

Sentiment Analysis at the Product Feature Level and Based on Users Gender

Shahriar Mohammadi 

MSc., Information Technology Engineering,
Islamic Azad University, Qazvin Branch,
Qazvin, Iran.

Eslam Nazemi * 

Assistant Professor, Faculty of Computer
Engineering and Science, Shahid Beheshti
University, Tehran, Iran

Abstract

Social media data is one of the most effective and accurate indicators of public sentiment so that analyzing this information can provide researchers with interesting results from users' sentiment about characters, subjects, products, and services. In this study, while reviewing users' opinions on Twitter about the various features of two competing mobile phone products on the market, Apple's Iphone X and Samsung's Galaxy S9, we examine their sentiment based on the gender of consumers of these two products. This study is performed using the relation-based method in the feature extraction step and Lexicon-Based in the polarity of opinions step. The results of this study show that the popularity of different product features varies between male and female users, and based on these results, business owners can produce products that focus on people's gender or design smart advertising plans according to their interests. These measures ultimately lead to increased business profitability and customer satisfaction.

Keywords: Sentiment Analysis, Twitter, Feature Extraction, Polarity Of Opinions.


* Corresponding Author: Nazemi@sbu.ac.ir

How to Cite: Mohammadi, S., Nazemi, E., (2021). Sentiment Analysis at the Product Feature Level and Based on Users Gender, *Journal of Business Intelligence Management Studies*, 10(37), 267-296.




تجزیه و تحلیل احساسات در سطح ویژگی محصول و مبتنی بر جنسیت کاربران

کارشناسی ارشد مهندسی فناوری اطلاعات، دانشکده مهندسی کامپیوتر و فناوری اطلاعات، دانشگاه آزاد اسلامی، واحد قزوین، قزوین، ایران.

شهریار محمدی 

دانشیار دانشکده مهندسی و علوم کامپیوتر دانشگاه شهید بهشتی، تهران، ایران.

اسلام ناظمی *

چکیده

داده‌های شبکه‌های اجتماعی، یکی از مؤثرترین و دقیق‌ترین شاخص‌های احساسات عمومی است؛ به طوری که تجزیه و تحلیل این اطلاعات می‌تواند نتایج جالبی از احساسات کاربران نسبت به هر شخصیت، موضوع، محصول و خدمات را برای محققین نمایان سازد. در این پژوهش، ضمن بررسی نظرات کاربران در شبکه اجتماعی توئیتر در مورد ویژگی‌های مختلف دو محصول رقیب تلفن همراه-شرکت‌های اپل و سامسونگ- در بازار، احساسات آنان را بر اساس جنسیت مصرف‌کنندگان این دو محصول مورد بررسی قرار گرفته است. این بررسی با استفاده از روش مبتنی بر رابطه در مرحله استخراج ویژگی و رویکرد مبتنی بر لغت‌نامه احساسی در مرحله تعیین قطبیت نظرات انجام گرفته است. نتایج این پژوهش، حاکی از آن بود که محبوبیت ویژگی‌های مختلف محصول بین کاربران مرد و زن متفاوت بوده و بر اساس این نتایج، صاحبان کسب و کار می‌توانند اقدام به تولید محصولاتی با تمرکز بر جنسیت افراد کرده و یا به طراحی برنامه‌های هوشمند تبلیغاتی با توجه به علایق آنان پردازند. این اقدامات در نهایت به افزایش سوددهی کسب و کار و رضایت‌مندی مشتریان منجر می‌شود.

کلیدواژه‌ها: تجزیه و تحلیل احساسات، توئیتر، استخراج ویژگی، تعیین قطبیت نظرات.

مقدمه

امروزه، استفاده از اینترنت و شبکه‌های اجتماعی برای مردم جهان به امری حیاتی تبدیل شده است. طبق آماری که توسط شاخص وب جهانی^۱ در سه‌ماهه سوم سال ۲۰۲۰ منتشر شده است، ۹۸٪ استفاده‌کنندگان از اینترنت در سراسر جهان، از شبکه‌های اجتماعی استفاده می‌کنند و به‌طور متوسط ۲ ساعت و ۲۲ دقیقه از یک شبانه‌روز را در شبکه‌های اجتماعی می‌گذرانند (شاخص وب جهانی، ۲۰۲۰).

محتوای ایجادشده در شبکه‌های اجتماعی، به دلیل اینکه حاوی سیلی از نظرات و مکالمات میلیون‌ها نفر از کاربران هستند، فرصت‌های بی‌سابقه‌ای برای سازمان‌ها فراهم می‌کند تا به‌طور مؤثر در محتوای رسانه‌های اجتماعی مربوط به سازمان خود دخالت کرده و آن‌ها را تجزیه و تحلیل نمایند (هه و همکاران^۲، ۲۰۱۳). به‌عنوان مثال، مدیران اجرایی ممکن است بخواهند بینش و دانش کسب‌وکار خود را بر اساس آنچه مشتریان درباره سازمانشان در فیس‌بوک یا توییتر می‌گویند، به دست آورند. به‌طوری‌که آن دانش و بینش، به بهبود تصمیم‌گیری و یا دیگر جنبه‌های کسب و کارشان منجر شود (هه و همکاران، ۲۰۱۵).

مطالعه‌ای که توسط شرکت بین^۳ انجام گرفت نشان داد که شرکت‌های بزرگی که قابلیت تجزیه و تحلیل داده‌های پیشرفته‌ای دارند، توانسته‌اند رقبای خود را به حاشیه برانند. آن‌ها دریافته‌اند که پذیرندگان تجزیه و تحلیل داده‌ها، پنج برابر سریع‌تر از رقبای خود در تصمیم‌گیری هستند و احتمالاً در چارت عملکرد مالی، دو برابر سریع‌تر از رقبای خود در صنعت ایستاده‌اند (پیرسون و وگنر^۴، ۲۰۱۳).

هدف کلی این پژوهش، بررسی نظرات کاربران در شبکه اجتماعی توییتر در مورد محصولات و تجزیه و تحلیل این نظرات است. همچنین درک درست از نظرات

-
1. Global Web Index
 2. He et al.
 3. Bain
 4. Pearson & Wegener

مصرف کنندگان و مقایسه ویژگی‌های هر محصول به‌طور جداگانه و بر اساس جنسیت، به بهبود تصمیم‌گیری‌های کلان سازمان در جهت افزایش رضایت مشتری و ایجاد محصولات جدید به تفکیک جنسیت و علایق مشتریان خود منجر می‌شود. علاوه بر این، از این مقایسه می‌توان به نقاط قوت و ضعف محصولات رقبا در بازار دست‌یافت تا شرکت‌ها بتوانند با رفع نقاط ضعف خود، به افزایش سهم بازار کمک کرده و با درک نقاط قوت رقبا، تلاش خود را در جهت رسیدن به سطح رقبای قوی‌تر افزایش دهند. اهداف جزئی‌تر این پژوهش عبارت‌اند از:

- ارائه روشی برای جمع‌آوری، پاک‌سازی و آماده‌سازی نظرات از شبکه‌های اجتماعی.

- پایش و تحلیل نظرات به‌وسیله استفاده ترکیبی از رویکرد مبتنی بر رابطه در استخراج ویژگی و استفاده از رویکرد مبتنی بر لغت‌نامه احساسی در شبکه‌های اجتماعی در تعیین قطبیت نظرات.

- بیان شدت احساسات مصرف کنندگان نسبت به ویژگی‌های متفاوت هر محصول بر اساس جنسیت مصرف کنندگان محصول.

مبانی نظری و پیشینه پژوهش

از آنجایی که حجم اطلاعات الکترونیکی و برخط روزبه‌روز بیشتر می‌شود، استفاده از این منبع اطلاعاتی بزرگ به یکی از بزرگ‌ترین دغدغه‌های سازمان‌ها تبدیل شده است. ارائه ابزارهایی که با بررسی متون بتواند تحلیلی روی آن‌ها انجام دهند، به شکل‌گیری زمینه‌ای در هوش مصنوعی منجر شده که به "متن‌کاوی" معروف است. این حوزه تمام فعالیت‌هایی که به‌نوعی به دنبال کسب دانش از متن هستند را شامل می‌شود. بر طبق تعریف آبدوس و هی^۱، متن‌کاوی شامل پیدا کردن مدل‌های سودمند، روندها، الگوها و یا قوانین از داده‌های

1. Abdous and He.

متنی بدون ساختار مانند فایل‌های متنی، فایل‌های فرا متن^۱، پیام‌ها، ایمیل‌ها و ... است (ابدوس و هه، ۲۰۱۱).

از شاخه‌های گسترش یافته متن کاوی، می‌توان به تجزیه و تحلیل احساسات اشاره کرد. تجزیه و تحلیل احساسات، شامل مجموعه‌ای از فرآیندهای مورد استفاده برای شناسایی احساسات، عاطفه، ذهنیت و دیگر حالات احساسی در محتوای متن است که می‌تواند به اطلاعاتی معنی‌دار تبدیل شده و در تصمیم‌گیری مورد استفاده قرار گیرد (کمبریا و همکاران^۲، ۲۰۱۳).

شروع و رشد مطالعه بر روی تجزیه و تحلیل احساسات با رشد شبکه‌های اجتماعی تحت وب همچون انجمن‌های بحث و گفتگو، وبلاگ‌ها، میکرو بلاگ‌ها، توییتر و سایر شبکه‌های اجتماعی دیگر هم‌زمان شد؛ زیرا برای اولین بار در تاریخ بشر، محققین با حجم عظیمی از اطلاعات ثبت شده در فرم‌های دیجیتالی مواجه شدند. از اوایل سال ۲۰۰۱، تجزیه و تحلیل احساسات، این فیلد از پردازش زبان طبیعی، به یکی از مهم‌ترین زمینه‌های تحقیقاتی محققین تبدیل شد (پیریانی و همکاران^۳، ۲۰۱۶).

در ادامه، کارهای انجام گرفته در زمینه تجزیه و تحلیل احساسات در سه سطح سند، جمله و ویژگی بررسی شده و روش‌ها، فنون و نقاط قوت و ضعف هر یک بیان می‌گردد.

تجزیه و تحلیل احساسات در سطح سند

آنچه در این سطح از تجزیه و تحلیل مفروض است این است که کل یک سند، نمایش‌دهنده نظر پیرامون یک موجودیت واحد و با احساسات مثبت و منفی است.

اگرچه رویکردهای متفاوتی برای تعیین قطبیت نظرات در سطح سند بکار برده شده‌اند، اما دو رویکرد یادگیری ماشین (یادگیری با ناظر) و رویکرد مبتنی بر لغت‌نامه (یادگیری بدون ناظر) رویکردهایی مهم در این زمینه هستند.

1. HTML

2. Cambria et al.

3. Piryani et al.

تجزیه و تحلیل احساسات در سطح سند با استفاده از رویکرد یادگیری ماشین یکی از اولین کارها با رویکرد یادگیری ماشین، تحقیقات پانگ و لی^۱ در سال ۲۰۰۲ بود که آزمایش‌های گسترده‌ای را روی نظرات مربوط به فیلم با استفاده از سه الگوریتم نایو بیس^۲، ماشین بردار پشتیبان^۳ و طبقه‌بندی کننده حداکثر آنتروپی^۴ انجام دادند (پانگ و لی، ۲۰۰۲). نتایج این پژوهش حاکی از آن بود که روش‌های یادگیری ماشین بهتر از کدگذاری‌های انسانی عمل می‌کنند.

در سال ۲۰۰۹، ژانگ و لاو^۵ به مقایسه عملکرد سه الگوریتم یادگیری ماشین شامل ماشین بردار پشتیبان، نایویز و مدل کاراکتری مبتنی بر چندجمله‌ای در حوزه کاوش نظرات موجود در وبلاگ‌های مسافرتی پرداختند (یه و همکاران^۶، ۲۰۰۹). نتایج تجربی به دست آمده نشان داد که رویکردهای ماشین بردار پشتیبان و مدل کاراکتری مبتنی بر چندجمله‌ای عملکرد بهتری نسبت به رویکرد نایویز دارند و زمانی که مجموعه داده آموزشی شامل نظرات زیادی است، دقت هر سه الگوریتم تقریباً ۸۰٪ است.

در سال ۲۰۲۰، انصاری و همکاران به مقایسه سرعت و دقت الگوریتم‌های یادگیری ماشین شامل حافظه طولانی کوتاه-مدت^۷، ماشین بردار پشتیبان، درخت تصمیم‌گیری، رگرسیون و جنگل تصادفی در حوزه تجزیه و تحلیل نظرات کاربران تویتر پیرامون انتخابات سراسری هند پرداختند (انصاری و همکاران^۸، ۲۰۲۰). نتایج پژوهش نشان داد که الگوریتم رگرسیون دارای کمترین زمان اجرا و الگوریتم حافظه طولانی کوتاه-مدت دارای بیشترین زمان اجرای الگوریتم است. همچنین الگوریتم جنگل تصادفی دارای زمان کمینه برای آموزش روی داده‌ها همراه با دقت بالا را ثبت کرده است، در صورتی که

-
1. Pang and Lee
 2. Naive Bayes
 3. Support Vector Machine (SVM)
 4. maximum Entropy Classifier
 5. Zhang and Law
 6. Ye et al.
 7. Long Short-Term Memory (LSTM)
 8. Ansari et al.

الگوریتم حافظه طولانی کوتاه-مدت با دقت ۷۴٪ دارای عملکرد بالاتری نسبت به الگوریتم جنگل تصادفی است، البته با یک هزینه محاسبه بالاتر.

روش‌هایی نیز برای بهبود الگوریتم‌های یادگیری ماشین همچون ماشین بردار پشتیبان ارائه شده است. بصری و همکاران^۱ در پژوهش خود که در سال ۲۰۱۳ منتشر شده است، روشی را با ترکیب دو الگوریتم ماشین بردار پشتیبان و بهینه‌سازی ازدحام ذرات پیشنهاد کردند که از الگوریتم بهینه‌سازی ازدحام ذرات برای بهبود پارامترهای ماشین بردار پشتیبان بهره می‌برد (بصری و همکاران، ۲۰۱۳). ویندهینی و چاندراسکاران^۲، ترکیب الگوریتم ماشین بردار پشتیبان و تحلیل مؤلفه اصلی را برای بهبود این الگوریتم به منظور دسته‌بندی نظرات پیشنهاد دادند (ویندهینی و چاندراسکاران، ۲۰۱۳). آن‌ها از تحلیل مؤلفه‌های اصلی به منظور کاهش ابعاد داده‌ها استفاده کردند و بدین ترتیب پیچیدگی ماشین بردار پشتیبان را کاهش دادند.

تجزیه و تحلیل احساسات در سطح سند با استفاده از رویکرد مبتنی بر لغت‌نامه در رویکردهای مبتنی بر لغت‌نامه، نیاز به تعیین قطبیت کلمات در سطح کلمه داریم و پس از تعیین قطبیت، از تجمیع قطبیت کلمات می‌توان به قطبیت جملات و قطبیت کل سند دست یافت.

هانگ و لین^۳ در سال ۲۰۱۳ از لغت‌نامه سنتی وُردنِت^۴ برای استخراج احساسات درون متن استفاده کردند (هانگ و لین، ۲۰۱۳). آن‌ها در روش خود کلماتی که قطبیت آن‌ها در سنتی وُردنِت نبود را بر اساس اینکه بیشتر در جمله‌های منفی یا مثبت حضور دارند، تعیین قطبیت کردند. در این حالت درستی الگوریتم نسبت به زمانی که از این کلمه‌ها استفاده نمی‌شود، بهبود می‌یابد.

-
1. Basari and et al.
 2. Vindohini and Chandrasekaran
 3. Hang and Lin
 4. SentiWordNet

ابراهیم و وانگ^۱ در سال ۲۰۱۹ از لغت‌نامه سنتی‌استرنیت^۲ برای تعیین گرایش، قدرت و شدت احساسات توییت‌های کاربران در مورد ۵ برند خرده‌فروشی آنلاین بریتانیا استفاده کردند که به هر توییت یک نمره در بازه (۵،-۵) اختصاص می‌دهد (ابراهیم و وانگ، ۲۰۱۹). نتایج این پژوهش نشان داد که استفاده از این لغت‌نامه به دلیل اینکه از قوانین زبانی شامل منفی‌کننده‌ها، تشدیدکننده‌ها و شکلک‌ها برای محاسبه قدرت و شدت قطبیت نظرات استفاده می‌کند، برای توییت‌ها که شامل داده‌های کوتاه و غیررسمی است به خوبی جواب می‌دهد.

ضعف اصلی سطح سند، کلی‌نگری آن نسبت به نظرهای داده‌شده در سند است، درحالی‌که یک سند، مجموعه‌ای از جملاتی است که قطبیت جملات (مثبت، منفی و خنثی) آن متغیر است. به همین دلیل دسته‌بندی در سطح جمله نسبت به دسته‌بندی در سطح سند مفیدتر است.

تجزیه و تحلیل احساسات در سطح جمله

وظیفه اصلی این سطح این است که تعیین کند یک جمله از یک نظر، می‌تواند مثبت، منفی و یا خنثی باشد.

هو و لئو^۳ یک متد ساده برای تشخیص گرایش جمله ارائه کردند (هو و لئو، ۲۰۰۴). مطابق نظریه آن‌ها اگر در یک جمله تعداد کلمات مثبت بیشتر از کلمات منفی باشد، جمله نیز مثبت خواهد بود و در غیر این صورت منفی است.

میکولا و ماچووا^۴ در پژوهش خود در سال ۲۰۱۵، گرایش جملات را بر اساس محتوای مکالمه‌ای و بر اساس میانگین شدت قطبیت تک‌تک کلمات آن جمله بیان کرده‌اند (میکولا و ماچووا، ۲۰۱۵). الگوریتم آنان مبتنی بر لغت‌نامه و بر اساس محاسبه وزن هر کلمه بر اساس وجود کلمات تشدیدکننده و منفی‌کننده است، به طوری که در

-
1. Ibrahim and Wang
 2. SentiStrength
 3. Hu and Liu
 4. Mikula and Machova

صورت وجود کلمات تشدیدکننده در جمله، کلمات با قطبیت قوی‌تر، بیشتر از کلمات با قطبیت ضعیف‌تر، تأثیر پذیرفته و تقویت می‌شود. مدل پیشنهادی آنان توانست به دقت ۸۶/۲٪ در تعیین میزان گرایش احساسی جملات دست یابد.

مشاهده می‌شود که در اکثر کارهایی که در سطح جمله انجام شده است، نتیجه بسیار به گرایش معنایی کلمات در آن جمله وابسته است و ممکن است که یک کلمه در جمله، با توجه به گرایش احساسی کلمات پیرامون آن، گرایش متفاوتی را پذیرا باشد.

تجزیه و تحلیل احساسات در سطح ویژگی

هیچ‌یک از دو سطح قبلی یعنی سند و جمله به‌طور دقیق آنچه را که مردم دوست دارند و آنچه دوست ندارند را کشف نمی‌نماید. بجای نگرش به ساختار زبان (سند، پارگراف، جمله و ...)، سطح ویژگی مستقیماً خود نظر را دنبال می‌کند. در حقیقت این نوع تجزیه و تحلیل بر این ایده استوار است که هر نظر شامل یک گرایش (مثبت یا منفی) و یک ویژگی است که کاربر نظر خود را در مورد آن ویژگی بیان می‌کند.

در این سطح، از روش‌های گوناگونی برای استخراج ویژگی استفاده می‌شود. روش‌هایی همچون روش‌های آماری (مانند بهره اطلاعاتی، مربع کای و ...)، مبتنی بر واژگان، هستی‌شناسی، روش‌های مبتنی بر تکرار و رابطه و ... که هر یک دارای مزایا و معایب خاص خود است.

لیانگ پووی و بی رو دای^۱ در سال ۲۰۱۳ یک روش جدید برای استخراج احساسات از میکرو بلاگ‌ها همانند توییت ارائه کردند (لیانگ پووی و بی رو دای، ۲۰۱۳). آن‌ها دریافتند که بعضی توییت‌ها معنی مثبت می‌دهد، اما دارای احساسی منفی است. برای غلبه بر این چالش، آن‌ها ترکیبی از یادگیری تحت نظارت که قادر به استخراج، یادگیری و طبقه‌بندی توییت‌ها است استفاده کردند. به‌منظور کاهش ویژگی‌های مجموعه، از روش‌های آماری اطلاعات متقابل و مربع کای به‌عنوان روش‌های انتخاب ویژگی استفاده

1. Liang Po-Wei and Bi-Ru Dai

شده است. همچنین برای طبقه‌بندی توییت‌ها از نایوبیز استفاده شد. دقت به‌دست آمده در هنگام استفاده از روش انتخاب ویژگی مربع کای، برابر ۹۱٪ و بهترین دقت به‌دست آمده نیز ۹۶/۶٪ بود.

روش مبتنی بر واژگان، ویژگی‌های محصول را بر اساس یک لغت‌نامه استخراج می‌کند. لی و همکارانش^۱ در پژوهش خود در سال ۲۰۰۹، لغت‌نامه‌ای را ایجاد کردند که حاوی لیستی از کلمات ویژگی و کلمات نظر بود که از آن برای تخصیص قطبیت به هر ویژگی استفاده می‌کردند (لی و همکاران، ۲۰۰۹). آن‌ها سپس با استفاده از روش‌های آماری و تکنیک‌های زبان طبیعی، به استخراج کلمات ویژگی و کلمات نظر بر اساس واژگان موجود در لغت‌نامه اقدام کردند.

وانگ و همکاران^۲ در سال ۲۰۱۷ به بررسی ۲۰,۰۰۰ نظر در مورد خدمات محلی ۱۰۰ رستوران فعال در شهر شانگهای چین پرداختند (وانگ و همکاران، ۲۰۱۷). در این پژوهش، از ترکیب قوانین توالی کلاس^۳ و روش‌های یادگیری ماشین برای طبقه‌بندی نظرات استفاده شده و برای بهبود طبقه‌بندی از یک لغت‌نامه برای معرفی بیشتر جملات مقایسه‌ای و شناسایی رابطه معادل و غیر معادل استفاده شده است. بر اساس مدل فوق، ویژگی‌های هر موجودیت (رستوران) و روابط تلفیقی برای تحلیل احساسات استخراج شده و به مقایسه ویژگی‌ها و شناسایی نقاط ضعف آن‌ها پرداخته می‌شود.

مارتینز و همکاران^۴ در سال ۲۰۱۱ یک روش تجزیه و تحلیل احساسات مبتنی بر تجزیه و تحلیل بردار را پیشنهاد داده و با استفاده از هستی‌شناسی، نتایج حاصل را بهبود بخشیده‌اند (مارتینز و همکاران، ۲۰۱۱). در انتخاب ارزش هر ویژگی، اکثر کارها ارزش یکسانی برای تمام ویژگی‌ها در نظر گرفته‌اند، ولی مسلماً برخی ویژگی‌ها دارای ارزش بیشتری نسبت به دیگر ویژگی‌ها هستند. مارتینز و همکارانش ارزش متفاوتی برای

-
1. Li and et al.
 2. Wang et al.
 3. Class Sequence Roles (CSR)
 4. Martinez and et al.

ویژگی‌ها با توجه به تعداد تکرار ویژگی از سوی کاربران و موقعیت آن‌ها در جمله تعریف کردند.

در سال ۲۰۱۷ روش جدیدی برای استخراج ویژگی از نظرات تویتر ارائه گردید که بجای استفاده از توییت‌های ساده و انفرادی، از مکالمات تویتر استفاده می‌کند (عثمان و همکاران^۱، ۲۰۱۷). برای ایجاد مکالمات، ابتدا توییت ریشه و توییت‌های مرتبط با آن را بر اساس لینک‌های پاسخ^۲، ریتوییت^۳ و منشن^۴ و همچنین مقایسه هر جفت توییت به روش شباهت کوزین^۵ استخراج کرده تا برای هر مکالمه، یک درخت پاسخ ایجاد شود. سپس بر اساس مرتبط بودن محتوای توییت و میزان نفوذ توییت، به فیلتر کردن مکالمات پرداخته می‌شود. در نهایت در فاز استخراج ویژگی، عبارات اسمی پرتکرار در درخت مکالمات استخراج گردیده و با استفاده از معیار PMI-IR، میزان معنی‌دار بودن عبارات اسمی بررسی می‌شود. این روش استخراج ویژگی بر اساس تکرار ویژگی در مکالمات، دارای دقتی در حدود ۷۹/۵۴٪ است که بالاتر از زمانی است که توییت‌ها به صورت انفرادی مورد بررسی قرار می‌گیرند.

روش استفاده‌شده در پژوهش عثمان و همکاران (۲۰۱۷)، بسیار شبیه به روشی است که در سال ۲۰۱۵ برای ایجاد مکالمات از صفحات فرامتن و پاسخ‌های کاربران توسط بلکارویی و فیض^۶ ارائه گردید. از آنجایی که در این روش، استخراج مکالمات از طریق لینک‌های پاسخ است، پیام‌های غیرمستقیم مانند ریتوییت و منشن و شباهت بین هر جفت توییت برای ایجاد درخت پاسخ در نظر گرفته نمی‌شود.

مشکل روش‌های مبتنی بر تکرار این است که با اینکه این روش‌ها، می‌توانند به صورت بدون نظارت یا با نظارت ضعیف پیاده‌سازی شوند، اما زمانی که از یک دامنه

-
1. Othman
 2. Reply
 3. Retweet
 4. Mention
 5. Cosine Similarity
 6. Belkaroui and Faiz

موضوع به دامنه موضوع دیگر برده می‌شوند، نیاز به تنظیم دوباره دستی بسیاری از پارامترها (مانند مقادیر آستانه) دارند (شاه طالبی و همکاران، ۲۰۱۵).

روش‌های مبتنی بر رابطه، در مقایسه با روش‌های مبتنی بر تکرار، قابلیت استخراج ویژگی‌های با تعداد تکرار کم را از متون دارند. ضمن اینکه از معایب این روش‌ها این است که همانند روش‌های مبتنی بر تکرار در فرآیند استخراج ویژگی، تعداد زیادی ویژگی زائد تولید می‌کنند.

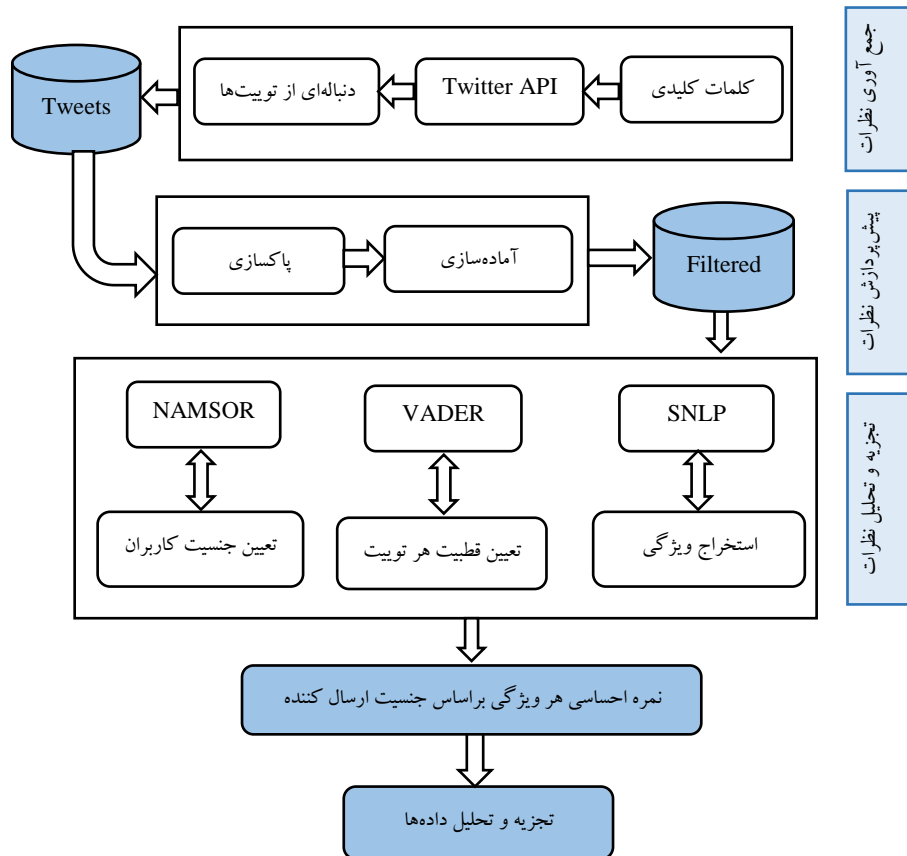
در سال ۲۰۱۰، مدلی با استفاده از روش ترکیبی مبتنی بر تکرار و مبتنی بر رابطه ارائه گردید. در این مدل از یک الگوریتم رتبه‌بندی برای ویژگی‌های استخراج شده استفاده می‌شود. این الگوریتم رتبه‌بندی بر اساس تعداد تکرار ویژگی‌ها و ارتباط آن‌ها عمل می‌کند. در این مدل از روش رتبه‌بندی تعریف شده برای صفحات وب استفاده می‌شود (ژانگ و همکاران^۱، ۲۰۱۰).

روش

این پژوهش از نظر هدف کاربردی است، زیرا برای حل مسئله‌ای عملی بکار می‌رود و از لحاظ روش‌شناسی، یک پژوهش توصیفی است. این پژوهش از چهار مرحله کلی برای رسیدن به نتیجه استفاده می‌کند: جمع‌آوری نظرات، پیش‌پردازش نظرات، تجزیه و تحلیل نظرات که خود شامل استخراج ویژگی نظرات، تعیین جنسیت کاربران ارسال‌کننده نظر و تعیین قطبیت نظرات است و تجزیه و تحلیل داده‌ها.

در شکل ۱ مراحل انجام کار و مدل پیشنهادی این پژوهش نمایش داده شده است.

شکل ۱. مدل پیشنهادی پژوهش



جمع‌آوری نظرات: در این پژوهش از شبکه اجتماعی توییت برای جمع‌آوری داده‌های موردنیاز استفاده شده است. جامعه پژوهش حاضر عبارت است از کلیه کاربران انگلیسی زبان که نظرات و احساسات خود را پیرامون دو محصول پرفروش تلفن همراه یعنی آیفون ایکس^۱ شرکت اپل و گلکسی اس ۹^۲ شرکت سامسونگ و ویژگی‌های آن در فاصله زمانی ۱۴ جون ۲۰۱۸ تا ۱۴ دسامبر ۲۰۱۸ به مدت شش ماه از طریق این شبکه اجتماعی با دیگر کاربران به اشتراک گذاشته‌اند. این مجموعه شامل ۱۵۸۵۳۲ توییت انگلیسی است که از

1. Iphone X
2. Galaxy S9

این تعداد، ۱۱۴۴۴۹ توییت مربوط به آیفون ایکس و ۴۴۰۸۳ توییت مربوط به گلکسی اس ۹ می‌باشد.

برای جمع‌آوری توییت‌ها از ابزاری بنام "Twitter API" استفاده می‌شود. همچنین برای فیلتر کردن، از کلمه کلیدی حاوی نام دو محصول مورد بررسی، ویژگی‌های مشترک هر دو محصول و هشتک‌های مرتبط با آن استفاده می‌شود. به‌طور مثال عبارت " iPhone X Camera" باعث می‌شود توییت‌هایی برگردانده شود که شامل هر دو کلمه " iPhone X" و "Camera" باشد که نتیجه آن، اظهار نظر در مورد عملکرد دوربین آیفون ایکس است. ویژگی‌های مورد جستجو عبارت‌اند از: دوربین^۱، باتری^۲، صفحه‌نمایش^۳، صدا^۴، سیستم‌عامل^۵، شناسایی صورت^۶، قیمت^۷، اندازه^۸، شارژر به یسیم^۹، دستیار صوتی^{۱۰} و پردازنده^{۱۱}.

پیش‌پردازش نظرات: برای پیش‌پردازش توییت‌ها ابتدا می‌بایست توییت‌های غیرضروری حذف گردیده و پاک‌سازی شوند. برای پاک‌سازی، از روش ۵ مرحله‌ای حذف توییت‌های تکراری و ریتوییت‌ها، حذف توییت‌های ایجاد شده توسط آژانس‌های تبلیغاتی و اینترنتی، حذف توییت‌های ایجاد شده توسط بخش خدمات پس از فروش و پشتیبانی شرکت، حذف توییت‌های حاوی کلمات تبلیغاتی و حذف توییت‌های حاوی لینک استفاده می‌شود. در پایان این مرحله، از مجموع کل ۱۵۸۵۳۲ توییت جمع‌آوری شده،

-
1. Camera
 2. Battery
 3. Screen
 4. Sound
 5. IOS & Android
 6. Face ID
 7. Price
 8. Size
 9. Wireless Charger
 10. Siri & Bixby
 11. processor

تعداد ۱۲۱۱۸ توییت به‌عنوان توییت معتبر و مفید شناخته می‌شود که از آن، ۹۷۶۰ توییت مربوط به آیفون ایکس و تعداد ۲۳۵۸ توییت مربوط به گلکسی اس ۹ است. سپس می‌بایست توییت‌ها به یک شکل استاندارد برای تجزیه و تحلیل تبدیل شوند.

آماده‌سازی اولیه: در این مرحله به برخی اصلاحات در متن توییت پرداخته می‌شود. مواردی همچون حذف کاراکترهای غیراسکی، حذف نام‌های کاربری که با علامت @ مشخص شده‌اند، حذف عبارت "RT" که مشخص‌کننده بازتوییت بودن متن توییت است، حذف هشتگ‌ها که با علامت # مشخص شده‌اند، حذف فاصله و خطوط جدید در ابتدا و انتهای متن و حذف چندین فاصله متوالی در متن توییت. به عنوان مثال توییت زیر را در نظر بگیرید:

Rt @darkxdom: @dylan_sann @yeisonjmv2 @fortnitegame my galaxy9 is better than any iphone ever #samsung #galaxys9

پس از عملیات آماده‌سازی اولیه توییت و حذف موارد غیر ضروری، توییت تصفیه شده به شکل زیر درمی‌آید:

my galaxy9 is better than any iphone ever

نرمال‌سازی: در فرآیند نرمال‌سازی، چندین عملیات مختلف قابل استفاده است تا یک متن به شکل یک متن استاندارد درآمده و در مراحل بعدی تجزیه و تحلیل از آن استفاده شود.

- جایگزینی اختصارات زبان انگلیسی با شکل کامل آن‌ها به طوری که به راحتی برای فرهنگ لغت قابل شناسایی و معنی دار باشد. بعضی مثال‌ها در جدول ۱ آورده شده است:

جدول ۱. جایگزینی اختصارات زبان انگلیسی با شکل کامل آن‌ها

Tweet	Tweet after replace contractions
just bought a 2017 ford fusion and can't get any apps to link from my galaxy9	just bought a 2017 ford fusion and cannot get any apps to link from my galaxy9

Tweet	Tweet after replace contractions
you've been a apple customer for 12 years	you have been a apple customer for 12 years
I'm making a pinhole camera out of a hollowed out iphonex.	I am making a pinhole camera out of a hollowed out iphonex.

- طولانی شدن یک کلمه زمانی اتفاق می‌افتد که کاربر برای تأکید بر احساسات بیشتر و یا اشتباهات تایپی، یک کاراکتر را مکرراً تکرار کند. کلمات در زبان انگلیسی حداکثر دارای دو کاراکتر تکراری و پشت سرهم می‌باشند، پس در مرحله نرمال‌سازی، نیاز است که کاراکترهای اضافی برداشته شوند. به‌عنوان مثال کلمه "finalllly" به "finally" و کلمه "coooooool" به "cool" تبدیل می‌شود.

اما فرایند اصلاح کلمات برای بعضی از کلمات به‌درستی کار نمی‌کند. به‌عنوان مثال برای کلمه "amazzzzzing"، فرآیند اصلاح آن را به "amazing" تبدیل می‌کند که از نظر املائی نادرست است. پس برای تصحیح خطاهای املائی از کتابخانه انگلیسی پای‌اسپل‌چکر^۱ و ماژول اسپل‌چکر^۲ آن در پایتون استفاده شده است.

استخراج ویژگی توییت: قبل از تعیین قطبیت نظرات، می‌بایست مشخص شود که هر نظر به کدام ویژگی مشترک تلفن همراه اشاره نموده و کاربر در مورد کدام ویژگی، احساسات خود را بیان کرده است. بر این اساس برای استخراج ویژگی موردنظر، توییت‌ها بر اساس رابطه گرامری بین کلمات و ویژگی و دیگر کلمات بررسی می‌شوند. اگر بین کلمات و ویژگی و دیگر کلمات بخصوص کلمات احساسی رابطه گرامری معناداری وجود داشته باشد، نشان‌دهنده وجود کلمه ویژگی در توییت و ابراز احساسات کاربر در مورد آن ویژگی است.

1. PySpellChecker
2. SpellChecker

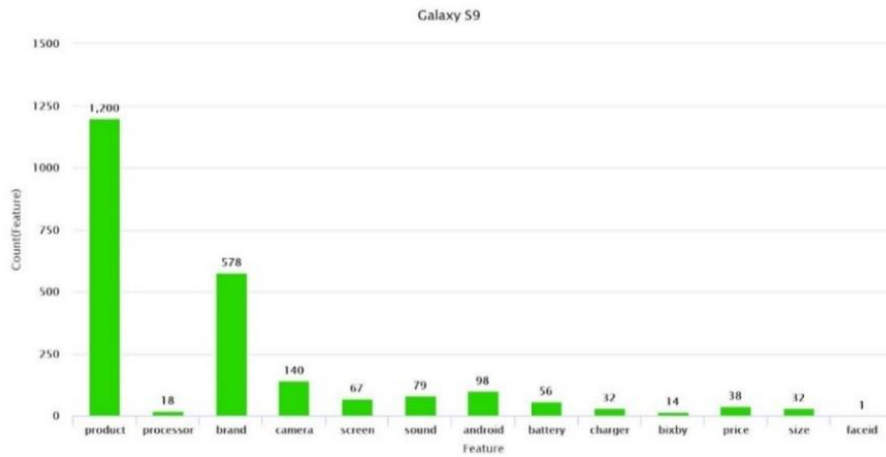
برای این منظور از ابزار پردازش زبان طبیعی استنفورد^۱ استفاده شده است. این ابزار دارای ۵۰ رابطه از پیش تعریف شده است که آن‌ها را وابستگی^۲ می‌نامند. در اینجا فقط از سه وابستگی `amod`، `nsubj` و `dobz` برای بیان رابطه گرامری بین کلمات ویژگی و دیگر کلمات استفاده می‌شود.

اگر یک توییت شامل یکی از ۳ وابستگی فوق باشد و این وابستگی‌ها، ارتباط گرامری بین یکی از ویژگی‌های مشترک تلفن همراه و دیگر کلمات توییت را بیان کنند، نشان‌دهنده این است که کاربر احساسات خود را در مورد آن ویژگی خاص از تلفن همراه بیان نموده است.

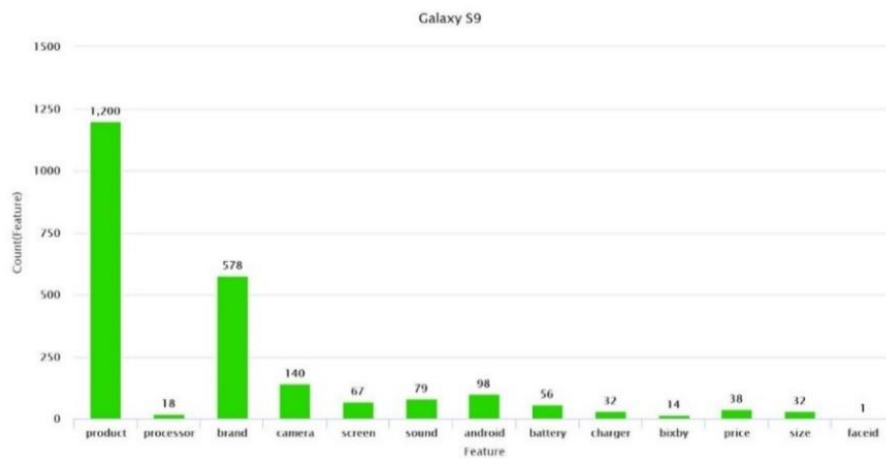
همچنین اگر وابستگی‌ها ارتباط گرامری بین کلمات "iphonex" یا "galaxys9" را با دیگر کلمات نشان دهند، ویژگی را با عنوان "محصول" و اگر با کلمات "apple" یا "samsung" باشد، با عنوان "برند" نشان‌گذاری می‌شوند. بدیهی است توییت‌هایی که حاوی بیش از یک وابستگی باشند، اولویت با وابستگی است که آرگومان‌های آن یکی از ویژگی‌های مشترک تلفن همراه باشد و نه "برند" و "محصول". تنها زمانی ویژگی یک توییت، "برند" یا "محصول" نشان‌گذاری می‌شود که هیچ‌کدام از دیگر ویژگی‌های تلفن همراه در آرگومان‌های وابستگی‌ها وجود نداشته باشد.

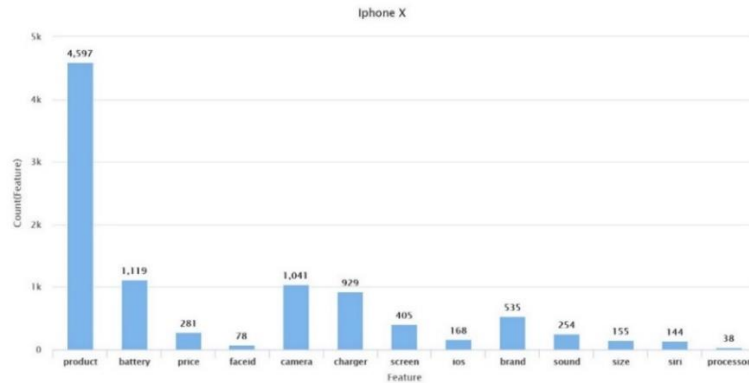
1. Stanford Natural Language Processing (SNLP)
2. Dependency

در



تعداد توییت‌ها به تفکیک هر ویژگی مشاهده می‌شود.

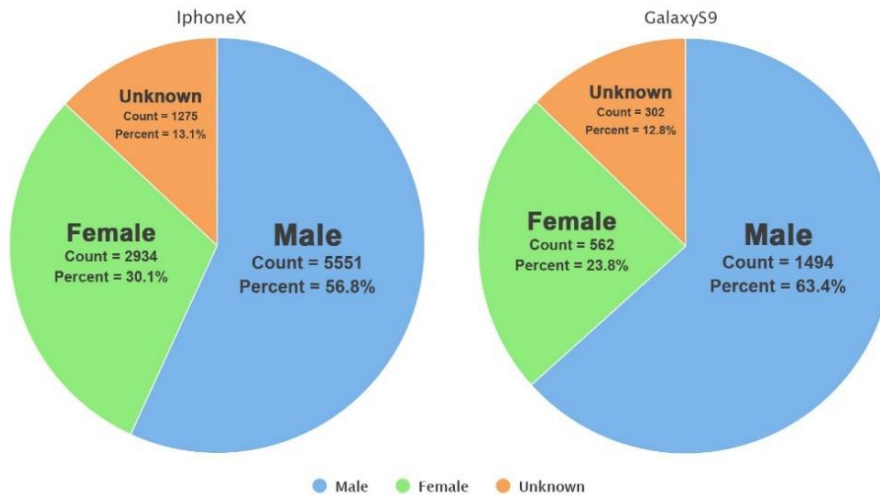




شکل ۲. تعداد توییت‌ها به تفکیک هر ویژگی

تعیین جنسیت کاربران: تویتر، در هنگام ساخت حساب کاربری، جنسیت را به‌عنوان یک پارامتر سؤال نمی‌کند، بنابراین جنسیت یک کاربر را نمی‌توان از داده‌های توییت آن به دست آورد. برای تعیین جنسیت کاربران، از یک ابزار هوش مصنوعی بنام نَمسور^۱ استفاده می‌شود. نَمسور، نام و نام‌خانوادگی کاربر را در فیله‌های جداگانه گرفته و با استفاده از پایگاه داده خود، جنسیت کاربر را تعیین می‌کند. نَمسور، برای بهبود دقت الگوریتم‌های زبان‌شناسی خود، از ترکیبی از الگوریتم‌های یادگیری ماشین، یادگیری تقویت‌شده و سایر تکنیک‌های هوش مصنوعی استفاده می‌کند و طبق ادعای سازندگان آن، دارای دقت ۹۶٪ است.

نتایج استفاده از این ابزار که شامل تعداد و درصد ارسال‌کنندگان توییت به تفکیک جنسیت برای هر کدام از محصولات است، در شکل ۳ قابل مشاهده است.



شکل ۳. جنسیت ارسال‌کنندگان توییت به تفکیک هر محصول

شکل ۳ نشان می‌دهد که جنسیت حدود ۱۳٪ از ارسال‌کنندگان توییت قابل تشخیص نبوده است. علت آن است که پر کردن فیلد نام و نام‌خانوادگی در پروفایل تویتر اجباری نبوده و کاربران می‌توانند نام خود را اعلام نکنند. همچنین بسیاری از کاربران بجای نام خود از اسامی مستعار، علامت‌ها، نشان‌ها و یا حروف زائد استفاده می‌کنند که این موضوع باعث عدم تشخیص جنسیت آنان می‌شود.

تعیین قطبیت توییت: پس از مرحله آماده‌سازی توییت‌ها، نیاز است تا قطبیت مثبت، منفی و خنثی توییت‌ها محاسبه گردیده و به شاخص‌های عددی تبدیل گردد. برای این کار، از کتابخانه پردازش زبان طبیعی و در^۱ در زبان برنامه‌نویسی پایتون استفاده می‌شود. کتابخانه و در^۱، یک ابزار تجزیه و تحلیل احساسات مبتنی بر قاعده و مبتنی بر لغت‌نامه احساسی^۲ است که به‌طور خاص برای بیان احساسات در رسانه‌های اجتماعی طراحی شده

1. Valence Aware Dictionary for Sentiment Reasoning (VADER)
2. Sentiment Lexicon

است. هوتو و گیلبرت^۱ در پژوهش خود در سال ۲۰۱۴، ضمن معرفی کتابخانه وِدر، آن را با ۷ لغت‌نامه دیگر مقایسه کرده‌اند. نتایج آن پژوهش نشان می‌دهد که دقت طبقه‌بندی وِدر با ۹۶٪ بسیار بهتر از دقت افراد انسانی مستقل با استفاده از دیگر لغت‌نامه‌ها با دقت طبقه‌بندی ۸۴٪ است (هوتو و گیلبرت^۲، ۲۰۱۴). علت این امر، عدم پوشش متناسب سایر لغت‌نامه‌ها با محتوای رسانه‌های اجتماعی است که اغلب منظور خود را با استفاده از شکلک‌ها، مخفف‌ها و اصطلاحات عامیانه و ابتدایی بیان می‌کنند.

وِدر مزایای زیادی نسبت به روش‌های سنتی تعیین قطبیت نظرات دارد. مزایایی همچون عدم نیاز به اطلاعات آموزشی، پردازش سریع جریان داده‌های برخط و تعیین نمره احساسی نظرات بر اساس نوع علامت‌گذاری‌ها، حروف بزرگ و کوچک، تشدیدکننده‌ها، کاهنده‌ها، منفی‌کننده‌ها، شکلک‌ها، مخفف‌ها و ایموجی‌ها.

برای به دست آوردن نمره احساسی هر توییت از تابع "Polarity_Scores()" استفاده می‌شود. خروجی این تابع شامل میزان قطبیت مثبت^۳، میزان قطبیت منفی^۴، میزان قطبیت خنثی^۵ و میزان قطبیت ترکیبی^۶ برای هر توییت است. میزان قطبیت ترکیبی، یک معیار هم وزن شده دقیق است که جمع امتیاز قطبیت کلمات را محاسبه می‌کند و بین -۱ تا +۱ نرمال شده است. اعداد مثبت بزرگ‌تر از ۰/۰۵ نشان‌دهنده احساسات مثبت، کوچک‌تر از ۰/۰۵ و بزرگ‌تر از -۰/۰۵ نشان‌دهنده احساسات خنثی و اعداد کوچک‌تر از -۰/۰۵ نشان‌دهنده احساسات منفی است.

تجزیه و تحلیل داده‌ها

-
1. Hutto and Gilbert
 2. Hutto and Gilbert
 3. Positive Score
 4. Negative Score
 5. Neutral Score
 6. Compound Score

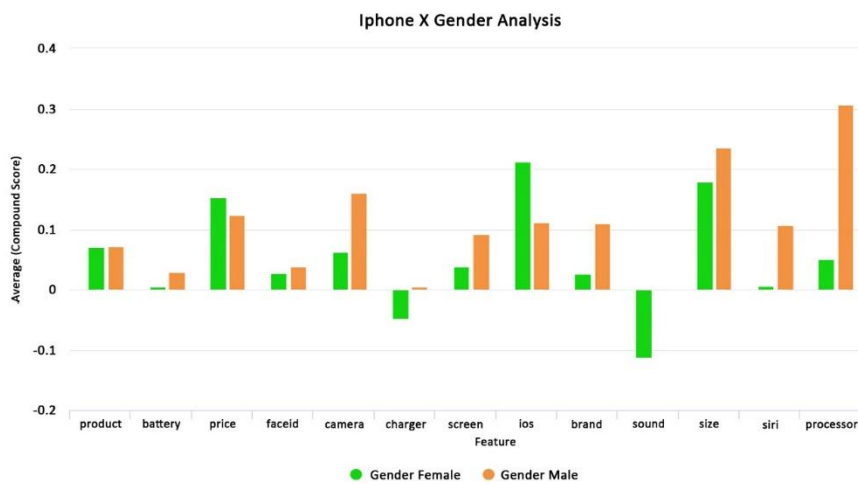
برای تجزیه و تحلیل داده‌ها، در ابتدا از دو پارامتر حجم توییت در هر روز و میانگین نمره ترکیبی احساسات در هر روز استفاده می‌شود. میانگین نمره ترکیبی احساسات در هر روز برابر است با جمع نمره ترکیبی توییت‌های هر روز بر تعداد توییت‌های جمع‌آوری شده در آن روز. حجم کلی توییت‌ها و روند احساسات کاربران در بازه زمانی شش‌ماهه را می‌توان در شکل ۴ مشاهده کرد.



شکل ۴. حجم و روند احساسات توییت‌های آیفون ایکس و گلکسی اس ۹

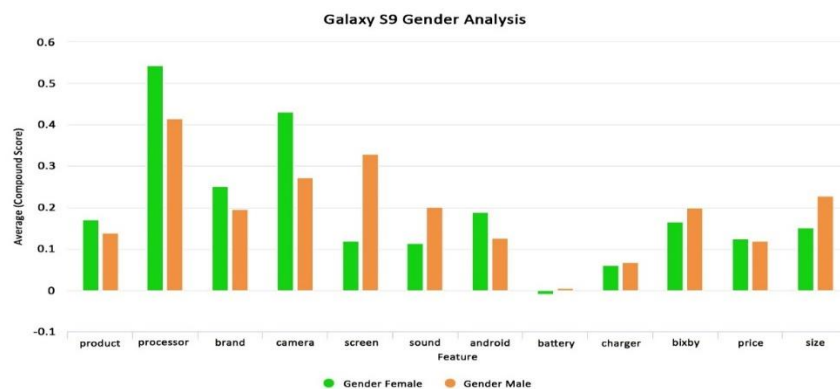
در ادامه به تجزیه و تحلیل داده‌ها بر اساس جنسیت کاربران ارسال‌کننده توییت پرداخته شده است. در این مرحله توییت‌هایی که دارای جنسیت ناشناخته^۱ هستند، مورد بررسی قرار نمی‌گیرند.

شکل ۵ نشان‌دهنده احساسات کاربران نسبت به ویژگی‌های مختلف آیفون ایکس است که به تفکیک جنسیت کاربران، به رنگ‌های نارنجی و سبز نمایش داده شده است. همان‌طور که در شکل ۵ نشان داده می‌شود، کاربران مرد احساسات مثبت بیشتری به ویژگی‌های "پردازنده"، "اندازه" و "دوربین" به ترتیب با شدت ۰/۳۱، ۰/۲۴ و ۰/۱۶ داشته و کاربران زن نسبت به ویژگی‌های "سیستم‌عامل"، "اندازه" و "قیمت" احساسات مثبت بیشتری به ترتیب با شدت ۰/۲۱، ۰/۱۸ و ۰/۱۵ دارند. ضمن اینکه کاربران زن احساسات منفی خود در مورد "صدا" و "شارژر" را به ترتیب با شدت ۰/۱۱- و ۰/۰۴- بیشتر از مردان به اشتراک گذاشته‌اند که می‌تواند بیانگر اهمیت بیشتر این دو ویژگی برای کاربران زن باشد.



شکل ۵. احساسات کاربران در مورد ویژگی‌های مختلف آیفون ایکس به تفکیک جنسیت آنان

در شکل ۶، ویژگی‌های "پردازنده" و "صفحه‌نمایش" گلکسی اس ۹ به ترتیب با شدت ۰/۴۱ و ۰/۳۲ برای مردان از اهمیت بالاتری برخوردار بوده و امتیاز بالاتری را کسب کرده است. همچنین احساسات کاربران زن با شدت ۰/۱۲ برای ویژگی "صفحه‌نمایش" در مقایسه با کاربران مرد، نشان از اختلاف معنادار این دو گروه در معیارهای انتخاب تلفن همراه خود است. نکته قابل توجه، هماهنگی هر دو گروه مردان و زنان در دادن نمره احساسی تقریباً صفر به ویژگی "باتری" گلکسی اس ۹ است.



شکل ۶. احساسات کاربران در مورد ویژگی‌های مختلف گلکسی اس ۹ به تفکیک جنسیت آنان

نتیجه‌گیری و پیشنهادها

تجزیه و تحلیل احساسات می‌تواند به عنوان مقیاسی برای ارزیابی یک محصول و یا محصول رقبا بر اساس معیارهای مختلفی از جمله جنسیت مصرف کنندگان در بازار مورد استفاده قرار گیرد.

در این پژوهش، ضمن جمع‌آوری نظرات کاربران در مورد دو محصول پرتعدادار تلفن همراه یعنی آیفون ایکس شرکت اپل و گلکسی اس ۹ شرکت سامسونگ، یک مدل ۳ مرحله‌ای از تجزیه و تحلیل احساسات کاربران در سطح ویژگی معرفی گردید. در مرحله چهارم، با به دست آوردن نمرات احساسی هر ویژگی به تفکیک جنسیت کاربران ارسال کننده توییت، به مقایسه هر ویژگی از تلفن همراه با ویژگی متناظر آن در محصول

رقیب پرداخته شد. همچنین در مرحله پاک‌سازی نظرات، از یک مدل ۵ مرحله‌ای برای پاک‌سازی توییت‌ها استفاده شد که منجر به حذف ۸۳/۶۲٪ توییت‌ها گردید که غیرمفید، تبلیغاتی و فاقد احساس بودند.

نتایج این پژوهش بیان می‌دارد که کاربران مرد احساسات مثبت بیشتری نسبت به خصوصیات سخت‌افزاری آیفون ایکس همچون "دوربین"، "پردازنده" و "اندازه" داشته و کاربران زن نسبت به خصوصیات نرم‌افزاری همچون "سیستم‌عامل"، "اندازه" و "قیمت" احساسات مثبت بیشتری دارند. این موضوع نشان‌دهنده توجه بیشتر کاربران مرد به خصوصیات سخت‌افزاری و توجه بیشتر کاربران زن به خصوصیات نرم‌افزاری آیفون ایکس است. همچنین نمره احساسی تقریباً صفر به ویژگی "باتری" گلکسی اس ۹ نشان‌دهنده نارضایتی هر دو گروه کاربران مرد و زن نسبت به باتری تلفن همراه خود است که لزوم توجه شرکت سامسونگ به طول عمر باتری و مدت‌زمان نگهداری شارژ تلفن همراه را بیان می‌دارد.

صاحبان کسب‌وکار با دانستن اینکه مشتریان مرد و مشتریان زن خود و حتی رقیب به کدام ویژگی محصول علاقه بیشتری داشته و از کدام ویژگی رضایت ندارند، می‌توانند اقدام به تولید محصولاتی با تمرکز بر جنسیت افراد کرده و یا به طراحی برنامه‌های هوشمند تبلیغاتی یا کمپین‌های بازاریابی با توجه به ویژگی‌های موردعلاقه آنان نمایند. این اقدامات در نهایت به افزایش سوددهی کسب‌وکار و رضایت‌مندی مشتریان منجر می‌شود.

با توجه به اینکه پژوهش حاضر بر روی داده‌های جمع‌آوری شده از توییتر انجام گرفته است، این شبکه اجتماعی تنها اجازه دسترسی به ۱٪ از داده‌ها آن‌هم در بازه زمانی ۱۰ روزه را می‌دهد که باعث می‌شود دسترسی به داده‌ها محدود شود. این محدودیت، منجر به عدم دسترسی به تمامی توییت‌ها و کاهش تعداد توییت‌های جمع‌آوری شده می‌شود. همچنین عدم وجود فیلد جنسیت برای کاربران توییتر، نیاز به استفاده از ابزار هوش مصنوعی نامسور را الزامی می‌کند. اگرچه این ابزار در طبقه‌بندی جنسیتی بسیار دقیق است، اما هنوز هم

دارای خطاست و در برخی موارد به علت عدم ارائه درست نام و نام خانوادگی از جانب کاربران، در تشخیص جنسیت آنان ناتوان است. همچنین این پژوهش از یک روش بدون نظارت (مبتنی بر لغت‌نامه) برای تعیین قطبیت نظرات استفاده می‌کند. برای جلوگیری از برخی مشکلات احتمالی، روش نظارت‌شده ممکن است نتایج بهتری ایجاد کند. پیشنهاد می‌شود در پژوهش‌های آتی، با استفاده از ترکیب روش‌های نظارت‌شده و بدون نظارت، برای دستیابی به نتایج بهتر تلاش کرد.

ORCID

Shahriar Mohammadi



<https://orcid.org/0000-0002-3696-1984>

Eslam Nazemi



<https://orcid.org/0000-0001-9084-3789>

منابع

شاه طالبی، ن.، کارگر، م.، میرزایی، ک. (۱۳۹۵). بررسی مدل‌های نظرکاوی و تجزیه و تحلیل احساسات کاربران در محیط وب. *دومین کنفرانس بین‌المللی وب پژوهی*. تهران: دانشگاه علم و فرهنگ.

- Abdous, M., & He, W. (2011). Using text mining to uncover students' technology-related problems in live video streaming. *British Journal of Educational Technology*, 40(5), 40-49.
- Ansari, M., Aziz, M., Siddiqui, M., Mehra, H., & Singh, K. (2020). Analysis of Political Sentiment Orientations on Twitter. *Procedia Computer Science*, 167, 1821-1828.
- Basari, A. S., Hussin, B., Ananta, G. B., & Zeniarja, J. (2013). Opinion Mining of Movie Review using Hybrid Method of Support Vector Machine and Particle Swarm Optimization. *Procedia Engineering*, 53, 453-462.
- Belkaroui, R., & Faiz, R. (2015). Towards events tweet contextualization using social in uence model and users conversations. *In Proceedings of the 5th International Conference on Web Intelligence, Mining and Semantics*, page3.ACM.
- Cambria, E., Schuller, B., Xia, Y., & Havasi, C. (2013). New avenues in opinion mining and sentiment analysis. *IEEE Intelligent Systems*, 28(2), 15-21. doi:0.1109/MIS.2013.30.
- Global Web Index. (2020). *Social H3 2020 report*. Global Web Index: <https://www.globalwebindex.com/reports/social>
- He, W., Shen, J., Li, Y., Akula, V., Yan, G., & Tao, R. (2015). Gaining competitive intelligence from social media data: Evidence from two largest retail chains in the world. *Industrial Management & Data Systems*, 115(9), 1622-1636.
- He, W., Zha, S., & Li, L. (2013). Social media competitive analysis and text mining: A case study in the pizza industry. *International Journal of Information Management*, 33, 464-472.
- Hu, M., & Liu, B. (2004). Mining opinion features in customer review. *In processing of 19th National Conference on Artificial intelligence*, AAAI Press, (. 755-760).
- Hung, C., & Lin, H. K. (2013). Using Objective Words in SentiWordNet to Improve Word-of-Mouth Sentiment Classification. *IEEE Intelligent Systems*, 28(2), 47-54.
- Hutto, C. J., & Gilbert, E. (2014). VADER: A Parsimonious Rule-Based Model for Sentiment Analysis of Social Media Text. *AAAI*

Publications, Eighth International AAAI Conference on Weblogs and Social Media.

- Ibrahim, N., & Wang, X. (2019). Decoding the sentiment dynamics of online retailing customers: Time series analysis of social media. *Computers in Human Behavior*(96), 32-45.
- Li, Z. C., Zhang, M., MA, S. P., Zhou, B., & Sun, Y. (2009). Automatic extraction for product feature words from comments on the web. In: *G.G. Lee (Ed.), Information Retrieval Technology, Springer-Verlag Berlin, Berlin*, 112-123.
- Liang, P. W., & Bi-Ru, D. (2013). Opinion Mining on Social Media Data. *2013 IEEE 14th International Conference on Mobile Data Management*. Milan, Italy: IEEE. doi:10.1109/MDM.2013.73
- Martinez, I., Valencia-Garcia, R., & Garcia-Sanchez, F. (2011). ntology-Guided Approach to Feature-Based Opinion Mining. *Natural Language Processing and Information Systems*, 193-200.
- Mikula, M., & Machova, K. (2015). Classification of opinions in conversational content. *IEEE 13th International Symposium on Applied Machine Intelligence and Informatics*, (. 22-24).
- Othman, R., Belkaroui, R., & Faiz, R. (2017). Extracting Product Features for Opinion Mining Using Public Conversations in Twitter. *International Conference on Knowledge Based and Intelligent Information and Engineering Systems, KES2017, Procedia Computer Science 112 (2017)*, (. 927-935).
- Pang, B., & Lee, L. (2002). Thumbs up ? Sentiment Classification using Machine Learning Techniques. In *ACL-02 conference on Empirical methods in natural language processing*, (79-86). doi:10.3115/1118693.1118704.
- Pearson, T., & Wegener, R. (2013). *Big data: the organizational challenge*. September 11, 2013 , www.bain.com/publications/articles/big_data_the_organizational_challenge.aspx
- Piryani, R., Madhavi, D., & Singh, V. K. (2016). Analytical mapping of opinion mining and sentiment analysis research during 2000–2015. *Information Processing and Management*, 53(1), 122-150.
- Vinodhini, G., & Chandrasekaran, R. M. (2013). Sentiment Mining Using SVM-Based Hybrid Classification Model. *Computational Intelligence, Cyber Security and Computational Models* (. 155-162). New Delhi: Springer. doi:10.1007/978-81-322-1680-3_18
- Wang, H., Gao, S., Yin, P., & Nga-Kwok, J. (2017). Competitiveness analysis through comparative relation mining: evidence from restaurant` online review. *Industrial Management & Data Systems*, 117(4), 672-687.

- Ye, Q., Zhang, Z., & Law, R. (2009). Sentiment Classification of Online Review to Travel destinations by supervised machine learning approaches. *Expert Systems with Applications*, 36(3), 267-307.
- Zhang, L., Biu, B., Lim, S. H., & O'Brien-Strain, E. (2010). Extracting and ranking product features in opinion documents. *In Proceedings of the 23rd International Conference on Computational Linguistics: Posters, COLING' 10, Stroudsburg, PA, USA*, (1462–1470).
- Zhuang, L., Jing, F., & Zhu, X. (2006). Movie review mining and summarization. *In processing of the ACM International Conference on Information and Knowledge Management (CIKM-2006)*, (. 43-50).

استناد به این مقاله: محمدی، شهریار، ناظمی، اسلام. (۱۴۰۰). تجزیه و تحلیل احساسات در سطح ویژگی محصول و مبتنی بر جنسیت کاربران، مطالعات مدیریت کسب و کار هوشمند، ۱۰(۳۷)، ۲۶۷-۲۹۶.

DOI: 10.22054/IMS.2021.52110.1723



Journal of Business Intelligence Management Studies is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License..

