

Sentiment Analysis of People's opinion about Iranian National Cars with BERT

Leila Gonbadi¹, Niloofar Ranjbar^{2*}

1. Master Student, Department of Computer Engineering, Persian Gulf University, Bushehr, Iran.
2. Instructor, Department of Computer Engineering, Persian Gulf University, Bushehr, Iran. (*Corresponding Author*), Email: nranjbar@pgu.ac.ir

Abstract:

Introduction: With the development of the internet and social media, people have been actively discussing political and economic issues, and sharing their opinions online. The vast amount of data generated from these online discussions can be analyzed through text mining methods to extract valuable information. One such method is aspect-based sentiment analysis, which allows for the analysis of people's opinions on various aspects of a topic. In this paper, we will focus on the analysis of people's opinions on the Iranian automobile industry and national cars using aspect-based sentiment analysis.

Methodology: To achieve our goal, we first introduce a method for extracting different aspects related to national cars. We then use the "BERT" language model to extract vectors for different sentences related to the various aspects and finally use a neural network to classify the sentiments expressed in these sentences as positive, negative, or neutral.

Results: The analysis of public opinions on various aspects of Iranian cars showed that the most discussed aspects were design, quality, and price. The sentiment expressed towards design was largely positive, with people expressing admiration for the unique and modern designs of Iranian cars. The sentiment expressed towards quality was mixed, with some people praising the improved quality of national cars, while others criticized the use of low-quality materials. The sentiment expressed towards price was largely negative, with people complaining about the high prices of Iranian cars compared to their foreign counterparts.

Discussion: The results of our analysis provide valuable insights into the level of public satisfaction with various aspects of Iranian cars. The mixed sentiment expressed towards the quality of Iranian cars highlights the need for manufacturers to focus on using high-quality materials to improve the overall quality of their products. The negative sentiment expressed toward the price of Iranian cars suggests that car manufacturers need to find ways to reduce production costs and offer more competitive pricing. In conclusion, aspect-based sentiment analysis can be used as an effective method to analyze public opinions on various aspects of a topic. Our analysis of public opinions on the Iranian automobile industry and national cars provides valuable insights for car manufacturers to improve the design, quality, and pricing of their products. By taking these insights into account, manufacturers can improve their performance and meet the needs and expectations of their customers.

Keywords: Iranian National Cars, BERT, Aspect-base Sentiment Analysis, Text mining, Machine Learning.

تحلیل احساس نظرات مردم در مورد خودروهای ملی ایران به کمک مدل زبانی BERT

دوره سوم، زمستان ۱۴۰۱
شماره چهارم، صص: ۵۱-۶۰

تاریخ دریافت: ۱۴۰۱/۰۷/۱۸
تاریخ پذیرش: ۱۴۰۱/۰۸/۲۶

لیلا گنبدی^۱، نیلوفر رنجبر^{۲*}

۱- دانشجوی کارشناسی ارشد، گروه مهندسی کامپیوتر، دانشگاه خلیج فارس، بوشهر، ایران. lgonbadi@gmail.com
۲- مربی، گروه مهندسی کامپیوتر، دانشگاه خلیج فارس، بوشهر، ایران. (نویسنده مسئول) nranjbar@pgu.ac.ir

چکیده: امروزه با پیشرفت شبکه جهانی اینترنت و توسعه شبکه‌های اجتماعی، افراد در مورد موضوعات مختلف سیاسی و اقتصادی به تبادل نظر و گفت‌وگو می‌پردازند. با استفاده از روش‌های متن‌کاوی می‌توان به استخراج اطلاعات از این نظرات پرداخت. یکی از روش‌های تحلیل نظرات، تحلیل نظرات مبتنی بر جنبه است که به کمک آن می‌توان نظرات مردم را در مورد جنبه‌های مختلف یک موضوع تحلیل و بررسی کرد. تحلیل نظرات مردم در مورد صنعت خودروسازی و خودروهای ملی یکی از کاربردهای مهم استفاده از روش‌های متن‌کاوی در موضوعات اقتصادی است. این کار باعث مطلع شدن مدیران صنعت خودروسازی از میزان رضایت واقعی مردم از خودروهای ملی است. برای دست‌یافتن به این مهم روش‌های مختلفی ارائه شده است. در این مقاله ما به معرفی روشی برای استخراج جنبه‌های مختلف مربوط به خودروهای ملی می‌پردازیم، سپس به کمک مدل زبانی «برت» برای جملات مختلف، بردار استخراج کرده و پس از آن به کمک شبکه عصبی به دسته‌بندی آن‌ها می‌پردازیم. سپس با داشتن جنبه‌های مختلف مهم برای خودروها و این که نظر مطرح‌شده به کدام یک از این جنبه‌ها می‌پردازد و حس مطرح شده در نظر مثبت، منفی یا خنثی است، به تحلیل میزان رضایت مردم از جنبه‌های مختلف خودروهای ایرانی می‌پردازیم. هدف نهایی این مقاله تحلیل نظرات مردمی در مورد انواع خودروهای ملی و اجزای آن با هدف بهبود عملکرد خودروسازان است.

واژه‌های کلیدی: خودروهای ملی ایران، برت، تحلیل احساس، متن‌کاوی، یادگیری ماشین.

۱. مقدمه

یکی از مهم‌ترین صنعت‌های ملی، صنعت خودروسازی است. بر همین اساس بهبود عملکرد این صنعت متناسب با نیاز مصرف‌کننده نیز بسیار مهم است. هدف ما تحلیل نظر و دیدگاه‌های مردم درباره خودروهایی داخلی است. در این مقاله قصد داریم روشی ارائه دهیم که به کمک آن می‌توان با یافتن نظرات مردم در مورد جنبه‌های مختلف خودروهایی داخلی، به خودروسازان کمک کرد.

برای این کار متن گفت‌وگوهای گروه‌های مختلف خودرویی در شبکه اجتماعی تلگرام را مورد بررسی قرار داده‌ایم. افراد مختلفی در این گروه‌های تلگرامی عضو بوده‌اند که می‌توان انتظار داشت که اغلب نظرات واقعی و بی‌طرفانه باشند.

پس از استخراج متن گفت‌وگوها، ابتدا سعی شده است به کمک روش‌های نیمه خودکار، جنبه‌های مختلف مربوط به خودروها و همچنین سلسله‌مراتب این جنبه‌ها شناسایی شود. برای این کار ابتدا پیش‌پردازش‌هایی شامل حذف علائم نگارشی، لینک‌ها و اعداد از متن‌ها انجام شده است. پس از آن برخی از کلمات بسیار پرتکرار و همچنین کلمات بدون معنی و بسیار خاص حذف شده‌اند. سپس به کمک قطعه‌بندی^۱ و واژگان تعریف‌شده برای مدل‌زبانی «پارس برت»^[1] جملات قطعه‌بندی شده و کلماتی که در لیست واژگان این مدل‌زبانی موجود نبودند، حذف شده‌اند. همچنین به کمک این مدل‌زبانی برای هر جمله یک بردار ساخته شده است. پس از آن با استفاده از روش خوشه‌بندی «k-means» و با در نظر گرفتن این بردارها، جملات خوشه‌بندی شده‌اند. سپس به کمک مدل «Latent Dirichlet Allocation (LDA)»^[2] که یک روش استخراج موضوع از متن است، کلمات کلیدی و موضوعات مهم خوشه‌های مختلف شناسایی شده‌اند. در نهایت با بررسی دستی کلمات کلیدی، مواردی که به جنبه‌هایی از خودروها اشاره داشتند، شناسایی شده‌اند و سلسله‌مراتب آن‌ها مشخص شده است. با این کار ۱۴ جنبه کلی که هر کدام شامل چند زیرمجموعه هستند، مشخص شده‌اند. سپس به کمک مدل از قبل آموزش داده شده برای تحلیل احساسات متن‌های فارسی^۳ که از «پارس برت» کمک می‌گیرد، حس هر جمله شناسایی شده است. جملاتی که حس خنثی داشتند، حذف شده و تنها جملات مثبت و منفی باقی‌ماندند. در نهایت با تحلیل این که در هر جمله به چه جنبه‌هایی از خودرو و با چه حسی اشاره شده است، به بررسی نظرات مردم پرداختیم.

لازم به ذکر است کلیه کدها و فایل‌ها و نتایج مربوط به این پژوهش در صفحه گیت‌هاب^۴ موجود است.

۲. پیشینه پژوهش

تحلیل احساسات یکی از کاربردهای متن‌کاوی می‌باشد که مورد توجه بسیاری از متخصصان علوم داده قرار گرفته است. تاکنون کارهای بسیاری در زمینه بررسی نظرات مردم در زمینه‌های مختلف سیاسی،

اقتصادی و اجتماعی به کمک تحلیل احساسات داده‌های تولید شده از طریق شبکه‌های اجتماعی انجام شده است. برای مثال پای و همکاران^[3] به کمک مدل‌های رگرسیون چندمتغیره، تحلیل نظرات در شبکه‌های اجتماعی و مدل‌های سری زمانی، سعی داشته‌اند فروش وسایل نقلیه به صورت ماهانه در ایالات متحده آمریکا را پیش‌بینی کنند. علی و همکاران^[4] به کمک تعبیه‌های زبانی و مدل‌کردن موضوعات و داده‌های شبکه‌های اجتماعی سعی کرده‌اند میزان رضایت مردم از سیستم‌های حمل‌ونقل را بررسی کنند. در برخی از پژوهش‌های اخیر از شبکه‌های عصبی عمیق مانند «CNN» و «RNN» و ترنسفورمرها و همچنین مدل‌زبانی «برت» برای تحلیل احساسات نظرات مردم کمک گرفته شده است. برای مثال دانگ و همکاران^[5] به کمک «CNN» و «RNN» سعی کرده‌اند نظراتی را که مردم در مورد فروش کالاهای مختلف به صورت آنلاین منتشر می‌کنند، مورد بررسی قرار دهند. در این زمینه کار مشابهی توسط ونگ و همکاران^[6] انجام شده است که از مدل‌زبانی «آلبرت»^۵ استفاده کرده‌اند. علاوه بر موارد نام‌برده شده، پژوهش‌های دیگری نیز در زمینه تحلیل نظرات در شبکه‌های اجتماعی توییتر و تلگرام انجام شده است^[7-12].

یکی از پژوهش‌هایی که در زمینه تحلیل احساسات مبتنی بر جنبه انجام شده است، پژوهش ارائه شده توسط محسنی و همکاران^[13] است. در این پژوهش تحلیل احساسات روی نظرات سایت تریپ‌ادوایزر درباره رستوران‌های ایرانی انجام شده است. در پژوهش ارائه شده توسط خلج و همکاران^[14] به بررسی اهمیت نظرات کاربران در صنعت گردشگری و هتل‌داری در بستر اینترنت و امکان استفاده از روش‌های متن‌کاوی برای کشف عقاید دیگران پرداخته شده است. در این راستا، روش ترکیبی جدیدی برای تحلیل احساسات و طبقه‌بندی بار احساسی نظرات با استفاده از واژگان با فرهنگ واژگان احساس عمومی و استاندارد لیو بینگ ارائه شده است. این واژگان با استفاده از روش‌های آماری و الگوریتم ژنتیکی ساخته شده‌اند تا دقت طبقه‌بندی افزایش یابد. در پژوهش دیگری، عباسی و همکاران^[15] است، مدلی جهت تحلیل ادراکات کاربران در خرید تلفن همراه از سایت دیجی کالا ارائه کرده‌اند.

از پژوهش‌هایی که روی داده‌های شبکه اجتماعی تلگرام انجام شده است، پژوهش زارع مهرجردی و همکاران^[16] است که در آن به بررسی تجزیه و تحلیل احساسات بر روی داده‌های تلگرام فارسی پرداخته شده است. برای این منظور چند روش استخراج ویژگی و روش‌های یادگیری ماشین کلاسیک و یادگیری عمیق بررسی شده و نتایج ارزیابی شده‌اند. بهترین کارایی با استفاده از روش استخراج ویژگی ماتریس تعبیه کلمات به همراه شبکه حافظه طولانی کوتاه مدت دوطرفه^۶ به دست آمده است. پژوهش طباطبایی و همکاران^[17] نیز پژوهش دیگری بر اساس داده‌های شبکه‌های اجتماعی است. این تحقیق با هدف تحلیل احساسات در حوزه داروسازی انجام شده و در آن دو شبکه عصبی عمیق کانولوشنی و شبکه عصبی بازگشتی GRU

یک معماری واحد ترکیب شده‌اند. استفاده از ترکیب مناسب لایه‌های این دو شبکه در یک معماری عمیق و همچنین استفاده از شبکه تعبیه کلمات پیش آموزش دیده «فست تکست» باعث شده‌است که معماری پیشنهادی نسبت به معماری‌های ترکیبی دیگر به نتایج موفق‌تری دست‌یابد.

در نهایت نزدیک‌ترین کار، مقاله لیو و همکاران [18] است. در این مقاله، تحلیل نظرات کاربران در مورد خودروها به کمک آموزش خصمانه و همچنین مدل «برت» انجام گرفته است. تفاوت اصلی این مقاله یادشده با روش ارائه‌شده در مقاله حاضر این است که ما سعی کردیم جنبه‌های مختلف خودروها نیز در نظر گرفته‌شود و مشخص‌شود نظرات مثبت یا منفی مردم روی کدام یک از جنبه‌های خودروها بوده‌است.

در این بخش به شرح روش ارائه شده می‌پردازیم. ابتدا نظرات مردم در مورد خودروهای ملی از طریق گروه‌های تلگرامی مربوط به هر خودرو استخراج می‌شود. لینک این گروه‌ها در بخش پیوسته‌ها موجود است. پس از استخراج متن از گروه‌های مرتبط در تلگرام، ابتدا پیش‌پردازش‌هایی روی آن‌ها صورت می‌گیرد تا آماده تحلیل احساسات در بخش‌های مختلف شود. این پیش‌پردازش‌ها شامل حذف جملات اضافی، حذف حروف اضافی، حذف کلمات اضافی است. سپس به منظور مشخص کردن جنبه‌های مختلفی که در خودروها اهمیت دارند، کلمات کلیدی نظرات استخراج می‌شوند. سپس جملات موجود در نظرات خوشه‌بندی می‌شوند تا جنبه‌های کلی و جزئی که در خوشه‌ها به صورت پرتکرار وجود داشته، شناسایی شوند. برای مثال یک خوشه از جملات ممکن است به‌طور کلی به بخش‌های فنی خودرو بپردازد و خوشه دیگر به بخش‌های ظاهری خودرو. با این کار و به کمک استخراج کلمات کلیدی هر خوشه می‌توان جنبه‌های مختلفی را که در هر خودرو اهمیت دارند و کاربران به آن‌ها اشاره کرده‌اند، استخراج کرد. پس از خوشه‌بندی، تحلیل احساسات جملات به کمک مدل «برت» انجام می‌شود. و در نهایت مشخص می‌شود چند درصد از مردم نظر مثبت یا منفی راجع به هر یک از جنبه‌های مختلف خودروها داشته‌اند.

۳. روش پیشنهادی

در این بخش به شرح روش ارائه شده می‌پردازیم. ابتدا نظرات مردم در مورد خودروهای ملی از طریق گروه‌های تلگرامی مربوط به هر خودرو استخراج می‌شود. پس از استخراج متن از گروه‌های مرتبط در تلگرام، پیش‌پردازش‌هایی روی آن‌ها صورت می‌گیرد تا آماده تحلیل احساسات در بخش‌های مختلف شود. این پیش‌پردازش‌ها شامل حذف جملات اضافی، حذف حروف اضافی، حذف کلمات اضافی است. سپس به منظور مشخص کردن جنبه‌های مختلفی که در خودروها اهمیت دارند، کلمات کلیدی نظرات استخراج می‌شوند. سپس جملات موجود در نظرات خوشه‌بندی می‌شوند تا جنبه‌های کلی و جزئی که در خوشه‌ها به صورت پرتکرار وجود داشته‌اند، شناسایی شوند. برای مثال یک خوشه از

جملات ممکن است به‌طور کلی به بخش‌های فنی خودرو بپردازد و خوشه دیگر به بخش‌های ظاهری خودرو. با این کار و به کمک استخراج کلمات کلیدی هر خوشه می‌توان جنبه‌های مختلفی که در هر خودرو اهمیت دارند و کاربران به آن‌ها اشاره کرده‌اند را استخراج کرد. پس از این خوشه‌بندی نوبت به تحلیل احساسات جملات می‌رسد که به کمک مدل «برت» انجام می‌شود. و در نهایت مشخص می‌شود چند درصد از مردم نظر مثبت یا منفی راجع به هر یک از جنبه‌های مختلف خودروها داشته‌اند. در ادامه ابتدا به صورت جزئی پیش‌پردازش‌های انجام شده را شرح می‌دهیم، سپس به روش به دست آوردن کلمات کلیدی و خوشه‌بندی جملات می‌پردازیم و در نهایت مدل نهایی را شرح می‌دهیم.

۳-۱- پیش‌پردازش‌ها

در این بخش پیش‌پردازش‌های انجام‌شده را شرح خواهیم داد:

- **حذف جملات اضافی:** با حذف علامت سوال جملات سوالی این جملات در ظاهر تبدیل به یک جمله خبری می‌شوند و سیستم را دچار اشتباه می‌کنند. مثلاً دو جمله «این ماشین خوب است؟» و «این ماشین خوب است.» را در نظر بگیرید. با حذف علامت سوال و نقطه هر دو جمله یکسان خواهند شد، در صورتی که محتوا و هدف آن‌ها یکسان نیست. در این پژوهش هدف ما تحلیل نظر بوده و سوال جزو نظر محسوب نمی‌شود، لذا جملات سوالی را حذف می‌کنیم.
- **حذف حروف اضافی:** برای یکدست شدن متن و سادگی پردازش، لازم است حروف اضافی حذف شوند. علاوه بر آن وجود این حروف کمکی در ادامه راه به ما نمی‌کند. حروف اضافی شامل « / . > < [] @ * + = ! » است. علاوه بر آن چون پردازش‌ها و پژوهش‌های ما روی متن فارسی بوده‌است، تمام حروف انگلیسی نیز حذف شده‌اند.
- **حذف کلمات اضافی:** در ادامه اقدام به حذف ایست‌واژه‌ها کردیم. ایست‌واژه‌ها در واقع کلماتی هستند که دارای بار معنایی نیستند و مفهومی را منتقل نمی‌کنند. بنابراین بهتر است در پیش‌پردازش متن، این کلمات حذف شوند.

۳-۲- یافتن جنبه‌های مختلف مطرح‌شده در نظرات

به منظور یافتن جنبه‌های مهم در مورد خودروها، نیاز است تا موضوعات مطرح‌شده در نظرات مختلف شناسایی شوند. این کار نیاز به یک تحلیل معنایی دارد. برای این کار ابتدا جملات را خوشه‌بندی می‌کنیم و سپس با روش LDA به یافتن موضوعات مختلف در هر خوشه و در نهایت بررسی و یافتن دستی جنبه‌ها لازم به ذکر است در این مرحله به منظور استخراج جنبه‌های مختلف خودروها ما به صورت تصادفی ۱۰٪ از کل نظراتی را که برای خودروی دنا ثبت شده‌بود، انتخاب و پردازش کرده‌ایم.

جدول ۱: جنبه‌های استخراج‌شده از متن نظرات برای خودروها

سطح اول	سطح دوم
مشخصات فنی	گیربکس، دیفرانسیل، انژکتور، بلبرینگ، توریو، ابروغن، اهرم، میل‌لنگ، تسمه‌تایم، تسمه، سوپاپ، سیلندر، شمع، وایر، جرّقه، احتراق، کمپرسور، پروانه، بازویی، دمنده، فنی، دریچه
عملکرد فنی	اسب‌بخار، گشتاور، سرعت، سوخت، بنزین، آگزوز، قدرت، نیوتون، شتاب، مصرف ترکیبی، روغن، رادیاتور، فیلتر، اتیلن، کاتالیزور، گاز، مقطر
ابعاد و اندازه	بدنه، شاسی، اتاق، محور، باک، لاستیک، ارتفاع، وزن، سدان، طول، عرض، ارتفاع، فاصله، سبک، سنگین، حجم، چرخ، عقب، جلو، استپ
سیستم دستیار راننده	کروز کنترل، تطبیقی، ایستایی، سربالایی، سرازیری، سرعت، کشش، پایداری، واژگونی
ترمز	ضدقفل، ای بی اس، چرخ، تایر، ترمز دستی، ترمز، پدال، برف، بارون، باران، بارندگی، توزیع، نیرو، اسفالت، کاسه نمد، اصطکاک، دیسک، استهلاک
سیستم روشنایی	دی‌لایت، مه شکن، نوربالا، چراغ، پروژکتور
فرمان	نیرو، کمکی
مولتی مدیا	صوتی، صدا، بلندگو، ضبط، آهنگ، مانیتور، تاج، دوربین
ایمنی کابین	ایمنی، ایربگ، کیسه هوا، کمربند
رفاهی کابین	رفاهی: ورود بدون کلید، کی لس، استارت دکمه‌ای؛ تهویه: کولر، بخاری، گرمکن، خنک‌کننده، ترموستات، فن؛ صندلی: پارچه، چرم، روکش؛ آپشن: سیگار، فنک، داشبورد، کنسول، چراغ مطالعه
آینه و شیشه	شیشه: ساتروف، دودی، شیشه شور؛ آینه: تاشو، پاناروما
مشخصات ظاهری	رنگ: کهکشانی، رنگ، زنگ، دوپوسته، متالیک؛ اسپورت: کوسه ای، انتن، دفی، ورز، رکاب، رینگ، لوتوس، قاب، لگو؛ بدنه: گلگیر، سپر، کاپوت، برف پاکن، شلگیر، دستگیره، پلاک
مولتی پکس	اژیر، اداپتور، باطری، باتری، اتصال، برق، اورو، سنسور، امپر
تجهیزات جانبی	قفل فرمون، زنجیر، صندلی کودک، قفل پدال، اچار، کمک های اولیه، جک، بکسل
خدمات	امداد خودرو، نمایندگی، تعمیرگاه، کارخانه

جدول ۱: تعداد نظرات استخراج‌شده برای هر خودرو

نام خودرو	تعداد نظرات
دنا	۵۰۲۵۲۸
پژو ۲۰۷	۱۸۲۱۰۸
پژو ۲۰۰۸	۳۳۱۴۵۰
پژو ۴۰۵	۵۳۸۰۱۹
ساینا	۲۵۸۰۱۷
سمند	۵۲۳۱۶۲
ساندرو	۴۰۳۸۳۵
تیبا	۴۳۱۵۶۷
تندر ۹۰	۴۹۰۳۵۴

استفاده می‌کردند و روی مجموعه‌داده‌های مختلف آموزش داده شده بودند، محاسبه شده است. در نهایت مدلی که بیشترین دقت را داشت، انتخاب شده است. لازم به ذکر است مدل مورد نظر شامل ۵ برچسب است و برای مطابقت با مجموعه‌داده‌ای که روی آن آزمایش شد، ۵ برچسب آن به سه برچسب کاهش داده شد. دقت ۷ مدل‌های بررسی شده روی مجموعه‌داده آزمایشی خودشان و همچنین مجموعه‌داده ساخته‌شده (۱۰۰۰ نظر تصادفی) در جدول ۳ قابل مشاهده است. لازم به ذکر است برای محاسبه دقت، تعداد نظراتی که درست برچسب خورده‌اند را تقسیم بر تعداد کل نظرات کردیم.

۴-۴. آزمایش‌ها و نتایج

در این بخش ابتدا مشخصات نظرات استخراجی و مدل استفاده شده برای تحلیل احساسات را بیان می‌کنیم. سپس نتایج حاصل از تحلیل احساسات را ارائه می‌کنیم.

۴-۱- مجموعه‌داده

لیست خودروهای مورد بررسی و تعداد نظرات استخراج شده برای هر خودرو در جدول ۲ نشان می‌دهد که بیشترین تعداد نظر مربوط به خودروی پژو ۴۰۵ و کمترین برای خودروی پژو ۲۰۷ است.

۴-۲- مدل استفاده‌شده برای تحلیل احساسات

همان‌گونه که در بخش ۳-۳ اشاره شد، از ماژول تحلیل احساسات گروه هوش‌واره برای تحلیل احساسات جملات استفاده شده است. به منظور انتخاب بهترین مورد برای تحلیل احساسات، ابتدا ۱۰۰۰ جمله به صورت تصادفی از مجموعه نظرات استخراجی، در نظر گرفته شده است. سپس این هزار جمله توسط دو نفر با مشورت یکدیگر برچسب زده شده‌اند. برچسب‌ها شامل مثبت، منفی و خنثی بوده‌اند. پس از آن دقت مدل‌های مختلفی که از «پارس‌برت» برای تحلیل احساسات

جدول ۲: دقت مدل‌های مختلف روی مجموعه داده آزمایش اصلی و

داده مربوط به خودروها

نام مدل	دقت داده آزمایش	دقت داده خودروها
albert-fa-base-v2-sentiment-snappfood	٪۹۳،۴۸	٪۲۵
albert-fa-base-v2-sentiment-multi	٪۷۳،۰۶۹	٪۴۶
bert-fa-base-uncased-sentiment-digikala	٪۸۴،۷۱۵	٪۲۷
albert-fa-base-v2-sentiment-deepsentipers-multi	٪۶۸،۷۱۶	٪۶۱
bert-fa-base-uncased-sentiment-deepsentipers-multi	٪۷۱،۸۹۹	٪۶۴

همان‌گونه که در جدول ۳ قابل مشاهده است، بیشترین دقت برای داده‌های مربوط به خودروها مربوط به مورد آخر است و سایر مدل‌ها اگرچه دقت خوبی روی مجموعه داده آزمایشی خودشان دارند، روی داده‌های مربوط به خودروها دقت مناسبی نداشته‌اند.

۴-۳- نتایج حاصل از آزمایش‌ها

در این بخش نتایج مربوط به برخی از خودروها را ارائه و بررسی می‌کنیم. نتایج مربوط به سایر خودروها در صفحه گیت‌هاب ما موجود است.

۴-۳-۱- نتایج مربوط به خودروی دنا

در شکل‌های ۲، ۳ و ۴ نتایج مربوط به خودروی دنا قابل مشاهده است. شکل ۲ نتایج مربوط به جنبه‌های اصلی خودروی دنا و شکل ۳ و ۴ نتایج مربوط به رفاهی کابین و سیستم روشنایی این خودرو است. در شکل ۲ دو حس منفی و مثبت به ترتیب با رنگ‌ها قرمز و سبز مشخص شده‌اند. همان‌گونه که قابل مشاهده است سیستم دستیار راننده، رفاهی کابین و تجهیزات جانبی مشخصاتی بوده‌اند که حس نظرات مثبت برای آن‌ها بیشتر از منفی بوده است. برای عملکرد فنی، نسبت نظرات مثبت و منفی تقریباً مساوی است و برای سایر جنبه‌ها

سیستم رو شنایی نسبت نظرات منفی بسیار بیشتر از نظرات مثبت است.

در شکل ۳ امکانات رفاهی کابین خودروی دنا به صورت جزئی‌تر بررسی شده‌اند. همان‌گونه که در شکل ۳ قابل مشاهده است، از نظر کاربران، برای خودروی دنا امکاناتی مانند چراغ مطالعه، سیگار، بدون کلید بودن، کنسول، استارت دکمه‌ای، داشبورد و فنک امکانات مناسب بوده‌اند و اغلب نظرات کاربران در مورد این امکانات مثبت بوده است اما نظر کاربران در مورد کولر، بخاری، فن، گرم‌کن، چرم، روکش، پارچه و ترموستات خودروی دنا اغلب منفی بوده است.

در شکل ۴ سیستم روشنایی خودروی دنا به صورت جزئی‌تر بررسی شده است. در شکل ۲ دیدیم که به طور کلی نظرات کاربران در مورد سیستم روشنایی خودروی دنا منفی بوده است و در شکل ۳ مشاهده می‌کنیم نظرات کاربران در مورد همه مشخصات سیستم روشنایی خودروی دنا منفی است. در مورد چراغ، نور بالا و پروژکتور نظرات منفی خیلی بیشتر از مثبت‌ها هستند ولی در مورد مه‌شکن، دی لایت و نور بالا این اختلاف کمتر است.

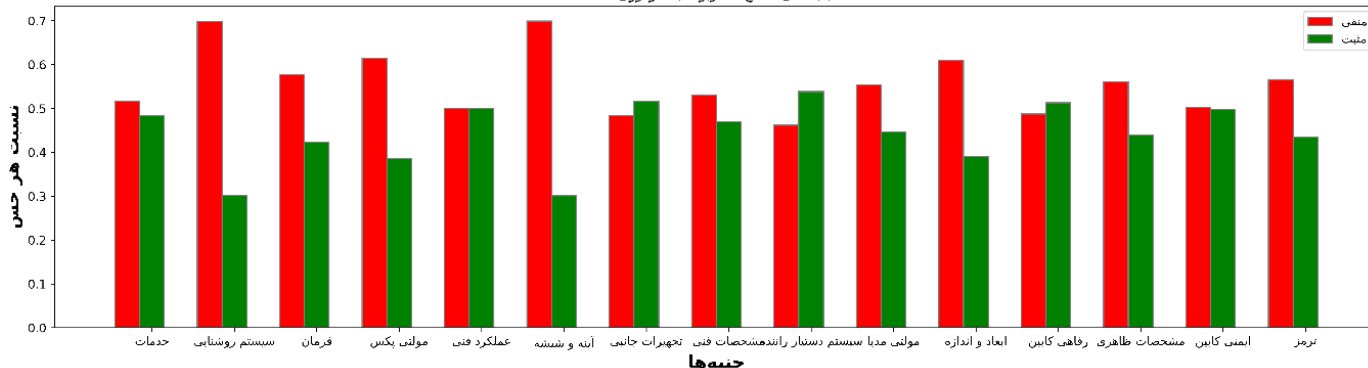
به طور کلی در مورد خودروی دنا نظر کاربران در مورد اغلب مشخصات منفی بوده است. سایر نمودارهای مربوط به این خودرو در صفحه گیت‌هاب ما قابل مشاهده است.

۴-۳-۲- نتایج مربوط به خودروی تیبا

در شکل‌های ۵، ۶ و ۷ نتایج مربوط به خودروی تیبا قابل مشاهده است. در شکل ۵ نتایج مربوط به جنبه‌های اصلی خودروی تیبا و در شکل ۶ و ۷ نتایج مربوط به مولتی مدیا و سیستم روشنایی این خودرو قابل مشاهده است.

همان‌گونه که در شکل ۵ قابل مشاهده است، نظرات کاربران در مورد تمامی مشخصات خودروی تیبا بیشتر منفی بوده است تا مثبت. به طور کلی می‌توان گفت نظرات منفی در مورد سیستم رو شنایی، آینه و شیشه و ترمز این خودرو خیلی بیشتر از نظرات مثبت بوده است. در مورد خدمات، سیستم دستیار راننده، مولتی

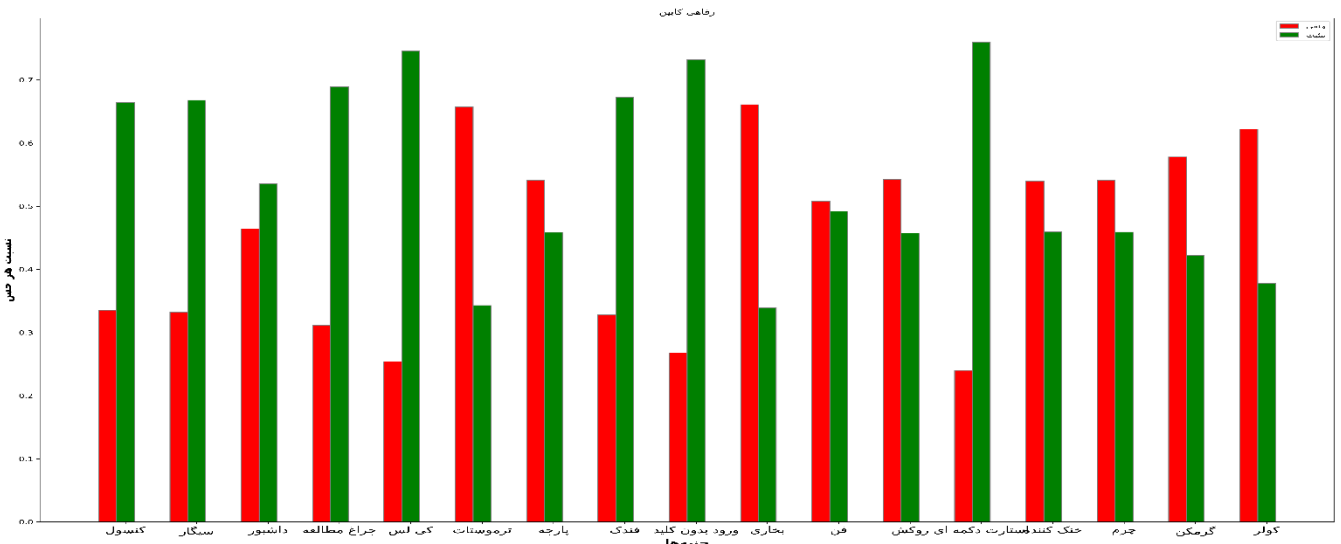
جنبه‌های سطح 1 مربوط به خودروی دنا



مدیا و رفاهی کابین این اختلاف کمتر است.

نظرات منفی بیشتر از مثبت بوده‌اند. به خصوص برای آینه و شیشه و

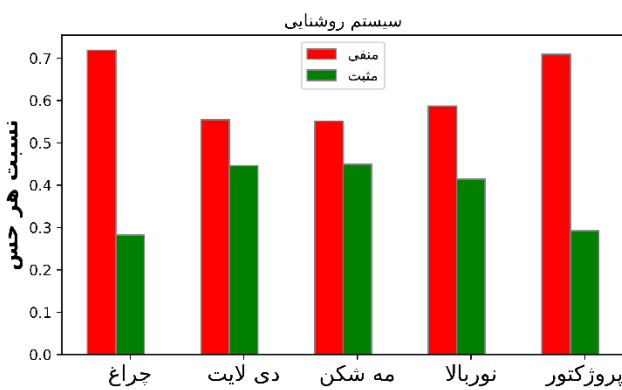
شکل ۲: حس مربوط به جنبه‌های اصلی مربوط به خودروی دنا



شکل ۳: حس مربوط به رفاهی کابین خودروی دنا گنار شکل نیست

بالاتری نیاز دارد. همچنین ممکن است کاربر در یک جمله از چند نظر یک خودرو را بررسی کرده و در هر مورد نظر مثبت یا منفی داشته باشد. شناسایی این که مثبت یا منفی بودن هر مورد در کدام جنبه از آن خودرو است، نیز چالش دیگر در این پژوهش است.

به منظور بهبود نتایج، می توان به حل چالش‌های گفته شده پرداخت و صرفاً نظراتی را بررسی کرد که مستقیماً نام آن خودروی خاص در آن‌ها آمده است. علاوه بر این، تحلیل شکلک‌هایی که در گروه‌ها ارسال می شود نیز تا حدی به تحلیل احساسات پیام‌ها کمک خواهد کرد. همچنین به منظور تحلیل احساسات با دقت بیشتر می توان مدلی را بر اساس مجموعه داده‌ای مخصوص نظرات مربوط به خودروها آموزش داد افزایش دقت مازول تحلیل احساسات را بالا برد تا نتایج قابل اطمینان تر باشند.



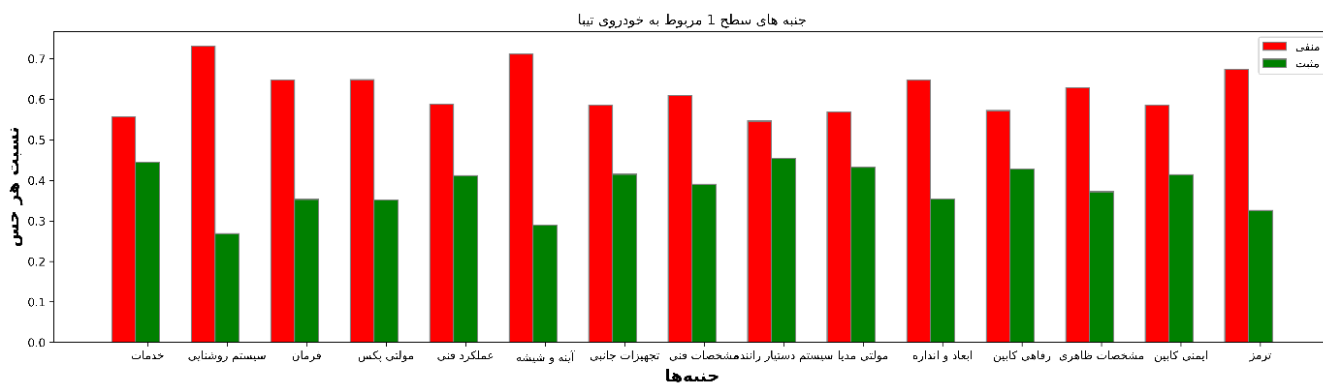
شکل ۴: حس مربوط به سیستم روشنایی خودروی دنا

مطابق شکل ۶، در مورد آهنگ، صوتی، تاج و مانیتور خودروی تیبا اغلب نظرات مثبت و در مورد دوربین، ضبط، بلندگو و صدا اغلب نظرات منفی بوده است.

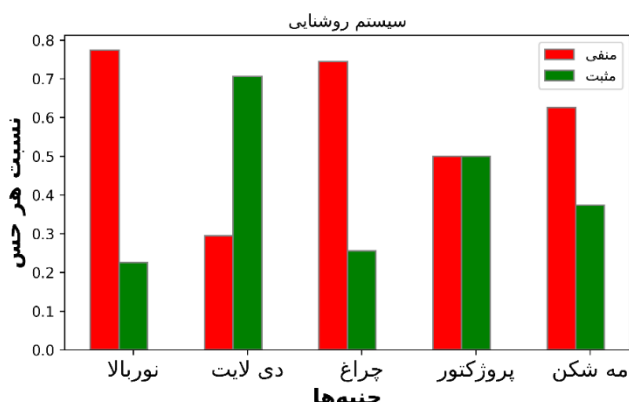
مطابق شکل ۷، نظرات کاربران در مورد، نوربالا، چراغ و مه شکن خودروی تیبا منفی، ولی در مورد دی لایت مثبت بوده است، در مورد پروژکتور نظرات مثبت و منفی تقریباً یکسان بوده اند. به طور کلی نظرات کاربران در مورد خودروی تیبا در اغلب مشخصات مربوط به این خودرو منفی بوده است.

۵- نتیجه گیری و کارهای آتی

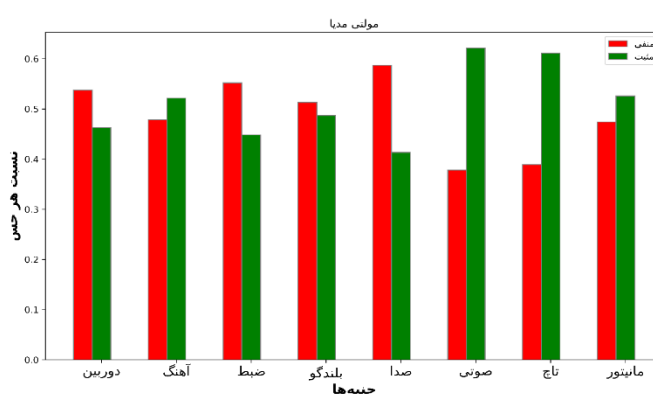
در این پژوهش روشی برای یافتن جنبه‌های مختلف مربوط به خودروها از نظرات مردم در شبکه اجتماعی تلگرام و سپس تحلیل احساسات این نظرات ارائه شد. در این روش ابتدا به کمک مدل زبانی "پارس برت" برای جملات مختلف بردار استخراج شد. سپس به کمک این بردارها خوشه بندی انجام شد. پس از آن به کمک روش "LDA" موضوعات مربوط به خوشه‌ها به دست آمد و در نهایت با بررسی این موضوعات، جنبه‌های مختلف مربوط به خودروها استخراج شد. پس از آن، حس هر جمله مشخص شد و در نهایت مثبت و منفی بودن نظرات کاربران و این که در مورد هر مشخصه از هر خودرو چه نظری داده اند، مشخص شد. یکی از چالش‌های اساسی این پژوهش امکان صحبت در مورد سایر خودروها در گروه تلگرامی مربوط به یک خودروی خاص است بدین ترتیب، تشخیص و جداسازی نظرات در مورد سایر خودروها صرفاً به کمک کلمات کلیدی دشوار است و به تحلیل‌های زبانی سطح



شکل ۵: حس مربوط به جنبه‌های اصلی مربوط به خودروی تیا



شکل ۷: حس مربوط به سیستم روشنایی خودروی تیا



شکل ۶: حس مربوط به مولتی‌مدیای خودروی تیا

Conference on Signal Processing and Intelligent Systems (ICSPIS) 2018 Dec 25 (pp. 77-82). IEEE.

- [8] Heydari M, Khazeni M, Soltanshahi MA. "Deep learning-based sentiment analysis in persian language". In *2021 7th International Conference on Web Research (ICWR) 2021* May 19 (pp. 287-291). IEEE.
- [9] Qin Z, Ronchieri E. "Exploring Pandemics Events on Twitter by Using Sentiment Analysis and Topic Modelling". *Applied Sciences*. 2022 Nov 22;12(23):11924.
- [10] Jain PK, Quamer W, Saravanan V, Pamula R. "Employing BERT-DCNN with sentic knowledge base for social media sentiment analysis". *Journal of Ambient Intelligence and Humanized Computing*. 2022 Jan 25:1-3.
- [11] Maathuis C, Kerkhof I. "The first two months in the war in Ukraine through topic modeling and sentiment analysis." *Regional Science Policy & Practice*. 2023 Feb;15(1):56-74.
- [12] Kokab ST, Asghar S, Naz S. "Transformer-based deep learning models for the sentiment analysis of social media data." *Array*. 2022 Jul 1;14:100157.
- [13] Mohseni S, Khadivar A, Abbasi F. "Sentiment Analysis of Trip Advisor Comments for Iranian Restaurants With a Deep Learning Approach." *Business Intelligence Management Studies*, 2022, Vol. 10, Issue 40, pp:17-41.
- [14] Khalaj.E, Mohammadi.Sh. "Presenting the model for opinion mining at the document feature level for hotel users' reviews." *Journal of Information and Communication Technology*. 2021 Nov;49(49):85.
- [15] Abbasi F, Khadivar A, Yazdinejad M. "Sentiment Analysis of Users for Buying on Cell Phone in Digikala." *BI Management Studies*. 2020 Aug 22;8(32):181-210.

References

- [1] Farahani M, Gharachorloo M, Farahani M, Manthouri M. "Parsbert: Transformer-based model for persian language understanding". *Neural Processing Letters*. 2021 Dec; 53:3831-47.
- [2] Blei DM, Ng AY, Jordan MI. "Latent dirichlet allocation". *Journal of machine Learning research*. 2003;3(Jan):993-1022.
- [3] Pai PF, Liu CH. "Predicting vehicle sales by sentiment analysis of twitter data and stock market values". *IEEE Access*. 2018 Oct 4; 6:57655-62.
- [4] Ali F, Kwak D, Khan P, El-Sappagh S, Ali A, Ullah S, Kim KH, Kwak KS. "Transportation sentiment analysis using word embedding and ontology-based topic modeling". *Knowledge-Based Systems*. 2019 Jun 15; 174:27-42.
- [5] Dong J, He F, Guo Y, Zhang H. "A commodity review sentiment analysis based on BERT-CNN model". In *2020 5th International conference on computer and communication systems (ICCCS)*, 2020 May 15 (pp. 143-147). IEEE.
- [6] Wang H, Hu X, Zhang H. "Sentiment analysis of commodity reviews based on ALBERT-LSTM". In *Journal of Physics: Conference Series*, 2020 Nov 1 Vol. 1651, No. 1, p. 012022. IOP Publishing.
- [7] Salari S, Sedighpour N, Vaezinia V, Momtazi S. "Estimation of 2017 Iran's presidential election using sentiment analysis on social media". In *2018 4th Iranian*

- [16] Zare Mehrjardi F, Yazdian-Dehkordi M, Latif A, “Evaluating Machine learning and Deep-learning methods in Sentiment Analysis on Persian Telegram Message”, *Soft Computing Journal*, 2022 September, Vol.11 , No. 1.
- [17] Tabatabaei S, Barati M, Bayati M. “Sentiment Analysis, Drug Industry, Hybrid Deep Learning, Convolutional And Recursive Architecture.” *Intelligent Information Systems Journal*. 2022 Dec 10;1(4):21-9.
- [18] Liu X, Li Y, Shao Y, Li A, Liang J. “A sentiment analysis model for car review texts based on adversarial training and whole word mask BERT.” In *Intelligent Networked Things: 5th China Conference, CINT 2022, Urumqi, China, August 7-8, 2022, Revised Selected Papers 2023 Jan 28* (pp. 107-121). Singapore: Springer Nature Singapore.
- [19] Sharami JP, Sarabestani PA, Mirroshandel SA. Deepsentipers: Novel deep learning models trained over proposed augmented persian sentiment corpus. *arXiv preprint arXiv:2004.05328*. 2020 Apr 11.

پی‌نوشت

1. Tokenizer
2. ParsBERT
3. <https://huggingface.co/HooshvareLab/bert-fa-base-uncased-sentiment-deepsentipers-multi>
4. <https://anonymous.4open.science/r/IranianCarsSentimentAnalysis-3C81/README.md>
5. ALBERT
6. BiLSTM
7. Accuracy