Archive of SID

فصلنامهٔ مطالعات مدیریت کسب و کار هوشمند سال ششم، شماره ۲۱، پاییز ۱۳۹٦
صفحات ٦٥ تا ۹۲

ارائه مدلی جهت دسته‌بندی احساسات خریداران کتاب با استفاده از رویکرد ترکیبی

فاطمه عباسی*
بابک سهرابی**
امیر مانیان***
آمنه خدیور****

چکیده

در سال‌های اخیر رشد شبکه‌های اجتماعی و به‌تبع آن افزایش فزاینده محتوای این شبکه‌ها باعث شده است تا افراد برای خرید و استفاده از محصولات، خدمات و یا حتی انتخاب‌های سیاسی خود از نظرات سایر افراد برای تصمیم‌گیری استفاده نمایند. با توجه به آنکه نظرات کاربران به‌صورت متنی است و خواندن و جمع‌بندی آن‌ها زمان‌بر و مشکل است، خودکارسازی استخراج عقاید و احساسات نظرات کاربران یکی از راهکارهای پیشنهادی برای سایت‌های فروش آنلاین جهت ارائه خدمات کاراتر به

* دانشجوی دکتری، مدیریت فناوری اطلاعات، دانشکده مدیریت، دانشگاه تهران، تهران. (نویسنده مسئول)
sati9634@yahoo.com
** استاد، گروه مدیریت فناوری اطلاعات، دانشکده مدیریت، دانشگاه تهران، تهران.
*** استاد، گروه مدیریت فناوری اطلاعات، دانشکده مدیریت، دانشگاه تهران، تهران.
**** استادیار، گروه مدیریت، دانشکده علوم اجتماعی و اقتصاد، دانشگاه الزهرا (س)، تهران.
تاریخ دریافت: ۱۳۹٦/۰۹/۰٤ تاریخ پذیرش: ۱۳۹٦/۱۰/۱٦

www.SID.ir

Archive of SID

مشتریان جهت تصمیم‌گیری آگاهانه‌تر است. تحلیل احساسات یا عقیده کاوی فرآیندی است که نظرات، احساسات و نگرش افراد در ارتباط با موضوعی خاص استخراج می‌شود و به‌عنوان شاخه‌ای از متن‌کاوی شناخته می‌شود. نتایج حاصل از تحلیل احساسات می‌تواند در سیستم‌های پیشنهاددهنده جهت ارائه پیشنهادهای کاراتر برای خرید مورد استفاده قرار گیرد. اطلاعات حاصل از عقیده کاوی می‌تواند در زمینه‌های مختلف ازجمله کتابخانه‌ها در انتخاب بهتر و خرید مبتنی بر نظرات واقعی کاربران کاربرد داشته باشد. در این پژوهش سیستمی جهت دسته‌بندی خودکار احساسات بیان‌شده در نظرات مربوط به خریداران کتاب سایت آمازون ارائه شده است. سیستم با استفاده از مدل‌های ترکیبی برای تحلیل احساسات نظرات کاربران سایت آمازون طراحی شده است. جهت کلیه تحلیل‌ها از پکیج‌های متن‌کاوی پایتون استفاده است. نتایج نشان می‌دهند سیستم پیشنهادی می‌تواند به‌صورت خودکار نظرات مثبت و منفی را با دقت بالای ۸۰٪ دسته‌بندی نماید.

کلیدواژگان: متن‌کاوی، تحلیل احساسات، عقیده کاوی، مدل ترکیبی.

www.SID.ir

Archive of SID

مقدمه

عقاید و نظرات دیگران در زمان تصمیم‌گیری و انتخاب یک گزینه از میان چند گزینه می‌تواند بسیار حیاتی باشد (کامبریا و حوسین[1]، ۲۰۱۵). به دلیل تأثیر عقاید و نظرات بر رفتار انسان می‌توان ادعا نمود، دیدگاه‌ها مرکز فعالیت‌های بشر هستند (لیو[2]، ۲۰۱۲). شبکه‌های اجتماعی مجازی نسل جدیدی از فضای روابط اجتماعی هستند که توانسته‌اند به‌خوبی در زندگی مردم جای باز کنند. با توجه به دامنه نفوذ روزافزون و تأثیرگذاری این شبکه‌ها، مسئله شناخت الگوهای رفتاری کاربران آن‌ها به یکی از موضوعات جذاب تحقیقی بدل شده است (جعفری و همکاران، ۱۳۹۵). انجمن‌ها، وبلاگ‌ها، شبکه‌های اجتماعی و سرویس‌های اشتراک محتوا به مردم در اشتراک‌گذاری اطلاعات مفید کمک می‌کند. دریافت عقاید عمومی درباره‌ی رویدادهای اجتماعی، فعالیت‌های بازاریابی و اولویت‌های محصول، توجه جوامع علمی و جهان تجارت را به خود جلب کرده است (کامبریا و حوسین، ۲۰۱۵). امروزه، اگر کسی بخواهد محصولی را خریداری نماید دیگر محدود به نظرات خانواده و دوستان نیست زیرا نظرات و بحث‌های کاربران در مورد محصولات مختلف در وب در دسترس است. برای یک سازمان نیز دیگر نیازی به انجام نظرسنجی و گروه کانونی نیست زیرا چنین اطلاعاتی به‌وفور در اینترنت در دسترس است. بااین‌حال پیدا کردن و ارزیابی نظرات به دلیل گستردگی و تنوع نظرات کار ساده‌ای نیست. هر سایت معمولاً دارای حجم زیادی نظرات متنی است که رمزگشایی از آن‌ها کار ساده‌ای نیست. به‌طورکلی استخراج و خلاصه‌سازی چنین اطلاعاتی کار ساده‌ای نیست، بنابراین سیستم‌های تحلیل احساسات خودکار مورد نیاز است (لیو، ۲۰۱۲). در حال حاضر پذیرش رشد محتوای تولیدشده توسط کاربران در وب‌سایت‌ها و شبکه‌های اجتماعی مانند توییتر[3]، آمازون[4] و تریپ ادوایزر[5] منجر به افزایش قدرت شبکه‌های اجتماعی برای بیان نظرات در مورد خدمات، محصولات و رویدادها شده است. امروزه افراد و سازمان‌ها به‌طور فزاینده‌ای از محتوای

1. Cambria & Hussain
2. Liu B
3. Twitter
4. Amazon
5. Trip Advisor

www.SID.ir

Archive of SID

این رسانه‌ها برای تصمیم‌گیری استفاده می‌نمایند (لیو[1] ۲۰۱۵). تحلیل احساسات در حوزه‌های مختلف ازجمله انتخاب و خرید کتاب برای کتابخانه‌های دانشگاهی بسیار مؤثر است. شواهد نشان می‌دهند میزان مراجعه و استفاده از کتاب‌های کتابخانه‌های دانشگاه‌ها نسبت به هزینه صورت گرفته برای خرید کتاب علی‌الخصوص 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، عدم‌کفایت و بروز نبودن منابع از علل مراجعه نکردن دانشجویان به کتابخانه‌های دانشگاهی است (تفرشی و دروگر کلخوران، ۱۳۸۸). فرآیند خرید کتاب برای اغلب کتابخانه‌های دانشگاهی بدین‌صورت است که بر اساس نیازهای گروه‌های علمی خرید انجام می‌شود. با توجه به پایین بودن میزان مراجعه دانشجویان برای استفاده از منابع کتابخانه‌ای نسبت به هزینه صورت گرفته، اصلاح فرآیند و رویه خرید کتاب برای کتابخانه‌های دانشگاه‌های کشور ضروری است. با توجه به آنکه تحلیل احساسات می‌تواند عقاید و نظرات واقعی کاربران را استخراج نماید می‌تواند به اصلاح روند کمک شایانی نماید. با توجه اهمیت توسعه و استفاده از رویکردهای متن‌کاوی و تحلیل احساسات برای جامعه علمی و نیز نقش مؤثر این موضوع از جهت کاربرد برای کسب‌وکارهای مختلف و خلأ شناسایی‌شده در این زمینه، مسئله اصلی این پژوهش ارائه مدلی برای تحلیل احساسات خریداران جهت اصلاح فرآیند خرید کتاب کتابخانه‌ها است.

پیشینه نظری پژوهش

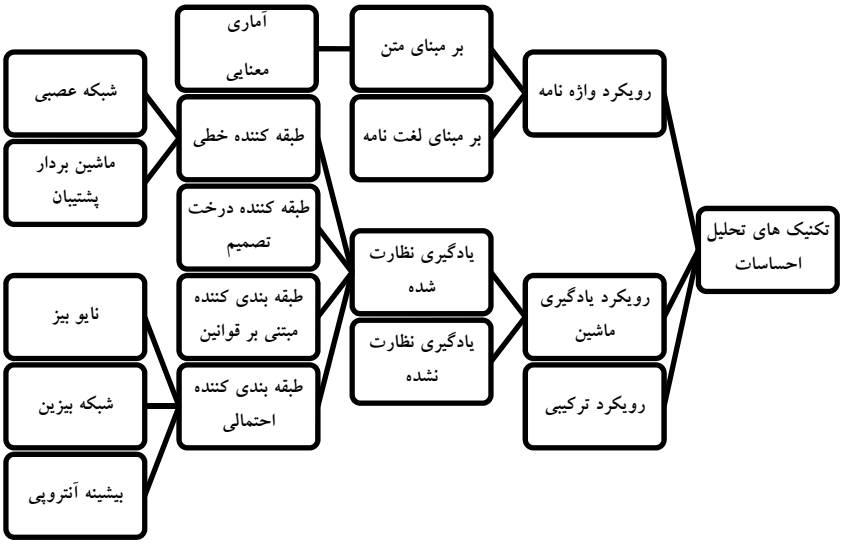
امروزه حجم کثیری از داده‌های متنی برای پردازش موجود است و هرساله بر میزان این داده‌ها افزوده می‌شود (نیکنام و نیک‌نفس، ۱۳۹۵). داده‌های متنی یکی از مهم‌ترین اشکال داده‌های بدون ساختار[1] می‌باشند. این نوع از داده‌ها، مخزن بزرگی از اطلاعات هستند که می‌توانیم محتوای مناسب را از آن استخراج نماییم. یکی از راه‌های استخراج اطلاعات از داده‌های متنی، تحلیل احساسات است (ژان و فنگ[2]، ۲۰۱۵). تحلیل

1. Unstructured Data
2. Fang & Zhan

www.SID.ir

Archive of SID

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 تجزیه‌وتحلیل می‌نماید (لیو، ۲۰۱۲). تکنیک‌های مختلفی جهت تحلیل احساسات ارائه گردیده است که در شکل ۱ خلاصه‌ای از این روش‌ها نشان داده شده است.



شکل ۱. تکنیک‌های تحلیل احساسات (پندی و همکاران[1]، ۲۰۱۷)

رویکرد یادگیری ماشین[2]: رویکردهای یادگیری ماشین را می‌توان به‌طورکلی در دو بخش تقسیم‌بندی نمود: ۱) رویکردهای یادگیری ماشین نظارت شده[3] و ۲) رویکردهای یادگیری ماشین بدون نظارت[4]. رویکرد یادگیری ماشین نظارت‌شده نوعی از یادگیری است که در آن ورودی و خروجی مشخص است و ناظری وجود دارد که اطلاعاتی را در اختیار یادگیرنده قرار می‌دهد و به‌این‌ترتیب سیستم سعی می‌کند تا تابعی را از ورودی به خروجی فراگیرد (ونگ[5]، ۲۰۱۷). رویکرد یادگیری ماشین بدون نظارت در

1. Pandey et al.
2. Machine Learning Approach
3. Supervised Machine learning Approaches
4. Unsupervised Machine learning Approaches
5. Wang

www.SID.ir

Archive of SID

یادگیری بدون نظارت برخلاف یادگیری نظارت‌شده، داده‌های مشخصی از قبل وجود ندارد و هدف، ارتباط ورودی و خروجی نیست، بلکه تنها دسته‌بندی آن‌ها مهم است و این یادگیرنده است که بایستی در داده‌ها به دنبال ساختاری خاص بگردد (لیو، ۲۰۱۵).

رویکرد واژه‌نامه[1]: رویکرد مبتنی بر واژه‌نامه، رویکردی اساسی برای تجزیه‌وتحلیل متن است. این رویکرد از یک لغت‌نامه یا متن استفاده می‌نماید که هر کلمه در متن با عقیده خاصی در ارتباط است. با استفاده از کلمات درون متن، میانگین وزنی از نمرات احساسات متن ارائه می‌شود (ونگ، ۲۰۱۷).

رویکرد ترکیبی[2]: در برخی از پژوهش‌ها از روش ترکیبی استفاده شده است. بدین ترتیب، داده‌هایی که در ابتدا با تکیه بر روش مبتنی بر لغت‌نامه بار احساسی‌شان تعیین می‌شوند، به‌عنوان داده‌های ورودی برای آموزش مدل در الگوریتم‌های یادگیری ماشین در روش تحت نظارت مورد استفاده قرار می‌گیرند (فیلهو و پاردو[3]، ۲۰۱۳).

پیشینه تجربی پژوهش

در سال‌های اخیر تحقیقات زیادی در زمینه تجزیه‌وتحلیل احساسات و نظرات شخصی صورت گرفته است. برهانی زردی و همکاران در سال ۱۳۹۲ یک سیستم عقیده کاوی که از تکنیک‌های پردازش زبان طبیعی و شبکه واژگان احساسی برای عقیده کاوی در مجموعه‌ای از نقدهای کالا استفاده می‌نماید را بررسی نموده است (برهانی زرندی و همکاران، ۱۳۹۲). پیکری و همکاران در سال ۱۳۹٤ با کاربرد تکنیک متن‌کاوی و تحلیل محتوا پدیده فوت مرتضی پاشایی را در شبکه اجتماعی توییتر مورد مطالعه و بررسی قرار داده و تمام توییت‌های انتشاریافته را در پنج مقوله بررسی نموده است (پیکری و همکاران، ۱۳۹٤). توکلی و رافع در سال ۱۳۹۵ به ارائه روشی برای تحلیل نظرات ذیل خبر در سایت‌های خبری با توجه به متن خبر پرداخته‌اند. در این راستا سعی شده با استفاده از ویژگی‌های دستوری متون مانند اسم و فعل و همچنین تحلیل بار احساسی جمالت، رابطه نظر با خبر و دیدگاه نویسنده نظر را با توجه به موضوع

1. Lexicon-based Approach
2. Hybrid Approach
3. Filho & Pardo

www.SID.ir

Archive of SID

متن خبر کشف گردد (توکلی گارماسه و رافع، ۱۳۹۵). کولتا و همکاران[1] در سال ۲۰۱٤ الگوریتم‌های دسته‌بندی و خوشه‌بندی را برای تحلیل احساسات توییت‌ها ترکیب نموده‌اند. در این پژوهش جهت دسته‌بندی الگوریتم‌های نایو بیز، حداکثر آنتروپی، ماشین بردار پشتیبان استفاده شده است (کولتا و همکاران[2]، ۲۰۱٤). ایسا و همکاران[3] در سال ۲۰۱٤ در پژوهشی چارچوبی را برای جمع‌آوری و تحلیل نقطه نظرات تجربیات کاربران از استفاده از دارو و محصولات آرایشی را با استفاده از یادگیری ماشین، متن‌کاوی و تحلیل احساسات را ارائه نموده است. در این مقاله چگونگی ایجاد واژه‌نامه و آموزش داده‌ها برای پیش‌بینی احساسات نسبت به محصولات دارویی و آرایشی شرح داده شده است (ایسا و همکاران، ۲۰۱٤). زانگ و همکاران[4] در سال ۲۰۱٤ مدل عامل صریح[5] برای ایجاد سیستم پیشنهاددهنده‌ای با دقت بالا را پیشنهاد نموده‌اند. در این پژوهش ابتدا ویژگی‌های محصول استخراج گردید و بر روی نظرات در سطح عبارت تحلیل احساسات صورت گرفته است. سپس بر اساس علایق کاربران پیشنهادهایی ایجاد شده است، بعلاوه توضیحات شهودی در سطح ویژگی در مورد اینکه چرا یک آیتم پیشنهاد شده است یا نشده است از مدل ارائه گردیده است (زانگ، ۲۰۱٤). چن و همکاران[6] در سال ۲۰۱٤ در پژوهشی با عنوان احساسات و ترجیحات هدایت‌کننده پیشنهادهای شبکه‌های اجتماعی بر دو منبع دانشی متمرکز شده‌اند: نظرات غنی از احساسات کاربران و آمارهای ترجیحات خرید کاربران که نظرات کاربران را با ترجیحات خریدشان یکپارچه نموده است (چن و همکاران، ۲۰۱٤). فنگ و زان[7] در سال ۲۰۱۵ به بررسی مسئله طبقه‌بندی قطبیت احساسات که از مشکلات اصلی تحلیل احساسات است، پرداخته است. در این پژوهش یک فرآیند کلی برای طبقه‌بندی قطبیت احساسات ارائه شده است و این پژوهش بر روی داده‌های گردآوری‌شده از سایت آمازون انجام شده است (فنگ و زان،۲۰۱۵). بات و همکاران[8] در سال ۲۰۱۵ تحلیل احساسات را بر روی نظرات مربوط به محصولات و خدمات سایت آمازون انجام داده

1. Coletta et al.
2. Coletta et al.
3. Isah et al.
4. Zhang et al.
5. Explicit Factor Models
6. Chen et al.
7. Fang & Zhan
8. Bhatt et al.

www.SID.ir

Archive of SID

است. در این مقاله سیستمی را بر دسته‌بندی نظرات مشتریان بر مبنای احساسات نظرات پیشنهاد نموده است و نتایج خود را به‌صورت خلاصه و تصویری ارائه نموده است (بات و همکاران،۲۰۱۵). نگوین و همکاران[1] در سال ۲۰۱۵ مدلی برای پیش‌بینی قیمت سهام با استفاده تحلیل احساسات شبکه‌های اجتماعی را ارائه نموده‌اند. جهت تحلیل داده‌ها از تکنیک یادگیری ماشینی استفاده شده است که تحلیل بر روی عنوان موضوعات صورت گرفته است (نگوین و همکاران، ۲۰۱۵). حمدان و همکاران در سال ۲۰۱٦ تحلیل احساسات را بر روی کتاب‌های علمی که از نظر سبک و اندازه واژگان متفاوت هستند انجام داده‌اند. در این مقاله ترکیبی از ویژگی‌های مختلف برای استخراج عقاید و قطبیت احساسات پیشنهاد شده است (حمدان و همکاران[2]، ۲۰۱٦). چیاوتا و همکاران در سال ۲۰۱٦ در پژوهشی سیستمی را جهت دسته‌بندی خودکار جهت‌گیری احساسات در مورد کتاب‌های به زبان ایتالیایی ارائه نموده‌اند. در این مقاله از رویکرد مبتنی بر واژه‌نامه و از روش‌های پردازش زبان طبیعی برای در نظر گرفتن ارتباط زبانی کلمات جهت تحلیل متن استفاده شده است (چیاوتا و همکاران، ۲۰۱٦). آلدوگان و یاسلان[3] در سال ۲۰۱۷ به مطالعه مقایسه‌ای در مورد رویکردهای ترکیبی یادگیری فعال در تحلیل احساسات پرداخته‌اند. در این مقاله چارچوبی را که متشکل از رویکردهای ترکیبی است برای تحلیل احساسات معرفی می‌نماید (آلدوگان و یاسلان، ۲۰۱۷). بوداد و همکاران[4] در سال ۲۰۱۷ تحلیل احساسات را با سه روش یادگیری نظارت‌شده، یادگیری نظارت‌نشده و روش ترکیبی در زبان عربی انجام داده‌اند (بوداد و همکاران،۲۰۱۷).

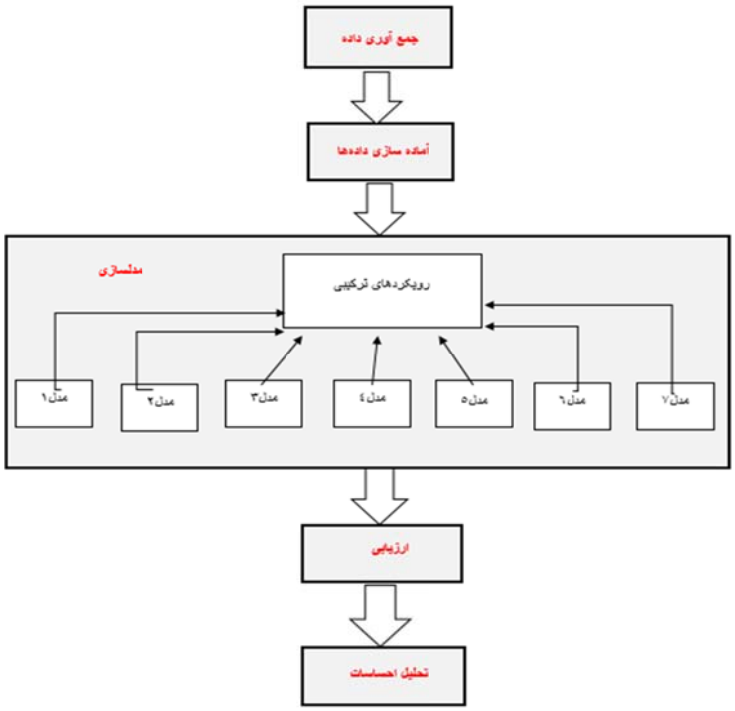
روش‌شناسی پژوهش

در شکل زیر مراحل و رویکردهای هر مرحله نشان داده شده است.

1. Nguyen et al.
2. Hamdan, Bellot, & Bechet
3. Aldogan & Yaslan
4. Boudad et al.

www.SID.ir

Archive of SID



شکل ۲. چهارچوب پژوهش

جمع‌آوری داده

جمع‌آوری داده شامل دو گام گردآوری داده از وب و تجمیع داده‌ها است. برای گردآوری داده از وب از یک نرم‌افزار خزنده[1] استفاده گردیده است. این نرم‌افزار اختصاصاً برای پیمایش صفحات سایت آمازون طراحی شده تا بتواند بر اساس یک موضوع جستجو شده تمامی کتاب‌های مرتبط را از بین تمامی اقلام سایت جدا کرده و سپس عنوان کتاب، عنوان کامنت، امتیاز کامنت، نام کامنت گذارنده و متن کامنت را در ارتباط با هر کتاب پیدا و جدا کند و در قالب فایل جی‌سن[2] ذخیره کند. جهت طراحی این خزنده از زبان برنامه‌نویسی جاوا استفاده شده است که با وارد نمودن کلمات کلیدی[3] کلیه اطلاعات مرتبط با موضوع جستجو شده استخراج و ذخیره می‌گردد. با

1. Crawler
2. JSON
3. Keywords

www.SID.ir

Archive of SID

وارد نمودن عنوان "هوشمندی کسب‌وکار" عنوان کتاب، عنوان کامنت، امتیاز کامنت، نام کامنت گذارنده و متن کامنت کتاب‌هایی که مرتبط با این کلمات هستند از سایت آمازون استخراج گردید و در ۱۹ فایل جی‌سن ذخیره گردید. با استفاده نرم‌افزار مذکور ۷۲۱۰ کامنت و سایر اطلاعات مرتبط با کتاب‌های با موضوع هوشمندی کسب‌وکار استخراج گردید. گام دوم در این مرحله تجمیع داده‌ها است که با توجه به آنکه داده-های در فایل‌های مختلف ذخیره شده‌اند لازم است پس از گردآوری تجمیع شوند تا جهت استفاده در مرحله بعد دیدی واحد در مورد داده‌ها ایجاد شود.

آماده‌سازی داده‌ها[1]

آماده‌سازی داده‌ها مهم‌ترین و زمان‌برترین بخش از فرآیند تحلیل احساسات است. در این پژوهش از تکنیک‌های زیر جهت آماده‌سازی داده‌ها از Case fold، نشانه‌گذاری کلمات[2]، جایگزینی کلمات بی‌معنی[3]، ایجاد n-grams، برچسب‌گذاری اجزای واژگانی کلام[4]، حذف ایست واژه‌ها[5] و ریشه‌یابی کلمات[6] استفاده شده است.

Case Fold: در این گام کلیه کلمات از نظر کوچک بودن یا بزرگ بودن حروف به یک شکل درمی‌آیند. این مرحله ازاین‌جهت انجام می‌شود که اگر کلمه‌ای چندین بار با صورت یکسان اما متفاوت در حروف بزرگ و کوچک تکرار شده باشد در مدل‌سازی یک‌بار در نظر گرفته شوند (تلوال و چیبلوشی[7]، ۲۰۰۹). در این پژوهش کلیه حروف بزرگ به حروف کوچک[8] تبدیل شده است. لازم به توضیح است این مرحله در مورد مستندات فارسی کاربرد ندارد و بیشتر در مورد متون به زبان انگلیسی کاربردی است.

نشانه‌گذاری کلمات: نشانه‌گذاری یا جداسازی متن، تقسیم‌بندی متن به واحدهای کوچک‌تری چون کلمات[9]، اصطلاحات[10]، نمادها[11] است که در اصطلاح نشانه[12] نامیده

1. Data Preprocessing
2. Tokenization
3. Noisy Data
4. Part of Speech Tagging (POS Tagging)
5. Stop-Word
6. Stemming
7. Thelwall & Chibelushi
8. Lower Case
9. words
10. Terms
11. Symbols
12. Tokens

www.SID.ir

Archive of SID

می‌شوند (کاتاریا و چادهاری، ۲۰۱۵). در این پژوهش کلیه نظرات به واحدهای کوچک‌تری تقسیم شدند و درنهایت کلمات به‌دست‌آمده از متن‌ها به‌صورت متغیرهای مجزا در نظر گرفته شده‌اند. در ضمن در این پژوهش کلماتی که کوچک‌تر از چهار حرف و بزرگ‌تر از بیست‌وپنج حرف هستند در نظر گرفته نشده‌اند، به دلیل آنکه کلمات زائد و بی‌تأثیر به‌عنوان متغیر در مدل در نظر گرفته نشوند.

جایگزینی کلمات بی‌معنی: کلمات بی‌معنی اشاره به اصطلاحات کوتاه و اختصارات[1] دارد. به‌عنوان‌مثال "TY" به‌عنوان مختصر برای کلمه "Thank You" به کار می‌رود (استاوریاتو و همکاران[2]، ۲۰۰۷). اغلب این کلمات دارای بار احساسی هستند و جهت افزایش دقت مدل‌سازی لازم است این کلمات و اختصارات کوتاه با کلمه اصلی جایگزین شوند. در این پژوهش جهت جایگزینی این کلمات بی‌معنی فهرستی از اختصارات تهیه شد تا در صورت وجود این کلمات در متن، کلمه اصلی جایگزین این کلمات شوند. در این مرحله کلیه کارکترهای بی‌معنی مانند شکلک[3]، ارقام و نشانه‌ها[4] حذف گردید.

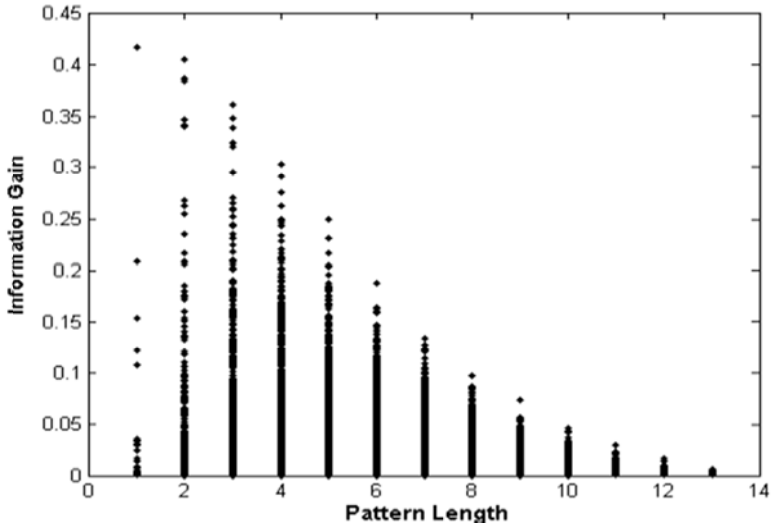
پیدا کردن n-grams: مدل n-grams به‌صورت گسترده‌ای در متن‌کاوی و پردازش زبان طبیعی[5] استفاده می‌شود. n-grams یک توالی به‌هم‌پیوسته از n عنصر یک متن یا سخنرانی است. در یک n-grams اندازه ۱ به unigram، اندازه ۲ به bigrams و اندازه ۳ به trigrams اشاره دارد. به‌عنوان‌مثال در جمله "This is a sentence" درصورتی‌که n=2 باشد (bigrams) در این صورت n-grams عبارت‌اند از: This is، is a، a sentence. یکی از کاربردهای n-grams ایجاد ویژگی‌هایی است که برای مدل‌های یادگیری ماشین نظارت‌شده استفاده می‌شود. n-grams در دسته‌بندی احساسات جهت انتخاب ویژگی‌ها مورد استفاده قرار می‌گیرد. این روش اشاره به تجزیه نمودن متن به بخش‌های مختلف اشاره دارد که n نشان‌دهنده تعداد کلمات هر بخش است (Wang, 2017).چنگ و همکاران در سال ۲۰۰۷ در یک آزمایش سامانمند ارتباط میان تعداد n و اطلاعات به‌دست‌آمده از متن را بررسی کرده‌اند که در شکل ۳ نتیجه تحقیقات آن‌ها نشان داده شده است که با افزایش n میزان اطلاعات به‌دست‌آمده کاهش می‌یابد و

1. abbreviations and acronyms
2. Stavrianou et al.
3. Emoji
4. digits and punctuations
5. Natural Language Processing (NLP)

www.SID.ir

Archive of SID

حداکثر میزان اطلاعات دریافتی در n=2 حاصل مشود (چنگ و همکار[1] ، ۲۰۰۷). با توجه نتایج پژوهش‌های پیشین در این bigram موردبررسی قرار گرفته است.



شکل ۳. ارتباط میان تعداد n و اطلاعات به‌دست‌آمده

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 ضمایر، حروف اضافه و حروف ربط (رامیا و راماسوبرامانیان[2]، ۲۰۱۳).

برچسب‌گذاری اجزای واژگانی کلام: برچسب‌گذاری اجزای واژگانی کلام، روش زبانی است که از سال ۱۹٦۰ استفاده می‌شود و اخیراً هم توجه محققان حوزه پردازش زبان طبیعی را برای استخراج ویژگی‌های محصول خود جلب کرده است. برچسب‌گذاری اجزای واژگانی کلام به هر کلمه‌ای بر اساس 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 (زوبیر و همکاران[3]، ۲۰۱٤). در این پژوهش با توجه به آنکه کلمات دارای نقش فعل، صفت و قید بار

1. Cheng et al.
2. Ramasubramanian & Ramya
3. Zubair et al.

www.SID.ir

Archive of SID

احساسی متن را مشخص می‌نماید تنها کلمات دارای نقش‌های مذکور جهت تحلیل در فاز مدل‌سازی انتخاب شده‌اند.

ریشه‌یابی کلمات: این روش جهت یافتن ریشه کلمات استفاده می‌شود. در این مرحله کلیه کلمات به فرمت ریشه اصلی خود درمی‌آیند. الگوریتم‌های ریشه‌یابی پسوند کلمه بر اساس قوانین گرامری حذف می‌کند. در این پژوهش از کتابخانه ریشه‌یاب snowball[1] که محبوب‌ترین و استانداردترین رویکرد است، استفاده شده است (روسکا و همکاران[2]، ۲۰۱۶). در جدول ۱ نمونه‌ای از مراحل پیش‌پردازش نشان داده شده است.

جدول ۱. نمونه‌ای از مراحل پیش‌پردازش

پیش‌پردازش	قبل از پیش‌پردازش	پس از پیش‌پردازش
جایگزینی کلمات بی‌معنی	I recommend this book to .anyone TMI	I recommend this book to Too Much Information...anyone
نشانه‌گذاری	I recommend this book to . anyone Too Much Information	I', 'recommend', 'this', 'book', 'to', ' 'anyone', '.', 'Too', 'Much', "Information
حذف ایست واژه‌ها	I', 'recommend', 'this', ' 'book', 'to', 'anyone', '.', "Too', 'Much', 'Information	I', 'recommend', 'book', 'anyone', ' '.', 'Too', 'Much', 'Information’
Case Fold	I', 'recommend', 'book', ' 'anyone', '.', 'Too', 'Much', 'Information’	‘i', 'recommend', 'book', 'anyone', "'.', 'too', 'much', 'information
ریشه‌یابی	‘i', 'recommend', 'book', 'anyone', '.', 'too', 'much', "information	i recommend book anyon . too much inform
برچسب‌گذاری اجزای واژگانی کلام	I recommend this book to . anyone Too Much Information	('I', 'PRP'), ('recommend', 'VBP'), ('this', 'DT'), ('book', 'NN'), ('to', 'TO'), ('anyone', 'NN'), ('.', '.'), ('Too', 'VB'), ('Much', 'JJ'), ('Information', 'NN')
Bi-gram	I recommend this book to .	('I', 'recommend'), ('recommend', 'this'), ('this', 'book'), ('book', 'to'),

1. Snowball stemmer library
2. Krouska et al.

www.SID.ir

Archive of SID

پیش‌پردازش	قبل از پیش‌پردازش	پس از پیش‌پردازش
	Too Much .anyone Information	('to', 'anyone'), ('anyone', '.'), ('.', 'Too'), ('Too', 'Much'), ('Much', 'Information')

یافته‌های پژوهش

مدل‌سازی

پس از تبدیل داده‌های متنی که بدون ساختار هستند به داده‌های ساختاریافته با استفاده از روش برچسب‌گذاری اجزای واژگانی کلام، کلمات و واژگان جهت استفاده در مدل‌سازی استخراج گردیدند. با توجه به هدف پژوهش که مدل‌سازی جهت تحلیل احساسات است در این مرحله کلمات دارای نقش فعل، قید و صفت جهت مدل‌سازی انتخاب شدند. پس‌ازاین مرحله واژگان و کلمات کلیدی استخراج‌شده برای مدل‌سازی مورد استفاده قرار گرفته‌اند. در این مرحله با استفاده از روش ترکیبی[۱]، مدل طبقه‌بندی کننده و پیش‌بینی کننده احساسات ایجاد و اجرا می‌شود. جهت مدل‌سازی از روش ترکیبی مبتنی بر چند الگوریتم طبقه‌بندی کننده استفاده شده است که این الگوریتم‌ها جهت دسته‌بندی نمودن هر نظر[۲] با یکدیگر همکار می‌کنند. در این پژوهش شش دسته‌بندی کننده بیز ساده[۳]، طبقه‌بندی کننده بردار پشتیبان[۴]، بیزین ساده چندجمله‌ای[۵]، پرسپترون چندلایه[۶]، بیز ساده برنولی[۷]، رگرسیون لجستیک[۸] و طبق بندی کننده بردار پشتیبان Nu[۹] از کتابخانه NLTK پایتون جهت ایجاد مدل استفاده شده‌اند. درنهایت نتیجه نهایی هر نظر با استفاده از دو روش رأی اکثریت[۱۰] و وزن دهی[۱۱] در رویکرد رای‌گیری[۱۲] مشخص می‌گردد. بر اساس مقایسه نتایج ارزیابی این دو مدل، مدلی که دقت بالاتری داشته باشد در گام نهایی جهت تحلیل احساسات نظرات گردآوری‌شده از

1. Ensemble Methods
2. Comment
3. Naïve Bayes
4. Linear Support Vector Classification (Linear SVC)
5. multinomial naive Bayes (MNB)
6. Multilayer perceptron (MLP)
7. Bernoulli Naive Bayes
8. Logistic Regression
9. Nu - Vector Classification (Nu - SVC)
10. Majority Voting
11. Weighted
12. Voting

www.SID.ir

Archive of SID

سایت آمازون انتخاب می‌شود.

رویکرد ترکیبی

هدف از یادگیری نظارت‌شده، دسته‌بندی نمونه‌ها به مجموعه‌ای از دسته‌ها است. به‌طورکلی دسته‌بندی بر اساس مدل‌های دسته‌بندی کننده که با فرآیند آموزش و آزمون ایجاد شده‌اند، انجام می‌شود. در محیط یادگیری نظارت‌شده مجموعه‌ای از داده‌ها که دارای برچسب هستند به‌عنوان داده‌های آموزشی جدا و مدل ایجاد می‌شود. ایده رویکرد ترکیبی ایجاد مدل پیش‌بینی با ادغام چندین مدل با هدف به دست آوردن مدلی با عملکرد بهتر است. به‌طورکلی، رویکرد ترکیبی برای بهبود عملکرد پیش‌بینی استفاده می‌شود (پیلکار،۲۰۰٦). روش ترکیبی که نتایج مدل‌های مختلف طبقه‌بندی را برای تولید خروجی یکپارچه ترکیب می‌نماید، یک روش طبقه‌بندی مؤثر برای حوزه‌های مختلف است (هو و همکاران[1]،۱۹۹٤؛ کیتلر[2]، ۱۹۹۸). رویکرد ترکیبی مبتنی بر ترکیب مجموعه‌ای از مدل‌های دسته‌بندی است که ترکیب آن‌ها صحت عملکرد بیشتری را نسبت به استفاده منفرد و مستقل از آن‌ها در اختیار کاربر خواهد گذاشت. مدل جمعی دسته‌بندی کننده، مجموعه‌ای از مدل‌های دسته‌بندی است که تصمیمات منفرد آن‌ها را با یک روش‌شناسی خاص ترکیب می‌شود (الهی و همکاران، ۱۳۹۳). رویکرد ترکیبی با پوشش نواحی که هر یک از دسته‌بندی کننده‌ها نتوانستند پیش‌بینی نمایند، دقت مدل پیش‌بینی کننده تحلیل احساسات افزایش باید. روش‌های رأی اکثریت و رتبه‌بندی از روش‌های متداول برای ترکیب نتایج در رویکردهای ترکیبی است. روش‌های رأی‌گیری زمانی که هر دسته‌بندی کننده یک برچسب منحصربه‌فرد برای هر کلاس را نمایش می‌دهد، استفاده می‌شود. روش رأی اکثریت بر این اساس است که کلاس برنده بیش از نیمی از آراء را به خود اختصاص می‌دهد و در این روش همه دسته‌بندی کننده‌ها دارای وزن یکسان هستند (سام ولم،۲۰۰۰). در روش رأی‌گیری مبتنی بر وزن، وزن رأی دسته‌بندی کننده‌های مختلف متفاوت است و وزن بالاتر به دسته‌بندی کننده‌ای که عملکرد بهتری داشته باشد تخصیص می‌یابد؛ بنابراین انتخاب وزن مناسب بسیار مهم

1. Ho et al.
2. Kittler

www.SID.ir

Archive of SID

است (زانگ و همکاران[1]،۲۰۱٤) . در این پژوهش بر اساس دقت هفت دسته‌بندی کننده وزن‌های متفاوتی تخصیص می‌یابد و مدل مبتنی بر رأی‌گیری بر اساس وزن‌های مختلف مورد ارزیابی قرار می‌گیرد. روش وزن دهی بر این اساس پیاده شده است که به دسته‌بندی کننده‌های با دقت بالاتر وزن بیشتری تخصیص داده شده است که بر اساس دقت، وزن‌های ۱، ۲ و ۳ جهت تخمین استفاده شده است. لازم به ذکر است حالت‌های مختلف وزن دهی مورد ارزیابی قرار گرفته است که وزن‌های ذکر شده عملکرد بهتری را نشان می‌دهند.

ارزیابی

پس از پیاده‌سازی مدل با استفاده از رویکرد ترکیبی لازم است دقت، بازخوانی و معیار ارزیابی F تخمین زده شود. دقت[2] و بازخوانی[3] معیارهای کاربردی در حوزه بازیابی اطلاعات هستند که میزان تناسب اسناد بازیابی شده توسط سیستم را با نیاز کاربر تعیین می‌کنند (رغوان و جانگ[4]، ۱۹۸۹). در این پژوهش جهت تخمین معیارهای فوق از ماتریس درهم‌ریختگی[5] استفاده شده است. ماتریس درهم‌ریختگی عملکرد الگوریتم‌های مربوطه را نشان می‌دهند. در جدول ۲ ماتریس درهم‌ریختگی به تصویر کشیده شده است که بر اساس اطلاعات این ماتریس ارزیابی صورت می‌گیرد.

جدول ۲. ماتریس درهم‌ریختگی الگوریتم‌های طبقه‌بندی

		مقدار پیش‌بینی‌شده[6]	
		دسته مثبت	دسته منفی
مقدار واقعی[7]	دسته مثبت	مثبت درست[8]	مثبت نادرست[9]
	دسته منفی	منفی نادرست[10]	منفی درست[1]

1. Zhang et al.
2. Precision
3. Recall
4. Raghavan & Jung
5. Confusion matrix
6. Predicted
7. Actual
8. True Positive(TP)
9. False Positive(FP)
10. False Negative(FN)

www.SID.ir

Archive of SID

مثبت درست (TP): تعدادی از داده‌ها که به درست به‌عنوان دسته مثبت شناسایی شده‌اند.
مثبت نادرست (FP): تعدادی از داده‌ها که به غلط به‌عنوان دسته مثبت شناسایی شده‌اند.
منفی نادرست (FP): تعدادی از داده‌ها که به غلط به‌عنوان دسته منفی شناسایی شده‌اند.
منفی درست (TN): تعدادی از داده‌ها که به درست به‌عنوان دسته منفی شناسایی شده‌اند.
مقادیر ماتریس همواره عددی بین ۰ و ۱ هستند که هر چه به ۱ نزدیک‌تر باشند عملکرد روش پیشنهادی بهتر خواهد بود (صنیعی آباده و محمودی، ۱۳۹٤).

معیارهای ارزیابی:

$$\text{بازخوانی} = \frac{TN}{FP+TN}$$

$$\text{دقت} = \frac{TN}{TN+FN}$$

$$\text{معیار F} = \frac{2*Recall*Precision}{Recall+Precision}$$

معیارهای ارزیابی الگوریتم دسته‌بندی کننده:

$$\text{نرخ دقت}^{۲} = \frac{TN+TP}{TN+FN+TP+FP}$$

$$\text{نرخ خطا}^{۳} = ۱ - \text{نرخ دقت} = \frac{FN+FP}{TN+FN+TP+FP}$$

یکی از روش‌های ارزیابی الگوریتم‌های دسته‌بندی کننده، اعتبار سنجی چند لایه‌ای[٤] است. برای کاهش خطا در مدل‌سازی جهت جداسازی داده‌های آموزش[٥] و آزمون[٦] از روش اعتبار سنجی چندلایه‌ای استفاده شده است. 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 بکار گرفته می‌شوند. بدین ترتیب فرآیند مذکور K بار تکرار می‌شود، به‌گونه‌ای که از هر کدام از K قسمت تنها یک‌بار برای ارزیابی استفاده شده و در هر مرتبه یک دقت برای مدل ساخته‌شده محاسبه شود. درنهایت میانگین نتیجه این K بار به‌عنوان تخمین نهایی برگزیده می‌شود (نبوتی, و غیره, ۱۳۹۲). معمول‌ترین مقداری که برای K در متون علمی در نظر گرفته می‌شود برابر با ۱است (ت(کوحاوی[۷]، ۱۹۹۵؛ برادفور و برادلی[۱]، ۱۹۹۵). با توجه به نتایج

1. True Negative(TN)
2. Accuracy-Rate
3. Error Rate
4. K-Fold Cross Validation
5. train
6. test
7. Kohavi

www.SID.ir

Archive of SID

پژوهش‌های پیشین در این تحقیق نیز مقدار K برابر با ۱۰ در نظر گرفته شده است و ۱۰ بار مدل اجرا گردید و درنهایت میانگین ارزیابی محاسبه گردید. جهت ارزیابی هر یک از هفت الگوریتم از معیارهای دقت، بازخوانی، معیارF، نرخ دقت و نرخ خطا استفاده شده است که در جدول ۳ دقت هر از الگوریتم‌ها ارائه شده است. نتایج جدول زیر از ده بار تکرار مدل هر یک از دسته‌بندی کننده‌ها حاصل شده است. نتایج نشان می‌دهند دسته‌بندی کننده بیزین ساده چندجمله‌ای با نرخ دقت ۰/۷۹۱۱ دارای دقت بالاتری نسبت به الگوریتم‌ها جهت مدل‌سازی است.

جدول ۳. ارزیابی الگوریتم‌های دسته‌بندی کننده

الگوریتم	دقت	بازخوانی	معیار F	نرخ دقت	نرخ خطا
Naïve Bayes	۰/٦۹	۰/٦۷۵	۰/٦۷۸	۰/۷۰٤	۰/۲۹۵
Linear Support Vector Classification (Linear SVC)	۰/۷٤٤	۰/۷٤۳	۰/۷٤۳	۰/۷۵۵	۰/۲٤۳
multinomial naive Bayes (MNB)	۰/۷۸۳	۰/۷۷٤	۰/۷۷۷	۰/۷۹۱	۰/۲۰۷
Multilayer perceptron (MLP)	۰/۷۵۳	۰/۷٤۸	۰/۷٤۷	۰/۷۵۹	۰/۲۳۹
Bernoulli Naive Bayes (BernoulliNB)	۰/٦٦۸	۰/٦٤۷	۰/٦۵۰	۰/٦۸۳	۰/۳۱۵
LogisticRegression	۰/۷۸۳	۰/۷۷٤	۰/۷۷۷	۰/۷۹۰	۰/۲۰۸
Nu - Vector Classification (Nu - SVC)	۰/۷٦۱	۰/۷۵۲	۰/۷۵۳	۰/۷٦۸	۰/۲۳۱

پس‌ازآنکه نتایج هر یک از الگوریتم‌ها مشخص شد با استفاده دو روش رأی اکثریت و رأی‌گیری بر مبنای وزن دهی جهت ترکیب دسته‌بندی کننده‌ها استفاده شده است. در جدول ۵ نتایج ارزیابی دو روش ترکیبی ارائه شده است. همان‌طور که نتایج جدول زیر نشان می‌دهند روش‌های ترکیبی دارای دقت بالاتری نسبت به هر یک از دسته‌بندی کننده‌ها هستند. 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 می‌نماید و مدلی با دقت بالاتر را توسعه می‌دهد. همان‌طور که نتایج جدول ٤ نشان می‌دهند نرخ دقت روش وزن دهی نسبت

1. BRADFORD & BRODLEY

www.SID.ir

Archive of SID

به روش رأی اکثریت با دقت ۰/۸۰۰۳ بالاتر است، بنابراین مدل ترکیبی منتج از روش وزن دهی جهت تحلیل احساسات انتخاب می‌شود.

جدول ٤. ارزیابی دو رویکرد ترکیب‌کننده

	روش رأی اکثریت	روش وزن دهی
دقت	۰/۷۸۵۹	۰/۷۹۵۸
بازخوانی	۰/۷٦۹۸	۰/۷۸۰٤
معیار F	۰/۷۷۵۰	۰/۷۸۵٤
نرخ دقت	۰/۷۹۱٦	۰/۸۰۰۳
نرخ خطا	۰/۲۰۷۸	۰/۱۹۸۹

تحلیل احساسات

در گام پایانی مدل پیشنهادی بر روی ۷۲۱۰ کامنت مربوط به ۲۰۰ عنوان کتاب مرتبط با موضوع هوشمندی کسب‌وکار منتج از سایت آمازون پیاده‌سازی می‌شود و نظرات مثبت و منفی هر نظر دهنده نسبت به هر کتاب مشخص می‌شود. در جدول ۵ نمونه‌ای از امتیاز حاصل از تحلیل احساسات یعنی میزان مثبت یا منفی نظر نسبت به هر کتاب و نیز نمره تخصیص داده شده توسط کاربر ارائه شده است. در این جدول بخش امتیاز نظر، نرخ مثبت یا منفی بودن نظر نسبت به کتاب را مشخص می‌نماید، به‌عنوان‌مثال عدد ۰/٦۱۵۳۸ – به این معنی است که ٦۱ درصد دسته‌بندی کننده‌ها نظر را منفی تشخیص داده‌اند.

جدول ۵. نمونه‌ای از نظرات مثبت و منفی

نظر	امتیاز نظر (مثبت / منفی)	امتیاز کاربر
This is an excellent and thorough overview of the dynamics of emotional maturity and responsibility.	+۱	۵
The reviews that degrade Goleman's work are rather petty. This book was not written to debunk the importance of IQ, but rather to explore the other side of humanity that we often overlook-emotionality.	-۰/٦۱۵۳۸	٤

www.SID.ir

Archive of SID

در گام نهایی دسته‌بندی مثبت و منفی نظرات نسبت به هر کتاب مشخص گردیده است. نظر نهایی بر اساس میانگین نظرات نظردهندگان هر کتاب محاسبه می‌گردد. در جدول ٦ میانگین نظرات نسبت به کتاب‌ها به‌عنوان نمونه ارائه گردیده است. اعداد ارائه شده در جدول زیر بین ۱+ تا ۱- می‌باشند که هر چه عدد به سمت ۱+ نزدیک‌تر باشد نشان‌دهنده محبوبیت کتاب و هر چه به ۱- نزدیک‌تر باشد به معنای عدم محبوبیت کتاب است.

جدول ٦. تحلیل احساسات نظرات خریداران در مورد کتاب‌ها

دسته‌بندی نظر	عنوان کتاب
+۰/۵۰٥٤	Business Intelligence Guidebook: From Data Integration to Analytics
-۱	Analyzing Data with Power BI and Power Pivot for Excel (Business Skills)

نتیجه‌گیری و پیشنهادها

کتابخانه‌های دانشگاهی به‌عنوان یکی از ارکان آموزشی و پژوهشی، نقش مهمی در کیفیت آموزش و پژوهش در دانشگاه‌ها را دارا می‌باشند. با وجود اهمیت کتابخانه‌های دانشگاهی در امر تحقیق و پژوهش میزان مراجعه و استفاده از کتاب‌های دانشگاهی با وجود هزینه‌های بسیار جهت تأمین کتاب بسیار پایین است. نتایج بررسی‌ها نشان می‌دهند عدم‌کفایت و بروز نبودن منابع از علل مراجعه نکردن دانشجویان به کتابخانه‌های دانشگاهی است. از علل این عدم‌کفایت و بروز نبودن را می‌توان در بی‌اطلاعی از نظرات و عقاید واقعی خوانندگان و استفاده‌کنندگان نسبت به کتاب‌ها دانست. با توجه به خلأ شناسایی‌شده در این حوزه، تحقیق پیش رو مدلی را جهت تحلیل احساسات برای تهیه و خرید کتاب برای کتابخانه-های دانشگاهی ارائه می‌نماید تا کتب خریداری‌شده با روند حرکت علم در دنیا همسو باشد. در این پژوهش تلاش شد تا با استفاده از رویکرد ترکیبی نظرات مثبت و منفی خریداران کتاب دسته‌بندی گردد و بدین ترتیب نظرات واقعی خریداران کتاب نسبت به کتاب‌ها استخراج گردد. نتایج پژوهش نشان می‌دهند روش وزن دهی در رویکرد ترکیبی با دقت ۰/۸۰۰۳ مدل بهتری را جهت تحلیل احساسات خریداران ارائه می‌نماید. در این تحقیق از رویکرد ترکیبی جهت تحلیل احساسات استفاده شده است که در تحقیقات آتی

www.SID.ir

Archive of SID

می‌توان از سایر روهای ترکیب جهت مقایسه نتایج و دستیابی به‌دقت بالاتر استفاده نمود. با توجه به کاربرد و اهمیت نتایج حاصل از تحلیل احساسات در حوزه‌های مرتبط با خرید آنلاین پیشنهاد می‌گردد مدل پیشنهادی برای تحلیل احساسات نظرات خریداران درباره محصولات و برندها و استفاده از نتایج برای مقاصد بازاریابی بکار گرفته شود. در ضمن نظرات کاربران منعکس‌کننده عقیده و نظر واقعی کاربران در مورد محصولات و خدمات می‌باشند و ازاین‌جهت منبع ارزشمندی برای پیشنهاد هستند. در محیط شبکه‌های اجتماعی معمولاً سیستم‌های پیشنهاددهنده جهت ارائه توصیه به کاربران استفاده می‌شود. اساس کار این رویکرد، تجربه و نظر سایر افراد برای خرید اقلام و محصولات است. این روش بر این فرض استوار است که کاربرانی که سلیقه‌های مشابهی دارند، رتبه شبیه به هم می‌دهند و ازاین‌جهت نتایج این پژوهش یعنی نظرات و عقاید استخراج‌شده می‌تواند به‌عنوان ورودی برای ارائه پیشنهاد به کاربران مورد استفاده قرار گیرد.

www.SID.ir

Archive of SID

منابع

الهی، ش، قدس الهی، ا.، و ناجی، ح. (۱۳۹۳). ارائه مدل ترکیبی شبکه‌های عصبی با بهره‌گیری از یادگیری جمعی به‌منظور ارزیابی ریسک اعتباری. *انجمن فناوری اطلاعات و ارتباطات ایران*، ۱۱-۲۸.

برهانی زرندی، س.، نیک‌نفس، ع.، و محمدی, م. (۱۳۹۲). عقیده کاوی در نقد کامل با استفاده از شبکه واژگان احساسی. *دومین کنفرانس ملی مهندسی صنایع و سیستمها. نجف‌آباد: دانشگاه آزاد واحد نجف‌آباد.*

پیکری, ن.، یعقوبی، س.، و طاهری, ح. (۱۳۹٤). تحلیل احساسات در شبکه اجتماعی توییتر با تکنیک متن‌کاوی. *کنفرانس بین‌المللی وب پژوهی*. تهران: دانشگاه علم و فرهنگ.

تفرشی، ش.، و دروگر کلخوران, س. (۱۳۸۸). بررسی میزان آشنایی و بهره‌گیری دانشجویان دانشگاه آزاد اسلامی واحد کرج از کتابخانه‌های. *فصلنامه دانش شناسی*, ۳۳-٤٤

توکلی گارماسه، & رافع، و. (۱۳۹۵). ارائه روشی برای آنالیز احساسات در متن نظرات. *نخستین کنفرانس ملی تحقیقات بین‌رشته‌ای در مهندسی کامپیوتر، برق، مکانیک و مکاترونیک*. قزوین: مرکز آموزش عالی فنی مهندسی بوئین زهرا، پارک علم و فناوری استان قزوین. بازیابی از https://www.civilica.com/Paper-IRCEM-۰۱-۰۱IRCEM_۱٤٤.html.

جعفری، س.، کریمی، آ.، و ابرقوی زاده، ز. (۱۳۹۵). عوامل تأثیرگذار بر تمایل به ادامه استفاده از وب‌سایت شبکه‌های اجتماعی. *فصلنامه مطالعات مدیریت فناوری اطلاعات*، ۱٤۷-۱۸۲.

صنیعی آباده، م.، و محمودی، س. (۱۳۹٤). *داده‌کاوی کاربردی*. تهران: نیاز دانش.

نبوتی، ا.، عزیزی، ا.، عباسی، ا.، وکیلی ارکی، ح.، زارعی، ج.، و رضوی، ا. (۱۳۹۲). کاربرد داده‌کاوی در پیش‌بینی مرگ بیماران سوختگی: مقایسه عملکرد چند الگوریتم. *مدیریت اطلاعات سلامت*، ۷۸۹- ۷۷۹.

نیکنام, ف.، و نیک‌نفس، ع. (۱۳۹۵). بهبود روش‌های متن‌کاوی در کاربرد پیش‌بینی بازار با استفاده از الگوریتم‌های انتخاب نمونه اولیه. *مدیریت فناوری اطلاعات*, ٤۱۵-٤۳٤

Aly, M & ,.Atiya, A.(2013) .LABR: A Large Scale Arabic Book Reviews Dataset .*Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics* ,498-494.

Liu, B .(2015) .*Sentiment analysis: Mining opinions, sentiments,and emotions* .

www.SID.ir

Archive of SID

Cambridge University Press.

Suen, C & ,Lam, L .(2000) .*Multiple Classifier Combination Methodologies* . MCS 2000, LNCS 66-25..

Aldogan, D & ,.Yaslan, Y.(2017) .A comparison study on active learning integrated ensemble approaches in sentiment analysis .*Computers and Electrical Engineering* ,323-311.

Bhatt, A ,.Patel, A., Chheda, H & ,.Gawande, K.(2015) .Amazon Review Classification and Sentiment Analysis .*International Journal of Computer Science and Information Technologies* ,5110-5107.

Boudad, N., Faizi, R., Haj Thami, R & ,.Chiheb, R.(2017) .Sentiment analysis in Arabic: A review of the literature .*Ain Shams Engineering Journal*.

Cambria, E & ,.Hussain, A.(2015) .*Sentic Computing A Common-Sense-Based Framework for Concept-Level Sentiment Analysis* .Switzerland: AG Switzerland.

Chen, Y., Ferrer, X ,.Wiratunga, N & ,.Plaza, E.(2014) .Sentiment and Preference Guided Social .L. Lamontagne and E. Plaza (Eds.): ICCBR 2014, LNCS94-79, 8765

Cheng, H ,.Yan, X ,.Han, J & ,.Hsu, C.-W .(2007) .Discriminative Frequent Pattern Analysis for Effective Classification 2007 .*IEEE 23rd International Conference on Data Engineering* .Istanbul: IEEE.

Chiavetta, F., Lo Bosco, G & ,.Pilato, G .(2016) .A Lexicon-based Approach for Sentiment Classification of Amazon Books Reviews in Italian Language .*In Proceedings of the 12th International Conference on Web Information Systems and Technologies* (WEBIST 2016), (170-159.)

Chibelushi, C & ,.Thelwall , M .(2009) .Text Mining for Meeting Transcript Analysis to Extract Key Decision Elements .*International MultiConference of Engineers and Computer Scientists* ,715-710.

www.SID.ir

Archive of SID

Coletta, L., da Silva, N., Hruschka, E & ,.Hruschka Jr, E.(2014) .*Combining classification and clustering for tweet sentiment analysis* .Biblioteca Digital da Produção Intelectual - BDPI. 215-210,

Das, S & ,.Chen, M.(2001) .*Yahoo! for Amazon: Extracting market sentiment from stock message boards* .in Proceedings of APFA-. 2001

Dave, K., Lawrence, S & ,.Pennock, D. M.(2003) .Mining the peanut gallery: Opinion extraction and semantic classification of product reviews .*International Conference on World Wide Web*(WWW-2003) .

Esuli , A & ,.Sebastiani, F .(2006) .Determining term subjectivity and term orientation for opinion mining .In Proceedings of EACL-06, *11th Conference of the European Chapter of the Association for Computational Linguistics*,(26-1) .Trento: Forthcoming.

Fang, X & ,.Zhan, u .(2015) .Sentiment analysis using product review data . *Journal of Big Data*. 14-1, (5)2, doi:DOI 10, 1186/s2-0015-015-40537

Filho, P & ,.Pardo, T.(2013) .NILC USP: A Hybrid System for Sentiment Analysis in Twitter Messages .*Second Joint Conference on Lexical and Computational Semantics* (*SEM). (572-568.) Association for Computational Linguistics.

Govindarajan, M .(2013) .Sentiment Analysis of Movie Reviews using Hybrid Method of Naive Bayes and Genetic Algorithm .*International Journal of Advanced Computer Research*. 145-139,

Gurusamy, V & ,.Kannan, S .(2014) .P*reprocessing Techniques for Text Mining* .RTRICS .Podi.

Haddi, E., Liu, X & ,.Shi, Y .(2013) .The Role of Text Pre-processing in Sentiment Analysis .*Procedia Computer Science*. 32-26,

Hamdan, H ,.Bellot, P & ,.Bechet, F .(۲۰۱٦) .*Sentiment Analysis in Scholarly* .Book Reviews.

Ho, T., Hull, J & ,.Srihari, S .(1994) .Decision combination in multiple

www.SID.ir

Archive of SID

classifier systems .*IEEE Transactions on Pattern Analysis and Machine Intelligence*. 75-66,

Isah, H., Trundle, P & ,.Neagu, D .(2014) .Social Media Analysis for Product Safety using Text Mining and Sentiment Analysis .*Institute of Electrical and Electronics Engineering*.

Katariya, N. P & ,.Chaudhari, M. S .(2015) .TEXT PREPROCESSING FOR TEXTMINING USING SIDE INFORMATION .*International Journal of Computer Science and Mobile Applications*. 5-1,

Kohavi, R .(1995) .A Study of CrossValidation for Accuracy Estimation and Model Selection .*Appears in the International Joint Conference on Articial Intelligence (IJCAI)*.

Krouska, A., Troussas, C & ,.Virvou, M .(2016) .*The effect of preprocessing techniques on Twitter* .Chalkidiki .Greece: IEEE.

Liu, B.(2010) .Sentiment Analysis and Subjectivity .*Invited Chapter for the Handbook of Natural Language Processing* .Boca: Taylor and Francis Group .http://gnode١.mib.man.ac.uk/tutorials/NLPhandbook-NLPhandbook–

Liu, B.(2012) .*Sentiment Analysis and Opinion Mining (Synthesis Lectures on Human Language Technologies)* .Williston: Morgan & Claypool Publishers.

Markić ,B., Bijakšić ,S & ,.Bevanda, A.(2016) .SENTIMENT ANALYSIS OF SOCIAL NETWORKS AS A CHALLENGE TO THE DIGITAL MARKETING .Ekonomski vjesnik/Econviews - Review of Contemporary Business, *Entrepreneurship and Economic Issues*. (1)29,

Medhat, W., Hassan, A & ,.Korashy, H .(2014) .Sentiment analysis algorithms and applications:A survey .*Ain Shams Engineering Journal*.

Morinaga, S., Yamanishi, K., Tateishi, K & ,.Fukus, T .(2002) .Mining

www.SID.ir

Archive of SID

product reputations on the web .Proceedings of the eighth ACM SIGKDD *international conference on Knowledge discovery and data mining* - KDD (349-341).

Morinaga, Satoshi, Yamanishi, K ,.Tateishi, K & ,Fukushima, T.(2002) . Mining product reputations on the web . ACM SIGKDD *International Conference on Knowledge Discovery and Data Mining* (KDD-2002) .

Nasukawa, T & ,.Yi, J .(2003) .Sentiment analysis: Capturing favorability using natural language processing .KCAP-03, *2nd Intl. Conf. on Knowledge Capture* .

Nayak, A. S., Kanive, A. P., Chandavekar, N & ,R, B.(2016) ,June .Survey on Pre-Processing Techniques for Text Mining .*International Journal Of Engineering And Computer Science*. 16879-16875, (6)5,

Nguyen, T., Shirai, K & ,.Velcin, J .(2015) .Sentiment analysis on social media for stock movement prediction .*Expert Systems With Applications*. 9611-9603,4

Oza, N .(2006) .*Ensemble Data Mining Methods* .*In Encyclopedia of Data Warehousing and Mining*. 453-448,

P. BRADFORD, J & ,.E. BRODLEY, C .(2001) .*The Effect of Instance-Space Partition on Significance* .Machine Learning. 286-269,

Pandey, A., Rajpoor, D & ,.Saraswat, M .(2017) .Twitter sentiment analysis using hybrid cuckoo search method .*Information Processing and Management*. 779-764,

Pang, B & ,.Lee, L .(2008) .*Opinion Mining and Sentiment Analysis* . *Foundations and Trends*. 135-1,(2-1)2,

Pang, B., Lee, L & ,.Vaithyanathan, S .(2002) .Thumbs up?: sentiment classification using machine learning techniques .*EMNLP 02 Proceedings of the ACL-02 conference on Empirical methods in natural language processing*. 86-79,

www.SID.ir

Archive of SID

Pang, B., Lee, L & ,.Vaithyanathan, S .(2002) .Thumbs up?:sentiment classification using machine learning techniques .*EMNLP '02 Proceedings of the ACL-02 conference on Empirical methods in natural language processing* . 86-79,

Polikar , R .(2006) .Ensemble based systems in decision making .*IEEE Circuits Syst Mag* ,45-21.

Rain, C .(2012) .Sentiment Analysis in Amazon Reviews Using Probabilistic Machine Learning .*Rain2012SentimentAI* .

Rain, C.(2013) .Sentiment Analysis in Amazon Reviews Using Probabilistic Machine Learning .*Swarthmore College Compter Society*.

Ramasubramanian, C & ,.Ramya, R .(2013) .Effective Pre-Processing Activities in Text Mining using Improved Porter's Stemming Algorithm . *International Journal of Advanced Research in Computer and Communication Engineering*. 4538-4536,

Rylander, R. G., Propst, D. B & ,.Mcmurtry, T. R.(1995) .Nonresponse and Recall Biases in a Survey of Traveler Spending .*Journal of Travel Research* ,45-39 .doi:https://doi.org/00472875950330040 6/10,1177

Sebaoun ,É ,.Rafrafi, A., Guigue, V & ,.Gallinari, P .(2013) .Cross-media sentiment classification and application to box-office forecasting .*OAIR '13 Proceedings of the 10th Conference on Open Research Areas in Information Retrieval* . 208-201,

Stavrianou, A ,.Andritsos, P & ,.Nicoloyannis, N .(2007) .Overview and semantic issues of text mining .*ACM SIGMOD Record*. 34-23,

Tong, R. M .(2001) .An operational system for detecting and tracking opinions in on-line discussion .*in Proceedings of SIGIR Workshop on Operational Text Classification* .

Turney, P. D .(2002) .Thumbs up or thumbs down?: semantic orientation

www.SID.ir

Archive of SID

Thumbs up or thumbs down?: semantic orientation .*in Proceedings of Annual Meeting of the Association for Computational Linguistics* (ACL-2002) .

V. RAGHAVAN a, V & ,.S. JUNG , G .(1989) .A Critical Investigation of Recall and Precision as Measures of Retrieval System Performance .*ACM Transactions on Information Systems*. 229-206.

Wang, Z.(2017) .*The Evaluation of Ensemble Sentiment Classification Approach on Airline Services Using Twitter* .Dublin : Dublin Institute of Technology.

Wiebe, J .(2000) .*Learning subjective adjectives from corpora* .in Proceedings of National Conf. on Artificial Intelligence (AAAI-2000) .

Zhang, Y., Lai, G ,.Zhang, M., Zhang, Y ,.Liu, Y & ,.Ma, S .(2014) . Explicit Factor Models for Explainable Recommendation based on Phrase-level Sentiment Analysis .*SIGIR '14Proceedings of the 37th international ACM SIGIR conference on Research & development in information retrieval*. (92-83.) Queensland: ACM.

Zhang, Y., Zhang, H., Cai, J & ,.Yang, B .(2014) .*A Weighted Voting Classifier Based on Differential Evolution* .Abstract and Applied Analysis .http://dx.doi.org/376950/2014/10,1155

Zhang, Z ,.Ye, Q., Zhang, Z & ,.Li, Y .(2011) .Sentiment classification of Internet restaurant reviews written in Cantonese .*Expert Systems with Applications: An International Journal*. 7682-7674,(6)38.

Zubair Asghar, M., Khan, A., Ahmad, S & ,.Kundi, F .(2014) .A Review of Feature Extraction in Sentiment Analysis .*Journal of Basic and Applied Scientific Research*. 186-181.

www.SID.ir