژه‌نامه‌ها تا حدی به صورت دستی و بر اساس دامنه‌ی مورد نظر انجام گرفته است. همچنین به منظور فراهم آوردن الگویی مناسب برای استخراج عقاید، نیاز است که بخش‌هایی از داده‌های متنی موجود به صورت دستی برچسب گذاری شده و سپس بر اساس این موارد، الگوها و قوانین زبانی موجود شناسایی گردند. این قوانین می‌توانند در شناسایی و برچسب گذاری خودکار بخش‌های عقیده‌ای سایر متون نظرات مورد استفاده قرار گیرند. در این پژوهش به منظور انجام پردازش‌های متنی خودکار از نرم‌افزار متن کاوی 7.0 Gate Developer استفاده شده است که یک پلتفرم نرم‌افزاری متن باز و رایگان برای پردازش زبان طبیعی می‌باشد. [۲۱]. شکل (۱) خلاصه‌ای از مراحل رویکرد ارائه شده را نشان می‌دهد.



مشخصات ظاهری، باتری، دوربین فیلمبرداری، بدنه، نمایشگر، حافظه، انتقال داده، پردازنده، نرم افزار، صدا و سایر مشخصات استخراج شد. نظرات هر کاربر بر اساس این ویژگی‌ها مورد تجزیه و تحلیل قرار می‌گیرد و گرایش احساسی هر بخش نظر و کل متن نظر به صورت کل برای هر ویژگی تعیین می‌گردد.

## ۲-۴- ایجاد واژه‌نامه‌ها

تحلیل نظرات مشتریان نیازمند این است که بتوان بخش‌هایی از متن نظر را که در رابطه با یک ویژگی خاص در مورد کالا است و توصیف مربوط به آن شناسایی نمود. منظور از توصیف، استفاده از کلماتی است که نشان دهنده‌ی یک گرایش احساسی در مورد یکی از ویژگی‌های کالایی است که نظرات آن مورد تحلیل قرار می‌گیرد. آنچه مد نظر است، این است که با داشتن مجموعه‌ای از نظرات، که هر کدام از آنها از چندین جمله تشکیل شده است و با این فرض که هر جمله می‌تواند حاوی توصیفی در مورد یک ویژگی کالا باشد، بتوان این جملات را شناسایی و به همراه توصیف مربوط به آنها برچسب گذاری کرد. این امر باعث می‌شود تا امکان یافتن قوانین زبانی موجود در بیان نظرات فراهم شود. بنابراین بر اساس نیاز رویکرد پیشنهادی در این پژوهش به سه نوع واژه‌نامه یا لیست واژگان احتیاج است:

- واژه‌نامه‌ی مرتبط با ویژگی‌ها: بر اساس لیست ویژگی‌هایی که در مرحله‌ی قبل شناسایی شد، لیست واژگان مرتبط با هر یک از آنها نیز تکمیل می‌گردد. در این حالت برای هر ویژگی یک واژه‌نامه‌ی مجزا تشکیل می‌شود و بدین ترتیب مشخص می‌گردد که هر واژه با کدام ویژگی در ارتباط است.
- واژه‌نامه‌ی توصیفی یا مرتبط با صفات: این لیست حاوی صفات عقیده-ای و گرایش احساسی هر یک (مثبت یا منفی) است. لازم به ذکر است که فقط صفاتی در نظر گرفته می‌شوند که یک گرایش احساسی را نیز بیان نمایند. از این رو، در این مرحله لیست این صفات نیز تهیه می‌گردد. علاوه بر این، برخی واژه‌های توصیفی نیز وجود دارند که گرایش احساسی آنها برای ویژگی‌های مختلف کالا متفاوت است. به عنوان مثال می‌توان صفت توصیفی «کم» را در نظر گرفت. این صفت اگر در مورد ویژگی «بدنه» تلفن همراه در نظر گرفته شود، گرایش احساسی مثبت دارد و در مقابل اگر در مورد ویژگی «باتری» آن در نظر گرفته شود، دارای گرایش احساسی منفی می‌باشد. البته تعیین این گرایش احساسی بستگی به نحوه‌ی قرار گیری این صفات احساسی در کنار ویژگی‌های کالا دارد.
- واژه‌نامه‌ی نفی کننده: این واژه‌نامه حاوی کلمات نفی کننده نظیر «نیست»، «نمی‌شود» و غیره است که در مرحله‌ی پس پردازش از آن استفاده می‌شود. نفی کننده‌ها با کلمه‌ی کلیدی Negation مشخص می‌گردند. هرگاه این کلمات در جمله یا عبارتی که برای توصیف بخش از کالا به دست آمده قرار گیرند، گرایش احساسی آن را تغییر می‌دهند. برای تشکیل واژه‌نامه یک لیست اولیه، حداقل حاوی ۱۰ واژه، در نظر گرفته می‌شود و لغات هم خانواده و هم گروه هر یک از آنها به صورت عمقی افزوده می‌شود. بدین ترتیب برای هر یک از ویژگی‌ها لیستی حدود حداقل ۴۰ واژه و برای صفات توصیف کننده احساس حداقل ۷۰ واژه به دست می‌آید. برخی از این واژه‌ها می‌تواند بر اساس تحلیل دستی متون نظرات استخراج گردند تا ارتباط آن با موضوع مورد بررسی حفظ گردد.

علاوه بر این باید در نظر داشت که نمی‌توان صرفاً به تهیه‌ی لیستی از اسامی برای ویژگی‌های کالا و یا لیستی از صفات برای واژه‌نامه‌ی توصیفی اکتفا نمود. همانند هر زبان دیگر، در زبان فارسی نیز برخی از کلمات در نقش‌های مختلف دستوری (نظیر اسم، صفت، قید، فعل و مانند آن) می‌توانند بیان کننده‌ی یک ویژگی خاص و یا توصیف کننده یک احساس و نظیر آن باشند. از آنجایی که در این تحقیق از پلتفرم گیت برای تحلیل نهایی الگوها استفاده می‌شود، باید واژه‌نامه‌ها را نیز بر اساس این پلتفرم تهیه نمود. در این ابزار، واژه‌نامه‌های مورد استفاده برای برچسب گذاری کلمات در ساختاری به نام فرهنگ واژگان<sup>۳</sup> نگهداری می‌شود. این ساختارها می‌توانند حاوی یک یا چند واژه‌نامه‌ی مختلف باشند. جدول (۳) نشان دهنده‌ی فرهنگ واژگان و زیرمجموعه‌ی آنهاست و اینکه چگونه هر یک از این واژه‌نامه‌ها تشکیل می‌گردند.

فرهنگ واژگان	واژه‌نامه‌های موجود	چگونگی تشکیل
فرهنگ واژگان ویژگی‌های کالا	واژه‌نامه‌ی مربوط به باتری، ظاهر، بدنه، انتقال داده، دوربین، نرم افزار، نمایشگر، پردازنده، حافظه، صدا و سایر مشخصات	اسامی ساده و مرکب افعال بیان کننده‌ی یک ویژگی بن‌های ماضی و مضارع مرتبط با ویژگی
فرهنگ واژه توصیفی	واژه‌نامه‌ی توصیفات مثبت عام	صفات و افعال توصیفی مثبت
	واژه‌نامه‌ی توصیفات منفی عام	صفات و افعال توصیفی منفی
	واژه‌نامه‌ی توصیفات مبهم	صفات و افعال توصیفی با گرایش احساسی دو جانبه برای ویژگی‌های مختلف
	واژه‌نامه‌ی لغات نفی کننده	افعال، صفات و اسامی نفی کننده

جدول (۳): لیست فرهنگ واژگان موجود به همراه واژه‌نامه‌های آنها

## ۲-۵- استخراج الگوهای نظرات

برای استخراج الگوهای موجود در متن نظرات مشتریان، از یک رویکرد نیمه خودکار استفاده شده است. در واقع روند کار بدین صورت است که ابتدا تمام ویژگی‌های کالا و توصیفات مرتبط به آنها در متون برچسب زده می‌شود و آنگاه بر اساس این موارد، الگوهای ممکن برای بیان عقاید مشتریان استخراج می‌گردد. در مرحله‌ی بعد می‌توان از این الگوها استفاده نمود تا هنگام تطبیق با هر کدام از بخش‌های متون جدید، ویژگی‌های کالا را مشخص کرده و به صورت خودکار توصیف مرتبط با آنها را به دست آورد.

استخراج الگوی نظرات از جملات شامل سه مرحله است:

۱. برچسب زدن ویژگی‌های کالا که در نظرات مشتریان به آن اشاره شده و در فرهنگ واژگان مربوطه آمده است؛
۲. برچسب زدن تمام توصیفات کننده‌ها و نفی کننده‌ها بر اساس فرهنگ واژگان توصیفی؛
۳. مشخص کردن توصیفات کننده‌ها و ویژگی‌های کالایی که با یکدیگر مرتبط هستند و نماینده‌ی یک نظر مثبت یا منفی در مورد یک ویژگی کالا هستند.

جدول (۴) مشخصات مجموعه داده‌های جمع آوری شده را نشان می‌دهد.

نام مدل تلفن همراه	تعداد الگوهای یافت شده	تعداد کل جملات	تعداد متون
Apple iPhone 5s	۳۵۸	۲۰۵۶	۵۷۹
Samsung Galaxy Note 3 N9000	۴۹۰	۲۷۹۵	۶۹۱
LG G2 - 32GB	۳۳۸	۱۰۰۲	۲۵۰
مجموع	۱۱۸۶	۵۸۵۳	۱۵۲۰

جدول (۴): مشخصات مجموعه داده‌های مورد استفاده

باید توجه داشت که تعداد الگوهای یافت شده نسبت به تعداد کل جملات و نیز تعداد کل متون نظرات مشتریان کمتر می‌باشد. دلیل این امر این است که مشتریان در هنگام بیان نظرات خود در رابطه با یک موضوع، به مسائل حاشیه‌ای نیز می‌پردازند. به عبارت دیگر تمام جملاتی که در یک متن نظر بیان شده‌اند لزوماً یک گرایش احساسی را نسبت به یکی از ویژگی‌های کالا بیان نمی‌کنند.

## ۲-۶- تعیین قطبیت یا گرایش احساسی بخش‌های نظرات

با استفاده از الگوهای به دست آمده در مرحله قبل و برچسب‌گذاری‌هایی که با استفاده از واژه‌نامه‌ها صورت می‌گیرد، می‌توان گرایش احساسی هر بخش عقیده‌ای یافت شده در متن نظرات را مشخص نمود. برای این کار، با توجه به قوانین یا الگوهای نوشته شده برای استخراج بخش عقیده‌ای، گرایش احساسی هر بخش با توجه به گرایش احساسی توصیف‌کننده‌ای که در آن وجود دارد مشخص می‌گردد. به عبارت دیگر، در صورتی که گرایش صفت موجود در الگو، مثبت باشد، ویژگی کالا مثبت و در غیر این صورت منفی ارزیابی می‌گردد. جدول (۵) نمونه‌ای از این برچسب‌گذاری و تعیین گرایش احساسی را نشان می‌دهد.

نحوه برچسب‌گذاری	الگوهای استخراج شده
$PartType = Battery, Polarity = Positive$	عملکرد خوب باتری
$PartType = Software, Polarity = Positive$	نرم‌افزار بسیار بسیار قوی
$PartType = Cpu, Polarity = Negative$	پردازنده کند
$PartType = Camera, Polarity = Positive$	کیفیت دوربین فوق العاده نیست

جدول (۵): نمونه‌ای از برچسب‌گذاری و تعیین گرایش احساسی الگوهای یافت شده

## ۲-۷- پس پردازش

در این مرحله به بررسی الگوهای پرداخته می‌شود که قطبیت یا گرایش احساسی آن‌ها با توجه به چیزی که در متن نظر بیان شده، اشتباه تشخیص داده شده است. بدین منظور، از واژه‌نامه‌ی نفی‌کننده‌ها استفاده می‌شود. در

صورتی که یک نفی‌کننده در الگوی یافت شده وجود داشته باشد، به سادگی گرایش الگو از منفی به مثبت و یا بالعکس تغییر می‌کند. برای نمونه، با دقت در جدول (۴) که در بخش قبلی عنوان شد، جمله‌ی «کیفیت دوربین فوق العاده نیست» قبل از پس پردازش به صورت زیر برچسب‌گذاری می‌شود:

$PartType = Camera, Polarity = Positive;$

همان‌طور که مشخص است، این الگو به صورت مثبت ارزیابی شده است. در حالی که مشتری عدم رضایت خود را از کیفیت دوربین فیلمبرداری تلفن همراه بیان نموده است. با اعمال پس پردازش، این برچسب‌گذاری به صورت زیر تغییر می‌کند:

$PartType = Camera, Polarity = Negative;$

در صورتی که نفی‌کننده در الگوی یافت شده وجود نداشته باشد، در این صورت کل جمله‌ی حاوی الگو، در نظر گرفته می‌شود. در صورتی که با احتساب کل جمله، نفی‌کننده در الگو قرار گیرد، گرایش آن متناسب با گرایش فعلی تغییر خواهد نمود.

## ۲-۸- جمع بندی و کمی سازی گرایش احساسی نظرات

در این گام، نظرات مشتریان به صورت گرایش احساسی مثبت، منفی یا خنثی در هر متن برای هر ویژگی کالا خلاصه‌سازی می‌شود. تفاوت خلاصه‌سازی روش معرفی شده با روش‌های پیشین در این است که گرایش احساسی تمام الگوهای مربوط به یک ویژگی تجمیع شده و نظر کلی مشتریان راجع به آن ویژگی به دست می‌آید. این گام شامل سه مرحله است:

**مرحله اول:** ابتدا تجمیع نظرات هر ویژگی، برای هر مشتری و سپس برای کل مشتریان صورت می‌گیرد. تعیین گرایش برای هر ویژگی  $i$  در هر متن  $k$ ، یا به عبارت دیگر برای هر مشتری، از رابطه‌ی (۱) به دست می‌آید.

$$Polarity(Part_{ik}) = \begin{cases} 1 & \sum_j Pattern_j(Part_{ik}) > 0 \\ 0 & \sum_j Pattern_j(Part_{ik}) = 0 \\ -1 & \sum_j Pattern_j(Part_{ik}) < 0 \end{cases} \quad (1)$$

که در این رابطه:

$Part_i$ : ویژگی نام کالا.

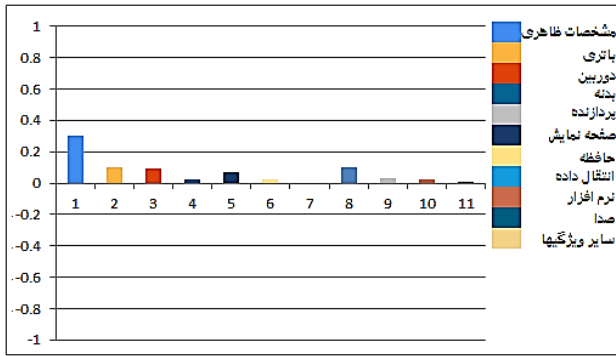
$Pattern_j(Part_{ik})$ : الگوی  $j$ ام مربوط به ویژگی  $i$  در متن نظر  $k$ ام

در صورتی که مجموع امتیازات مثبت شود، مقدار ۱ به معنی رضایت فرد، اگر منفی شود، مقدار -۱ به معنی عدم رضایت در کل و در صورتی که مجموع امتیازات برابر با صفر شود، مقدار ۰ که نشان دهنده‌ی خنثی بودن نظر فرد است، به ویژگی مربوطه تعلق می‌گیرد.

**مرحله دوم:** پس از اینکه گرایش احساسی یا به عبارت دیگر امتیاز هر ویژگی کالا برای هر متن نظر به دست آمد، میانگین امتیازات هر ویژگی در کل متون (تمام مشتریان)، با استفاده از رابطه‌ی (۲) محاسبه می‌گردد.

$$AveragedPolarity(Part_i) = \frac{\sum_k Polarity(Part_{ik})}{n} \quad (2)$$

که در آن،  $n$  تعداد متون نظرات برای یک کالا، یا تعداد مشتریان است. اگر میانگین امتیازات به دست آمده برای یک ویژگی صفر شود، بدین معنی است که یا مجموع نظرات مثبت و منفی در مورد آن یکسان بوده و یا اینکه



شکل (۴): نمودار نتایج ارزیابی ویژگی‌های تلفن همراه - LG G2 32GB

در جدول (۶) مقادیر هر یک از ویژگی‌ها، بعد از اعمال عقیده‌کاوی، برای ۹ مدل تلفن همراه ذکر شده، آمده است.

ویژگی	LG G2 - 32GB	Samsung Galaxy Note 3 N9000	Apple iPhone 5s
Apperance	0.304	0.178	0.169
Battery	0.100	0.029	-0.007
Camera	0.092	0.048	0.021
Case	0.020	0.013	0.014
Display	0.068	0.074	0.014
Memory	0.024	0.007	0.005
Network	-0.008	0.004	0.000
Other	0.104	0.050	0.081
Processor (CPU)	0.028	0.038	0.029
Software	0.002	0.014	0.041
Voice	0.004	0.001	0.000
جمع کل امتیاز	0.756	0.456	0.367

جدول (۶): نتایج کمی سازی نظرات فارسی برای نه مدل تلفن همراه

یافته‌ها نشان می‌دهد از بین گوشی‌های مورد بررسی، گوشی LG G2- 32GB با مجموع امتیاز ۰.۷۵۶ بالاترین و Apple iPhone 5s با مجموع امتیاز ۰.۳۶۷ کمترین ارزش را از نظر کاربران این سایت داشتند. این ارزش، حاصل مجموع امتیاز ۱۱ ویژگی مورد بررسی می‌باشد و این در حالی است که هر کدام از گوشی‌ها ممکن است در ویژگی خاصی بالاترین و یا کمترین امتیاز را به دست آورده باشند.

نکته‌ای که در مورد نتایج این مرحله باید در نظر داشت، این است که این مرحله، ویژگی‌ها یک کالا را از دید مشتریان ارزیابی می‌کند. این نتایج ممکن است با آنچه که مدیران صنایع در مورد تولیدات خود تصور می‌کنند، تفاوت داشته باشد. ممکن است یک ویژگی از نظر مدیران در وضعیت مطلوبی قرار داشته باشد، در حالی که مشتریان آن را ضعیف ارزیابی کرده‌اند. بنابراین این

افراد در مورد آن ویژگی نظری بیان نکرده‌اند. هر چه عدد به دست آمده به +۱ نزدیکتر باشد، نشان می‌دهد که افراد بیشتری نسبت به ویژگی خاص گرایش مثبت داشته‌اند و در واقع آن را مثبت ارزیابی نموده‌اند. هر چه این عدد به -۱ نزدیکتر باشد، عکس مطلب بیان شده را نشان می‌دهد.

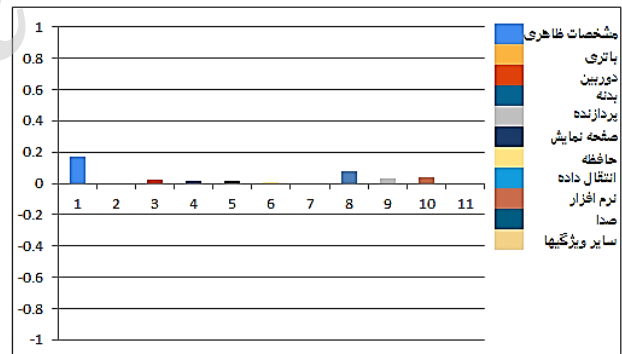
**مرحله سوم:** با یافتن میانگین امتیازات برای هر ویژگی در این مرحله، این امتیازات به صورت یک نمودار نشان داده می‌شوند و در این نمودار هر ویژگی با یک رنگ مجزا مشخص می‌گردد. علت انجام این مرحله این است که کاربران نمایش گرافیکی را بهتر از آمار عددی درک می‌کنند و بیان به صورت بصری تأثیر بیشتری بر تصمیم‌گیری‌های افراد خواهد گذاشت.

### ۳- یافته‌ها

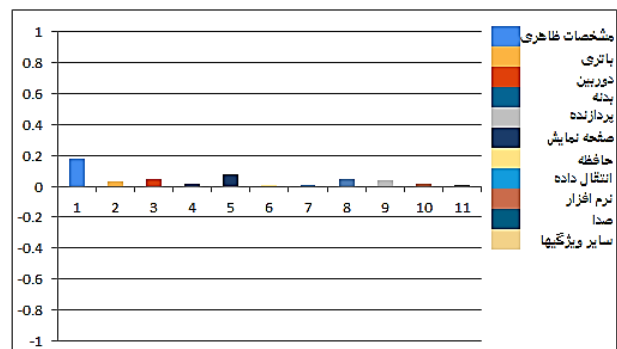
با توجه به آنچه که در بخش‌های قبلی بیان شد، یازده ویژگی برای کالای تلفن همراه بر اساس مشخصات فنی آن که در سایت دیجی کالا بیان شده و بر مبنای نظرات مشتریان استخراج شد. با ایجاد فرهنگ‌واژگان مناسب هم برای ویژگی‌ها و هم برای صفات با گرایش احساسی مثبت و منفی، الگوهای به دست آوردن بخش‌های عقیده‌ای در متن نظرات مشتریان مشخص و از آنها برای برچسب‌گذاری چنین بخش‌هایی در سایر متون نظرات به کار گرفته شد.

شکل (۲) تا شکل (۴) نمایش گرافیکی ارزیابی ویژگی‌های تلفن همراه برای سه مدل تلفن همراه است.

نتایج به دست آمده پس از استخراج الگوهای تعیین شده در نظرات مشتریان و انجام عملیات برچسب‌گذاری و تعیین گرایش احساسی هر یک و نیز انجام پس‌پردازش، برای هر مدل تلفن همراه به دست آمد.



شکل (۲): نمودار نتایج ارزیابی ویژگی‌های تلفن همراه Apple iPhone 5s



شکل (۳): نمودار نتایج ارزیابی ویژگی‌های تلفن همراه Samsung Galaxy Note 3 N9000



نتایج به مدیران کمک می‌کند تا به دنبال یافتن منشأ این اختلاف پذیرش در میان مشتریان باشند. علاوه بر این، سایر مشتریانی که تمایل به خرید یک کالای خاص را دارند، راهنمایی می‌نماید. نکته دیگر اینکه امتیاز پایین در بعضی ویژگی‌ها می‌تواند ناشی از این باشد که این ویژگی از نظر مشتریان اهمیت کمتری داشته و در نتیجه کمتر در مورد آن نظر داده‌اند.

## ۴- اعتبار سنجی

رایج‌ترین روش استفاده شده برای بررسی اعتبار نتایج به دست آمده از یک تکنیک عقیده‌کاوی بدون ناظر، تعیین استانداردهای طلایی<sup>۷</sup> و یا به عبارت دیگر برچسب‌گذاری دستی عبارات عقیده‌ای مورد نظر و تعیین گرایش احساسی هر یک و سپس مقایسه‌ی آن با نتایج به دست آمده از سیستم خودکار است [۱۰، ۱۱، ۱۳، ۱۷، ۲۰، ۲۲]. برای اعتبار سنجی مجموعه‌ی محدودی از داده‌ها در نظر گرفته شده است. این مجموعه حدود یک سوم از کل مجموعه‌ی متون برای سه مدل تلفن همراه انتخابی است. از آنجا که در این مرحله نکته‌ی مهم استخراج درست الگوها و نیز تعیین صحیح بودن گرایش احساسی هر یک است، با تعیین استاندارد طلایی، از سه معیار دقت<sup>۸</sup>، بازخوانی<sup>۹</sup> و معیار F<sup>۱۰</sup> برای بررسی میزان صحت و تطبیق نتایج با استاندارد استفاده شده است.

**دقت:** تعداد الگوهایی که به صورت درست شناسایی شده‌اند، به عنوان درصدی از تعداد کل الگوهای شناسایی شده بیان می‌کند. به عبارت دیگر این معیار بیان می‌کند صرف نظر از اینکه برخی الگوهای درست شناسایی نشده‌اند، چه تعدادی واقعاً توسط سیستم تشخیص داده شده‌اند.

**بازخوانی:** این معیار نسبت الگوهایی که به صورت درست شناسایی شده‌اند را به کل الگوهای درست نشان می‌دهد. در واقع این معیار بیان می‌کند که از بین الگوهایی که باید شناسایی شوند، چه تعداد شناسایی شده‌اند.

**معیار F:** در ارتباط با بازخوانی و دقت قرار دارد. به عبارت دیگر میانگین

وزنی دو معیار بیان شده را نشان می‌دهد و از رابطه‌ی (۳) به دست می‌آید.

$$F\_measure = \frac{(\beta^2 + 1)P * R}{(\beta^2 P) + R} \quad (3)$$

در این رابطه  $\beta$ ، وزن دقت (P) در مقابل بازخوانی (R) است. به طور معمول وزن دقت و بازخوانی برابر با یکدیگر در نظر گرفته می‌شود. در اینجا  $\beta=1$  در نظر گرفته شده است.

با محاسبه‌ی این معیارها طبق خروجی سیستم، جدول (۸) به دست می‌آید.

معیار F	بازخوانی	دقت	کل الگوها	الگوهای جعلی	الگوهای مفقودی	الگوهای نیمه کامل	الگوهای صحیح	مدل تلفن همراه	
۰.۹۱	۰.۹۲	۰.۹۱	۹۶	۷	۶	۳	۸۰	Apple iPhone 5s	
۰.۸۹	۰.۹۱	۰.۸۸	۱۲۰	۱۰	۷	۶	۹۷	Samsung Galaxy Note3 N9000	
۰.۸۷	۰.۸۹	۰.۸۶	۱۷۵	۱۶	۱۱	۱۳	۱۳۵	Sony Xperia Z1	
۰.۸۹	۰.۹۱	۰.۸۸	میانگین						

**جدول (۸): نتایج کمی سازی نظرات فارسی برای نه مدل تلفن همراه**  
نتایج نشان می‌دهد که روش پیشنهادی برای ارزیابی ویژگی‌های یک کالا می‌تواند نتایج قابل قبولی را ارائه نماید. همان‌طور که مشخص است؛ تعداد الگوهای جعلی تشخیص داده شده، نسبت به الگوهای دیگر بالاست.

البته این نسبت با تعداد کل الگوهای یافت شده در ارتباط است. یعنی هرچه تعداد الگوهای یافت شده بالاتر باشد، تعداد کل الگوهای جعلی یا مفقودی نیز افزایش می‌یابد. اما این مسئله لزوماً بدین معنی نیست که با افزایش تعداد متون نظرات برای بررسی، خطا نیز افزایش می‌یابد. علاوه بر این، لازم به ذکر است که این نتایج تا حد زیادی وابسته به نحوه‌ی بیان نظرات توسط افراد است که در رویکرد پیشنهادی تلاش شده است تا تمام الگوهای بیان نظرات استخراج و به کار گرفته شود.

## ۵- نتیجه گیری

هدف این پژوهش، معرفی یک روش مبتنی بر واژه نامه برای استخراج و کمی سازی گرایش احساسی مشتریان در مورد ویژگی‌های یک کالا با بررسی و تحلیل نظرات آنها به زبان فارسی بود. بنابراین، سیستمی معرفی شد که با کاوش در نظرات بیان شده توسط مشتریان، چگونگی ارزیابی آن‌ها را از ویژگی‌های یک کالا به دست می‌آورد. روش عقیده‌کاوی معرفی شده به صورت مبتنی بر واژه‌نامه و بدون ناظر عمل نموده و سعی در استخراج واژه‌هایی دارد که به طور ضمنی یا صریح به ویژگی‌های کالا دلالت می‌کنند. نتایج تجربی استفاده از این سیستم در مورد مدل‌های مختلف تلفن همراه نشان می‌دهد که این سیستم قابلیت به کارگیری در حوزه‌های کاربردی مختلفی را دارد.

اگرچه در این پژوهش تلاش بر این بود تا تمام جنبه‌های کمی سازی نظرات فارسی مشتریان، الگوریتم‌ها و روش‌های پیاده‌سازی در نظر گرفته شود، اما به دلیل برخی از محدودیت‌ها، همه‌ی این جنبه‌ها پوشش داده نشده است. از این‌رو، در ادامه پیشنهاداتی برای کارهای آتی و بهبود سیستم ارائه شده‌ی فعلی بیان شده است:

- در این تحقیق تنها بر صفات از نظر مثبت و منفی تمرکز شده بود و به قوت احساسی آن‌ها توجهی نشده بود. در عمل برخی از صفات در یک خانواده از قدرت احساسی بیشتری نسبت به سایر اعضا برخوردار هستند (مثلاً صفت «خوب» و مشتقات قوی‌تر آن مثل «فوق‌العاده» و «عالی»). می‌توان با در نظر گرفتن قوت احساسی هر یک از صفات، میزان ارزیابی مشتریان از ویژگی‌های یک کالا را به صورت دقیق‌تر و با دید بهتری به دست آورد.
- مرتبط نمودن ویژگی‌های کالا با عوامل تاثیر گذار در شهرت یک برند در بازار و ارائه‌ی راهبردهایی برای بهبود شهرت.
- بخش‌بندی مشتریان به گروه‌های مختلف مثلاً از نظر دموگرافیک، میزان استفاده و غیره، سپس تحلیل نظرات آن‌ها و تعیین ارزش ویژگی‌های یک کالا از دید گروه‌های مختلف.
- بررسی کالاهای در یک محدوده قیمت از تولیدکنندگان مختلف و یا بررسی کالاهای یک تولیدکننده خاص
- توجه به قدرت نفوذ مشتریان در سایت‌های دریافت‌کننده‌ی انتقاد و اثر بخشیدن در نتیجه‌ی ارزیابی بر اساس اهمیت مشتری در بین مشتریان دیگر. در برخی از سایت‌ها مثل دیجی کالا، تعداد افرادی که به نظر یک مشتری اهمیت می‌دهند و یا در واقع آن را Like می‌زنند، نشان دهنده‌ی میزان تاثیر نظر آن فرد بر سایرین می‌باشد.

- Empirical Methods in Natural Language Processing (HLT/EMNLP)*, Vancouver, 2005, pp. 339–346.
- [17] K. Xu, S. S. Liao, J. Li, and Y. Song, "Mining comparative opinions from customer reviews for Competitive Intelligence," *Decision Support Systems*, vol. 50, pp. 743–754, 2011.
- [18] H. Liu, J. He, T. Wang, W. Song, and X. Du, "Combining user preferences and user opinions for accurate recommendation," *Electronic Commerce Research and Applications*, vol. 12, pp. 14-23, 2013.
- [19] D. Wang, S. Zhu, and T. Li, "SumView: A Web-based engine for summarizing product reviews and customer opinions," *Expert Systems with Applications*, vol. 40, pp. 27-33, 2013.
- [20] S.-M. Kim and E. Hovy, "Determining the Sentiment of Opinions," in *COLING, Geneva, 2004*.
- [21] H. Cunningham, D. Maynard, K. Bontcheva, V. Tablan, N. Aswani, I. Roberts, G. Gorrell, A. Funk, and e. al., "Developing Language Processing Components with GATE Version 7 (a User Guide)," ed, 2012.
- [22] J. M. Wiebe, R. F. Bruce, and T. P. O'Hara, "Development and use of a gold-standard data set for subjectivity classifications," in *Proceedings of the 37th Annual Meeting of the Association for Computational Linguistics*, College Park, Maryland, USA, 1999, pp. 246–253.
- [23] N. Jahangiri, M. Kahani, R. Ahamdi, and M. Sazvar, "A Study on Part of Speech Tagging," *5th Symposium on Advances in Science and Technology (SASTech 2011)*, Mashhad, Iran.
- [1] B. Liu, M. Hu, and J. Cheng, "Opinion Observer: Analyzing and Comparing Opinions on the Web," in *WWW, Chiba, Japan, 2005*.
- [2] J. K. Smith, "Web Mining Tools to Actualize the Semantic Web And Future Natural Language Processing," *MCIS Capstone Course, University of Denver, 2009*.
- [3] P.-D. Turney, "Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews," in *The 40th Annual Meeting of the Association for Computational Linguistics (ACL)*, Philadelphia, 2002, pp. 417-424.
- [4] B. Pang, L. Lee, and S. Vaithyanathan, "Thumbs up? Sentiment Classification using Machine Learning Techniques," in *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, Philadelphia, PA, USA, 2002, pp. 79–86.
- [5] R. Xia, C. Zong, and S. Li, "Ensemble of feature sets and classification algorithms for sentiment classification," *Information Sciences*, vol. 181, pp. 1138–1152, 2011.
- [6] H. Kang, S. J. Yoo, and D. Han, "Senti-lexicon and improved Naïve Bayes algorithms for sentiment analysis of restaurant reviews," *Expert Systems with Applications*, vol. 39, pp. 6000–6010, 2012.
- [7] T. Mullen and N. Collier, "Sentiment analysis using support vector machines with diverse information sources," in *Proceedings of the Empirical Methods in Natural Language Processing (EMNLP) Conference*, Barcelona, Spain, 2004, pp. 412–418.
- [8] R. Moraes, J. F. Valiati, and W. P. G. Neto, "Document-level sentiment classification: An empirical comparison between SVM and ANN," *Expert Systems with Applications*, vol. 40, pp. 621–633, 2013.
- [9] C. Whitelaw, N. Garg, and S. Argamon, "Using appraisal groups for sentiment analysis," in *Proceedings of the 14th ACM Conference on Information and Knowledge Management*, New York, NY, USA, 2005, pp. 625-631.
- [10] M. R. Gholamian and S. Mazidi SharafAbadi, "Internal Factor Evaluation: An Application of Opinion Mining," in *3rd World Conference on Information Technology (WCIT-2012)*, Barcelona, Spain, 2013, pp. 915-921.
- [11] M. R. Gholamin and S. Mazidi SharafAbadi, "Resource Evaluation: A Novel Application of Opinion Mining for Competitive Intelligence," *International Journal of Business Information Systems*, 2014 vol.16, No.3, pp.339 - 358.
- [12] C. W.-k. Leung, S. C.-f. Chan, F.-l. Chung, and G. Ngai, "A probabilistic rating inference framework for mining user preferences from reviews," *World Wide Web*, vol. 14, pp. 187–215, 2011.
- [13] M. Hu and B. Liu, "Mining and Summarizing Customer Reviews," in *KDD'04, Seattle, Washington, USA, 2004*.
- [14] J. Yi, T. Nasukawa, R. Bunescu, and W. Niblack, "Sentiment Analyzer: Extracting Sentiments about a Given Topic using Natural Language Processing Techniques," in *Proceedings of the 3rd IEEE International Conference on Data Mining*, 2003, pp. 427–434.
- [15] K. Dave, S. Lawrence, and D. M. Pennock, "Mining the Peanut Gallery: Opinion Extraction and Semantic Classification of Product Reviews," in *WWW, Budapest, Hungary, 2003*.
- [16] A.-M. Popescu and O. Etzioni, "Extracting Product Features and Opinions from Reviews," in *Human Language Technology Conference and Conference on*

## زیر نویس ها

<sup>1</sup> Pros and cons

<sup>2</sup> Free format

<sup>3</sup> Sequential Minimal Optimization

<sup>4</sup> www.digikala.com

<sup>5</sup> Web Crawler

<sup>6</sup> Gazetteer

<sup>7</sup> Gold Standard

<sup>8</sup> Precision

<sup>9</sup> Recall

<sup>10</sup> F-Measure