



Paper Type: Original Article

Analyzing stakeholder review of an industrial lubricant refinery using a combination of topic modeling and keyword co-occurrence analysis

Abstract

Purpose: Analyzing customer feedback is crucial for developing effective market strategies. Given the research gap in analyzing stakeholder reviews for Persian data, this study aims to present a combined approach for analyzing stakeholder opinions of a lubricant refinery using topic modeling and keyword co-occurrence analysis.

Methodology: This research utilizes the LDA algorithm and keyword co-occurrence network to analyze the opinions of three stakeholder groups (sales agents, company website users, and Digikala buyers). The main steps include Persian text preprocessing, LDA implementation with specified parameters (optimal number of topics based on c_v

 Corresponding Author:



Licensee System Analytics. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0>).

coherence score, passes=40, iterations=200, alpha='symmetric'), co-occurrence matrix construction, and key phrase extraction from high-frequency edges using a scoring method.

Findings: The highest prevalence rate of extracted phrases from sales agents' comments (31.01%) was related to the weakness of chelik (steel drum) container packaging. Among website users' comments, the ambiguity about the container material of motor oil No. 2 had the highest prevalence rate of 36.67%. Digikala users' comments indicated satisfaction with appropriate price and quality of products. Based on the findings, recommendations such as packaging improvement (product development strategy) and creating an online recommender system (market penetration strategy) are proposed.

Originality/Value: The proposed approach provides a practical method for systematic analysis of organizational stakeholder opinions and increases the accuracy of key phrase extraction by combining two text analysis methods.

Keywords: Market Development, Stakeholders, Keyword Co-occurrence

آماره انتشار



تحلیل نظرات ذینفعان یک پالایشگاه روانکار با استفاده از ترکیب مدل سازی موضوع و تحلیل

هم‌رخدادی واژگان

چکیده

هدف: تحلیل نظرات مشتریان برای توسعه استراتژی‌های موثر بازار ضروری است. با توجه به خلا تحقیقاتی در تحلیل نظرات ذینفعان برای داده‌های فارسی، این پژوهش با هدف ارائه رویکردی ترکیبی برای تحلیل نظرات ذینفعان یک پالایشگاه روانکار با استفاده از مدل سازی موضوع و تحلیل هم‌رخدادی واژگان انجام شده است.

روش‌شناسی پژوهش: پژوهش با بهره‌گیری از الگوریتم LDA و شبکه هم‌رخدادی واژگان، نظرات سه گروه ذینفع (عاملین فروش، کاربران سایت شرکت و خریداران دیجی‌کالا) را تحلیل کرده است. مراحل اصلی شامل پیش‌پردازش متن فارسی، اجرای LDA با پارامترهای مشخص (تعداد بهینه موضوعات بر اساس امتیاز انسجام c_v ، تعداد $\alpha='symmetric'$ ، iterations=200، passes=40)، تشکیل ماتریس هم‌رخدادی و استخراج عبارات کلیدی از یال‌های پرتکرار با استفاده از روش امتیازدهی است.

یافته‌ها: بیشترین نرخ فراوانی عبارت استخراج شده از نظرات عاملین فروش (۳۱.۰۱٪) مربوط به ضعف بسته‌بندی با ظرف چلیک است. میان نظرات کاربران سایت، ابهام در جنس مظروف روغن موتور شماره ۲ با نرخ ۳۶.۶۷٪ بیشترین فراوانی را دارد. نظرات کاربران دیجی‌کالا نشان‌دهنده رضایت از قیمت و کیفیت مناسب محصولات است.

بر اساس یافته‌ها، پیشنهادهایی مانند بهبود بسته‌بندی (استراتژی توسعه محصول) و ایجاد سیستم توصیه‌گر آنلاین (استراتژی نفوذ در بازار) ارائه شده است.

اصالت/ارزش افزوده علمی: رویکرد ارائه‌شده، روشی کاربردی برای تحلیل سیستماتیک نظرات ذینفعان سازمان‌ها ارائه می‌دهد و با استفاده از ترکیب دو روش تحلیل متن، دقت استخراج عبارات کلیدی را افزایش می‌دهد. کلیدواژه‌ها: توسعه بازار، ذینفعان، هم‌رخدادی واژگان کلیدی

۱- مقدمه

در نظرات برخط دهان به دهان الکترونیکی^۱ که با نام (eWOM) شناخته می‌شوند مشتریانی که خدمات را تجربه کرده‌اند، این امکان را دارند که نمرات را انتخاب کرده یا تجربه خود را توصیف کنند که قابل اعتماد تر و عینی تر از اطلاعات ارائه شده توسط شرکت ها به نظر می‌رسد. به همین دلیل نظرات برخط نقش مهمی در شکل دهی به تصویر پیش از تجربه خدمت برای مشتریان دارند. در مقایسه با نظرات عمومی، ویژگی های نظرات برخط شامل دسترسی ۲۴ ساعته و فراهم کردن امکان ذخیره مداوم اطلاعات به صورت متن یا تصویر است. این نظرات به مشتریان نه تنها پوشش گسترده تری ارائه می‌دهد، بلکه انتشار اطلاعات نیز به سرعت انجام می‌شود. بنابراین، نظرات برخط تأثیر قابل توجهی بر مشتریان دارای خدمات فعال دارد، زیرا می‌توانند نظرات خود را بدون محدودیت زمانی و مکانی ارائه و ویرایش کنند. از طریق نظرات برخط، مشتریان به صورت آزاد و بی طرف نظرات خود را در جامعه برخط با دیگران به اشتراک می‌گذارند. در مقابل تبلیغات یک طرفه، نظرات برخط به دلیل انتقال نظرات واقعی مشتری یا تجربیات خاصی که تجربه می‌کنند، در کنترل انتخاب های مشتریان نقش فعال دارند.[1]

بسیاری از مطالعات قبلی مربوط به داده کاوی بر روی داده های ساخت یافته متمرکز شده اند در حالی که ۸۰ درصد از داده های جهان به شکل بدون ساختار هستند[۲]. امروزه حجم قابل توجهی از داده ها به صورت متنی در دسترس است. داده های ذخیره شده در پایگاه های داده متنی به سرعت در حال رشد هستند.[3]

تجزیه و تحلیل نظرات مشتریان برای توسعه استراتژی های موثر بازار بسیار مهم است [۴]. نظرات مشتریان که از طریق بررسی ها و رتبه بندی ها به دست می آیند، بینش های ارزشمندی را برای تجزیه و تحلیل محصول و بهبود استراتژی بازاریابی ارائه می دهند که در نهایت منجر به افزایش سود می شود. درک ترجیحات مصرف کننده از طریق بررسی های برخط به مناسب سازی محصولات برای برآوردن نیازها و انتظارات مشتری کمک می کند[5].

گنجاندن بازخورد مشتری در تصمیمات قیمت و کیفیت بر اساس بررسی های برخط می تواند به استراتژی های بهینه برای شرکت ها در بازار های رقابتی منجر شود.

در تلاش برای درک و استفاده از نظرات مشتریان در توسعه بازار، روش ها و استراتژی های متنوعی ارائه شده اند. یکی از این رویکردها بررسی برخط مشتریان برای تصمیم گیری در مورد کیفیت و قیمت گذاری است. در این روش، شرکت ها با استفاده از

ارزیابی های مشتری برای بهینه سازی عملکرد مالی خود، استراتژی های خود را تنظیم می کنند [۵]. علاوه بر این، تحلیل احساسات روی نظرات مشتریان با استفاده از رویکرد های یادگیری ماشین به طبقه بندی احساسات برای افزایش فروش و به دست آوردن مزیت رقابتی در بازار تجارت الکترونیک کمک می کند [۶]. این مدل ها در کل به درک و استفاده از نظرات مشتریان برای توسعه استراتژی های مؤثر در بازار کمک می کنند.

به علاوه، مدل سازی موضوع نقش مهمی در تجزیه و تحلیل نظرات مشتریان برای توسعه استراتژی های بازار ایفا می کند. چندین مدل و رویکرد در این زمینه ارائه شده است. احمدوف و همکاران مدل جمله سند موضوعی^۱ (TDS) را معرفی کرد که ترکیبی از موضوع احساسات مشترک^۲ (JST) و تخصیص دیریکله پنهان^۳ (LDA) برای تجزیه و تحلیل احساسات در هر دو سطح سند و کلمه است. [7]

عثمانی و محصفی مدل JST را با معرفی پارامتر دیدگاه نویسنده گسترش دادند که دقت طبقه بندی احساسات را در سطح متن بهبود می بخشد [۸]. ششاج و همکاران مدل LDA را با رویکرد تنظیم موضوعی مبتنی بر معنا برای تجزیه و تحلیل تجربه مشتری تقویت کردند. [9]

دریافت نظرات مشتریان در مورد محصولات و خدمات ارائه شده توسط شرکت ها یکی از نیاز های اصلی مدیران شرکت برای بهبود ارائه خدمات است. امروزه استفاده از شبکه های اجتماعی، پیام رسان ها و وب سایت های بررسی برای دریافت اطلاعات از نظرات مشتریان رایج است اما تجزیه و تحلیل داده های دریافتی از این منابع به دلیل حجم زیاد آن ها نیازمند یک ابزار کارآمد است. [10]

با تکامل هوش مصنوعی و رشته های مرتبط مانند یادگیری ماشین، کاربرد روش های پردازش به وسیله کامپیوتر با استفاده از الگوریتم های ریاضی برای درک الگوهای پنهان در داده ها افزایش یافته است که به داده کاوی معروف است. از جمله داده های قابل پردازش، داده های متنی است که می توان با استفاده از روش های مختلف پردازش متن دانش پنهان در متون را شناسایی کرد. [10]

این روش ها اغلب رایگان بوده، در پاسخ به داده های جدید خود انطباق پذیر هستند، پیاده سازی آن ها آسان است و کاملاً سازگار می باشند. لذا استفاده از روش های داده کاوی و یادگیری ماشین در فرآیند استخراج دانش از داده های متنی مشتریان مقرون به صرفه است. همان طور که پیش تر هم اشاره شد، با کمک این تکنیک ها می توان دانش ضمنی و بالقوه مفید را از حجم زیادی داده استخراج کرد و در فرآیند تصمیم گیری از آن ها کمک گرفت.

¹ Topic/Document/Sentence Model

² Joint Sentiment Topic

³ Latent Dirichlet Allocation

با توجه به مشخص شدن اهمیت و ضرورت تحلیل نظرات ذینفعان با استفاده از روش های سریع و کارآمد و تمرکز مقالات اخیر بر این مسئله، تحلیل نظرات ذینفعان جهت توسعه بازار برای داده های فارسی به عنوان یک خلا تحقیقاتی در این حوزه شناسایی شد. لذا در این پژوهش ما به دنبال ارائه رویکردی ترکیبی برای تحلیل نظرات ذینفعان یک پالایشگاه روانکار با استفاده از روش های کارآمد نظیر مدل سازی موضوع و تحلیل هم رخدادی واژگان به منظور استخراج عبارات کلیدی و پشتیبانی از تصمیم های بازاریابی هستیم.

ذینفعان در این تحقیق به عنوان عاملین فروش، سایت های فروش برخط و پیشنهادات مطرح شده در نظرسنجی های شش ماهه این پالایشگاه تعریف شده اند.

تجزیه و تحلیل نظرات ذینفعان برای توسعه بازار بسیار مهم است زیرا بازخورد ارزشمندی را برای بهبود استراتژی ها و تجاری سازی ارائه می دهد [11][12][13].

تحلیل نظرات ذینفعان به درک احساسات نسبت به محصولات یا خدمات و هدایت استراتژی های بهبود کمک می کند [۱۱]. علاوه بر این، ذینفعان متنوع به طور قابل توجهی در تجاری سازی و ایجاد بازار نوآوری های ریشه ای، بهبود موقعیت در بازار و در دسترس بودن محصول کمک می کنند [۱۳]. به طور کلی، تجزیه و تحلیل نظرات ذینفعان با استفاده از بازخورد، بهبود موقعیت محصول و همسویی استراتژی ها با انتظارات ذینفعان، رشد بازار را تقویت می کند. بنابراین، تجزیه و تحلیل نظرات ذینفعان به شکل دهی استراتژی های بازار، تقویت نوآوری و هدایت موفقیت کلی بازار کمک می کند و از این رو این امر برای شرکت ها دارای اهمیت است. با توجه به سرعت و حجم تولید داده در بستر اینترنت، واضح است که شرکت ها علاقه مند به استفاده از راه های کارآمد تر با سرعت و دقت بیشتر هستند.

مقالات اخیر نیز در زمینه تجزیه و تحلیل نظرات ذینفعان بر استفاده از روش های داده کاوی، یادگیری ماشین و مدل سازی موضوع تمرکز کرده اند [۱۴][۱۵][۸]. این مقالات تکنیک های مختلفی مانند طبقه بندی احساسات را با استفاده از مدل های یادگیری ماشین مانند بیز ساده^۱، ماشین بردار پشتیبان^۲ و درخت های تصمیم بررسی می کنند [۱۶]. این مطالعات بر اهمیت استفاده از فناوری های پیشرفته مانند یادگیری عمیق برای تجزیه و تحلیل نظرات و احساسات، به ویژه در زمینه حجم وسیع داده های اجتماعی موجود در اینترنت، تأکید می کنند.

مدل سازی موضوع امکان استخراج مضامین کلیدی را از حجم وسیعی از داده های متنی فراهم کرده و به جمع بندی و درک احساسات ذینفعان کمک می کند [۱۷]. درک پویایی نگرانی های عمومی از طریق تجزیه و تحلیل ذینفعان برای موفقیت پروژه های زیرساختی و عملکرد ذینفعان حیاتی است [۱۸]. مدل سازی موضوع، می تواند مسائل و موضوعات کلیدی مورد بحث توسط ذینفعان را آشکار کند و به شناسایی اولویت ها و تأثیر آن ها کمک کند [۱۸]. با استفاده از مدل سازی موضوع در تجزیه و تحلیل ذینفعان، ترسیم علایق، نفوذ و اهمیت ذینفعان مختلف آسان تر می شود و بررسی متوازن دیدگاه های آن ها را تسهیل می کند و

¹ Naïve Bayes

² Support Vector Machine

کیفیت فرآیند های تصمیم گیری را افزایش می دهد [۱۹]. در نهایت، این رویکرد تضمین می کند که تمام دیدگاه های ذینفعان در نظر گرفته شده و منجر به نتایج دارای تاثیر و فراگیری بیشتر در پروژه می شود.

ترکیب مدل سازی موضوع با شبکه های هم رخدادی واژگان می تواند کشف موضوع را با استفاده از روابط کلمات تقویت کند [۲۰]. در حالی که مدل های سنتی موضوع مانند تخصیص دیریکله پنهان استقلال بین کلمات را فرض می کنند، شبکه های هم رویداد روابط کلمات را برای بهبود کیفیت موضوع جذب می کنند [۲۱]. مدل هایی مانند CWIBTD از شبکه های هم رخدادی کلمه استفاده می کنند تا موضوعات نادر را به طور مؤثر شناسایی کنند و تراکم معنایی و حساسیت را در کشف موضوع افزایش دهند [۲۲]. به علاوه، الگوریتم هایی مانند موضوع جامعه، جوامع را از شبکه های هم روند استخراج می کنند تا ساختارهای موضوعی سلسله مراتبی ایجاد کنند و کاوش موضوعات فرعی و فوق العاده را با روابط منسجم تسهیل می کنند [۲۳]. با ادغام مدل سازی موضوع با شبکه های هم رخدادی واژگان، محققان می توانند روابط عمیق تر و سلسله مراتب موضوعی را برای تجزیه و تحلیل موضوعی روشن تر و جامع تر کشف کنند. لذا استفاده از ترکیب شبکه هم رخدادی واژگان با مدل سازی موضوع موجب افزایش دقت نتایج به دست آمده می شود.

این پژوهش با هدف ارائه رویکردی ترکیبی برای تحلیل نظرات ذینفعان جهت استخراج عبارات کلیدی و پشتیبانی از تصمیم های بازاریابی برای یک پالایشگاه روانکار با استفاده از تحلیل هم رخدادی واژگان کلیدی انجام شده است. لذا پژوهش حاضر در تلاش برای یافتن پاسخ این پرسش که «چگونه می توان با ترکیب مدل سازی موضوع و تحلیل هم رخدادی واژگان، عبارات کلیدی را از نظرات ذینفعان یک پالایشگاه روانکار استخراج کرد؟» انجام شده است.

علیرغم تمرکز مقالات اخیر در زمینه تجزیه و تحلیل نظرات ذینفعان بر استفاده از روش های داده کاوی، یادگیری ماشین و مدل سازی موضوع [۱۵] [۱۴] [۸]، تحلیل نظرات ذینفعان جهت توسعه بازار برای داده های فارسی به عنوان یک خلا تحقیقاتی در این حوزه شناسایی شده است.

در برخی از مقالات، تحلیل نظرات ذینفعان با استفاده از مدل سازی موضوع انجام شده است، اما در تحقیقات تا به حال اتصال نظرات ذینفعان به محصولات انجام نشده و مواردی از بررسی نظرات مشتریان به تفکیک دسته های مختلف مشتریان مشاهده شده است.

در ادامه این پژوهش، بررسی مبانی نظری و پیشینه پژوهشی مرتبط، روش شناسی پژوهش، یافته های پژوهش و ارائه پیشنهادها و نتایج کاربردی ارائه شده است.

۲- مروری بر مبانی نظری و پیشینه پژوهش

۲-۱- تحلیل نظرات ذینفعان

فریمن در سال ۱۹۸۴، ذینفعان را هر فرد یا گروهی می داند که در دستیابی سازمان به اهدافش اثر می گذارد یا از این موضوع اثر می پذیرد. اما او در آخرین تعریف خود از ذینفعان که در سال ۲۰۰۴ عنوان کرد، ذینفعان را به عنوان گروه هایی تعریف کرد که برای بقا و موفقیت شرکت ها لازم و ضروری هستند. فریمن و همکارانش در کتاب «تئوری ذینفعان» خود که در سال ۲۰۱۰ چاپ شده است، تعریف دیگری برای واژه ذینفع ارائه داده اند. او در این کتاب بیان کرده است که ذینفع به معنی افراد و گروه هایی است که به نوعی در موفقیت یا شکست یک کسب و کار، سهمیه هستند. [۲۴]

در کتاب «مدیریت استراتژیک» روترمل، بیان شده است که ذینفعان؛ افراد، گروه ها یا سازمان هایی هستند که از اقدام های یک بنگاه اقتصادی تاثیر می پذیرند یا بر روی اقدام های یک بنگاه اقتصادی تاثیر می گذارند. [۲۵]

تعریفی که هانسون و همکاران در کتاب خود با عنوان «مدیریت استراتژیک: رقابت پذیری و جهانی شدن» (۲۰۱۶) از ذینفعان ارائه داده اند بدین شرح است: ذینفعان؛ افراد، گروه ها و سازمان هایی هستند که می توانند روی چشم انداز و مأموریت یک بنگاه اقتصادی تاثیر بگذارند. هم چنین دستاوردها و خروجی های استراتژیک بنگاه، روی آن ها تاثیر گذار است. ضمناً، ذینفعان می توانند در زمینه عملکرد بنگاه، انتظارات و ادعاهایی داشته باشند و این موارد را به وسیله اهرم های مختلفی که در اختیار دارند اعمال کنند [۲۶].

کلمه ذینفعان برای اولین بار در سال ۱۹۶۳ در یک همایش بین المللی در موسسه تحقیقات استنفورد مطرح شد اما مطرح شدن مفهوم ذینفعان برای نخستین بار به صورت یک ساختار و تعریف منسجم به سال ۱۹۸۴ برمی گردد [۲۷].

تصدیقی (۱۳۹۰) در مقاله خود نقشه ادبیات متمرکز بر حوزه ذینفعان را بدین صورت بیان می کند.

ارائه مفهوم ذینفعان برای نخستین بار در موسسه تحقیقات استنفورد (۱۹۶۳)			
کاربرد مفهوم ذینفعان در طرح ریزی سازمانی	کاربرد مفهوم ذینفعان در تئوری سیستم ها	کاربرد مفهوم ذینفعان در حوزه مسئولیت های اجتماعی سازمان	کاربرد مفهوم ذینفعان در تئوری سازمان
مدیریت راهبردی، یک رویکرد مبتنی بر ذینفعان توسط فریمن (۱۹۸۴)			
جنبه های هنجاری	جنبه های ابزاری	جنبه های توصیفی/تجربی	
تئوری ذینفعان شرکت توسط دونالدسون و پرستون (۱۹۹۵)			
تئوری پویایی و خوشه بندی ذینفعان (۲۰۱۰)			
سایر تئوری های ذینفعان (۱۹۹۶ تا ۲۰۱۱)			

شکل ۱: نقشه پیشینه حوزه ذینفعان (منبع شکل : [۲۸])

Figure 1: Stakeholder domain background map (Source: [28])

تاکنون ذینفعان به روش های مختلفی از جنبه های متفاوت توسط محققین این حوزه طبقه بندی شده اند. از جمله این طبقه بندی ها می توان به طبقه بندی کمبل، طبقه بندی شولر و وندی ون، طبقه بندی کلارکسون، طبقه بندی حیث و نورمن و طبقه بندی فریدمن اشاره کرد که در تمامی این طبقه بندی ها مشتریان به عنوان گروهی از ذینفعان سازمان مطرح شده اند [۲۹].

با توجه به پیشینه موجود، مشتریان مهم ترین گروه ذینفعان برای سازمان ها محسوب می شوند. لذا می توان گفت نظرات و نیازهای آن ها برای مدیران سازمان ها حائز اهمیت است. برای کسب دانش در مورد نیازهای مشتریان، می توان این داده ها را جمع آوری کرد. اما با توجه به گسترش استفاده از اینترنت و امکاناتی نظیر خرید اینترنتی و اعلام نظر در مورد محصولات در سایت های مختلف و فروم ها، می توان گفت نظرات برخط محصول یک راه جایگزین برای جمع آوری داده ها است. این نظرات اطلاعات مفیدی را در اختیار مدیران قرار می دهد. بنابراین، مدیران ممکن است بینش ارزشمندی را از نظرات برخط استخراج نموده و بر

اساس آن عمل کنند. محتوای تولید شده توسط مصرف کننده، به ویژه نظرات مشتریان برخط در مورد برند های مختلف، منبع غنی اطلاعات است چرا که این نظرات فاقد جانب داری است و به عنوان یک پیشنهاد دوستانه بین مشتریان در نظر گرفته می شود [۲].

با ظهور وب ۲.۰، حجم عظیمی از داده ها تولید و در وب ارسال می شود. بازخوردها در قالب نظرات مشتریان برخط به سرعت در حال رشد است که شامل افکار، نظرات و تجربیات مشتریان در ارتباط با با برندها می شود. این موضوع به مدیران کمک می کند تا ترجیحات مشتریان را درک کنند و نظرات برخط مربوط به محصول یا نام تجاری خود را ردیابی و نظارت کنند.

در سال های اخیر، مطالعات مختلفی با استفاده از تجزیه و تحلیل احساسات در محصولات کلی، سیاست و جامعه، از جمله صنعت خدمات، مانند فیلم ها، گردشگری، رستوران ها و هتل ها انجام شده است که در آن ها بر اهمیت بررسی های برخط تأکید شده است [۱].

همچنین به گفته داهیا و همکارانش که مطالعات مربوط به نظرات مشتریان برخط منتشر شده از ۲۰۱۵-۲۰۲۰ را بررسی کرده اند، نتایج نشان می دهد که مطالعات مربوط به تجزیه و تحلیل بررسی مشتریان برخط در طول سال های اخیر افزایش یافته است [۲].

بسیاری از محققان به اهمیت بررسی های برخط پی بردند و با استفاده از آن ها تحقیق کردند. دلاروکاس و همکاران نشان داده اند که معیارهای بررسی برخط می تواند درآمد فیلم را به طور دقیق پیش بینی کند. سوتیریادیس و ون زیل دریافتند که بررسی ها و توصیه های برخط بر فرآیند تصمیم گیری گردشگران نسبت به خدمات گردشگری تأثیر می گذارد و تبلیغات دهان به دهان تأثیر قابل توجهی بر هنجارها و نگرش های ذهنی نسبت به یک شرکت هواپیمایی و تمایل مشتری به توصیه دارد. سیرینگ و همکاران، بررسی کرده اند که آیا محتوای تولید شده توسط کاربر در قالب بررسی های برخط می تواند برای توضیح و پیش بینی تصمیم توصیه مورد استفاده قرار گیرد یا خیر. علاوه بر این، آن ها دریافتند که احساسات مربوط به جنبه های مختلف خدمات نیز به طور قابل توجهی بر تصمیم توصیه تأثیر می گذارد. بنابراین، بررسی برخط برای خطوط هوایی برای درک مشتریان متنوع خود به منظور اتخاذ استراتژی های بهبود خدمات بسیار مفید خواهد بود، زیرا خطوط هوایی یک صنعت بسیار رقابتی است [۱].

۲-۲- توسعه بازار

ریشه واژه استراتژی، کلمه یونانی Strategema به معنای فرمانده ارتش است. این واژه از ترکیب واژه Stratos به معنای ارتش و ago به معنای رهبر پدید آمده است. مفهوم استراتژی ابتدا در علوم نظامی به معنای فن هدایت، تطبیق و هماهنگ سازی نیروها جهت دست یافتن به اهداف جنگ استفاده شد.

استراتژی مجموعه ای از اهداف اصلی و سیاست ها و برنامه های کلی به منظور نیل به اهداف است به گونه ای که قادر به تبیین این موضوعات باشد که در چه کسب و کار و چه نوع سازمانی فعالیت می کنیم یا قصد انجام فعالیت را داریم.

به بیان دیگر، استراتژی یک برنامه واحد، همه جانبه و تلفیقی است که نقاط قوت اصلی سازمان را به تغییرات محیط مربوط می سازد و به گونه ای طراحی می شود که حصول اطمینان از دست یافتن به اهداف اصلی سازمان با اجرای درست آن صورت گیرد.

ایگر آنسوف^۱، با ارائه ماتریسی با دو بعد محصول و بازار، برای اولین بار به بیان جهت راهبردی برای رشد فعالیت های واحد کسب و کار پرداخت.

به اعتقاد آنسوف (۱۹۸۷)، توسعه بازار به معنای یافتن بازارهای جدید از طریق روش هایی نظیر باز کردن کانال های بازاریابی و ورود به بازارهای جدید جهت عرضه محصولات فعلی است. با این حال، در بعضی منابع، توسعه محصول و نفوذ در بازار را نیز موجب توسعه بازار موجود می داند و آن را اصلاح بازار می نامد. او معتقد است زمانی که بازار در حالت اشباع قرار دارد، روش های اصلاح بازار سودمندی زیادی به همراه دارند [۳۰].

یک شرکت با استفاده از استراتژی توسعه بازار می تواند از طریق اشباع بازار و نفوذ در بازار، سهم بزرگی از بازار فعلی را در اختیار بگیرد. همچنین می تواند برای محصولات یا خدمات فعلی بازارهای جدید بیافریند.

۲-۳- مدل سازی موضوع

موضوع، مسئله یا ایده خاصی است که به عنوان موضوع یک پاراگراف، مقاله، گزارش یا سخنرانی عمل می کند. مدل سازی موضوع یکی از روش های یادگیری بدون نظارت است که توانایی گرفتن ساختار معنایی پنهان در یک سند را دارد. ایده اصلی مدل سازی موضوع این است که هر سند مجموعه ای از موضوعات است و یک موضوع از مجموعه ای از کلمات ساخته شده است. مدل سازی موضوع با موفقیت برای تسهیل بازاریابی اطلاعات، طبقه بندی اسناد و تجزیه و تحلیل اکتشافی مجموعه های بزرگی از متون مورد استفاده قرار گرفته است [۱۶].

مدل سازی موضوع در ابتدا در دهه ۱۹۸۰ توسعه یافت و از حوزه موضوعی «مدل سازی احتمالی مولد»^۲ منشعب شد. این نوع مدل سازی فرض می کند که متغیرهای مشاهده شده با پارامترهای مشاهده نشده یا نهفته در یک رابطه احتمالی خاص تعامل دارند که سپس داده ها را در یک مجموعه داده تولید می کند. توسعه این فرآیند ها از نیاز به توصیف مختصر عناصر در مجموعه های بزرگ داده ها بدون به خطر انداختن روابط آماری مورد نیاز برای تکمیل تحلیل های ساده تر، مانند طبقه بندی و خلاصه سازی نشئت گرفت. اولین روش توسعه یافته برای این کار، طرح کاهش tf-idf^۳ نام داشت و توسط دو محقق در زمینه بازاریابی اطلاعات، سالتون و مک گیل^۴، در سال ۱۹۸۳ ارائه شد. در این روش، هر ردیف داده به طور مجزا در یک مجموعه با واژگان خود در نظر گرفته می شود و تعداد تکرارهای هر کلمه برای تشکیل یک مقدار شمارش محاسبه می شود تا تعداد فراوانی کلمه یا اصطلاح (tf) مخصوص آن ردیف داده برای آن کلمه ایجاد شود. مجموع نمونه های یک کلمه در کل فایل داده ها، که شمارش فرکانس سند معکوس (idf) نامیده می شود، نیز محاسبه می شود. پس از نرمال کردن این مقادیر به مجموعه داده در صورت نیاز،

¹ Igor Ansoff

² Generative Probabilistic Modeling

³ Term Frequency – Inverse Document Frequency Reduction Scheme

⁴ Salton & McGill

دو مقدار با یکدیگر مقایسه می شوند و یک ماتریس متن به ردیف داده حاوی مقادیر tf-idf برای تمام ردیف های داده درون ستون ها را تشکیل می دهند.

اولین تلاش برای توسعه مدل سازی موضوع توسط پاپادیمیتریو، تاماکی، راگاوون و ومپالا^۱ (۱۹۹۸) و هافمن^۲ (۱۹۹۹) انجام شد [۳۱]. پاپادیمیتریو و همکارانش «مدل سازی موضوع» را براساس جبر خطی پیشنهاد دادند، ولی هافمن (۱۹۹۹) برای این هدف یک مدل احتمالاتی ارائه کرد [۳۲] [۴۰].

روش های رایج فعلی در مدل سازی موضوع، تحلیل معنایی پنهان احتمالی و تخصیص دیریکله نهفته است که از یک مدل مخلوط احتمال و مدل لفظی/اصطلاحی برای استخراج موضوعات از داده های متنی استفاده می کند [۳۳].

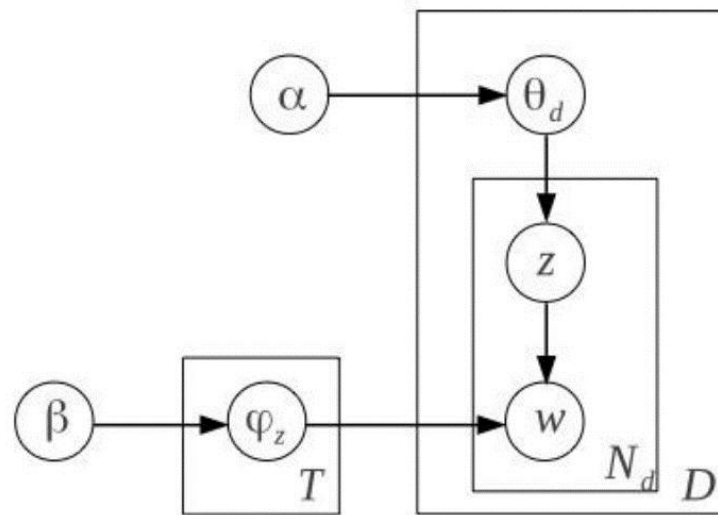
الگوریتم تخصیص دیریکله پنهان که در این پژوهش به کار می رود، الگوریتمی از نوع مدل سازی احتمالی است که اولین بار توسط بلای و جردن^۳ در سال ۲۰۰۳ معرفی شد، این الگوریتم محبوب ترین مدل موضوع و الگوریتم تجزیه و تحلیل موضوع در حال حاضر در نظر گرفته می شود. LDA یک روش مدل سازی احتمالی مولد بدون نظارت برای کشف ساختارهای معنایی پنهان در مجموعه ای از اسناد متنی است. LDA با هر سند به عنوان مخلوطی از موضوعات برخورد می کند و هر موضوع را ترکیبی از کلمات می داند [۱۶].

در شکل ۲ مدل «تخصیص دیریکله پنهان» نشان داده شده است که از دو ماتریس ϕ و θ تشکیل شده است. در این مدل ϕ ماتریسی است که توزیع موضوع T بر روی واژه های W را بر اساس توزیع اولیه دیریکله با پارامتر β بیان می کند. β نشان دهنده تراکم موضوع-کلمه است. θ ماتریسی است که توزیع متن d بر روی موضوع های T را براساس توزیع اولیه دیریکله با پارامتر α بیان می کند. α نشان دهنده چگالی سند-موضوع است. برای زایش هر واحد واژگانی w در متن d ، یک موضوع z از توزیع موضوعی مربوط به متن d به دست می آید؛ درحالی که خود آن واژه w از توزیع واژه های موضوع انتخاب شده ϕ_z به دست می آید [۳۲].

¹ Papadimitrio, tamaki, Raghavan, & Vempala

² Hofmann

³ Blei & Jordan



شکل ۲: نحوه کار الگوریتم تخصیص دیریکله پنهان (منبع شکل: [۳۲])

Figure 2: How the Latent Dirichlet Allocation algorithm works (Source: [32])

در این پژوهش، با استفاده از کلمات ظاهر شده در نتیجه پیاده سازی الگوریتم به عنوان گره های گراف هم رخدادی داده ها استفاده کرده و هم رخدادی این کلمات را برای نیل به موضوعات موجود در داده ها با دقت بیشتر بررسی کرده ایم.

۴-۲- تحلیل هم رخدادی واژگان

اسمال و گریفیث تحلیل هم رخدادی واژگان را یک نوع تحلیل هم استنادی می دانند که روشی مهم برای تصویرسازی روابط میان مفاهیم، ایده ها و مسائل علمی محسوب می شود. همچنین، آن ها تحلیل هم رخدادی واژگان را یک شیوه تحلیل کمی مهم به منظور کشف شبکه مفاهیم حوزه های علمی گوناگون می دانند [۳۴].

ادواردز^۱ (۱۹۹۵) بیان می دارد که تحلیل هم رخدادی واژگان یک تکنیک تجزیه و تحلیل محتوای خودکار است که در ساخت نقشه برداری علمی از طریق بررسی روابط داده موثر است. همچنین، این روش از فنون مدل سازی گرافیکی برای مصورسازی ارتباطات بین مفاهیم و تحلیل آن ها استفاده می کند.

هی^۲ (۱۹۹۹) تحلیل هم رخدادی واژگان را به عنوان روشی از تحلیل محتوا توصیف می نماید که الگوهایی را بررسی می کند که در آن جفت کلمات یا عبارات با هم در یک بدنه متن ظاهر می شوند. هدف این تکنیک ایجاد ارتباطات بین ایده ها در حوزه های موضوعی خاص است.

¹ Edwards

² He

به بیان دینگ، فو و چودهری^۱ (۲۰۰۱)، تحلیل هم رخدادی واژگان شیوه ای نو از تحلیل محتوا است که قادر است داده ها را با حفظ اطلاعات ضروری آن ها در قالب نمایش های بصری ارائه دهد.

تعاریف ارائه شده مفهوم تحلیل هم رخدادی واژگان را به شیوه زیر روشن می کند [۳۵]:

تحلیل هم رخدادی واژگان به عنوان یک رویکرد ارزشمند برای بررسی محتوای متن عمل می کند و مکانیسم مفیدی برای نشان دادن روابط و ارتباطات بین عناصر اطلاعاتی در مجموعه داده های متنی ارائه می دهد. این تکنیک تحلیلی متکی به فراوانی واژگان، عبارات و توصیف گرهای موجود در متون علمی است و بنابراین چارچوب مفهومی اساسی حوزه موضوع را با مشخص کردن و برقراری پیوند ها بین مفاهیم اصلی آن رونمایی می کند. این رویکرد به طور معمول برای نگاشت، ردیابی و تجسم الگو ها و گرایش های موضوعی در یک حوزه تحقیقاتی مشخص و دست یابی به واژگان اساسی حوزه موضوع مورد استفاده قرار می گیرد.

نخستین تلاش ها در زمینه واژگان درون متنی توسط زیف^۲ در سال ۱۹۴۹ انجام شد. زیف با مطالعه واژه هایی که در متون انگلیسی به کار رفته اند به مصادیقی برای اصل کمترین کوشش دست یافت.

اسمال و گریفیث در سال ۱۹۷۴ برای اولین بار برای مشخص کردن مفهوم کلی مدرک استناد شده از تحلیل بسامد واژگان در عناوین مقالات استفاده کردند.

تحلیل هم رخدادی واژگان نخستین بار در مرکز جامعه شناسی خلاقیت کول^۳ فرانسه در دهه ۱۹۸۰ به کار گرفته شد [۳۶]. آن ها سیستم طراحی شده بر اساس این تکنیک را LEXIMAPPE نامیدند. این سیستم اولین نرم افزار مورد استفاده در پژوهش های مرتبط با تحلیل هم رخدادی واژگان در دهه ۱۹۸۰ میلادی بود [۳۵]. این تحلیل به عنوان شاخصی در علم سنجی در این دهه معرفی شد و بعد از آن توسعه پیدا کرد و اغلب به عنوان جایگزینی برای رویکرد های استنادی و هم استنادی برای ترسیم علم استفاده شد [۳۶].

کالون^۴ (۱۹۸۳)، اولین فردی بود که به توصیف روش تحلیل هم رخدادی واژگان پرداخت و از آن در پژوهش های خود استفاده کرد. به اعتقاد او، تحلیل هم رخدادی واژگان نسبت به تحلیل های هم استنادی شیوه مناسب تری برای شناسایی ساختار علمی محسوب می شود.

¹ Ding, Foo & Chowdhury

² Zipf

³ Ecole

⁴ Collan

کالون، لاو^۱ و ریپ^۲ (۱۹۸۶)، کتابی با عنوان «ترسیم پویایی علم و تکنولوژی» تالیف کردند که نقطه عطف تحلیل هم رخدادی واژگان بود. پس از انتشار این کتاب، تحلیل هم رخدادی واژگان به سایر نقاط جهان گسترش یافت و بعد از آن، تعاریف، مراحل و مقیاس های هم رخدادی واژگان تحت تاثیر مطالعات بعدی تا اندازه زیادی توسعه یافت و بهبود پیدا کرد [۲۷].

احمدی و عصاره [۳۶] در پژوهش خود با بررسی مطالعات مختلف در حوزه تحلیل شبکه هم رخدادی واژگان، کاربردهای زیر را برای این روش معرفی کرده اند:

۱. طبقه بندی مدارک بر اساس موضوعات (کالون و دیگران، ۱۹۸۶)
۲. ترسیم حرکت و پویایی علم (کالون، لاو و ریپ، ۱۹۸۶؛ کورتایل و لاو، ۱۹۸۹)
۳. ترسیم ساختار حوزه ها و زمینه های علمی (وایتاکر^۳، ۱۹۸۹)
۴. ترسیم روابط میان پژوهش های بنیانی و پژوهش های فناورانه (کالون، کورتایل و لاویل، ۱۹۹۱)
۵. کشف الگوهای برجسته و رویدادهای در حال ظهور
۶. ارزیابی درون داد/برون داد روابط در یک شبکه پژوهش (ترنر و راجان^۴، ۱۹۹۱)
۷. دیداری سازی شبکه مفهومی علم و فناوری (سالمی و کوشا، ۱۳۹۲)
۸. تحلیل سیر تحول مفهومی در طول دوره های زمانی
۹. تعیین و تحلیل نواحی مورد پژوهش
۱۰. خوشه بندی مفاهیم حوزه های علمی
۱۱. برنامه ریزی و سیاست گذاری در حوزه های علمی
۱۲. درک و کشف روابط پنهان در حوزه های علوم مختلف
۱۳. تعیین روابط سلسله مراتبی مفاهیم یک حوزه یا یک زمینه خاص علمی

با توجه به کاربردهای ذکر شده، می توان از ترکیب تحلیل هم رخدادی واژگان و تکنیک های متن کاوی بهره گرفت و جهت بهبود کیفیت و دقت نتایج استفاده از تکنیک هایی نظیر مدل سازی موضوع، آن ها را با تحلیل هم رخدادی واژگان ترکیب کرد. ترکیب مدل سازی موضوع با شبکه های هم رخدادی واژگان می تواند کشف موضوع را با استفاده از روابط کلمات تقویت کند [۲۰].

با ادغام مدل سازی موضوع با شبکه های هم رخدادی واژگان، محققان می توانند روابط عمیق تر و سلسله مراتب موضوعی را برای تجزیه و تحلیل موضوعی روشن تر و جامع تر کشف کنند.

¹ Law

² Rip

³ Whittaker

⁴ Turner & Rojoun

مقالات پژوهشی ادغام مدل سازی موضوع با شبکه های هم رخدادی واژگان را برای افزایش کیفیت موضوعات استخراج شده بررسی کرده اند. سیموتی و همکاران، روشی مبتنی بر شبکه های معتبر آماری^۱ (SVN) برای ارزیابی انسجام موضوع با استفاده از هم رخدادی کلمه پیشنهاد کرده اند [۳۷]. زو و همکاران، از روابط قضاوت صحیح بین کلمات در متون برای فیلتر کردن لبه های نوین در شبکه های هم رخدادی استفاده کرده و کیفیت استخراج موضوع را افزایش دادند [۲۱]. آستین و همکاران، الگوریتم موضوع انجمن را معرفی کردند که جوامع را از شبکه های هم رخدادی واژگان استخراج می کند تا ساختارهای موضوعی سلسله مراتبی با انسجام و روابط بین موضوعات بهبود یافته را نشان دهد [۲۲]. موراکامی و چاکرابورتی، استفاده از بردارهای تعبیه کلمات از مجموعه های بزرگی مانند ویکی پدیا را برای جبران اطلاعات محدود هم رخدادی کلمه در متون کوتاه، با هدف افزایش ارزیابی انسجام موضوع بررسی کردند [۳۸].

با توجه به مقالات بررسی شده، می توان به نوآور بودن این پژوهش در زمینه به کارگیری ترکیب مدل سازی موضوع و تحلیل هم رخدادی واژگان جهت استخراج عبارت های موجود در نظرات ذینفعان پی برد. این کار باعث کاهش خطای مدل سازی موضوع و افزایش کیفیت یافتن موضوعات موجود در متن می شود.

با توجه به مطالب فوق الذکر و در جهت بهبود کیفیت نتایج مدل، در این پژوهش پس از پیاده سازی الگوریتم LDA، با در نظر گرفتن کلمات ظاهر شده در نتایج اجرای الگوریتم به عنوان کلمات کلیدی یا به عبارت دیگر گره های شبکه هم رخدادی واژگان، اقدام به تشکیل ماتریس هم رخدادی واژگان و رسم این شبکه کرده ایم.

۳- روش پژوهش

با توجه به مطالب ذکر شده در راستای ضرورت و اهمیت تحلیل داده های متنی با روش های کارآمد برای انجام تحلیل بازخوردهای بازار بر مبنای تحلیل نظرات ذینفعان، پس از دریافت داده های متنی نظرات گروه های مختلف ذینفعان مورد بررسی، مدل پیشنهادی پژوهش روی این داده ها پیاده سازی شده است. پس از توضیح چگونگی جمع آوری اطلاعات و معرفی ذینفعان مورد مطالعه، مدل پیشنهادی به تفصیل شرح داده شده است.

۳-۱- جمع آوری اطلاعات و معرفی ذینفعان

مورد مطالعه این پژوهش، یک پالایشگاه روانکار می باشد و مدل پیشنهادی روی داده های نظرات گروهی از ذینفعان این شرکت پیاده سازی شده است. ذینفعان مورد مطالعه در این پژوهش، عاملین فروش پالایشگاه، کاربران سایت پالایشگاه و مصرف کنندگان برخی محصولات این پالایشگاه هستند.

بخش فروش در این پالایشگاه به دو بخش کلی فروش عاملین و فروش صنایع تقسیم می شود. از آنجایی که این شرکت محصولاتی متناسب با نیازهای صنایعی نظیر صنایع فولادی، دریایی و هوایی را تولید می کند، مشتریان مرتبط با این صنایع در بخش فروش

¹ Statistically Validated Networks

صنایع قرار می گیرند. همچنین محصولات خودرویی شرکت نظیر انواع روغن موتور و ضد یخ توسط عاملین فروش در دسترس عموم قرار می گیرد. کانال های فروش دیگری از جمله فروشگاه اینترنتی دیجی کالا برای برخی از این محصولات که مصرف عمومی دارند هم فعال هستند.

با همکاری واحد مهندسی فروش پالایشگاه، داده های مربوط به چالش ها و نقطه نظرات عاملین فروش از بخش پیشنهادات در نظرسنجی های شش ماهه شرکت استخراج شد و داده های مربوط به کاربران سایت پالایشگاه که همان نظرات کاربران در سایت بود دریافت شده است. به منظور بررسی نظرات گروهی از مصرف کنندگان محصولات پالایشگاه، نظرات موجود در صفحه مربوط به چهار محصول «روغن موتور شماره ۱»، «روغن موتور شماره ۴»، «ضد یخ شماره ۱» و «ضد یخ شماره ۲» که از نظر تعداد قابلیت اجرای مدل را داشتند، با استفاده از افزونه Instant Data Scraper مرورگر کروم استخراج شدند.

جدول ۱: وضعیت داده های نظرسنجی های شش ماهه پالایشگاه از نظر تعداد داده موجود در قسمت پیشنهادات

Table 1: Status of refinery's semi-annual survey data regarding the number of entries in the suggestions section

سال	نیمه اول / دوم	گروه پاسخ دهنده	تعداد داده در بخش پیشنهادات
۱۴۰۰	نیمه اول	صنایع	0
		عاملین فروش	0
	نیمه دوم	صنایع	5
		عاملین فروش	30
۱۴۰۱	نیمه اول	صنایع	0
		عاملین فروش	85
	نیمه دوم	صنایع	14
		عاملین فروش	67

با توجه به تعداد پیشنهادات موجود در نظرسنجی نیمه اول و دوم سال ۱۴۰۱ برای عاملین فروش که در جدول ۱ قابل مشاهده است، جهت پیاده سازی مدل داده های نظرسنجی مربوط به عاملین فروش در سال ۱۴۰۱ به عنوان یک سری داده در نظر گرفته شد.

جدول ۲: وضعیت داده های نظرات سایت پالایشگاه از نظر تعداد داده موجود

Table 2: Status of refinery website comment data in terms of the number of available data points

گروه	دسته	شماره	تعداد نظر	تعداد نظر	تعداد نظر
		محصول			
گروه ۱	روغن های خودرویی (بنزینی)	1	2	3	6
		2	14	9	16
		3	0	1	2
		4	46	54	30
		5	61	52	37
		6	9	8	2
	روغن های خودرویی (دنده)	7	2	2	0
		8	6	1	3
		9	4	3	6
		10	3	3	0
		11	1	1	0
		12	0	1	1
	روغن های خودرویی (دیزلی)	13	3	6	7
		14	6	3	7
		15	1	2	1
		16	0	0	1
		17	0	2	3
		18	1	1	0
تعداد داده ها			159	152	122
روغن های صنعتی	1	0	0	1	
	2	0	0	1	
روغن های صنعتی (کمپرسور)	3	1	0	0	
	تعداد داده ها			1	0
ضد یخ	ضد یخ	1	1	5	7
	2	0	0	1	1
تعداد داده ها			2	5	9
گریس	گریس	1	0	1	0
	2	0	1	1	1
تعداد داده ها			0	2	1
وکس	وکس	1	0	0	1
	تعداد داده ها			0	0
-	-	1	0	1	1
	2	0	0	1	1
تعداد داده ها			0	0	1
تعداد داده ها			0	1	3

با توجه به تعداد نظرات روغن های خودروبی در سه بازه زمانی سال های ۱۳۹۸ تا ۱۳۹۹، ۱۳۹۹ تا ۱۴۰۰ و ۱۴۰۰ تا ۱۴۰۲ موجود در جدول ۲، این سه سری داده برای پیاده سازی مدل مناسب شناخته شدند.

جدول ۳: معرفی محصولات مورد بررسی و وضعیت داده های نظرات کاربران دیجی کالا مربوط به چهار محصول پالایشگاه از نظر تعداد

Table 3: Overview of the products under study and the status of user review data from Digikala for the four selected products, in terms of quantity

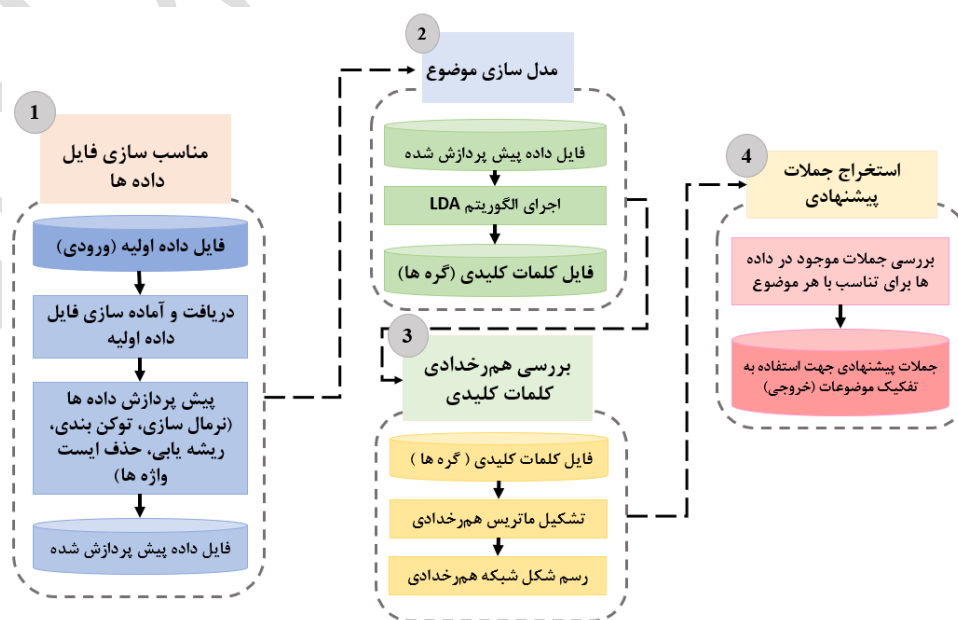
محصول منتخب	لینک صفحه محصول در دیجی کالا	تعداد نظرات
روغن موتور شماره ۱	B2n.ir/j40743	1523
روغن موتور شماره ۴	B2n.ir/g89488	771

تعداد نظرات	لینک صفحه محصول در دیجی کالا	محصول منتخب
263	B2n.ir/h41100	ضد یخ شماره ۱
75	B2n.ir/u62949	ضد یخ شماره ۲

از بین محصولاتی که در فروشگاه اینترنتی دیجی کالا عرضه می شد، چهار محصول تعداد نظرات بیشتر از ۵۰ عدد داشتند که قابلیت پیاده سازی مدل روی آن ها وجود داشت. لذا نظرات این چهار محصول از سایت دیجی کالا استخراج شده و مورد مطالعه قرار گرفت. همان طور که در جدول ۳ مشاهده می شود، روغن موتور شماره ۱ بیشترین تعداد نظرات و ضد یخ شماره ۲ کمترین تعداد نظرات را بین محصولات انتخاب شده دارند.

۳-۲- مدل پیشنهادی پژوهش

شکل زیر، مدل پیشنهادی جهت شناسایی چالش ها و نقطه نظرات ذینفعان با استفاده از تحلیل هم رخدادی واژگان کلیدی را نشان می دهد. این مدل به طور کلی از چهار بخش تشکیل شده و مراحل هر بخش به طور مختصر در شکل ۳ توضیح داده شده است.



شکل ۳: مدل پیشنهادی پژوهش

Figure 4: Proposed research model

- مناسب سازی فایل داده ها

پیش پردازش داده ها شامل مراحل مختلفی است که متناسب با داده ها و روش مورد استفاده، می توان برخی مراحل آن را انجام داد.

برای انجام پیش پردازش داده ها، تعدادی از ابزارهای پیش پردازش متن فارسی نظیر پارسی ور^۱، فارسی یار^۲، دادماتولز^۳ و هضم^۴ مورد بررسی قرار گرفتند که با توجه به ادبیات فارسی موجود در حوزه متن کاوی و آزمایش های انجام شده در استفاده از این نرم افزارها، ابزار هضم برای نرمال سازی و جداسازی کلمات و ابزار پرشین استمر^۵ برای ریشه یابی کلمات استفاده شده است. همچنین حذف ایست واژه ها با استفاده از لیست ایست واژه های فارسی آقای وحید خرازی و گروه *Awesome Persian NLP* و افزودن برخی موارد به صورت دستی پس از اجرای متعدد الگوریتم روی داده ها انجام شده است.

مراحل مختلف پیش پردازش عبارتند از :

۱. نرمال سازی^۶ داده ها: هدف اصلی در بخش نرمال سازی، یکپارچه سازی و متحدالشکل کردن کاراکتر ها به منظور کنترل بیشتر بر نتایج پردازش های پیشرفته تر بر متن است که در این پژوهش برای نرمال سازی داده ها از کتابخانه هضم در محیط *IDLE* پایتون استفاده شده است. (<https://data-hub.ir>)

نمودار مواردی که در ابزار نرمال سازی هضم انجام می شود در شکل زیر آمده است.

¹ Parsivar

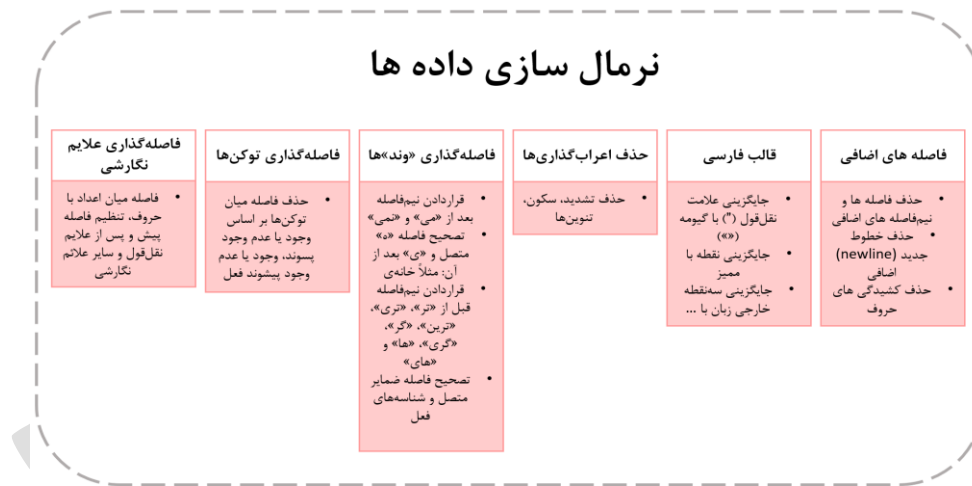
² Farsiyar

³ Dadma tools

⁴ Hazm

⁵ Persian Stemmer

⁶ Normalization



شکل ۴: فرایند نرمال سازی داده ها با استفاده از کتابخانه هضم

Figure 5: Data normalization process using the Hazm library

توکن بندی^۱ داده ها: در کتابخانه هضم این امکان فراهم شده است که به کمک کلاس *Token Splitter*، ساده یا مرکب بودن کلمات را بررسی نمود و در صورت مرکب بودن، کلمه را به کلمات ساده تشکیل دهنده آن شکست. متد *split_token_words* ابتدا کلمات را بر اساس نیم فاصله (در صورت وجود) می شکند و لیستی از کاندیدها را (برای استخراج کلمات ساده از کلمه اصلی) به وجود می آورد. سپس تمامی زیررشته های ناهمپوشان (با حداقل طول ۱ و حداکثر طول ۱) را بن یابی کرده و با حذف حالت های تکراری، حالت هایی را که در دادگان کلمات ابزار هضم وجود دارند به عنوان کاندیدها به عنوان خروجی نشان می دهد [۳۹].

ریشه یابی^۲ کلمات: ریشه جزئی از کلمه است که پس از حذف تمام عناصر الحاقی یا وندهای کلمه «پسوند، پیشوند، میانوند» بر جای می ماند. در این پژوهش، ریشه یابی کلمات با استفاده از کتابخانه *persian stemmer* در محیط *IDLE* پایتون انجام شده است.

حذف ایست واژه ها: جهت کاهش حجم و افزایش کیفیت داده ها و همچنین جلوگیری از افزایش احتمال خطا در الگوریتم، کلماتی مثل حروف اضافه (در، به، با و) و علائم نگارشی حذف می شوند. برای انجام این کار با دقت بالاتر، در این پژوهش اقدام به استفاده از دو لیست ایست واژه های فارسی به صورت همزمان نموده و به وسیله کد پایتون این کلمات را از لیست داده ها حذف کردیم.

با توجه به اینکه ابزار کارآمدی برای اصلاح اشتباهات تایپی و املائی برای زبان فارسی یافت نشد، پس از انجام چهار مرحله فوق، با استفاده از کتابخانه *genism* در محیط *IDLE* پایتون یک دیکشنری از کلمات تشکیل داده و دیکشنری را به صورت فایل *Excel* از

¹ Tokenization

² Stemming

کد خروجی گرفتیم. همچنین داده های پیش پردازش شده نیز به صورت فایل *Excel* خروجی گرفته شد. سپس فایل دریافت شده چک شده و مواردی که به دلیل وجود کلمات مترادف کلمه یا اشتباه تایپی بیش از یک بار در دیکشنری ظاهر شده اند با یک کلمه واحد یا شکل صحیح تایپی و املایی جایگزین شده اند. همچنین مواردی که علیرغم نرمال سازی متن و استفاده از لیست ایست واژه ها موجب کاهش دقت الگوریتم بودند حذف شدند. در نهایت، برای هر سری داده دو فایل *Excel* شامل موارد تغییر یافته و موارد حذف شده تشکیل شده و در قسمت دوم کد در محیط *Jupyter Notebook* روی فایل داده پیش پردازش شده که از مرحله اول کد خروجی گرفته شده بود، اعمال شدند.

پس از اعمال تغییرات ثانویه، مجدداً دیکشنری تشکیل شده و در باقی مراحل از این دیکشنری استفاده شده است.

• مدل سازی موضوع

در این فاز، روی داده های پیش پردازش شده که به عنوان ورودی از فاز قبل دریافت شده است، الگوریتم *LDA* با استفاده از زبان برنامه نویسی پایتون اجرا شده و نتایج الگوریتم در غالب فایل اکسل به صورتی که برای ورود به نرم افزار گفی¹ مطلوب است، ذخیره شده است. همچنین از این فایل در فاز بعدی برای تشکیل ماتریس هم رخدادی واژگان، استفاده خواهد شد. پارامترهای الگوریتم بر اساس جدول زیر تنظیم شده است.

¹ Gephi

جدول ۴: پارامترهای الگوریتم LDA
Table 4: LDA algorithm parameters

پارامتر	مقدار تنظیم شده	شرح
تعداد مرور کل مجموعه داده ها	40	به عنوان پارامتر ثابت آموزش در تمام مدل ها تنظیم شد تا شرایط آموزشی مدل ها یکسان باشد.
بازه مورد بررسی برای تعداد موضوع بهینه	1 تا 20 موضوع	به منظور تعیین تعداد مناسب موضوعات، مدل در بازه ۱ تا ۲۱ موضوع اجرا و برای هر اجرا مقدار معیار C-V Coherence محاسبه شد.
تعداد موضوعات	20	بر اساس بیشترین انسجام (Coherence) تعداد 20 موضوع انتخاب شده است.
مقدار اولیه مولد اعداد تصادفی ^۱	100	به منظور تضمین تکرارپذیری نتایج، مقدار Random State برابر با 100 برای مقداردهی اولیه مولد اعداد شبه تصادفی در اجرای مدل LDA تنظیم شد.
حداکثر تعداد تکرارها ^۲	200	به منظور دستیابی به همگرایی مناسب مدل، حداکثر تعداد تکرارهای الگوریتم (Iterations) برابر با 200 تنظیم شد.
پارامتر α	Symmetric	پارامتر Alpha (α) به صورت Symmetric تنظیم شد؛ به این معنا که در آغاز آموزش، احتمال اولیه یکسانی برای تمامی موضوعات در نظر گرفته شد.
پارامتر β	مقدار پیش فرض کتابخانه Gensim	برای پارامتر Beta (β) از مقدار پیش فرض کتابخانه Gensim استفاده شد.
اندازه دسته پردازش ^۳	10000	اندازه دسته پردازش (Chunksize) برابر با 10000 در نظر گرفته شد. با توجه به اینکه تعداد اسناد کمتر از این مقدار بود، کل داده ها در یک دسته پردازش شدند.
تناوب به روز رسانی ^۴	1	مقدار پارامتر Update Every برابر با 1 تنظیم شد؛ بدین معنا که مدل پس از پردازش هر دسته از داده ها به روزرسانی می شود.
اختصاص موضوع به هر واژه ^۵	True	گزینه Per Word Topics فعال شد تا علاوه بر توزیع موضوع در سطح اسناد، اطلاعات مربوط به ارتباط واژه ها با موضوع ها نیز محاسبه شود.
معیار انسجام موضوع	C-V	برای ارزیابی کیفیت موضوع های استخراج شده از معیار C-V Coherence استفاده شد. به این صورت که برای هر سری داده، این امتیاز برای تعداد ۱ تا ۲۱ موضوع محاسبه شده و تعداد موضوعی که بیشترین امتیاز انسجام را داشته انتخاب شده است. در صورتی که روند امتیازها در بازه ۱ تا ۲۱ همچنان صعودی بوده باشد، امتیاز انسجام برای تعداد موضوع بیشتر تا پیدا کردن امتیاز بیشینه ادامه یافته است.
تعداد واژه های نمایش داده شده در هر موضوع	6	برای تفسیر هر موضوع، شش واژه با بیشترین وزن در خروجی مدل گزارش شد.

¹ Random-state

² Iterations

³ Chunksize

⁴ Update-every

⁵ Per-word-topics

• بررسی هم‌رخدادی کلمات کلیدی

در این فاز، با استفاده از فایل کلمات کلیدی به عنوان گره های شبکه هم‌رخدادی و فایل داده های پیش پردازش شده، ماتریس هم‌رخدادی را تشکیل داده ایم. در این ماتریس، مقادیر درایه ها صفر یا یک می باشد. هر ستون ماتریس به یک کلمه کلیدی و هر ردیف ماتریس به یک ردیف از داده ها مربوط است. در صورتی که کلمه مربوط به ستون زام در ردیف نام وجود داشته باشد، درایه a_{ij} ماتریس مقدار یک می گیرد و در غیر این صورت، مقدار درایه صفر خواهد بود. با استفاده از این ماتریس، یال های شبکه هم‌رخدادی به دست می آید و می توان شکل این شبکه را رسم کرد. پس از تشکیل ماتریس، فایل یال های گراف هم‌رخدادی به صورتی که برای ورود به نرم افزار گفی مطلوب باشد، ذخیره شده است. سپس با استفاده از فایل های گره ها و یال ها برای هر سری داده ها، شکل شبکه هم‌رخدادی واژگان کلیدی با استفاده از نرم افزار گفی که یک نرم افزار منبع باز برای تجسم و تجزیه و تحلیل نمودارهای شبکه های بزرگ است، رسم شده است و به عنوان خروجی فاز سوم مدل در نظر گرفته شده است.

• استنتاج عبارت ها و محاسبه درصد رخداد آن ها

با توجه به یال هایی که وزن بیشتری در شبکه دارند، می توان هم‌رخدادی هایی که بیشتر از بقیه اتفاق افتاده اند را مشاهده کرد. با ایجاد ارتباط بین یال های با وزن بیشتر و گره های مرتبط با آنها، می توان عبارات استنتاجی را به دست آورد.

هر عبارت از طریق وجود ارتباط بین تعدادی کلمه مشخص می شود. برای محاسبه درصد رخداد هریک از این عبارات در فایل داده ها، پس از قرار دادن یک نام نماینده برای عبارت، هم‌رخدادی هایی که نشان دهنده عبارت بوده اند را به عنوان حالت های رخداد عبارت تعریف می کنیم و یک ماتریس ثانویه تشکیل می دهیم که در صورت بروز هریک از این حالت ها در هر ردیف ماتریس هم‌رخدادی، در ستون آخر ماتریس ثانویه نام نماینده عبارت مورد نظر نوشته می شود. در نهایت با شمارش تعداد هریک از عبارات و محاسبه درصد رخداد عبارات استنتاجی به دست می آید. در این پژوهش، این کار با استفاده از زبان پایتون و در محیط *Jupyter Notebook* انجام شده است. در ادامه به عنوان مثال برای عبارت « بسته بندی با ظرف چلیک ضعیف است » که یکی از عبارات مستخرج از داده های « چالش ها و نقطه نظرات عاملین پالایشگاه در سال ۱۴۰۱ » می باشد، مراحل محاسبه درصد رخداد بین داده های پیش پردازش شده را توضیح می دهیم. نام نماینده این عبارت، "A1401.2" قرار داده شده است.

۱. ابتدا کلماتی که وجود هم‌رخدادی های پرتکرار بین آنها نشان دهنده این عبارت است در جدول ۵ به همراه شماره تخصیص یافته به هر کلمه مشخص می شود. شماره تخصیص یافته به هر کلمه در لیست کلیدواژه های ایجاد شده مشخص می شود. کلماتی که نشان دهنده عبارت « بسته بندی با ظرف چلیک ضعیف است » می باشند عبارت است از: بسته بندی، ضعیف، ظرف و چلیک.

جدول ۵: شماره تخصیص یافته به کلمات مثال

Table 5: Numbers assigned to example words

کلمه	شماره تخصیص یافته
ضعیف	0

کلمه	شماره تخصیص یافته
چلیک	1
بسته بندی	4
ظرف	8

۲. یال‌های ایجادشده بین کلمات فوق در جدول ۶ در سه ستون « مبدا » ، « مقصد » و « وزن » مشخص شده اند که به ترتیب مربوط به نقطه شروع یال‌ها، نقطه پایان یال‌ها و وزن یال‌ها هستند.

جدول ۶: یال‌های موجود بین کلمات مثال

Table 6: Edges between the example words

مبدا	مقصد	وزن
1	8	8
0	4	6
1	4	6
0	1	5
0	8	4
4	8	4

۳. با استفاده از قطعه شبه کد شکل ۵، درصد رخداد عبارت بین داده‌های پیش‌پردازش شده محاسبه می‌شود. همچنین اینکه عبارت موردنظر در کدام ردیف‌های داده رخ داده است نیز در ماتریس ثانویه^۱ مشخص می‌شود که در صورت نیاز قابل دسترسی است.

¹ Secondary Matrix

```

Initialize matrix as input

Copy matrix to copied_matrix

Define states dictionary with key-value pairs where keys are state names and
values are lists of tuples representing conditions

Convert copied_matrix to object type

For each row index i in range of copied_matrix rows:

    For each state and its conditions in states dictionary:

        If any condition in the state's conditions matches copied_matrix[i, col-1]
        equals 1 for any column col in the condition tuple:

            Set copied_matrix[i, last column] to state

            Break loop to move to next row

Set secondary_matrix equal to copied_matrix

Extract last column from secondary_matrix

Count occurrences of "A1401.2" in last column

Calculate total members in last column

Calculate percentage of "A1401.2" occurrences by dividing count_A1401_2
by total_members and multiplying by 100

```

شکل ۵: قطعه شبه کد تشکیل ماتریس هم رخدادی واژگان

Figure 5: Pseudocode snippet for constructing the word co-occurrence matrix

۴- یافته های پژوهش

در این پژوهش، رویکرد پیشنهادی روی داده های نظرات و چالش های عاملین فروش پالایشگاه، کاربران سایت پالایشگاه و نظرات موجود در صفحه مربوط به تعدادی از محصولات پالایشگاه در فروشگاه اینترنتی دیجی کالا اجرا شده و نتایج آن تحلیل شده است. این پالایشگاه یکی از شرکت های فعال در تولید انواع روغن های پایه، موتور، صنعتی و ... است. این شرکت، نخستین مجموعه روانکاری در صنعت کشور است که فناوری های مربوط به تولید فرآورده های روانکار را به کشور منتقل کرده و صد ها مهندس و کارشناس را در طول خدمت در این شرکت با فناوری های به روز و ساخت و تولید روانکارها آشنا کرده است.

معیار انتخاب محصول برای اجرای مدل روی نظرات آن در فروشگاه اینترنتی دیجی کالا، کفایت تعداد نظرات موجود بوده است. همچنین داده های مربوط به نظرسنجی های شش ماهه شرکت نیز با همکاری واحد مهندسی فروش پالایشگاه دریافت شد که بر اساس تعداد داده های موجود در قسمت پیشنهادات، نظرسنجی عاملین در سال ۱۴۰۱ (نیمه اول و نیمه دوم سال) برای پیاده سازی مدل انتخاب شد. علاوه بر این، داده های نظرات کاربران در سایت شرکت در سه بازه زمانی بین سال های ۱۳۹۸ تا ۱۴۰۲ نیز با

همکاری واحد مهندسی فروش دریافت شده و با توجه به دسته بندی نظرات بر اساس محصولات مربوطه، نظرات مربوط به روغن های خودرویی شرکت از نظر تعداد برای پیاده سازی مدل از دسته های دیگر شرایط بهتری داشت و جهت پیاده سازی مدل انتخاب شد. در ادامه، نتایج اجرای فاز های مختلف مدل روی داده ها را در قالب کلمات کلیدی و تعداد تکرار آنها، شبکه هم رخدادی واژگان کلیدی و استخراج عبارت های استنتاجی به تفکیک برای داده های مورد بررسی ارائه شده است.

۴-۱- کلمات کلیدی و تعداد تکرار آن ها

کلمات کلیدی استخراج شده برای داده های مورد نظر، همان کلماتی هستند که در نتیجه اجرای الگوریتم LDA روی داده ها ذیل موضوعات مشاهده شدند. این کلمات به همراه تعداد تکرارشان در داده مربوطه برای هر سری از داده های مورد بررسی به تفکیک در پیوست الف آمده اند.

۴-۲- شبکه هم رخدادی واژگان کلیدی

پس از مشخص شدن کلمات کلیدی، ماتریس هم رخدادی این کلمات در سراسر متن به دست آمده و یال های گراف به دست می آیند. شکل گراف هم رخدادی واژگان رسم شده برای داده های مورد نظر در پیوست الف به تفکیک داده ها آمده است.

۴-۳- استخراج عبارت های استنتاجی

از مشاهده نتایج مربوط به پیاده سازی الگوریتم LDA روی داده های مورد نظر و رسم شبکه هم رخدادی برای آنها، عباراتی استنتاج شدند که برای هر سری داده در جداول مجزا به همراه کلمات مربوطه و درصد وجود آنها بین داده های مربوطه، در پیوست الف مشاهده می شوند.

لازم به ذکر است که برای حصول اطمینان بیشتر از درستی نتایج مدل، در مواردی که تعداد داده ها کمتر از ۱۰۰ ردیف بوده است، درستی نتایج و موضوعات بدست آمده به صورت دستی چک شده است.

• عبارات استخراج شده از چالش ها و نقطه نظرات عاملین در سال ۱۴۰۱

با توجه به اینکه تکرار زیاد هم رخدادی بین برخی کلمات از جمله «بسته بندی، ضعیف، ظرف، چلیک، تنوع، ضعیف، کارتن، رنگ بندی، اجبار، پکیج، جنس، خرید، رقیب، تبلیغ، لباس تبلیغاتی، تابلو تبلیغاتی» در شکل الف- ۱ نشان دهنده عباراتی است که تکرار بیشتری داشته اند، چهار عبارت که در

عبارت های استخراج شده از شبکه هم رخدادی واژگان برای سری های مختلف داده ها

جدول الف- ۹ آمده اند، استنتاج شدند.

همان طور که در

عبارت های استخراج شده از شبکه هم رخدادی واژگان برای سری های مختلف داده ها

جدول الف- ۹ مشاهده می شود، عبارت «بسته بندی با ظرف چلیک ضعیف است» بیشترین درصد رخداد را دارد. در واقع، بیشترین چالش مطرح شده توسط عاملین پالایشگاه در سال ۱۴۰۱، مربوط به ضعف ظرف چلیک استفاده شده در بسته بندی محصولات پالایشگاه است. در این راستا، پالایشگاه می تواند برای افزایش سطح رضایت عاملین و تلاش در راستای افزایش تعداد عاملین فروش خود به وسیله تبلیغات دهان به دهان، این موضوع را بررسی کرده و از راهکاری نظیر تغییر مظروف مورد استفاده در بسته بندی یا تقویت جنس بسته بندی چلیک استفاده نماید. در واقع، استفاده از استراتژی توسعه محصول در این مورد توصیه می شود.

- عبارات استخراج شده از نظرات کاربران سایت شرکت در مورد روغن های خودرویی در بازه زمانی ۱۳۹۸ تا ۱۳۹۹

با توجه به اینکه تکرار زیاد هم رخدادی بین برخی کلمات از جمله «پیشنهاد، 10W40، 20W50، روغن موتور شماره ۴، روغن موتور شماره ۵، فلومکس، کیفیت، شماره ۴، تفاوت» در شکل الف- ۳ نشان دهنده عباراتی است که تکرار بیشتری داشته اند، دو عبارت که در

جدول الف- ۱۰ آمده اند، استنتاج شدند.

با توجه به درصد رخداد محاسبه شده، عبارت «تقاضای ارائه محصول پیشنهادی مناسب خصوصا از نظر گرانروی برای انواع خودرو ها» از بیشترین درصد رخداد برخوردار است. لذا شرکت می تواند برای پیشرفت عملکرد خود در بخش فروش و در راستای کمک به انتخاب صحیح محصول از سوی خریداران، نسبت به راه اندازی یک سیستم توصیه گر بر مبنای نوع و سال تولید خودرو یا اضافه کردن چنین مواردی در بخش فیلترها در سایت، برای ارائه راهنمایی بهتر به مشتریان در سایت پالایشگاه اقدام نماید. راهکار دیگری که در این مورد توصیه می شود این است که در رابطه با اینکه هر کدام از محصولات خودرویی این پالایشگاه برای چه خودروهایی مناسب هستند، اطلاعاتی در بخش مقالات سایت اضافه شود. به عبارت دیگر، ارائه راهنمایی و اطلاعات در این حوزه یا ایجاد سیستم توصیه محصولات بر اساس ویژگی های انتخاب شده برای خودرو توسط مشتریان، موجب افزایش رضایت مشتریان از خرید شده و از دو طریق تبلیغات دهان به دهان و بازاریابی دیجیتال، موجب افزایش حجم فروش خواهد شد. به عبارتی استراتژی های تکنولوژی اطلاعات یا نفوذ در بازار به دلیل افزایش تلاش های فروش برای این مورد مناسب است.

- عبارات استخراج شده از نظرات کاربران سایت شرکت در مورد روغن های خودرویی در بازه زمانی ۱۳۹۹ تا ۱۴۰۰

با توجه به اینکه تکرار زیاد هم رخدادی بین برخی کلمات از جمله «اصالت، بازار، تقلبی، فلومکس، روغن موتور شماره ۴، کیفیت، روغن موتور شماره ۲، euro 4، پراید، راهنمایی، TU5» در شکل الف- ۴ نشان دهنده عباراتی است که تکرار بیشتری داشته اند، دو عبارت که در

جدول الف- ۱۱ آمده اند، استنتاج شدند.

بیشتر نظرات کاربران سایت پالایشگاه در بازه زمانی ۱۳۹۹ تا ۱۴۰۰، به نگرانی در مورد اصالت کالا و وجود کالای تقلبی در بازار معطوف است. لذا پالایشگاه می تواند از طریق ایجاد رمزین^۱ه یا رمزین^۲ه پاسخ سریع^۳ بر روی کالاهای خود و اطلاع رسانی این موضوع از طریق کانال های تبلیغاتی شرکت به مصرف کنندگان، به آنها اطمینان دهد که می توانند از اصالت محصولی که خریداری کرده اند مطمئن شوند.

- عبارات استخراج شده از نظرات کاربران سایت شرکت در مورد روغن های خودرویی در بازه زمانی ۱۴۰۰ تا ۱۴۰۲

با توجه به اینکه تکرار زیاد هم رخدادی بین برخی کلمات از جمله « روغن موتور شماره ۲، قوطی، تعویض، فلزی، پلاستیکی، گالن، کیفیت، تیب، دوگانه سوز، پیشنهاد، پراید، راهنمایی، اصالت، محصول» در شکل الف-۵ نشان دهنده عباراتی است که تکرار بیشتری داشته اند، سه عبارت که در جدول الف-۱۲ آمده اند، استنتاج شدند.

عبارت «پلاستیکی یا فلزی بودن مظروف محصول روغن موتور شماره ۲» بیشترین درصد رخداد را به خود اختصاص داده است که با توجه به اینکه این عبارت از داده های نظرات سایت پالایشگاه استخراج شده است، نشان دهنده ابهام در مورد مظروف این محصول است. در صورتی که اخیرا تغییری در بسته بندی این محصول صورت گرفته است، شرکت می تواند از طریق کانال های ارتباطی و تبلیغاتی خود این موضوع را به اطلاع مشتریان برساند. وجود چنین ابهام هایی ممکن است موجب شود مشتریان به اصالت کالا مطمئن نباشند و به همین دلیل از خرید خود منصرف شوند.

- عبارات استخراج شده در مورد روغن موتور شماره ۱

با توجه به اینکه تکرار زیاد هم رخدادی بین برخی کلمات از جمله « قیمت، بصرفه، پایین، ارزان، اقتصادی، نشتی، قوطی، بسته، ضربه، ریخت، خودرو، قدیمی، پراید، کاربراتور، دیجی کالا، بازار، منصفانه» در شکل الف-۶ نشان دهنده عباراتی است که تکرار بیشتری داشته اند، چهار عبارت که در

جدول الف-۱۳ آمده اند، استنتاج شدند.

بین عبارات استخراج شده از نظرات کاربران سایت دیجی کالا در مورد روغن موتور شماره ۱، عبارت «قیمت مناسب و اقتصادی روغن» بیشترین درصد رخداد را دارد. اما دومین عبارت پرتکرار، به ضربه خورده بودن قوطی و نشت مواد از قوطی اشاره دارد که

با توجه به همکاری دیجی کالا در خرید انجام شده توسط این دسته از مشتریان، این موضوع باید علت یابی و رفع شود. چرا که این موضوع هم موجب نارضایتی مشتریان شده، هم ممکن است بر تصمیم خرید مشتریان بالقوه تاثیر منفی بگذارد.

- عبارات استخراج شده در مورد روغن موتور شماره ۴

با توجه به اینکه تکرار زیاد هم‌رخدادی بین برخی کلمات از جمله « قیمت، بازار، دیجی کالا، دولتی، بصرفه، کیفیت، مناسب، روغن» در شکل الف- ۷ نشان دهنده عباراتی است که تکرار بیشتری داشته اند، دو عبارت که در

جدول الف- ۱۴ آمده اند، استنتاج شدند.

در مورد محصول روغن موتور پارس روغن موتور شماره ۴، بیشترین درصد رخداد مربوط به عبارت «تفاوت قیمت محصول در بازار و دیجی کالا و قیمت بصرفه محصول در دیجی کالا» می باشد. لذا شرکت با تامین موجودی این محصول در دیجی کالا می تواند این کانال فروش موثر را نگه دارد که خریداران بیشتری بتوانند با قیمت مناسب از این محصول بهره مند شوند و در نتیجه میزان فروش محصول افزایش یابد.

- عبارات استخراج شده در مورد ضد یخ شماره ۱

با توجه به اینکه تکرار زیاد هم‌رخدادی بین برخی کلمات از جمله « کیفیت، ضد یخ، قیمت، مناسب، روغنی، محصول، قوطی، تعویض، فلزی، بازار، دیجی کالا» در شکل الف- ۹ نشان دهنده عباراتی است که تکرار بیشتری داشته اند، سه عبارت که در جدول الف- ۱۵ آمده اند، استنتاج شدند.

بیشترین عبارت تکرار شده بین نظرات کاربران سایت دیجی کالا در مورد ضد یخ شماره ۱، «کیفیت و قیمت مناسب ضد یخ» می باشد که نشان دهنده رضایت مصرف کنندگان از این محصول است.

- عبارات استخراج شده در مورد ضد یخ شماره ۲

با توجه به اینکه تکرار زیاد هم‌رخدادی بین برخی کلمات از جمله « کیفیت، ضد یخ، قیمت، روغنی، محصول، قوطی، فلزی» در شکل الف- ۱۰ نشان دهنده عباراتی است که تکرار بیشتری داشته اند، دو عبارت که در

جدول الف- ۱۶ آمده اند، استنتاج شدند.

درصد رخداد نزدیک به ۷۰٪ برای عبارت «کیفیت خوب، روغنی بودن و قیمت ضد یخ» به دست آمده است که نشان دهنده مزایای این محصول در نگاه مصرف کنندگان آن است.

۵- بحث و نتیجه گیری

در این پژوهش، مدلی برای استخراج موضوعات مطرح شده در داده های متنی نظرات ذینفعان سازمان ارائه شد. اجزاء اصلی این مدل، مدل سازی موضوع و شبکه هم رخدادی واژگان هستند که با استفاده از نتایج به دست آمده از آنها عبارات مطرح شده در مجموعه داده ها مشخص می شوند و درصد رخداد آنها با توجه به کلمات و یال های مربوط به این عبارات در داده ها محاسبه می شود. این مدل روی نظرات گروهی از ذینفعان پالایشگاه شامل عاملین فروش پالایشگاه، کاربران سایت پالایشگاه و خریدارانی که از طریق فروشگاه های اینترنتی (فروشگاه دیجی کالا) اقدام به خرید محصولات پالایشگاه نموده اند پیاده سازی شده است که نتایج آن را به تفصیل در فصل چهارم شرح داده ایم. خلاصه نتایج به دست آمده در جدول ۷ آمده است.

جدول ۷: خلاصه نتایج به دست آمده از پژوهش

Table 23: Summary of results obtained from the research

عنوان داده	عبارت پر تکرار	درصد رخداد عبارت	پیشنهاد مدیریتی ارائه شده
چالش ها و نقطه نظرات عاملین فروش در سال ۱۴۰۱	بسته بندی با ظرف چلیک ضعیف است.	31.01%	بهبود جنس ظرف بسته بندی یا تغییر نوع ظرف / استراتژی توسعه محصول
نظرات کاربران سایت پالایشگاه در بازه زمانی ۱۳۹۸-۱۳۹۹	تقاضای ارائه محصول پیشنهادی مناسب خصوصا از نظر گرانروی برای انواع خودرو ها	32.26%	ارائه راهنمایی و اطلاعات در مورد اینکه هر محصول برای چه نوع و مدل خودروهایی مناسب است یا ایجاد سیستم توصیه محصول مناسب براساس ویژگی های خودروی مورد نظر خریدار/استراتژی های تکنولوژی اطلاعات یا استراتژی نفوذ در بازار
نظرات کاربران سایت پالایشگاه در بازه زمانی ۱۳۹۹-۱۴۰۰	نگرانی در مورد اصالت کالا و وجود کالای تقلبی در بازار	35.57%	ایجاد رمزینه یا رمزینه پاسخ سریع روی کالاها و اطلاع رسانی به مشتریان از طریق کانال های ارتباطی و تبلیغاتی شرکت
نظرات کاربران سایت پالایشگاه در بازه زمانی ۱۴۰۰-۱۴۰۲	پلاستیکی یا فلزی بودن مظروف روغن موتور شماره ۲	36.67%	اطلاع رسانی تغییرات بسته بندی از طریق کانال های ارتباطی و تبلیغاتی شرکت
نظرات کاربران سایت دیجی کالا در مورد روغن موتور شماره ۱	قیمت مناسب و اقتصادی روغن ضربه خورده بودن قوطی و نشت مواد از قوطی	21.99%	کنترل هزینه های تولید این محصول جهت ثبات قیمت
نظرات کاربران سایت دیجی کالا در مورد روغن موتور پارس روغن موتور شماره ۴	تفاوت قیمت محصول در بازار و دیجی کالا	17.44%	پیگیری برای علت یابی و رفع این مسئله
نظرات کاربران سایت دیجی کالا در مورد روغن موتور شماره ۴	کیفیت و قیمت مناسب ضد یخ	44.48%	فعال نگه داشتن کانال فروش اینترنتی از طریق دیجی کالا
نظرات کاربران سایت دیجی کالا در مورد ضد یخ شماره ۱	کیفیت و قیمت مناسب ضد یخ	34.39%	کنترل کیفیت و قیمت ضد یخ جهت تداوم رضایت مشتریان

عنوان داده	عبارت پر تکرار	درصد رخداد	پیشنهاد مدیریتی ارائه شده
		عبارت	
نظرات کاربران سایت	کیفیت خوب، روغنی بودن و قیمت ضد	60.94%	کنترل کیفیت و قیمت ضدیخ جهت تداوم
دیجی کالا در مورد ضد یخ			رضایت مشتریان
یخ شماره ۲			

برای روشن تر شدن جدول فوق، به عنوان مثال ردیف اول جدول که در مورد داده های چالش ها و نقطه نظرات عاملین در سال ۱۴۰۱ است، به طور مفصل توضیح داده می شود.

با توجه به کلمات کلیدی که به وسیله اجرای الگوریتم *LDA* روی سری داده های مذکور مشخص شد (جدول کلمات کلیدی استخراج شده و تعداد تکرار آنها برای سری های مختلف داده ها

جدول الف- ۱) و روابطی که بین این کلمات در شبکه هم رخدادی قابل مشاهده است (شکل الف- ۱ Error! Reference source not found.)، چند عبارت از این سری داده ها استخراج شد (

عبارت های استخراج شده از شبکه هم رخدادی واژگان برای سری های مختلف داده ها

جدول الف- ۹). همچنین بر اساس تکرار هم رخدادی کلمات کلیدی در سراسر سری داده ها، درصد رخداد هر عبارت مشخص شده است. عبارتی که بیشترین درصد رخداد را داشت، عبارت « بسته بندی با ظرف چلیک ضعیف است» می باشد که می توان با بهبود ظرف بسته بندی یا تغییر جنس ظرف که در دسته استراتژی های توسعه محصول قرار می گیرد، این عامل نارضایتی را از بین برد. جدول فوق شامل خلاصه نتایج بدست آمده است لذا برای هر سری داده ها عبارت هایی که بیشترین درصد رخداد را داشته اند، در این جدول قرار دارند.

تشکر و قدردانی

محققان از اساتید محترم دانشکده مهندسی صنایع و سیستم ها و همچنین پرسنل واحد مهندسی فروش پالایشگاه روانکار که در نگارش این مقاله همراهی نموده اند، کمال تشکر و قدردانی را اعلام می دارند.

منابع مالی

این پژوهش، با حمایت مالی پالایشگاه روانکار مورد مطالعه انجام شده است.

تعارض با منافع

هیچ تضادی در منافع در مورد انتشار این نسخه از پژوهش وجود ندارد. همه نویسندگان، نسخه نهایی ارسال شده را مطالعه و تایید کرده اند. نویسندگان تضمین می کنند که این پژوهش، اثر اصلی آن ها بوده، قبلا چاپ نشده و در حال حاضر تحت انتشار نمی باشد.

- [1] Kwon, H.-J., Ban, H.-J., Jun, J.-K., & Kim, H.-S. (2021). Topic Modeling and Sentiment Analysis of Online Review for Airlines. *Information*, 12(2), 78. <https://doi.org/10.3390/info12020078>
- [2] Dahiya, A., Gautam, N., & Gautam, P. K. (2021). Data mining methods and techniques for online customer review analysis: A literature review. *Journal of System and Management Sciences*, 11(3), 1-26. DOI:10.33168/JSMS.2021.0301
- [3] Konchady, M. (2006). *Text Mining Application Programming*. Charles River Media. https://www.academia.edu/107979269/Text_Mining_Application_Programming
- [4] Tanwar, G., Tripathi, S., Gaur, D., & Shakya, D. (2022). An approach to improve analysis of product reviews using sentiment analysis. In 2022 4th International Conference on Advances in Computing, Communication Control and Networking (ICAC3N) (pp. 1-6). IEEE. <https://ieeexplore.ieee.org/abstract/document/10074049>
- [5] Zhao, C., Peng, X., & Li, Z. (2023). The influence of online customer reviews on two-stage product strategy in a competitive market. *Annals of Operations Research*, 326(1), 411-503. <https://doi.org/10.1007/s10479-021-04352-2>
- [6] Sunil, N., & Shirazi, F. (2023). Customer Review Classification Using Machine Learning and Deep Learning Techniques. In *International Conference on Human-Computer Interaction*, Springer. https://doi.org/10.1007/978-3-031-35915-6_42
- [7] Farkhod, A., Abdusalomov, A., Makhmudov, F., & Cho, Y. I. (2021). LDA-based topic modeling sentiment analysis using topic/document/sentence (TDS) model. *Applied Sciences*, 11(23), 11091. <https://doi.org/10.3390/app112311091>
- [8] Osmani, A., & Mohasefi, J. B. (2023). Opinion Mining Using Enriched Joint Sentiment-Topic Model. *International Journal of Information Technology & Decision Making*, 22(01), 313-375. <https://doi.org/10.1142/S0219622022500584>
- [9] Shashaj, A., Stirparo, D., & Kazemi, M. (2021). CEA-TM: A Customer Experience Analysis Framework Based on Contextual-Aware Topic Modeling Approach. In *Artificial Intelligence Applications and Innovations: 17th IFIP WG 12.5 International Conference, AIAI 2021* (pp. 17-32). Springer. https://doi.org/10.1007/978-3-030-79150-6_52
- [10] Farzadnia, S., & Vanani, I. R. (2022). Identification of opinion trends using sentiment analysis of airlines passengers' reviews. *Journal of Air Transport Management*, 103, 102232. <https://doi.org/10.1016/j.jairtraman.2022.102232>
- [11] Engez, A., & Aarikka-Stenroos, L. (2023). Stakeholder contributions to commercialization and market creation of a radical innovation: bridging the micro-and macro levels. *Journal of Business & Industrial Marketing*, 38(13), 31-44. <https://doi.org/10.1108/JBIM-03-2022-0136>
- [12] Ratniasih, N. L., & Jayanti, N. W. N. (2022). Sentiment Analysis of Stakeholder Satisfaction Measurement. *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*, 16(2), 217-226. DOI: 10.22146/ijccs.72245, <https://pdfs.semanticscholar.org/aea2/ca9890d1026d409a73ec39b38d312111264d.pdf>
- [13] Belyaeva, Z., Petrosyan, A., & Shams, S. R. (2022). Stakeholder data analysis in the video gaming industry: implications for regional development. *EuroMed Journal of Business*, 17(3), 333-349. <https://doi.org/10.1108/EMJB-10-2021-0150>
- [14] Kumar, S., & Kumar, N. (2022). Recent Trends in Opinion Mining using Machine Learning Techniques. In *International Conference on Innovative Computing and Communications: Proceedings of ICICC 2022, Volume 3*. Springer. https://doi.org/10.1007/978-981-19-3679-1_31
- [15] Chen, J., Du, S., & Yang, S. (2022). Mining and evolution analysis of network public opinion concerns of stakeholders in hot social events. *Mathematics*, 10(12), 2145. <https://doi.org/10.3390/math10122145>
- [16] Helan, A., & Sultani, Z. N. (2023). Topic modeling methods for text data analysis: A review. *AIP Conference Proceedings*, 2457(1). <https://doi.org/10.1063/5.0118679>
- [17] Soni, A., Rasool, A., Dubey, A., & Khare, N. (2022). Visualizing Commenters Opinions Through Topic Analysis. In *ICT Analysis and Applications: Proceedings of ICT4SD 2022* (pp. 249-257). Springer. https://doi.org/10.1007/978-981-19-5224-1_27
- [18] Xue, J., Shen, G. Q., Li, Y., Wang, J., & Zafar, I. (2020). Dynamic stakeholder-associated topic modeling on public concerns in megainfrastructure projects: Case of Hong Kong–Zhuhai–Macao bridge. *Journal of Management in Engineering*, 36(6), 04020078. [https://doi.org/10.1061/\(ASCE\)ME.1943-5479.000084](https://doi.org/10.1061/(ASCE)ME.1943-5479.000084)
- [19] Han, S., Li, X., & Hwang, H. (2022). Analysis of news data on 'super app' using topic modeling. In *International Conference on Human-Computer Interaction*, Springer. https://doi.org/10.1007/978-3-031-19682-9_5
- [20] Ma, C., Du, J., Shao, Y., Li, A., & Guan, Z. (2022). A Rare Topic Discovery Model for Short Texts Based on Co-occurrence word Network. *arXiv preprint, arXiv:2207.00432*. <https://doi.org/10.48550/arXiv.2207.00432>
- [21] Zhu, B., Cai, Y., & Ren, H. (2023). Graph neural topic model with commonsense knowledge. *Information Processing & Management*, 60(2), 103215. [10.1016/j.ipm.2022.103215](https://doi.org/10.1016/j.ipm.2022.103215)
- [22] Austin, E., Trabelsi, A., Largeron, C., & Zaiane, O. R. (2022). Hierarchical Topic Model Inference by Community Discovery on Word Co-occurrence Networks. In *Australasian Conference on Data Mining*. Springer. https://doi.org/10.1007/978-981-19-8746-5_11
- [23] Lin, W., Wu, X., Wang, Z., Wan, X., & Li, H. (2022). Topic network analysis based on Co-occurrence time series clustering. *Mathematics*, 10(16), 2846. <https://doi.org/10.3390/math10162846>
- [24] Freeman, R. E., Harrison, J. S., Wicks, A. C., Parmar, B. L., & De Colle, S. (2010). *Stakeholder theory: The state of the art*. ISBN 978-0-521-13793-5, https://sirdi-csirsu.com/wp-content/uploads/2023/07/Stakeholder_Theory_The_State_of_the_Art.pdf
- [25] Rothaermel, F. T. (2019). *Strategic management*. McGraw-Hill., https://books.google.com/books/about/Strategic_Management.html?id=RkXluQEACAAJ
- [26] Hanson, D., Hitt, M. A., Ireland, R. D., & Hoskisson, R. E. (2016). *Strategic management: Competitiveness and globalisation*. Cengage AU. https://books.google.com/books/about/Strategic_Management_Competitiveness_and.html?id=sy5MEAAAQBAJ
- [27] Rabiei, A., Ebrahimzadeh, S., & Ansari Mahyari, Z. (2012). Market development strategies of Iran Water & Power Contracting Company in the Middle East [Master's thesis, Payame Noor University]. (In Persian). <https://ganj.irandoc.ac.ir/#/articles/6580a198d8bad710cab0b706ee9bb195>

- [28] Tasdighi, H. (2012). Identification of stakeholders in intelligent vehicle communication system project through clustering of stakeholder instances based on K-Means method. *Journal of Industrial Technology Development*, 9(18), 29-36. (In Persian). https://www.sid.ir/fa/VEWSSID/J_pdf/13000013901803.pdf
- [29] Ghalenovi, M., Motahari, F. N., & Bahaderestani, A. (2016). Designing a model of key stakeholder expectations in metro construction projects: A case study of Mashhad urban train lines [Master's thesis, Ferdowsi University of Mashhad]. (In Persian). <https://ganj.irandoc.ac.ir/#/articles/1a4a0945cd71235fa517bc1ae432ba94>
- [30] Niki, F., Saadatmehr, A., & Aldozi, M. (2019). Prioritization of factors affecting market development strategy selection using fuzzy AHP method: A case study of Tabriz nuts and dried fruits cluster [Master's thesis, Al-Ghadir Institute of Higher Education]. (In Persian). <https://ganj.irandoc.ac.ir/#/articles/290dde3b55c1e1b9bd6d738df404d0a3>
- [31] Papadimitriou, C. H., Tamaki, H., Raghavan, P., & Vempala, S. (1998, May). Latent semantic indexing: A probabilistic analysis. In *Proceedings of the seventeenth ACM SIGACT-SIGMOD-SIGART symposium on Principles of database systems* (pp. 159-168). <https://dl.acm.org/doi/pdf/10.1145/275487.275505>
- [32] Gheyouni, M. (2021). Computational analysis of the evolution of Persian-language and linguistics articles. *Zaban va Zaban-shenasi*, 17(33), 99-130. (In Persian). https://journals.iics.ac.ir/article_8089_f7c50bfb93fd12a1a5b3f5404ccc1ed2.pdf
- [33] Grajzl, P., & Murrell, P. (2022). Using Topic-Modeling in Legal History, with an Application to Pre-Industrial English Case Law on Finance. *Law and History Review*, 40(2), 189-228. <https://doi.org/10.1017/S0738248022000153>
- [34] Small, H., & Griffith, B. C. (1974). The structure of scientific literatures I: Identifying and graphing specialties. *Science Studies*, 4(1), 17-40. <https://doi.org/10.1177/030631277400400102>
- [35] Asare, F., Farajpehloo, A., & Katbi, F. (2014). Co-occurrence analysis of knowledge management vocabulary in Web of Science database from 1993 to 2012 for visualization of vocabulary infrastructure [Master's thesis, Shahid Chamran University of Ahvaz]. (In Persian). <https://ganj.irandoc.ac.ir/#/articles/18cbce8386c8606b1d948ffb53709439>
- [36] Ahmadi, M., & Asare, F. (2017). A review of co-word analysis functions. *Motale'at Ketabdarie va Sazmandehie Etela'at*, 28(1), 125-145. (In Persian). https://nastinfo.nlai.ir/article_1132.html
- [37] Simonetti, A., Albano, A., Plaia, A., & Tumminello, M. (2023). Ranking coherence in topic models using statistically validated networks. *Journal of Information Science*, 01655515221148369. <https://doi.org/10.1177/01655515221148369>
- [38] Murakami, R., & Chakraborty, B. (2021). Neural topic models for short text using pretrained word embeddings and its application to real data. In *2021 IEEE 4th International Conference on Knowledge Innovation and Invention (ICKII)*, IEEE. <https://doi.org/10.1109/ICKII51822.2021.9574752>
- [39] Data-Hub. (n.d.). Data Hub. <https://data-hub.ir>
- [40] Hofmann, T. (1999, August). Probabilistic latent semantic indexing. In *Proceedings of the 22nd annual international ACM SIGIR conference on Research and development in information retrieval* (pp. 50-57). <https://dl.acm.org/doi/pdf/10.1145/312624.312649>

پیوست الف

- جدول کلمات کلیدی استخراج شده و تعداد تکرار آنها برای سری های مختلف داده ها

جدول الف- ۱: کلمات کلیدی استخراج شده و تعداد تکرار آنها برای داده های چالش ها و نقطه نظرات عاملین در سال ۱۴۰۱

Table A-1: Extracted keywords and their frequencies for data on challenges and stakeholder viewpoints in the year 1401

کلمه	تعداد تکرار	کلمه	تعداد تکرار	کلمه	تعداد تکرار
پکیج	24	سطل	6	حالت	3
ضعیف	17	همکاری	6	سیستم	3
بسته بندی	17	صادر	6	بشکه	2
حواله	16	افزایش	6	قوی	2
تبلیغ	16	ضد یخ	6	وجه	2
کارتن	16	سه میه	6	کشاورز	2

کلمه	تعداد تکرار	کلمه	تعداد تکرار	کلمه	تعداد تکرار
جنس	14	فاکتور	5	لوازم	2
ظرف	14	پلاستیکی	5	سرویس	2
منطقه	13	دیزل	5	مقاومت	2
نیاز	13	پیشنهاد	5	جایگاه	2
چلیک	12	موجودی	5	ارتقا	2
رقيب	12	مصرف کننده	5	موجود	2
کیفیت	11	حذف	5	مظروف	2
اجبار	11	تقاضا	5	راننده	2
عرضه	11	طول	4	صلاح دید	2
مصرف	10	قیمت	4	عاملین	1
خرید	10	لباس تبلیغاتی	4	تجارت	1
رنگ بندی	10	تعداد	4	معتبر	1
بازار	10	محصولات	4	سررسید	1
تولید	9	انبار	4	راضی کننده	1
تنوع	9	کارکنان	4	گلستان	1
تابلو تبلیغاتی	7	عملکرد	4	اختیار	1
بارگیری	7	پایین	4	مجدد	1
توزیع	7	اصلاح	3	آموزش	1
استان	7	ارسال	3	منظم	1
صنعتی	7	سمینار	3	راه اندازی	1
پرداخت	6	بنزین	3	نظرسنجی	1
شیوه	6	خارج	3	تعداد کلمات کلیدی	83

جدول الف- ۲: کلمات کلیدی استخراج شده و تعداد تکرار آنها برای داده های نظرات کاربران سایت پالایشگاه در مورد روغن های خودرویی در بازه

زمانی ۱۳۹۸ تا ۱۳۹۹

Table A-2: Extracted keywords and their frequencies based on user comments regarding automotive oils on the refinery's website(1398-1399)

کلمه	تعداد تکرار	کلمه	تعداد تکرار	کلمه	تعداد تکرار
پیشنهاد	38	دوگانه سوز	11	پژو پارس	7
روغن موتور	38	پراید	10	گرید	7
شماره ۴					
10W40	22	گیربکس	10	تابستان	6

کلمه	تعداد تکرار	کلمه	تعداد تکرار	کلمه	تعداد تکرار
20W50	22	تعویض	10	XU7 سمند	5
روغن موتور شماره ۵	19	تیبیا	9	زمستان	5
راهنمایی	16	ال نود	9	منطقه	5
کیفیت	12	بنزین	8	یورو ۴	4
روغن موتور شماره ۲	12	۲۰۶ تیپ ۲	7	XU7 پژو پارس	4
تفاوت	12	تک سوز	7	تعمیر	3
فلومکس	11	بازار	7	تعداد کلمات کلیدی	29

جدول الف-۳: کلمات کلیدی استخراج شده و تعداد تکرار آنها برای داده های نظرات کاربران سایت پالایشگاه در مورد روغن های خودرویی در بازه

زمانی ۱۳۹۹ تا ۱۴۰۰

Table A-3: Extracted keywords and their frequencies based on user comments regarding automotive oils on the refinery's website(1399-1400)

کلمه	تعداد تکرار	کلمه	تعداد تکرار	کلمه	تعداد تکرار
فلومکس	20	تقلبی	10	پژو	6
روغن موتور شماره ۴	19	پراید	10	خرید	6
راهنمایی	15	بازار	9	راضی	4
کیفیت	11	گرید	8	ریو	4
یورو ۴	11	دفترچه	7	گرانروی	4
روغن موتور شماره ۲	11	تیوفایو	7	توربوشارژ	3
اصالت	10	قو طی	6	تعداد کلمات کلیدی	20

جدول الف-۴: کلمات کلیدی استخراج شده و تعداد تکرار آنها برای داده های نظرات کاربران سایت پالایشگاه در مورد روغن های خودرویی در بازه

زمانی ۱۴۰۰ تا ۱۴۰۲

Table A-4: Extracted keywords and their frequencies based on user comments regarding automotive oils on the refinery's website (1400-1402).

کلمه	تعداد تکرار	کلمه	تعداد تکرار	کلمه	تعداد تکرار
روغن موتور شماره ۲	22	گیربکس	9	یورو	4
پراید	19	قو طی	8	بازار	4

کلمه	تعداد تکرار	کلمه	تعداد تکرار	کلمه	تعداد تکرار
پیشنهاد	18	کیفیت	8	گرید	4
روغن	14	بهران	8	نیمه سنتتیک	3
موتور					
شماره ۴					
دوگانه سوز	13	خرید	8	استاندارد	3
راهنمایی	13	اصالت	7	انقضا	2
تولید	12	فلزی	6	قوت	2
کارکرد	11	پلاستیکی	6	ماکسیما	2
محصول	11	سمند	6	دنا	2
تعویض	11	گالن	5	رنگ	1
فلو مکس	11	صدا	5	بادامک	1
تیبا	10	تهیه	5	تعداد کلمات	35
				کلیدی	

جدول الف- ۵: کلمات کلیدی استخراج شده و تعداد تکرار آنها برای داده های نظرات کاربران دیجی کالا در مورد روغن موتور شماره ۱

Table A-5: Extracted keywords and their frequencies for Digikala user reviews regarding engine oil No. 1

کلمه	تعداد تکرار	کلمه	تعداد تکرار	کلمه	تعداد تکرار
قیمت	425	محصول	18	هزینه	8
پراید	160	گران	18	تبخیر	8
خودرو	96	ارسال	17	وانت	8
دیجی کالا	84	سالم	17	پژو	8
بازار	63	نشتی	17	کارکرد	7
موتور	62	باکیفیت	17	استاندارد	6
توصیه	61	اصل	16	حمل	6
بصرفه	60	برند	19	تحويل	6
پایین	58	اقتصادی	15	گرم	5
بسته	54	سرریز	14	مغازه	5
تعویض	44	جنس	14	تقلبی	5
ارزان	43	موجود	13	مارک	4
قوטי	27	مشکل	13	منصفانه	4
پیکان	27	دقت	12	مصوب	4
راضی	25	کاربراتور	12	رنگ	4
ریخت	24	سفارش	12	کارآمد	4
موتورسیکلت	21	متوسط	12	دولتی	4
برند	19	روآ	11	تعداد کلمات	55
قدیمی	19	هوندا	10	کلیدی	

جدول الف- ۶: کلمات کلیدی استخراج شده و تعداد تکرار آنها برای داده های نظرات کاربران دیجی کالا در مورد روغن موتور شماره ۴

Table A-6: Extracted keywords and their frequency counts for Digikala user reviews regarding engine oil No.4

کلمه	تعداد تکرار	کلمه	تعداد تکرار	کلمه	تعداد تکرار
قیمت	215	توصیه	15	نرم	7
روغن	189	(XU7)یکس یوسون	14	رنگ	6
کیفیت	151	تولید	13	ارزش	6
مناسب	113	موجود	13	کارکرد	6
دیجیکالا	58	خارجی	11	داخلی	6
تعویض	56	موتورسیکلت	11	تاریخ	6
موتور	54	ریخت	11	دوگانه	6
خودرو	46	جنس	11	قدیمی	6
راضی	41	قوطی	10	ماه	6
پراید	39	سمند	10	کاربردی	5
خرید	37	بو	10	دولتی	5
پالایشگاه	36	پلمپ	10	روشن	5
بسته	30	ارسال	9	قابل قبول	5
روغن موتور	29	درجه	9	سیاه	5
اصل	29	پایین	8	تحویل	5
ایران	26	حد	8	کارخانه	4
بازار	26	سال	8	تخفیف	3
بصرفه	21	تقلبی	8	اطمینان	3
مدل	19	دویست و شش	8	اقتصادی	2
محصول	18	ایرانول	8	راضی کننده	2
پژو	18	فلزی	7	شدت	2
صدا	18	ارزان	7	بی انصاف	1
روغن موتور شماره ۴	17	فوق العاده	7	استعلام	1
برند	15	فلومکس	7	تعداد کلمات کلیدی	71

جدول الف- ۷: کلمات کلیدی استخراج شده و تعداد تکرار آنها برای داده های نظرات کاربران دیجی کالا در مورد ضد یخ شماره ۱

Table A-7: Extracted keywords and their frequency counts for Digikala user reviews regarding Antifreeze No.1

کلمه	تعداد تکرار	کلمه	تعداد تکرار	کلمه	تعداد تکرار
ضد یخ	62	خودرو	10	شک	3
قیمت	54	غلاظت	9	باکیفیت	3
کیفیت	47	تقلبی	9	بصرفه	3
قوطی	25	مکانیک	9	تعریف	3
توصیه	24	اطمینان	8	درب	3
محصول	21	مطمین	8	رقیق	3
بازار	18	کاسپین	8	راضی کننده	2

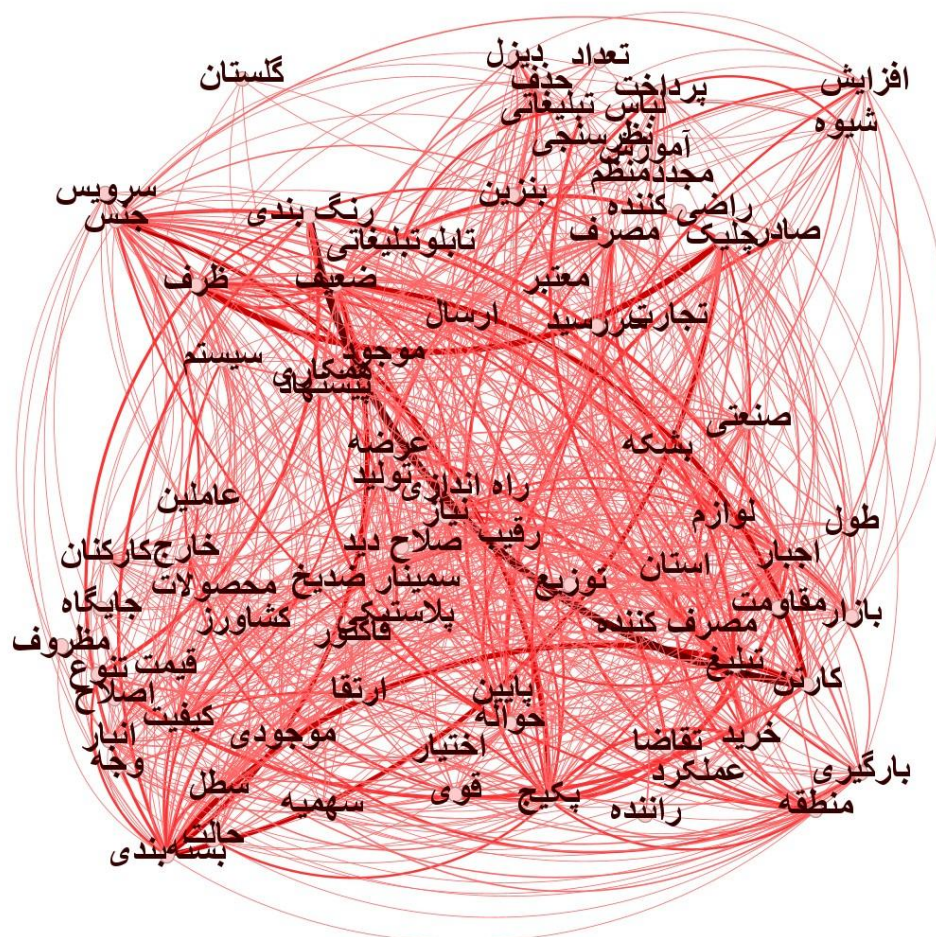
کلمه	تعداد تکرار	کلمه	تعداد تکرار	کلمه	تعداد تکرار
برند	17	تعویض	8	سیلندر	2
فلزی	16	جوش	6	کارآمد	2
مناسب	16	پالایشگاه	6	ترجیح	1
رادیاتور	14	زمستان	5	تسمه	1
تاریخ	14	جواب	5	دانگیل	1
دیجی کالا	13	پاک کردن	5	ساخت	1
روغنی	12	گران	5	تعداد کلمات کلیدی	45
تولید	12	امتحان	4		
راضی	11	خارجی	4		

جدول الف- ۸: کلمات کلیدی استخراج شده و تعداد تکرار آنها برای داده های نظرات کاربران دیجی کالا در مورد ضد یخ شماره ۲

Table A-8: Extracted keywords and their frequency counts for Digikala user reviews regarding Antifreeze No.2

کلمه	تعداد تکرار	کلمه	تعداد تکرار	کلمه	تعداد تکرار
ضد یخ	24	فلزی	5	یخ	2
کیفیت	20	تولید	4	توصیه	2
محصول	13	تاریخ	4	تصفیه	2
قیمت	10	مقطر	4	دستگاه	2
خودرو	9	رادیاتور	3	پلاستیکی	2
قوطی	9	مناسب	3	کاسپین	2
روغنی	6	مخلوط	3	خانگی	1
دیجی کالا	6	تقلبی	3	روشن	1
سفارش	5	پالایشگاه	3	امپر	1
برند	5	اسیدی	3	تعداد کلمات کلیدی	29

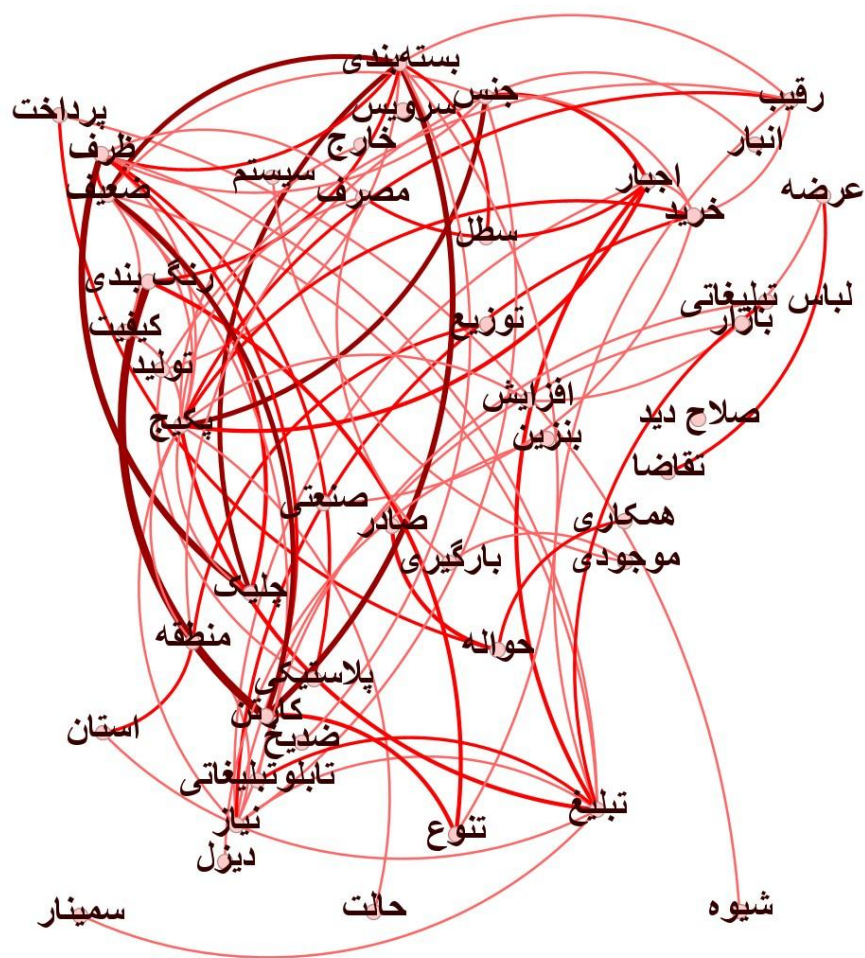
- شبکه هم رخدادی کلمات کلیدی برای سری های مختلف داده ها



شکل الف- ۱: شبکه هم‌رخدادی واژگان برای داده‌های چالش‌ها و نقطه‌نظرات عاملین در سال ۱۴۰۱

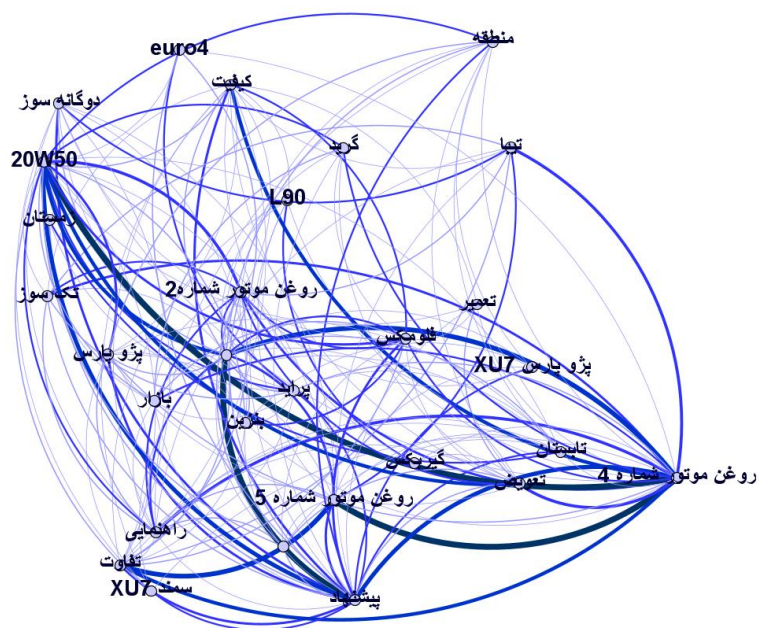
Figure A-1: Word co-occurrence network for data on stakeholders' challenges and viewpoints in the year 1401

برای این سری از داده‌ها، همان‌طور که در شکل الف-۱ مشاهده می‌شود، به دلیل حجم بالای کلمات کلیدی استخراج شده، تعداد یال‌ها در شکل شبکه هم‌رخدادی زیاد است. به همین دلیل جهت سهولت در درک شکل و یافتن هم‌رخدادی‌های قوی‌تر، یک شکل کاهش یافته که تنها یال‌های مربوط به هم‌رخدادی‌هایی با تکرار بیشتر را نمایش می‌دهند نیز رسم شده است که در شکل الف-۲ قابل مشاهده است.



شکل الف- ۲: شکل کاهش یافته شبکه هم‌رخدادی واژگان برای داده های چالش ها و نقطه نظرات عاملین در سال ۱۴۰۱

Figure A-2: Reduced word co-occurrence network for data regarding stakeholders' challenges and viewpoints in the year 1401



شکل الف- ۳: شبکه هم‌رخدادی واژگان برای داده‌های نظرات کاربران سایت پالایشگاه در مورد روغن‌های خودروبی در بازه زمانی ۱۳۹۸ تا ۱۳۹۹

Figure A-3: Word co-occurrence network based on user reviews of automotive oils from the refinery's website (1398-1399)

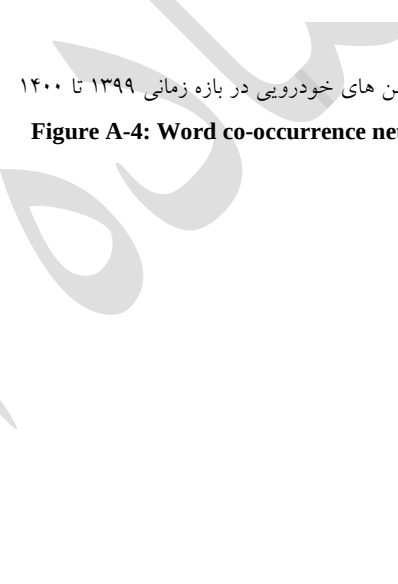
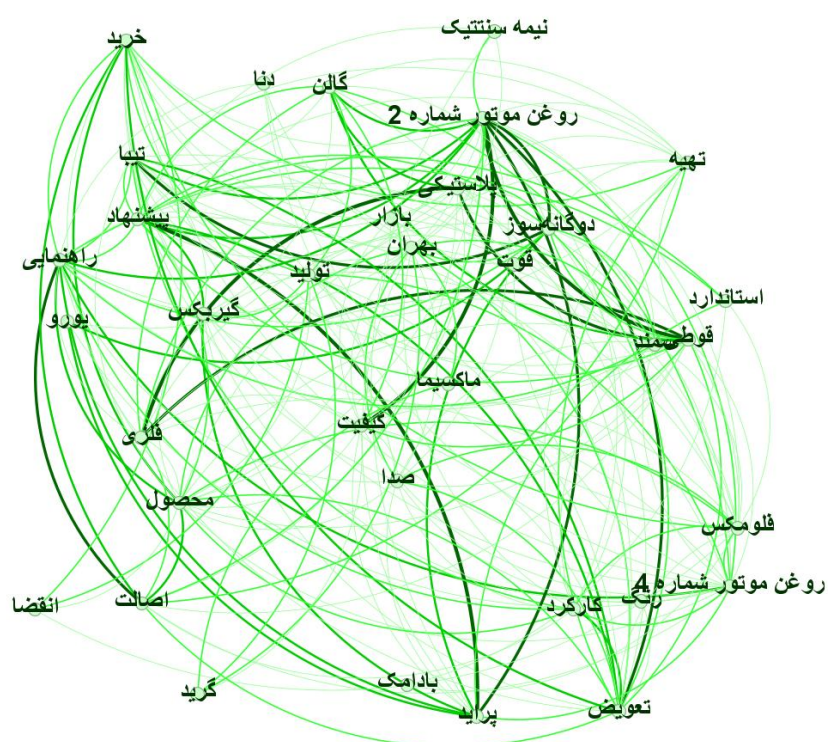


Figure A-4: Word co-occurrence network based on user reviews from the refinery's website regarding automotive oils(1399-1400)



شکل الف- ۵: شبکه هم‌رخدادی واژگان برای داده‌های نظرات کاربران سایت پالایشگاه در مورد روغن‌های خودرویی در بازه زمانی ۱۴۰۰ تا ۱۴۰۲

Figure A-5: Word co-occurrence network based on user reviews from the Palayeshgah website regarding automotive oils(1400-1402)

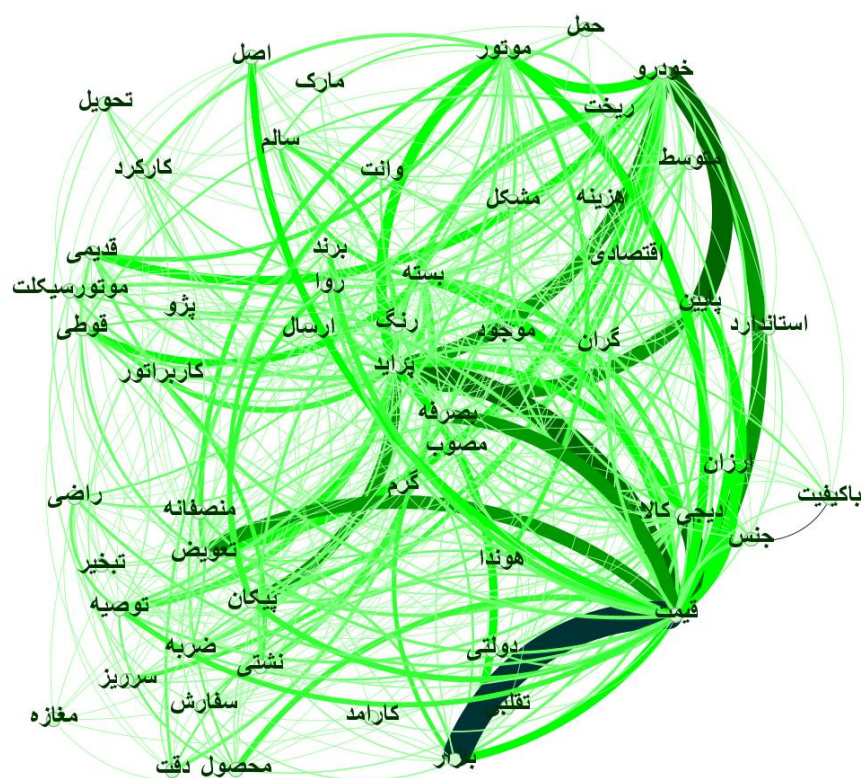
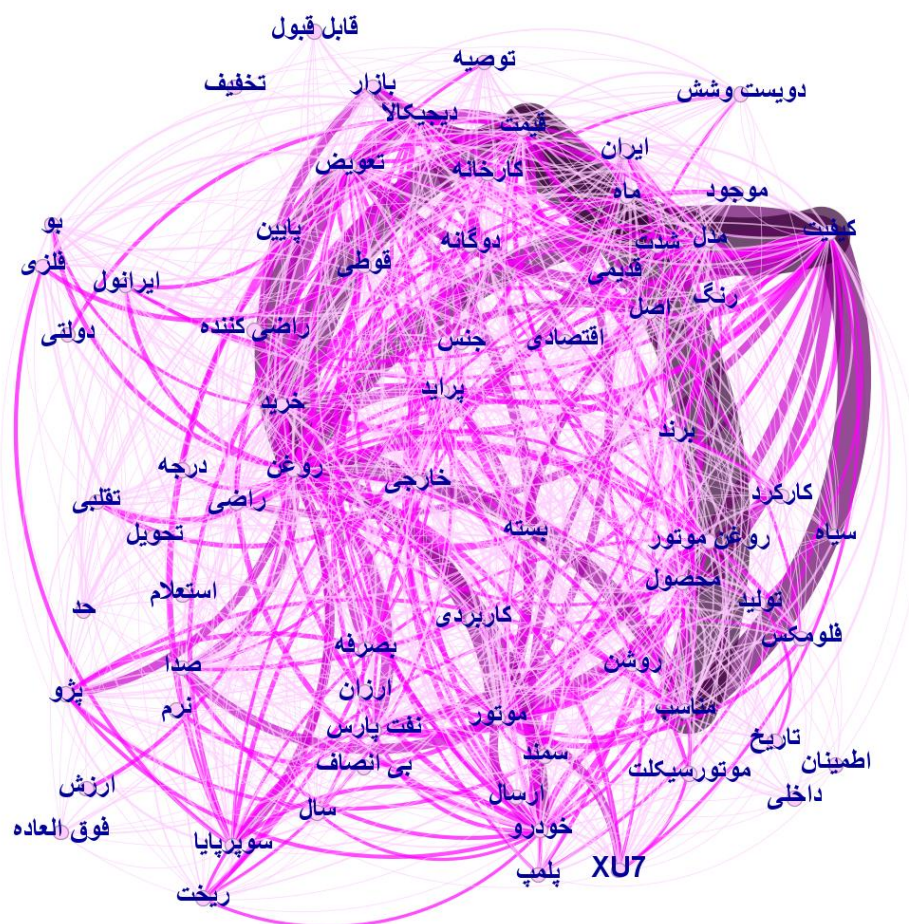


Figure A-6: Word co-occurrence network for Digikala user reviews regarding engine oil No. 1



شکل الف- ۷: شبکه هم‌رخدادی واژگان برای داده‌های نظرات کاربران دیجی کالا در مورد روغن موتور شماره ۴

Figure A-7: Word co-occurrence network for Digikala user reviews regarding engine oil No. 4

جهت تشخیص بهتر مبدا و مقصد یال‌های دارای وزن زیاد در شکل، شکل دیگری با یال‌های مستقیم (غیر منحنی) (شکل الف- ۸) رسم شده است.

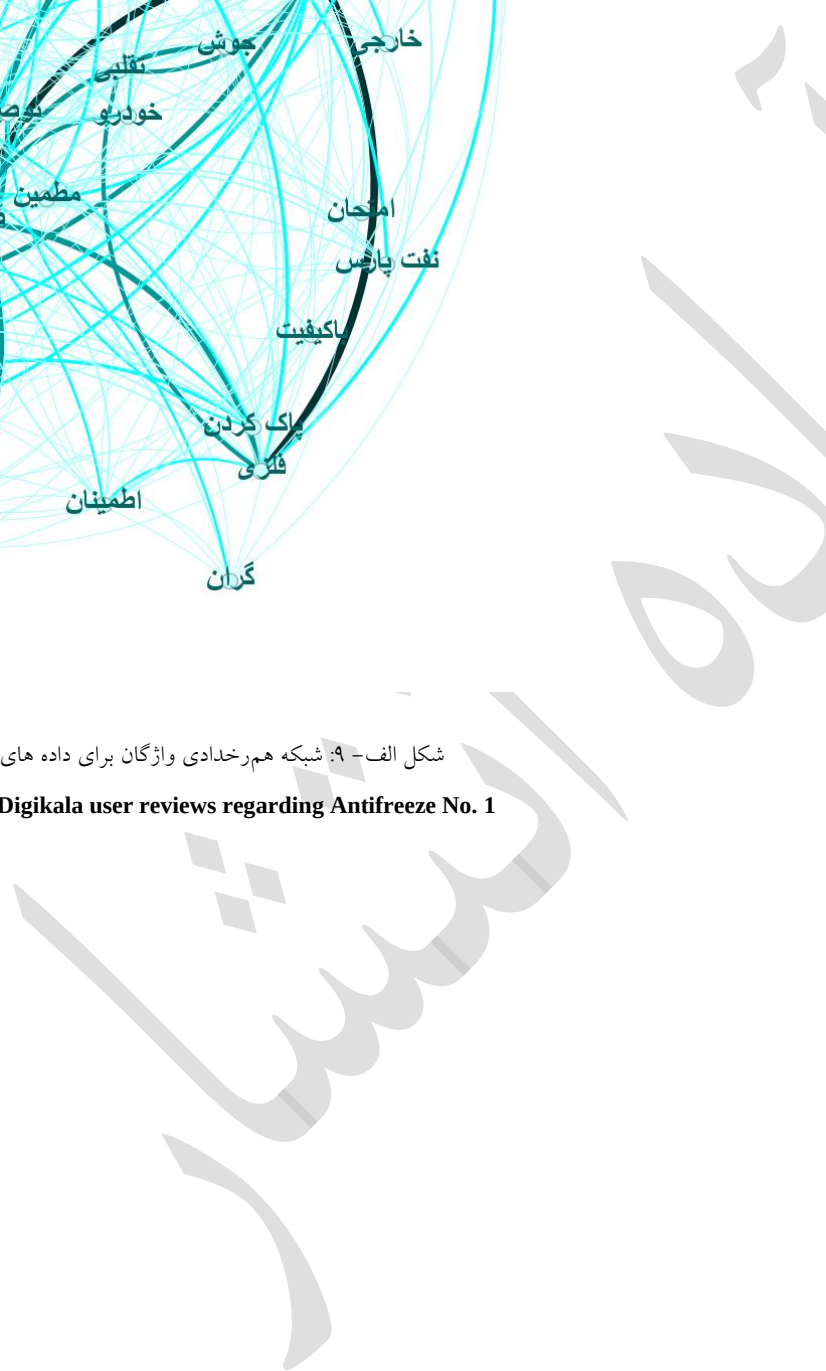
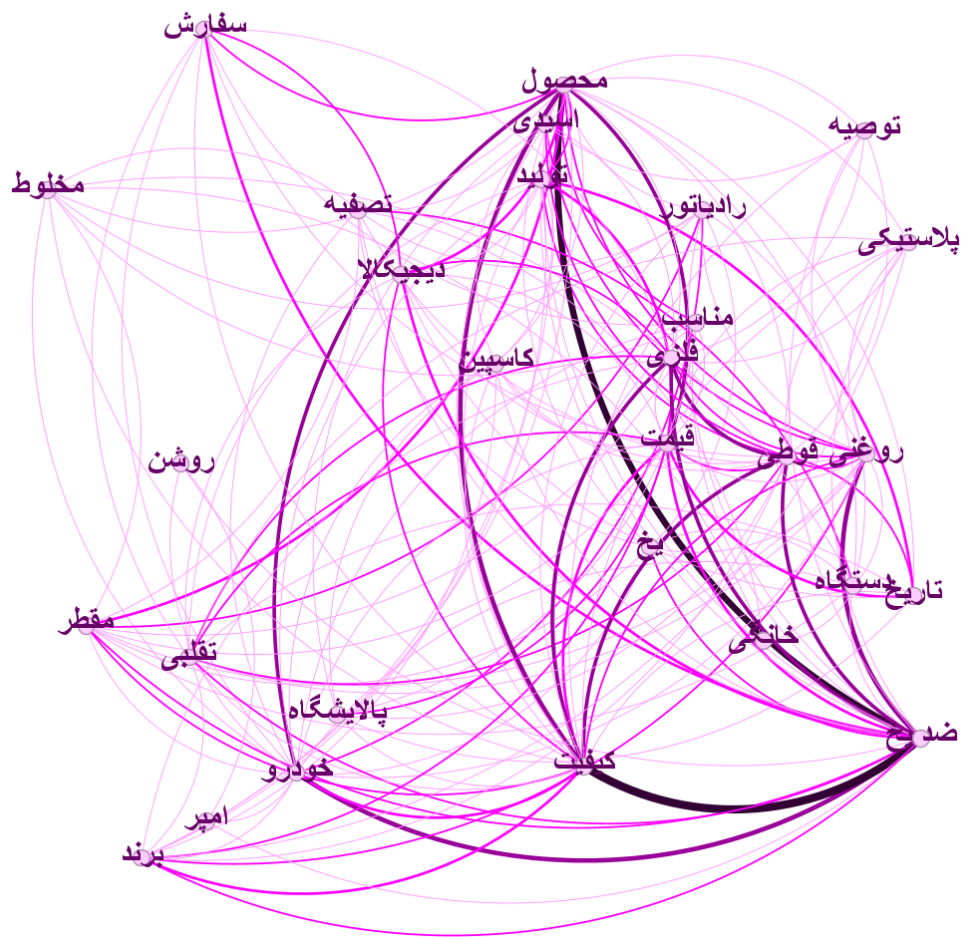


Figure A-9: Word co-occurrence network for Digikala user reviews regarding Antifreeze No. 1



شکل الف - ۱۰: شبکه هم‌رخدادی واژگان برای داده‌های نظرات کاربران دیجی کالا در مورد ضد یخ شماره ۲

Figure A-10: Word co-occurrence network for Digikala user reviews regarding antifreeze No. 2

• عبارت های استخراج شده از شبکه هم رخدادی واژگان برای سری های مختلف داده ها

جدول الف- ۹: عبارات استخراج شده از شبکه هم رخدادی واژگان داده های چالش ها و نقطه نظرات عاملین در سال ۱۴۰۱

Table A-9: Phrases extracted from the word co-occurrence network based on data regarding the challenges and viewpoints of stakeholders in the year 1401

توضیح	عبارت	درصد رخداد
کلمات «بسته بندی»، «ضعیف»، «ظرف» و «چلیک»	بسته بندی با ظرف چلیک ضعیف است	31.01٪
کلمات «تنوع»، «ضعیف»، «کارتن» و «رنگ بندی»	ضعیف بودن تنوع رنگ بندی کارتن	15.5٪
کلمات «اجبار»، «پکیج»، «جنس»، «خرید» و «رقیب»	فروش محصولات به صورت پکیج اجباری	13.95٪
کلمات «تبلیغ»، «لباس تبلیغاتی» و «تابلو تبلیغاتی»	عاملین خواستار تبلیغ به وسیله لباس و تابلو تبلیغاتی هستند	3.88٪

جدول الف- ۱۰: عبارات استخراج شده از شبکه هم رخدادی واژگان داده های نظرات کاربران سایت شرکت در مورد روغن های خودرویی در بازه زمانی

۱۳۹۸ تا ۱۳۹۹

Table A-10: Terms extracted from the word co-occurrence network based on user comments regarding automotive oils on the company's website(1398-1399)

توضیح	عبارت	درصد رخداد
کلمات «پیشنهاد»، «10W40»، «20W50» و «روغن موتور شماره ۴»	تقاضای ارائه محصول پیشنهادی مناسب خصوصا از نظر گرانروی برای انواع خودرو	32.26٪
«روغن موتور شماره ۵»، «فلومکس»، «کیفیت» و «شماره ۴»	ها	
کلمات «تفاوت»، «روغن موتور شماره ۴» و «روغن موتور شماره ۵»	تفاوت بین روغن موتور شماره ۴ و شماره ۵ برای خریداران مبهم است	18.06٪

جدول الف- ۱۱: عبارات استخراج شده از شبکه هم رخدادی واژگان داده های نظرات کاربران سایت شرکت در مورد روغن های خودرویی در بازه زمانی

۱۳۹۹ تا ۱۴۰۰

Table A-11: Phrases extracted from the word co-occurrence network based on user comments regarding automotive oils on the company's website (1399-1400)

توضیح	عبارت	درصد رخداد
کلمات «اصالت»، «بازار»، «تقلبی»، «فلومکس»، «روغن موتور شماره ۴» و «کیفیت»	نگرانی در مورد اصالت کالا و وجود کالای تقلبی در بازار	35.57٪

توضیح	عبارت	درصد رخداد
کلمات «euro 4»، «پراید»، «روغن موتور شماره ۲»، «روغن موتور شماره ۴» و «راهنمایی»، «TU5»، «فلومکس»	تقاضای راهنمایی برای خرید محصول مناسب با توجه به استاندارد و نوع خودرو	32.21٪

جدول الف- ۱۲: عبارات استخراج شده از شبکه هم‌رخدادی واژگان داده‌های نظرات کاربران سایت شرکت در مورد روغن‌های خودرویی در بازه زمانی

۱۴۰۰ تا ۱۴۰۲

Table A-12: Phrases extracted from the word co-occurrence network based on user comments regarding automotive oils on the company's website(1400-1402)

توضیح	عبارت	درصد رخداد
کلمات «روغن موتور شماره ۲»، «پلاستیکی یا فلزی بودن مظروف»، «قوطی»، «تعویض»، «فلزی»، «محول روغن موتور شماره ۲»	پلاستیکی یا فلزی بودن مظروف	36.67٪
کلمات «روغن موتور شماره ۲»، «کیفیت»، «تیبا»، «دوگانه سوز»، «پیشنهاد» و «پراید»	تقاضای راهنمایی برای خرید محصول مناسب با توجه به استاندارد و نوع خودرو	21.67٪
کلمات «راهنمایی»، «اصالت» و «محول»	تقاضای راهنمایی در مورد نحوه حصول اطمینان از اصالت کالا	8.33٪

جدول الف- ۱۳: عبارات استخراج شده از شبکه هم‌رخدادی واژگان داده‌های نظرات کاربران دیجی کالا در مورد روغن موتور شماره ۱

Table A-13: Phrases extracted from the word co-occurrence network based on Digikala user reviews regarding Engine Oil No. 1

توضیح	عبارت	درصد رخداد
کلمات «قیمت»، «بصرفه»، «پایین»، «ارزان»، «اقتصادی»	قیمت مناسب و اقتصادی روغن	21.99٪
کلمات «نشتی»، «قوطی»، «بسته»، «ضربه»، «ریخت»	ضربه خورده بودن قوطی و نشت مواد از قوطی	17.44٪
کلمات «خودرو»، «قدیمی»، «پایین»، «قیمت»، «پراید»، «کاربراتور»	مناسب بودن محصول برای ماشین‌های کاربراتوری و مدل پایین یا قدیمی	15.45٪
کلمات «قیمت»، «دیجی کالا»، «بازار»، «منصفانه»، «ارزان»، «بصرفه»، «اقتصادی»	تفاوت قیمت محصول در دیجی کالا و بازار و قیمت منصفانه دیجی کالا	15.26٪

جدول الف- ۱۴: عبارات استخراج شده از شبکه هم‌رخدادی واژگان داده های نظرات کاربران دیجی کالا در مورد روغن موتور شماره ۴

Table A-14: Phrases extracted from the word co-occurrence network based on Digikala user reviews regarding engine oil No. 4

توضیح	عبارت	درصد رخداد
کلمات «قیمت»، «بازار»، «دیجی کالا»، «دولتی»، «بصرفه»	تفاوت قیمت محصول در بازار و دیجی کالا و قیمت بصرفه محصول در دیجی کالا	44.48٪
کلمات «قیمت»، «کیفیت»، «مناسب» و «روغن»	قیمت و کیفیت مناسب روغن	7.84٪

جدول الف- ۱۵: عبارات استخراج شده از شبکه هم‌رخدادی واژگان داده های نظرات کاربران دیجی کالا در مورد ضد یخ شماره ۱

Table A-15: Phrases extracted from the word co-occurrence network of Digikala user reviews regarding Antifreeze No. 1

توضیح	عبارت	درصد رخداد
کلمات «کیفیت»، «ضد یخ»، «قیمت»، «مناسب»، «روغنی» و «محصول»	کیفیت و قیمت مناسب ضد یخ	34.39٪
کلمات «قوطی»، «تعویض»، «فلزی»، «ضد یخ»	رضایت از فلزی بودن قوطی ضد یخ	14.03٪
کلمات «قیمت»، «بازار»، «مناسب»، «دیجی کالا»	تفاوت قیمت بازار و قیمت دیجی کالای ضد یخ	11.76٪

جدول الف- ۱۶: عبارات استخراج شده از شبکه هم‌رخدادی واژگان داده های نظرات کاربران دیجی کالا در مورد ضد یخ شماره ۲

Table A-16: Phrases extracted from the word co-occurrence network of Digikala user reviews regarding Antifreeze No. 2

توضیح	عبارت	درصد رخداد
کلمات «کیفیت»، «ضد یخ»، «قیمت»، «روغنی» و «محصول»	کیفیت خوب، روغنی بودن و قیمت ضد یخ	60.94٪
کلمات «قوطی»، «فلزی»، «ضد یخ»، «محصول»	رضایت از فلزی بودن قوطی ضد یخ	20.31٪