Journal of Quality Engineering and Management

www.pqprc.ir

J. Qual. Eng. Manag. Vol. 16, No. 2 (2026) 108–140.

Paper Type: Original Article

Analyzing Stakeholder Review of an Industrial Lubricant Refinery Using a Combination of Topic Modeling and Keyword Co-occurrence Analysis

Zahra Bayat[1], Bakhtiar Ostadi[1,*], Ehsan Nikbakhsh[2]

[1] Department of Systems Management and Productivity, Faculty of Industrial and Systems Engineering, Tarbiat Modares University, Tehran, Iran; bayat.zahra@modares.ac.ir; bostadi@modares.ac.ir.

[2] Department of Economic and Social Systems Engineering, Faculty of Industrial and Systems Engineering, Tarbiat Modares University, Tehran, Iran; nikbakhsh@modares.ac.ir.

Citation:

Received: 06 November 2025 *Revised: 09 March 2026* *Accepted: 05 April 2026*	Bayat, Z., Ostadi, B., & Nikbakhsh, E. (2026). Analyzing stakeholder review of an industrial lubricant refinery using a combination of topic modeling and keyword co-occurrence analysis. *Journal of Quality Engineering and Management*, *16*(2), 108-140.

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 Latent Dirichlet Allocation (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 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.

Corresponding Author: bostadi@modares.ac.ir

10.48313/jqem.2026.557985.1582

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).

مهندسی و مدیریت کیفیت

www.pqprc.ir

دوره ۱۶، شماره (۲)، (۱۴۰۵)، ۱۴۰-۱۰۸

نوع مقاله: پژوهشی

تحلیل نظرات ذی‌نفعان یک پالایشگاه روان‌کار با استفاده از ترکیب مدلسازی موضوع و تحلیل هم‌رخدادی واژگان

زهرا بیات[۱]، بختیار استادی[۱،*]، احسان نیک‌بخش[۲]

[۱]گروه مدیریت سیستم و بهره‌وری، دانشکده مهندسی صنایع و سیستم‌ها، دانشگاه تربیت مدرس، تهران، ایران.

[۲]گروه مهندسی سیستم‌های اقتصادی و اجتماعی، دانشکده مهندسی صنایع و سیستم‌ها، دانشگاه تربیت مدرس، تهران، ایران.

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 ارایه رویکردی ترکیبی برای تحلیل نظرات ذی‌نفعان یک پالایشگاه روان‌کار با استفاده از مدلسازی موضوع و تحلیل هم‌رخدادی واژگان انجام‌شده است.

روش‌شناسی پژوهش: پژوهش با بهره‌گیری از الگوریتم تخصیص دیریکله پنهان[۱] 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 تخصیص دیریکله پنهان با پارامترهای مشخص (تعداد بهینه موضوعات بر اساس امتیاز انسجام *c_v*، تعداد *passes=40*، *iterations=200*، *alpha='symmetric'*)، تشکیل ماتریس هم‌رخدادی و استخراج عبارات کلیدی از یال‌های پرتکرار با استفاده از روش امتیازدهی است.

یافته‌ها: بیشترین نرخ فراوانی عبارت استخراج‌شده از نظرات عاملین فروش (۳۱/۰۱%) مربوط به ضعف بسته‌بندی با ظرف چلیک است. میان نظرات کاربران سایت، ابهام در جنس مظروف روغن‌موتور شماره ۲ با نرخ ۳۶/۶۷% بیشترین فراوانی را دارد. نظرات کاربران دیجی‌کالا نشان‌دهنده رضایت از قیمت و کیفیت مناسب محصولات است.

اصالت/ارزش‌افزوده علمی: رویکرد ارایه‌شده، روشی کاربردی برای تحلیل سیستماتیک نظرات ذی‌نفعان سازمان‌ها ارایه 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 دهان‌به‌دهان الکترونیکی[۲] مشتریانی که خدمات را تجربه کرده‌اند، این امکان را دارند که نمرات را انتخاب کرده یا تجربه خود را توصیف کنند که قابل‌اعتمادتر و عینی‌تر از اطلاعات ارایه‌شده توسط شرکت‌ها به نظر می‌رسد. به همین دلیل نظرات برخط نقش مهمی در شکل‌دهی به تصویر پیش از تجربه خدمت برای مشتریان دارند. در مقایسه با نظرات عمومی، ویژگی‌های نظرات برخط شامل دسترسی ۲۴ ساعته و فراهم

[1] Latent Dirichlet Allocation (LDA)

[2] Electronic Word of Mouth (eWOM)

کردن امکان ذخیره مداوم اطلاعات به‌صورت متن یا تصویر است. این نظرات به مشتریان نه‌تنها پوشش گسترده‌تری ارایه می‌دهد، بلکه انتشار اطلاعات نیز به‌سرعت انجام می‌شود؛ بنابراین، نظرات برخط تاثیر قابل‌توجهی بر مشتریان دارای خدمات فعال دارد، زیرا می‌توانند نظرات خود را بدون محدودیت زمانی و مکانی ارایه و ویرایش کنند. از طریق نظرات برخط، مشتریان به‌صورت آزاد و بی‌طرف نظرات خود را در جامعه برخط با دیگران به اشتراک می‌گذارند. در مقابل تبلیغات یک‌طرفه، نظرات برخط به دلیل انتقال نظرات واقعی مشتری یا تجربیات خاصی که تجربه می‌کنند، در کنترل انتخاب‌های مشتریان نقش فعال دارند [1]. بسیاری از مطالعات قبلی مربوط به داده‌کاوی بر روی داده‌های ساخت‌یافته متمرکز شده‌اند درحالی‌که ۸۰% از داده‌های جهان به شکل بدون ساختار هستند [2]. امروزه حجم قابل‌توجهی از داده‌ها به‌صورت متنی در دسترس است. داده‌های ذخیره‌شده در پایگاه‌های داده متنی به‌سرعت در حال رشد هستند [3].

تجزیه‌وتحلیل نظرات مشتریان برای توسعه استراتژی‌های موثر بازار بسیار مهم است [4]. نظرات مشتریان که از طریق بررسی‌ها و رتبه‌بندی‌ها به‌دست می‌آیند، بینش‌های ارزشمندی را برای تجزیه‌وتحلیل محصول و بهبود استراتژی بازاریابی ارایه می‌دهند که درنهایت منجر به افزایش سود می‌شود. درک ترجیحات مصرف‌کننده از طریق بررسی‌های برخط به مناسب‌سازی محصولات برای برآوردن نیازها و انتظارات مشتری کمک می‌کند [5]. گنجاندن بازخورد مشتری در تصمیمات قیمت و کیفیت بر اساس بررسی‌های برخط می‌تواند به استراتژی‌های بهینه برای شرکت‌ها در بازارهای رقابتی منجر شود. در تلاش برای درک و استفاده از نظرات مشتریان در توسعه بازار، روش‌ها و استراتژی‌های متنوعی ارایه شده‌اند. یکی از این رویکردها بررسی برخط مشتریان برای تصمیم‌گیری در مورد کیفیت و قیمت‌گذاری است. در این روش، شرکت‌ها با استفاده از ارزیابی‌های مشتری برای بهینه‌سازی عملکرد مالی خود، استراتژی‌های خود را تنظیم می‌کنند [5]. علاوه بر این، تحلیل احساسات روی نظرات مشتریان با استفاده از رویکردهای یادگیری ماشین به طبقه‌بندی احساسات برای افزایش فروش و به‌دست آوردن مزیت رقابتی در بازار تجارت الکترونیک کمک می‌کند [6]. این مدل‌ها در کل به درک و استفاده از نظرات مشتریان برای توسعه استراتژی‌های موثر در بازار کمک می‌کنند.

به‌علاوه، مدلسازی موضوع نقش مهمی در تجزیه‌وتحلیل نظرات مشتریان برای توسعه استراتژی‌های بازار ایفا می‌کند. چندین مدل و رویکرد در این زمینه ارایه‌شده است. فرخود و همکاران [7] مدل جمله سند موضوعی[1] را معرفی کرد که ترکیبی از موضوع احساسات مشترک[2] و تخصیص دیریکله پنهان برای تجزیه‌وتحلیل احساسات در هر دو سطح سند و کلمه است. عثمانی و محصفی [8] مدل موضوع احساسات مشترک را با معرفی پارامتر دیدگاه نویسنده گسترش دادند که دقت طبقه‌بندی احساسات را در سطح متن بهبود می‌بخشد. ششاج و همکاران [9] مدل تخصیص دیریکله پنهان را با رویکرد تنظیم موضوعی مبتنی بر معنا برای تجزیه‌وتحلیل تجربه مشتری تقویت کردند. دریافت نظرات مشتریان در مورد محصولات و خدمات ارایه‌شده توسط شرکت‌ها یکی از نیازهای اصلی مدیران شرکت برای بهبود ارایه خدمات است. امروزه استفاده از شبکه‌های اجتماعی، پیام‌رسان‌ها و وب سایت‌های بررسی برای دریافت اطلاعات از نظرات مشتریان رایج است؛ اما تجزیه‌وتحلیل داده‌های دریافتی از این منابع به دلیل حجم زیاد آن‌ها نیازمند یک ابزار کارآمد است [10]. با تکامل هوش مصنوعی و رشته‌های مرتبط مانند یادگیری ماشین، کاربرد روش‌های پردازش به‌وسیله کامپیوتر با استفاده از الگوریتم‌های ریاضی برای درک الگوهای پنهان در داده‌ها افزایش یافته است که به داده‌کاوی معروف است. ازجمله داده‌های قابل‌پردازش، داده‌های متنی است که می‌توان با استفاده از روش‌های مختلف پردازش متن دانش پنهان در متون را شناسایی کرد [10]. این روش‌ها اغلب رایگان بوده، در پاسخ به داده‌های جدید خود انطباق پذیر هستند، پیاده‌سازی آن‌ها آسان است و کاملا سازگار می‌باشند. لذا استفاده از روش‌های داده‌کاوی و یادگیری ماشین در فرایند استخراج دانش از داده‌های متنی مشتریان مقرون‌به‌صرفه است. همان‌طور که پیش‌تر هم اشاره شد، با کمک این تکنیک‌ها می‌توان دانش ضمنی و بالقوه مفید را از حجم زیادی داده استخراج کرد و در فرایند تصمیم‌گیری از آن‌ها کمک گرفت. با توجه به مشخص شدن اهمیت و ضرورت تحلیل نظرات ذی‌نفعان با استفاده از روش‌های سریع و کارآمد و تمرکز مقالات اخیر بر این مساله، تحلیل نظرات ذی‌نفعان جهت توسعه بازار برای داده‌های فارسی به‌عنوان یک خلا تحقیقاتی در این حوزه شناسایی شد. لذا در این پژوهش ما به دنبال ارایه رویکردی ترکیبی برای تحلیل نظرات ذی‌نفعان یک پالایشگاه روان‌کار با استفاده از روش‌های کارآمد نظیر مدلسازی موضوع و تحلیل هم‌رخدادی واژگان به‌منظور استخراج عبارات کلیدی و پشتیبانی از تصمیم‌های بازاریابی هستیم. ذی‌نفعان در این تحقیق به‌عنوان عاملین فروش، سایت‌های فروش برخط و پیشنهادات مطرح‌شده در نظرسنجی‌های شش‌ماهه این پالایشگاه تعریف شده‌اند. تجزیه‌وتحلیل نظرات ذی‌نفعان برای توسعه بازار

[1] Topic Document Sentence (TDS) model

[2] Joint Sentiment Topic (JST)

بسیار مهم است زیرا بازخورد ارزشمندی را برای بهبود استراتژی‌ها و تجاری‌سازی ارایه می‌دهد [13–11]. تحلیل نظرات ذی‌نفعان به درک احساسات نسبت به محصولات یا خدمات و هدایت استراتژی‌های بهبود کمک می‌کند [11]. علاوه بر این، ذی‌نفعان متنوع به‌طور قابل‌توجهی در تجاری‌سازی و ایجاد بازار نوآوری‌های ریشه‌ای، بهبود موقعیت در بازار و در دسترس بودن محصول کمک می‌کنند [13]. به‌طورکلی، تجزیه‌وتحلیل نظرات ذی‌نفعان با استفاده از بازخورد، بهبود موقعیت محصول و هم‌سویی استراتژی‌ها با انتظارات ذی‌نفعان، رشد بازار را تقویت می‌کند؛ بنابراین، تجزیه‌وتحلیل نظرات ذی‌نفعان به شکل‌دهی استراتژی‌های بازار، تقویت نوآوری و هدایت موفقیت کلی بازار کمک می‌کند و ازاین‌رو این امر برای شرکت‌ها دارای اهمیت است. با توجه به‌سرعت و حجم تولید داده در بستر اینترنت، واضح است که شرکت‌ها علاقه‌مند به استفاده از راه‌های کارآمدتر با سرعت و دقت بیشتر هستند.

مقالات اخیر نیز در زمینه تجزیه‌وتحلیل نظرات ذی‌نفعان بر استفاده از روش‌های داده‌کاوی، یادگیری ماشین و مدلسازی موضوع تمرکز کرده‌اند [8]، [14]، [15]. این مقالات تکنیک‌های مختلفی مانند طبقه‌بندی احساسات را با استفاده از مدل‌های یادگیری ماشین مانند بیز ساده[1]، ماشین بردار پشتیبان[2] و درخت‌های تصمیم بررسی می‌کنند [16]. این مطالعات بر اهمیت استفاده از فناوری‌های پیشرفته مانند یادگیری عمیق برای تجزیه‌وتحلیل نظرات و احساسات، به‌ویژه در زمینه حجم وسیع داده‌های اجتماعی موجود در اینترنت، تاکید می‌کنند. مدلسازی موضوع امکان استخراج مضامین کلیدی را از حجم وسیعی از داده‌های متنی فراهم کرده و به جمع‌بندی و درک احساسات ذی‌نفعان کمک می‌کند [17]. درک پویایی نگرانی‌های عمومی از طریق تجزیه‌وتحلیل ذی‌نفعان برای موفقیت پروژه‌های زیرساختی و عملکرد ذی‌نفعان حیاتی است [18]. مدلسازی موضوع، می‌تواند مسایل و موضوعات کلیدی موردبحث توسط ذی‌نفعان را آشکار کند و به شناسایی اولویت‌ها و تاثیر آن‌ها کمک کند [18]. با استفاده از مدلسازی موضوع در تجزیه‌وتحلیل ذی‌نفعان، ترسیم علایق، نفوذ و اهمیت ذی‌نفعان مختلف آسان‌تر می‌شود و بررسی متوازن دیدگاه‌های آن‌ها را تسهیل می‌کند و کیفیت فرایندهای تصمیم‌گیری را افزایش می‌دهد [19]. درنهایت، این رویکرد تضمین می‌کند که تمام دیدگاه‌های ذی‌نفعان در نظر گرفته شده و منجر به نتایج دارای تاثیر و فراگیری بیشتر در پروژه می‌شود.

ترکیب مدلسازی موضوع با شبکه‌های هم رخدادی واژگان می‌تواند کشف موضوع را با استفاده از روابط کلمات تقویت کند [20]. درحالی‌که مدل‌های سنتی موضوع مانند تخصیص دیریکله پنهان استقلال بین کلمات را فرض می‌کنند، شبکه‌های هم رویداد روابط کلمات را برای بهبود کیفیت موضوع جذب می‌کنند [21]. مدل‌هایی مانند *CWIBTD* از شبکه‌های هم رخدادی کلمه استفاده می‌کنند تا موضوعات نادر را به‌طور موثر شناسایی کنند و تراکم معنایی و حساسیت را در کشف موضوع افزایش دهند [22]. به‌علاوه، الگوریتم‌هایی مانند موضوع جامعه، جوامع را از شبکه‌های هم‌روند استخراج می‌کنند تا ساختارهای موضوعی سلسله مراتبی ایجاد کنند و کاوش موضوعات فرعی و فوق‌العاده را با روابط منسجم تسهیل می‌کنند [23]. با ادغام مدلسازی موضوع با شبکه‌های هم رخدادی واژگان، محققان می‌توانند روابط عمیق‌تر و سلسله‌مراتب موضوعی را برای تجزیه‌وتحلیل موضوعی روشن‌تر و جامع‌تر کشف کنند. لذا استفاده از ترکیب شبکه هم رخدادی واژگان با مدلسازی موضوع موجب افزایش دقت نتایج به‌دست‌آمده می‌شود. این پژوهش با هدف ارایه رویکردی ترکیبی برای تحلیل نظرات ذی‌نفعان جهت استخراج عبارات کلیدی و پشتیبانی از تصمیم‌های 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 می‌توان با ترکیب مدلسازی 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 تجزیه‌وتحلیل نظرات ذی‌نفعان بر استفاده از روش‌های داده‌کاوی، یادگیری ماشین و مدلسازی موضوع [8]، [14]، [15]، تحلیل نظرات ذی‌نفعان جهت توسعه بازار برای داده‌های فارسی به‌عنوان یک خلا تحقیقاتی در این حوزه شناسایی شده است. در برخی از مقالات، تحلیل نظرات ذی‌نفعان با استفاده از مدلسازی موضوع انجام شده است؛ اما در تحقیقات تابه‌حال اتصال نظرات ذی‌نفعان به محصولات انجام نشده و مواردی از بررسی نظرات مشتریان به تفکیک دسته‌های مختلف مشتریان مشاهده شده است.

[1] Naïve Bayes

[2] Support Vector Machine (SVM)

در ادامه این پژوهش، بررسی مبانی نظری و پیشینه پژوهشی مرتبط، روش‌شناسی پژوهش، یافته‌های پژوهش و ارایه پیشنهادها و نتایج کاربردی ارایه‌شده است.

۲- مروری بر مبانی نظری و پیشینه پژوهش

۲-۱- تحلیل نظرات ذی‌نفعان

فریمن [24] در سال ۱۹۸۴، ذی‌نفعان را هر فرد یا گروهی می‌داند که در دستیابی سازمان به اهدافش اثر می‌گذارد یا از این موضوع اثر می‌پذیرد؛ اما او در آخرین تعریف خود از ذی‌نفعان که در سال ۲۰۰۴ عنوان کرد، ذی‌نفعان را به‌عنوان گروه‌هایی تعریف کرد که برای بقا و موفقیت شرکت‌ها لازم و ضروری هستند [25]. فریمن و همکارانش [26] در کتاب "تئوری ذی‌نفعان" خود که در سال ۲۰۱۰ چاپ شده است، تعریف دیگری برای واژه ذینفع ارایه داده‌اند. او در این کتاب بیان کرده است که ذی‌نفع به معنی افراد و گروه‌هایی است که به‌نوعی در موفقیت یا شکست یک کسب‌وکار، سهیم هستند. در کتاب "مدیریت استراتژیک" روترمل [27]، بیان‌شده است که ذی‌نفعان؛ افراد، گروه‌ها یا سازمان‌هایی هستند که از اقدام‌های یک بنگاه اقتصادی تاثیر می‌پذیرند یا بر روی اقدام‌های یک بنگاه اقتصادی تاثیر می‌گذارند. تعریفی که هانسون و همکاران [28] در کتاب خود با عنوان "مدیریت استراتژیک: رقابت‌پذیری و جهانی‌شدن" از ذی‌نفعان ارایه داده‌اند بدین شرح است: ذی‌نفعان؛ افراد، گروه‌ها و سازمان‌هایی هستند که می‌توانند روی چشم‌انداز و ماموریت یک بنگاه اقتصادی تاثیر بگذارند. هم‌چنین دستاوردها و خروجی‌های استراتژیک بنگاه، روی آن‌ها تاثیرگذار است. ضمنا، ذی‌نفعان می‌توانند در زمینه عملکرد بنگاه، انتظارات و ادعاهایی داشته باشند و این موارد را به‌وسیله اهرم‌های مختلفی که در اختیار دارند اعمال کنند. کلمه ذی‌نفعان برای اولین بار در سال ۱۹۶۳ در یک همایش بین‌المللی در موسسه تحقیقات استنفورد مطرح شد [24]؛ اما مطرح شدن مفهوم ذی‌نفعان برای نخستین بار به‌صورت یک ساختار و تعریف منسجم به سال ۱۹۸۴ برمی‌گردد [29]. تصدیقی [30] در مقاله خود نقشه ادبیات متمرکز بر حوزه ذی‌نفعان را بدین‌صورت بیان می‌کند.

<table>
<tr><td colspan="4">ارائه مفهوم ذینفعان برای نخستین بار در موسسه تحقیقات استنفورد (۱۹۶۳)</td></tr>
<tr><td>کاربرد مفهوم ذینفعان در تئوری سازمان</td><td>کاربرد مفهوم ذینفعان در حوزه مسئولیت‌های اجتماعی سازمان</td><td>کاربرد مفهوم ذینفعان در تئوری سیستم‌ها</td><td>کاربرد مفهوم ذینفعان در طرح‌ریزی سازمانی</td></tr>
<tr><td colspan="4">مدیریت راهبردی، یک رویکرد مبتنی بر ذینفعان توسط فریمن (۱۹۸۴)</td></tr>
<tr><td>جنبه های هنجاری</td><td colspan="2">جنبه های ابزاری</td><td>جنبه های توصیفی/تجربی</td></tr>
<tr><td colspan="4">تئوری ذینفعان شرکت توسط دونالدسون و پرستون (۱۹۹۵)</td></tr>
<tr><td colspan="4">تئوری پویایی و خوشه‌بندی ذینفعان (۲۰۱۰)</td></tr>
<tr><td colspan="4">سایر تئوری های ذینفعان (۱۹۹۶ تا ۲۰۱۱)</td></tr>
</table>

شکل ۱- نقشه پیشینه حوزه ذی‌نفعان ([30]).
Figure 1- Stakeholder domain background map ([30]).

تاکنون ذی‌نفعان به روش‌های مختلفی از جنبه‌های متفاوت توسط محققین این حوزه طبقه‌بندی شده‌اند. ازجمله این طبقه‌بندی‌ها می‌توان به طبقه‌بندی کمبل، طبقه‌بندی شولر و وندی ون، طبقه‌بندی کلارکسون، طبقه‌بندی حیث و نورمن و طبقه‌بندی فریدمن اشاره کرد که در تمامی این طبقه‌بندی‌ها مشتریان به‌عنوان گروهی از ذی‌نفعان سازمان مطرح شده‌اند [31]. با توجه به پیشینه موجود، مشتریان مهم‌ترین گروه ذی‌نفعان برای سازمان‌ها محسوب می‌شوند. لذا می‌توان گفت نظرات و نیازهای آن‌ها برای مدیران سازمان‌ها حایذ اهمیت است. برای کسب دانش در مورد نیازهای مشتریان، می‌توان این داده‌ها را جمع‌آوری کرد؛ اما با توجه به گسترش استفاده از اینترنت و امکاناتی نظیر خرید اینترنتی و اعلام نظر در مورد محصولات در سایت‌های مختلف و فروم‌ها، می‌توان گفت نظرات برخط محصول یک‌راه جایگزین برای جمع‌آوری داده‌ها است. این نظرات اطلاعات مفیدی را در اختیار مدیران قرار می‌دهد؛ بنابراین، مدیران ممکن است بینش ارزشمندی را از نظرات برخط استخراج نموده و بر اساس آن عمل کنند. محتوای تولیدشده توسط مصرف‌کننده، به‌ویژه نظرات مشتریان برخط در مورد برندهای مختلف، منبع غنی اطلاعات است چرا که این نظرات فاقد جانب‌داری است و به‌عنوان یک پیشنهاد دوستانه بین مشتریان در نظر گرفته می‌شود [2]. با ظهور وب ۲٫۰، حجم عظیمی از داده‌ها تولید و در وب ارسال می‌شود. بازخوردها در قالب نظرات مشتریان برخط به‌سرعت در حال رشد است که شامل افکار، نظرات و تجربیات مشتریان

در ارتباط با برندها می‌شود. این موضوع به مدیران کمک می‌کند تا ترجیحات مشتریان را درک کنند و نظرات برخط مربوط به محصول یا نام تجاری خود را ردیابی و نظارت کنند. در سال‌های اخیر، مطالعات مختلفی با استفاده از تجزیه‌وتحلیل احساسات در محصولات کلی، سیاست و جامعه، ازجمله صنعت خدمات، مانند فیلم‌ها، گردشگری، رستوران‌ها و هتل‌ها انجام شده است که در آن‌ها بر اهمیت بررسی‌های برخط تاکید شده است [1]. همچنین به گفته داهیا و همکارانش [2] که مطالعات مربوط به نظرات مشتریان برخط منتشرشده از ۲۰۲۰-۲۰۱۵ را بررسی کرده‌اند، نتایج نشان می‌دهد که مطالعات مربوط به تجزیه‌وتحلیل بررسی مشتریان برخط در طول سال‌های اخیر افزایش یافته است. بسیاری از محققان به اهمیت بررسی‌های برخط پی بردند و با استفاده از آن‌ها تحقیق کردند. دلاروکاس و همکاران [32] نشان داده‌اند که معیارهای بررسی برخط می‌تواند درآمد فیلم را به‌طور دقیق پیش‌بینی کند. سوتیریادیس و ون زیل [33] دریافتند که بررسی‌ها و توصیه‌های برخط بر فرایند تصمیم‌گیری گردشگران نسبت به خدمات گردشگری تاثیر می‌گذارد و تبلیغات دهان‌به‌دهان تاثیر قابل‌توجهی بر هنجارها و نگرش‌های ذهنی نسبت به یک شرکت هواپیمایی و تمایل مشتری به توصیه دارد. سیرینگ و همکاران [34]، بررسی کرده‌اند که آیا محتوای تولیدشده توسط کاربر در قالب بررسی‌های برخط می‌تواند برای توضیح و پیش‌بینی تصمیم توصیه مورداستفاده قرار گیرد یا خیر. علاوه بر این، آن‌ها دریافتند که احساسات مربوط به جنبه‌های مختلف خدمات نیز به‌طور قابل‌توجهی بر تصمیم توصیه تاثیر می‌گذارد؛ بنابراین، بررسی برخط برای خطوط هوایی برای درک مشتریان متنوع خود به‌منظور اتخاذ استراتژی‌های بهبود خدمات بسیار مفید خواهد بود، زیرا خطوط هوایی یک صنعت بسیار رقابتی است [1].

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 هماهنگ‌سازی نیروها جهت دست یافتن به اهداف جنگ استفاده شد. استراتژی مجموعه‌ای از اهداف اصلی و سیاست‌ها و برنامه‌های کلی به‌منظور نیل به اهداف است به‌گونه‌ای که قادر به تبیین این موضوعات باشد که در چه کسب‌وکار و چه نوع سازمانی فعالیت می‌کنیم یا قصد انجام فعالیت را داریم. به‌بیان‌دیگر، استراتژی یک برنامه واحد، همه‌جانبه و تلفیقی است که نقاط قوت اصلی سازمان را به تغییرات محیط مربوط می‌سازد و به‌گونه‌ای طراحی می‌شود که حصول اطمینان از دست یافتن به اهداف اصلی سازمان با اجرای درست آن صورت گیرد. آنسوف [35]، با ارایه ماتریسی با دو بعد محصول و بازار، برای اولین بار به بیان جهت راهبردی برای رشد فعالیت‌های واحد کسب‌وکار پرداخت. به اعتقاد آنسوف [35]، توسعه بازار به معنای یافتن بازارهای جدید از طریق روش‌هایی نظیر باز کردن کانال‌های بازاریابی و ورود به بازارهای جدید جهت عرضه محصولات فعلی است. بااین‌حال، در بعضی منابع، توسعه محصول و نفوذ در بازار را نیز موجب توسعه بازار موجود می‌داند و آن را اصلاح بازار می‌نامد. او معتقد است زمانی که بازار در حالت اشباع قرار دارد، روش‌های اصلاح بازار سودمندی زیادی به همراه دارند. یک شرکت با استفاده از استراتژی توسعه بازار می‌تواند از طریق اشباع بازار و نفوذ در بازار، سهم بزرگی از بازار فعلی را در اختیار بگیرد. همچنین می‌تواند برای محصولات یا خدمات فعلی بازارهای جدید بیافریند.

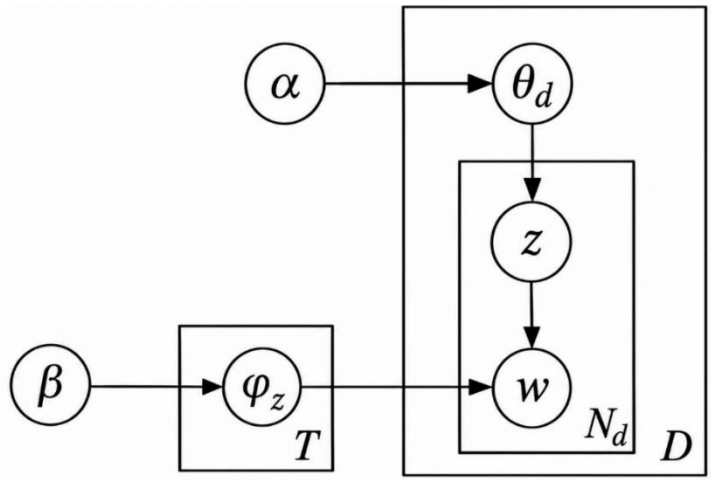
۲-۳- مدلسازی موضوع

موضوع، مساله یا ایده خاصی است که به‌عنوان موضوع یک پاراگراف، مقاله، گزارش یا سخنرانی عمل می‌کند. مدلسازی موضوع یکی از روش‌های یادگیری بدون نظارت است که توانایی گرفتن ساختار معنایی پنهان در یک سند را دارد. ایده اصلی مدلسازی موضوع این است که هر سند مجموعه‌ای از موضوعات است و یک موضوع از مجموعه‌ای از کلمات ساخته شده‌است. مدلسازی موضوع با موفقیت برای تسهیل بازیابی اطلاعات، طبقه‌بندی اسناد و تجزیه‌وتحلیل اکتشافی مجموعه‌های بزرگی از متون مورداستفاده قرار گرفته است [16]. مدلسازی موضوع در ابتدا در دهه ۱۹۸۰ توسعه یافت و از حوزه موضوعی "مدلسازی احتمالی مولد[1]" منشعب شد. این نوع مدلسازی فرض می‌کند که متغیرهای مشاهده‌شده با پارامترهای مشاهده‌نشده یا نهفته در یک رابطه احتمالی خاص تعامل دارند که سپس داده‌ها را در یک مجموعه داده تولید می‌کند. توسعه این فرایندها از نیاز به توصیف مختصر عناصر در مجموعه‌های بزرگ داده‌ها بدون به خطر انداختن روابط آماری موردنیاز برای تکمیل تحلیل‌های ساده‌تر، مانند طبقه‌بندی

[1] Generative Probabilistic Modeling (GPM)

و خلاصه‌سازی نشات گرفت. اولین روش توسعه‌یافته برای این کار، طرح کاهش[1] *tf-idf* نام داشت و توسط دو محقق در زمینه بازیابی اطلاعات، سالتون و مک گیل [36]، در سال ۱۹۸۳ ارایه شد. در این روش، هر ردیف داده به‌طور مجزا در یک مجموعه با واژگان خود در نظر گرفته می‌شود و تعداد تکرارهای هر کلمه برای تشکیل یک مقدار شمارش محاسبه می‌شود تا تعداد فراوانی کلمه یا اصطلاح (*tf*) مخصوص آن ردیف داده برای آن کلمه ایجاد شود. مجموع نمونه‌های یک کلمه در کل فایل داده‌ها که شمارش فرکانس سند معکوس (*idf*) نامیده می‌شود، نیز محاسبه می‌شود. پس از نرمال کردن این مقادیر به مجموعه داده در صورت نیاز، دو مقدار با یکدیگر مقایسه می‌شوند و یک ماتریس متن به ردیف داده حاوی مقادیر-*tf idf* برای تمام ردیف‌های داده درون ستون‌ها را تشکیل می‌دهند.

اولین تلاش برای توسعه مدلسازی موضوع توسط پاپادیمیتریو و همکاران [37] و هافمن [38] انجام شد. پاپادیمیتریو و همکارانش "مدلسازی موضوع" را بر اساس جبر خطی پیشنهاد دادند، ولی هافمن برای این هدف یک مدل احتمالاتی ارایه کرد. روش‌های رایج فعلی در مدلسازی موضوع، تحلیل معنایی پنهان احتمالی و تخصیص دیریکله نهفته است که از یک مدل مخلوط احتمال و مدل لفظی/اصطلاحی برای استخراج موضوعات از داده‌های متنی استفاده می‌کند [39]. الگوریتم تخصیص دیریکله پنهان که در این پژوهش به کار می‌رود، الگوریتمی از نوع مدلسازی احتمالی است که اولین بار توسط بلای و جردن [40] در سال ۲۰۰۳ معرفی شد، این الگوریتم محبوب‌ترین مدل موضوع و الگوریتم تجزیه‌وتحلیل موضوع در حال حاضر در نظر گرفته می‌شود. تخصیص دیریکله پنهان یک روش مدلسازی احتمالی مولد بدون نظارت برای کشف ساختارهای معنایی پنهان در مجموعه‌ای از اسناد متنی است. تخصیص دیریکله پنهان با هر سند به‌عنوان مخلوطی از موضوعات برخورد می‌کند و هر موضوع را ترکیبی از کلمات می‌داند [16]. 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 واژه‌های W را بر اساس توزیع اولیه دیریکله با پارامتر β بیان می‌کند. β نشان‌دهنده تراکم موضوع-کلمه است. θ ماتریسی است که توزیع متن d بر روی موضوع‌های T را بر اساس توزیع اولیه دیریکله با پارامتر α بیان می‌کند. α نشان‌دهنده چگالی سند-موضوع است. برای زایش هر واحد واژگانی w در متن d ، یک موضوع z از توزیع موضوعی مربوط به متنθ_d به‌دست می‌آید؛ درحالی‌که خود آن واژه w از توزیع واژه‌های موضوع انتخاب‌شده φ_z به‌دست می‌آید [41].



شکل ۲- نحوه کار الگوریتم تخصیص دیریکله پنهان ([41]).
Figure 2- How the latent dirichlet allocation algorithm works ([41]).

در این پژوهش، با استفاده از کلمات ظاهرشده درنتیجه پیاده‌سازی الگوریتم به‌عنوان گره‌های گراف هم رخدادی داده‌ها استفاده کرده و هم رخدادی این کلمات را برای نیل به موضوعات موجود در داده‌ها با دقت بیشتر بررسی کرده‌ایم.

[1] Term Frequency–Inverse Document Frequency (TF-IDF) reduction scheme

۴-۲- تحلیل هم رخدادی واژگان

اسمال و گریفیث [42] تحلیل هم رخدادی واژگان را یک نوع تحلیل هم استنادی می‌دانند که روشی مهم برای تصویرسازی روابط میان مفاهیم، ایده‌ها و مسایل علمی محسوب می‌شود. همچنین، آن‌ها تحلیل هم رخدادی واژگان را یک شیوه تحلیل کمی مهم به‌منظور کشف شبکه مفاهیم حوزه‌های علمی گوناگون می‌دانند. ادواردز [43] بیان می‌دارد که تحلیل هم رخدادی واژگان یک تکنیک تجزیه‌وتحلیل محتوای خودکار است که در ساخت نقشه‌برداری علمی از طریق بررسی روابط داده موثر است. همچنین، این روش از فنون مدلسازی گرافیکی برای مصورسازی ارتباطات بین مفاهیم و تحلیل آن‌ها استفاده می‌کند.

هی [44] تحلیل هم رخدادی واژگان را به‌عنوان روشی از تحلیل محتوا توصیف می‌نماید که الگوهایی را بررسی می‌کند که در آن جفت کلمات یا عبارات باهم در یک بدنه متن ظاهر می‌شوند. هدف این تکنیک ایجاد ارتباطات بین ایده‌ها در حوزه‌های موضوعی خاص است. به بیان دینگ و همکاران [45]، تحلیل هم رخدادی واژگان شیوه‌ای نو از تحلیل محتوا است که قادر است داده‌ها را با حفظ اطلاعات ضروری آن‌ها در قالب نمایش‌های بصری ارایه دهد. تعاریف ارایه‌شده مفهوم تحلیل هم رخدادی واژگان را به شیوه زیر روشن می‌کند [46]:

تحلیل هم رخدادی واژگان به‌عنوان یک رویکرد ارزشمند برای بررسی محتوای متن عمل می‌کند و مکانیسم مفیدی برای نشان دادن روابط و ارتباطات بین عناصر اطلاعاتی در مجموعه داده‌های متنی ارایه می‌دهد. این تکنیک تحلیلی متکی به فراوانی واژگان، عبارات و توصیف‌گرهای موجود در متون علمی است؛ بنابراین، چارچوب مفهومی اساسی حوزه موضوع را با مشخص کردن و برقراری پیوندها بین مفاهیم اصلی آن رونمایی می‌کند. این رویکرد به‌طور معمول برای نگاشت، ردیابی و تجسم الگوها و گرایش‌های موضوعی در یک حوزه تحقیقاتی مشخص و دست‌یابی به واژگان اساسی حوزه موضوع مورداستفاده قرار می‌گیرد. نخستین تلاش‌ها در زمینه واژگان درون‌متنی توسط زیف [47] در سال ۱۹۴۹ انجام شد. زیف با مطالعه واژه‌هایی که در متون انگلیسی به کار رفته‌اند به مصادیقی برای اصل کمترین کوشش دست‌یافت. اسمال و گریفیث [42] 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 جامعه‌شناسی خلاقیت کول[1] فرانسه در دهه ۱۹۸۰ به‌کار گرفته شد [48]. آن‌ها سیستم طراحی‌شده بر اساس این تکنیک را *LEXIMAPPE* نامیدند. این سیستم اولین نرم‌افزار مورداستفاده در پژوهش‌های مرتبط با تحلیل هم رخدادی واژگان در دهه ۱۹۸۰ میلادی بود [46]. این تحلیل به‌عنوان شاخصی در علم‌سنجی در این دهه معرفی شد و بعد از آن توسعه پیدا کرد و اغلب به‌عنوان جایگزینی برای رویکردهای استنادی و هم استنادی برای ترسیم علم استفاده شد [48]. کالون [49]، اولین فردی بود که به توصیف روش تحلیل هم رخدادی واژگان پرداخت و از آن در پژوهش‌های خود استفاده کرد. به اعتقاد او، تحلیل هم رخدادی واژگان نسبت به تحلیل‌های هم‌استنادی شیوه مناسب‌تری برای شناسایی ساختار علمی محسوب می‌شود. کالون و همکاران [50]، 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 مقیاس‌های هم رخدادی واژگان تحت تاثیر مطالعات بعدی تا اندازه زیادی توسعه یافت و بهبود پیدا کرد [29]. احمدی و عصاره [51] در پژوهش خود با بررسی مطالعات مختلف در حوزه تحلیل شبکه هم رخدادی واژگان، کاربردهای زیر را برای این روش معرفی کرده‌اند:

۱. طبقه‌بندی مدارک بر اساس موضوعات [50]
۲. ترسیم حرکت و پویایی علم [50]، [52]
۳. ترسیم ساختار حوزه‌ها و زمینه‌های علمی [53]
۴. ترسیم روابط میان پژوهش‌های بنیانی و پژوهش‌های فناورانه [54]
۵. کشف الگوهای برجسته و رویدادهای در حال ظهور
۶. ارزیابی درون داد/برون داد روابط در یک شبکه پژوهش [55]

[1] Ecole

۷. دیداری‌سازی شبکه مفهومی علم و فناوری [56]
۸. تحلیل سیر تحول مفهومی در طول دوره‌های زمانی
۹. تعیین و تحلیل نواحی موردپژوهش
۱۰. خوشه‌بندی مفاهیم حوزه‌های علمی
۱۱. برنامه‌ریزی و سیاست‌گذاری در حوزه‌های علمی
۱۲. درک و کشف روابط پنهان در حوزه‌های علوم مختلف
۱۳. تعیین روابط سلسله مراتبی مفاهیم یک حوزه یا یک زمینه خاص علمی

با توجه به کاربردهای ذکرشده، می‌توان از ترکیب تحلیل هم رخدادی واژگان و تکنیک‌های متن‌کاوی بهره گرفت و جهت بهبود کیفیت و دقت نتایج استفاده از تکنیک‌هایی نظیر مدلسازی موضوع، آن‌ها را با تحلیل هم رخدادی واژگان ترکیب کرد. ترکیب مدلسازی موضوع با شبکه‌های هم رخدادی واژگان می‌تواند کشف موضوع را با استفاده از روابط کلمات تقویت کند [20]. با ادغام مدلسازی موضوع با شبکه‌های هم رخدادی واژگان، محققان می‌توانند روابط عمیق‌تر و سلسله‌مراتب موضوعی را برای تجزیه‌وتحلیل موضوعی روشن‌تر و جامع‌تر کشف کنند. مقالات پژوهشی ادغام مدلسازی موضوع با شبکه‌های هم رخدادی واژگان را برای افزایش کیفیت موضوعات استخراج شده بررسی کرده‌اند. سیمونتی و همکاران [57]، روشی مبتنی بر شبکه‌های معتبر آماری[1] برای ارزیابی انسجام موضوع با استفاده از هم رخدادی کلمه پیشنهاد کرده‌اند.

زو و همکاران [21]، از روابط قضاوت صحیح بین کلمات در متون برای فیلتر کردن لبه‌های نویز در شبکه‌های هم رخدادی استفاده کرده و کیفیت استخراج موضوع را افزایش دادند. آستین و همکاران [22]، الگوریتم موضوع انجمن را معرفی کردند که جوامع را از شبکه‌های هم رخدادی واژگان استخراج می‌کند تا ساختارهای موضوعی سلسله مراتبی با انسجام و روابط بین موضوعات بهبودیافته را نشان دهد. موراکامی و چاکرابورتی [58]، استفاده از بردارهای تعبیه کلمات از مجموعه‌های بزرگی مانند ویکی‌پدیا را برای جبران اطلاعات محدود هم رخدادی کلمه در متون کوتاه، با هدف افزایش ارزیابی انسجام موضوع بررسی کردند. با توجه به مقالات بررسی‌شده، می‌توان به نوآور بودن این پژوهش در زمینه به‌کارگیری ترکیب مدلسازی موضوع و تحلیل هم رخدادی واژگان جهت استخراج عبارت‌های موجود در نظرات ذی‌نفعان پی برد. این کار باعث کاهش خطای مدلسازی موضوع و افزایش کیفیت یافتن موضوعات موجود در متن می‌شود. با توجه به مطالب فوق‌الذکر و در جهت بهبود کیفیت نتایج مدل، در این پژوهش پس از پیاده‌سازی الگوریتم تخصیص دیریکله پنهان، با در نظر گرفتن کلمات ظاهرشده در نتایج اجرای الگوریتم به‌عنوان کلمات کلیدی یا به‌عبارت‌دیگر گره‌های شبکه هم رخدادی واژگان، اقدام به تشکیل ماتریس هم رخدادی واژگان و رسم این شبکه کرده‌ایم.

۳- روش پژوهش

با توجه به مطالب ذکرشده در راستای ضرورت و اهمیت تحلیل داده‌های متنی با روش‌های کارآمد برای انجام تحلیل بازخوردهای بازار بر مبنای تحلیل نظرات ذی‌نفعان، پس از دریافت داده‌های متنی نظرات گروه‌های مختلف ذی‌نفعان موردبررسی، مدل پیشنهادی پژوهش روی این داده‌ها پیاده‌سازی شده است. پس از توضیح چگونگی جمع‌آوری اطلاعات و معرفی ذی‌نفعان موردمطالعه، مدل پیشنهادی به تفصیل شرح داده شده است.

۳-۱- جمع‌آوری اطلاعات و معرفی ذی‌نفعان

موردمطالعه این پژوهش، یک پالایشگاه روان‌کار می‌باشد و مدل پیشنهادی روی داده‌های نظرات گروهی از ذی‌نفعان این شرکت پیاده‌سازی شده است. ذی‌نفعان موردمطالعه در این پژوهش، عاملین فروش پالایشگاه، کاربران سایت پالایشگاه و مصرف‌کنندگان برخی محصولات این پالایشگاه هستند. بخش فروش در این پالایشگاه به دو بخش کلی‌فروش عاملین و فروش صنایع تقسیم می‌شود. ازآنجایی‌که این شرکت محصولاتی متناسب با نیازهای صنایعی نظیر صنایع فولادی، دریایی و هوایی را تولید می‌کند، مشتریان مرتبط با این صنایع در بخش فروش صنایع قرار می‌گیرند. همچنین محصولات خودرویی شرکت نظیر انواع روغن‌موتور و ضد یخ توسط عاملین فروش در دسترس عموم قرار می‌گیرد. کانال‌های فروش دیگری ازجمله

[1] Statistically Validated Networks (SVN)

فروشگاه اینترنتی دیجی‌کالا برای برخی از این محصولات که مصرف عمومی دارند هم فعال هستند. با همکاری واحد مهندسی فروش پالایشگاه، داده‌های مربوط به چالش‌ها و نقطه نظرات عاملین فروش از بخش پیشنهادات در نظرسنجی‌های شش‌ماهه شرکت استخراج شد و داده‌های مربوط به کاربران سایت پالایشگاه که همان نظرات کاربران در سایت بود دریافت شده است. به‌منظور بررسی نظرات گروهی از مصرف‌کنندگان محصولات پالایشگاه، نظرات موجود در صفحه مربوط به چهار محصول "روغن‌موتور شماره ۱"، "روغن‌موتور 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 داده‌های نظرسنجی‌های شش‌ماهه 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 پاسخ‌دهنده	تعداد داده در بخش پیشنهادات
1400	نیمه اول	صنایع	0
		عاملین فروش	0
	نیمه دوم	صنایع	5
		عاملین فروش	30
1401	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** قابل‌مشاهده است، جهت پیاده‌سازی مدل داده‌های نظرسنجی مربوط به عاملین فروش در سال ۱۴۰۱ به‌عنوان یک سری داده در نظر گرفته شد.

جدول ۲- وضعیت داده‌های 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 محصول	1398-1399	1399-1400	1400-1402
			تعداد نظر	تعداد نظر	تعداد نظر
روغن‌های خودرویی	روغن‌های 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
		7	2	2	0
		8	6	1	3
		9	4	3	6
	روغن‌های خودرویی (دنده)	10	3	3	0
		11	1	1	0
		12	0	1	1
		13	3	6	7
		14	6	3	7
	روغن‌های خودرویی (دیزلی)	15	1	2	1
		16	0	0	1
		17	0	2	3
		18	1	1	0
تعداد داده‌ها			159	152	122
روغن‌های صنعتی	روغن‌های صنعتی (سایر)	1	0	0	1
		2	0	0	1
	روغن‌های صنعتی (کمپرسور)	3	1	0	0
تعداد داده‌ها			1	0	1

جدول ۲- ادامه.
Table 2- Continued.

گروه	دسته	شماره محصول	1398-1399	1399-1400	1400-1402
			تعداد نظر	تعداد نظر	تعداد نظر
ضد یخ	ضد یخ	1	1	5	7
		2	0	0	1
تعداد داده‌ها			2	5	9
گریس	گریس	1	0	1	0
		2	0	1	1
تعداد داده‌ها			0	2	1
وکس	وکس	1	0	0	1
تعداد داده‌ها			0	0	1
_	_	1	0	1	1
		2	0	0	1
		3	0	0	1
تعداد داده‌ها			0	1	3

با توجه به تعداد نظرات روغن‌های خودرویی در سه بازه زمانی سال‌های ۱۳۹۸ تا ۱۳۹۹، ۱۳۹۹ تا ۱۴۰۰ و ۱۴۰۰ تا ۱۴۰۲ موجود در **جدول ۲**، این سه سری داده برای پیاده‌سازی مدل مناسب شناخته شدند.

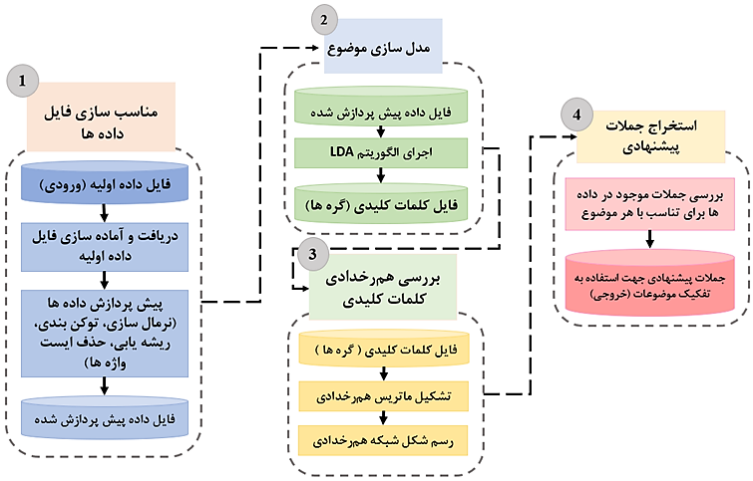
جدول ۳- معرفی محصولات موردبررسی و وضعیت داده‌های نظرات کاربران دیجی‌کالا مربوط به چهار محصول پالایشگاه از نظر تعداد.
Table 3- Overview of the products under study and the status of user review data from Digikala for the four selected products, in terms of quantity.

محصول منتخب	لینک صفحه محصول در دیجی‌کالا	تعداد نظرات
روغن‌موتور شماره ۱	B2n.ir/j40743	1523
روغن‌موتور شماره ۴	B2n.ir/g89488	771
ضد یخ شماره ۱	B2n.ir/h41100	263
ضد یخ شماره ۲	B2n.ir/u62949	75

از بین محصولاتی که در فروشگاه اینترنتی دیجی‌کالا عرضه می‌شد، چهار محصول تعداد نظرات بیشتر از ۵۰ عدد داشتند که قابلیت پیاده‌سازی مدل روی آن‌ها وجود داشت. لذا نظرات این چهار محصول از سایت دیجی‌کالا استخراج شده و موردمطالعه قرار گرفت. همان‌طور که در **جدول ۳** مشاهده می‌شود، روغن‌موتور شماره ۱ بیشترین تعداد نظرات و ضد یخ شماره ۲ کمترین تعداد نظرات را بین محصولات انتخاب‌شده دارند.

۲-۳- مدل پیشنهادی پژوهش

شکل زیر، مدل پیشنهادی جهت شناسایی چالش‌ها و نقطه نظرات ذی‌نفعان با استفاده از تحلیل هم رخدادی واژگان کلیدی را نشان می‌دهد. این مدل به‌طورکلی از چهار بخش تشکیل شده و مراحل هر بخش به‌طور مختصر در **شکل ۳** توضیح داده شده است.



شکل ۳- مدل پیشنهادی پژوهش.

Figure 3- Proposed research model.

مناسب‌سازی فایل داده‌ها: پیش‌پردازش داده‌ها شامل مراحل مختلفی است که متناسب با داده‌ها و روش مورداستفاده، می‌توان برخی مراحل آن را انجام داد. برای انجام پیش‌پردازش داده‌ها، تعدادی از ابزارهای پیش‌پردازش متن فارسی نظیر پارسی ور[1]، فارسی یار[2]، دادماتولز[3] و هضم[4] موردبررسی قرار گرفتند که با توجه به ادبیات فارسی موجود در حوزه متن‌کاوی و آزمایش‌های انجام‌شده در استفاده از این نرم‌افزارها، ابزار هضم برای نرمال‌سازی و جداسازی کلمات و ابزار پرشین استمر[5] برای ریشه‌یابی کلمات استفاده شده است. همچنین حذف ایست واژه‌ها با استفاده از لیست ایست واژه‌های فارسی آقای وحید خرازی و گروه *Awesome Persian NLP* و افزودن برخی موارد به‌صورت دستی پس از اجرای متعدد الگوریتم روی داده‌ها انجام شده است. مراحل مختلف پیش‌پردازش عبارت‌اند از:

نرمال‌سازی[6] داده‌ها: هدف اصلی در بخش نرمال‌سازی، یکپارچه‌سازی و متحدالشکل کردن کاراکترها به‌منظور کنترل بیشتر بر نتایج پردازش‌های پیشرفته‌تر بر متن است که در این پژوهش برای نرمال‌سازی داده‌ها از کتابخانه هضم در محیط *IDLE* پایتون استفاده شده است. نمودار مواردی که در ابزار نرمال‌سازی هضم انجام می‌شود در شکل زیر آمده است.

شکل ۴- فرایند نرمال‌سازی داده‌ها با استفاده از کتابخانه هضم.

Figure 4- Data normalization process using the Hazm library.

[1] Parsivar
[2] Farsiyar
[3] Dadma tools
[4] Hazm
[5] Persian stemmer
[6] Normalization

توکن بندی[1] داده‌ها: 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 تشکیل‌دهنده آن شکست. متد *split_token_words* ابتدا کلمات را بر اساس نیم‌فاصله (در صورت وجود) می‌شکند و لیستی از کاندیدها را (برای استخراج کلمات ساده از کلمه اصلی) به وجود می‌آورد. سپس تمامی زیررشته‌های ناهمپوشان (با حداقل طول ۱ و حداکثر طول ۱) را بن‌یابی کرده و با حذف حالت‌های تکراری، حالت‌هایی را که در دادگان کلمات ابزار هضم وجود دارند به‌عنوان کاندیدها به‌عنوان خروجی نشان می‌دهد[2].

ریشه یابی[3] کلمات: ریشه جزئی از کلمه است که پس از حذف تمام عناصر الحاقی یا وندهای کلمه "پسوند، پیشوند، میانوند" بر جای می‌ماند. در این پژوهش، ریشه‌یابی کلمات با استفاده از کتابخانه *persian stemmer* در محیط *IDLE* پایتون انجام شده است.

حذف ایست واژه‌ها: جهت کاهش حجم و افزایش کیفیت داده‌ها و همچنین جلوگیری از افزایش احتمال خطا در الگوریتم، کلماتی مثل حروف اضافه (در، به، با و ....) و علایم نگارشی حذف می‌شوند. برای انجام این کار با دقت بالاتر، در این پژوهش اقدام به استفاده از دو لیست ایست واژه‌های فارسی به‌صورت هم‌زمان نموده و به‌وسیله کد پایتون این کلمات را از لیست داده‌ها حذف کردیم.

با توجه به اینکه ابزار کارآمدی برای اصلاح اشتباهات تایپی و املایی 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 به‌صورت فایل *Excel* از کد خروجی گرفتیم. همچنین داده‌های پیش‌پردازش شده نیز به‌صورت 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 یک‌بار در دیکشنری ظاهرشده‌اند با یک کلمه واحد یا شکل صحیح تایپی و املایی جایگزین شده‌اند. همچنین مواردی که علیرغم نرمال‌سازی متن و استفاده از لیست ایست واژه‌ها موجب کاهش دقت الگوریتم بودند حذف شدند. درنهایت، برای هر سری داده دو فایل *Excel* شامل موارد تغییریافته و موارد حذف‌شده تشکیل‌شده و در قسمت دوم کد در محیط *Jupyter Notebook* روی فایل داده پیش‌پردازش شده که از مرحله اول کد خروجی گرفته‌شده بود، اعمال شدند. پس از اعمال تغییرات ثانویه، مجددا دیکشنری تشکیل‌شده و در باقی مراحل از این دیکشنری استفاده شده است.

مدلسازی موضوع: در این فاز، روی داده‌های پیش‌پردازش شده که به‌عنوان ورودی از فاز قبل دریافت شده است، الگوریتم تخصیص دیریکله پنهان با استفاده از زبان برنامه‌نویسی پایتون اجرا شده و نتایج الگوریتم در غالب فایل اکسل به صورتی که برای ورود به نرم‌افزار گفی[4] 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.

جدول ۴- پارامترهای الگوریتم تخصیص دیریکله پنهان.
Table 4- LDA algorithm parameters.

پارامتر	مقدار تنظیم شده	شرح
تعداد مرور کل مجموعه داده‌ها	40	به‌عنوان پارامتر ثابت آموزش در تمام مدل‌ها تنظیم شد تا شرایط آموزشی مدل‌ها یکسان باشد.
بازه موردبررسی برای تعداد موضوع بهینه	۱ تا ۲۰ 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 ۲	100	به‌منظور تضمین تکرارپذیری نتایج، مقدار Random State برابر با ۱۰۰ برای مقداردهی اولیه مولد اعداد شبه‌تصادفی در اجرای مدل تخصیص دیریکله پنهان تنظیم شد.

[1] Tokenization
[2] https://data-hub.ir
[3] Stemming
[4] Gephi

جدول ۴- ادامه.
Table 4- Continued.

پارامتر	مقدار تنظیم شده	شرح
حداکثر تعداد تکرارها ۳	200	به‌منظور دستیابی به همگرایی مناسب مدل، حداکثر تعداد تکرارهای الگوریتم (Iterations) برابر با ۲۰۰ تنظیم شد.
پارامتر α	Symmetric	پارامتر Alpha (α) به‌صورت Symmetric تنظیم شد؛ به این معنا که در آغاز آموزش، احتمال اولیه یکسانی برای تمامی موضوعات در نظر گرفته شد.
پارامتر β	مقدار پیش‌فرض کتابخانه Gensim	برای پارامتر Beta (β) از مقدار پیش‌فرض کتابخانه Gensim استفاده شد.
اندازه دسته پردازش ۲	10000	اندازه دسته پردازش (Chunksize) برابر با ۱۰۰۰۰ در نظر گرفته شد. با توجه به اینکه تعداد اسناد کمتر از این مقدار بود، کل داده‌ها در یک دسته پردازش شدند.
تناوب به‌روزرسانی ۳	1	مقدار پارامتر Update Every برابر با ۱ تنظیم شد؛ بدین معنا که مدل پس از پردازش هر دسته از داده‌ها به‌روزرسانی می‌شود.
اختصاص موضوع به هر واژه ۴	True	گزینه Per Word Topics فعال شد تا علاوه بر توزیع موضوع در سطح اسناد، اطلاعات مربوط به ارتباط واژه‌ها با موضوع‌ها نیز محاسبه شود.
معیار انسجام موضوع	C-V	برای ارزیابی کیفیت موضوع‌های استخراج‌شده 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. درصورتی‌که 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 واژه‌های نمایش داده‌شده در هر موضوع	6	برای تفسیر هر موضوع، شش واژه با بیشترین وزن در خروجی مدل گزارش شد.

بررسی هم‌رخدادی کلمات کلیدی: در این فاز، با استفاده از فایل کلمات کلیدی به‌عنوان گره‌های شبکه هم رخدادی و فایل داده‌های پیش‌پردازش شده، ماتریس هم رخدادی را تشکیل داده‌ایم. در این ماتریس، مقادیر درایه‌ها صفر یا یک می‌باشد. هر ستون ماتریس به یک کلمه کلیدی و هر ردیف ماتریس به یک ردیف از داده‌ها مربوط است. درصورتی‌که کلمه مربوط به ستون jth در ردیف ith وجود داشته‌باشد، درایه a_{ij} ماتریس مقدار یک می‌گیرد و در غیر این صورت، مقدار درایه صفر خواهد بود. با استفاده از این ماتریس، یال‌های شبکه هم رخدادی به‌دست می‌آید و می‌توان شکل این شبکه را رسم کرد. پس از تشکیل ماتریس، فایل یال‌های گراف هم رخدادی به صورتی که برای ورود به نرم‌افزار گفی مطلوب باشد، ذخیره شده است. سپس با استفاده از فایل‌های گره‌ها و یال‌ها برای هر سری داده‌ها، شکل شبکه هم رخدادی واژگان کلیدی با استفاده از نرم‌افزار گفی که یک نرم‌افزار منبع باز برای تجسم و تجزیه‌وتحلیل نمودارهای شبکه‌های بزرگ است، رسم شده است و به‌عنوان خروجی فاز سوم مدل در نظر گرفته شده است.

استنتاج عبارت‌ها و محاسبه درصد رخداد آن‌ها: با توجه به یال‌هایی که وزن بیشتری در شبکه دارند، می‌توان هم‌رخدادی‌هایی که بیشتر از بقیه اتفاق افتاده‌اند را مشاهده کرد. با ایجاد ارتباط بین یال‌های با وزن بیشتر و گره‌های مرتبط با آن‌ها، می‌توان عبارات استنتاجی را به دست آورد.

هر عبارت از طریق وجود ارتباط بین تعدادی کلمه مشخص می‌شود. برای محاسبه درصد رخداد هریک از این عبارات در فایل داده‌ها، پس از قرار دادن یک نام نماینده برای عبارت، هم‌رخدادی‌هایی که نشان‌دهنده عبارت بوده‌اند را به‌عنوان حالت‌های رخداد عبارت تعریف می‌کنیم و یک ماتریس ثانویه تشکیل می‌دهیم که در صورت بروز هریک از این حالت‌ها در هر ردیف ماتریس هم‌رخدادی، در ستون آخر ماتریس ثانویه نام نماینده عبارت موردنظر نوشته می‌شود. درنهایت با شمارش تعداد هریک از عبارات و محاسبه درصد رخداد عبارات استنتاجی به دست می‌آید. در این پژوهش، این کار با استفاده از زبان پایتون و در محیط *Jupyter Notebook* انجام شده است. در ادامه به‌عنوان مثال برای عبارت " بسته‌بندی با ظرف چلیک ضعیف است " که یکی از عبارات مستخرج از داده‌های " چالش‌ها و نقطه نظرات عاملین پالایشگاه در سال ۱۴۰۱ " می‌باشد، مراحل محاسبه درصد رخداد بین داده‌های پیش‌پردازش شده را توضیح می‌دهیم. نام نماینده این عبارت، " *A1401.2* " قرار داده شده است.

ابتدا کلماتی که وجود هم‌رخدادی‌های پرتکرار بین آن‌ها نشان‌دهنده این عبارت است در **جدول ۵** به همراه شماره تخصیص‌یافته به هر کلمه مشخص می‌شود. شماره تخصیص‌یافته به هر کلمه در لیست کلیدواژه‌های ایجاد شده مشخص می‌شود. کلماتی که نشان‌دهنده عبارت " بسته‌بندی با ظرف چلیک ضعیف است" می‌باشند عبارت است از: بسته‌بندی، ضعیف، ظرف و چلیک.

جدول ۵- شماره تخصیص‌یافته به کلمات مثال.

Table 5- Numbers assigned to example words.

کلمه	شماره تخصیص‌یافته
ضعیف	0
چلیک	1
بسته‌بندی	4
ظرف	8

یال‌های ایجادشده بین کلمات فوق در **جدول ۶** در سه ستون مبدا، مقصد و وزن مشخص شده‌اند 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

با استفاده از قطعه شبه‌کد **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 قابل‌دسترسی است.

```
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- یافته‌های پژوهش

در این پژوهش، رویکرد پیشنهادی روی داده‌های نظرات و چالش‌های عاملین فروش پالایشگاه، کاربران سایت پالایشگاه و نظرات موجود در صفحه مربوط به تعدادی از محصولات پالایشگاه در فروشگاه اینترنتی دیجی‌کالا اجرا شده و نتایج آن تحلیل شده است. این پالایشگاه یکی از شرکت‌های فعال در تولید انواع روغن‌های پایه، موتوری، صنعتی و ... است. این شرکت، نخستین مجموعه روان‌کاری در صنعت کشور است که فناوری‌های مربوط به تولید فرآورده‌های روان‌کار را به کشور منتقل کرده و صدها مهندس و کارشناس را در طول خدمت در این شرکت با فناوری‌های به‌روز و ساخت و تولید روان‌کارها آشنا کرده است.

معیار انتخاب محصول برای اجرای مدل روی نظرات آن در فروشگاه اینترنتی دیجی‌کالا، کفایت تعداد نظرات موجود بوده است. همچنین داده‌های مربوط به نظرسنجی‌های شش‌ماهه شرکت نیز با همکاری واحد مهندسی فروش پالایشگاه دریافت شد که بر اساس تعداد داده‌های موجود در قسمت پیشنهادات، نظرسنجی عاملین در سال ۱۴۰۱ (نیمه اول و نیمه دوم سال) برای پیاده‌سازی مدل انتخاب شد. علاوه بر این، داده‌های نظرات کاربران در سایت شرکت در سه بازه زمانی بین سال‌های ۱۳۹۸ تا ۱۴۰۲ نیز با همکاری واحد مهندسی فروش دریافت شده و با توجه به دسته‌بندی نظرات بر اساس محصولات مربوطه، نظرات مربوط به روغن‌های خودرویی شرکت ازنظر تعداد برای پیاده‌سازی مدل از دسته‌های دیگر شرایط بهتری داشت و جهت پیاده‌سازی مدل انتخاب شد. در ادامه، نتایج اجرای فازهای مختلف مدل روی داده‌ها را در قالب کلمات کلیدی و تعداد تکرار آن‌ها، شبکه هم رخدادی واژگان کلیدی و استخراج عبارت‌های استنتاجی به تفکیک برای داده‌های موردبررسی ارایه‌شده است.

۴-۱- کلمات کلیدی و تعداد تکرار آن‌ها

کلمات کلیدی استخراج‌شده برای داده‌های موردنظر، همان کلماتی هستند که درنتیجه اجرای الگوریتم تخصیص دیریکله پنهان روی داده‌ها ذیل موضوعات مشاهده شدند. این کلمات به همراه تعداد تکرارشان در داده مربوطه برای هر سری از داده‌های موردبررسی به تفکیک در **پیوست الف** آمده‌اند.

۴-۲- شبکه هم رخدادی واژگان کلیدی

پس از مشخص شدن کلمات کلیدی، ماتریس هم رخدادی این کلمات در سراسر متن به‌دست آمده و یال‌های گراف به دست می‌آیند. شکل گراف هم رخدادی واژگان رسم شده برای داده‌های موردنظر در **پیوست الف** به تفکیک داده‌ها آمده است.

۴-۳- استخراج عبارت‌های استنتاجی

از مشاهده نتایج مربوط به پیاده‌سازی الگوریتم تخصیص دیریکله پنهان روی داده‌های موردنظر 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 داده‌های مربوطه، در **پیوست الف** مشاهده می‌شوند. لازم به ذکر است که برای حصول اطمینان بیشتر از درستی نتایج مدل، در مواردی که تعداد داده‌ها کمتر از ۱۰۰ ردیف بوده است، درستی نتایج و موضوعات به‌دست آمده به‌صورت دستی چک شده است.

عبارات استخراج‌شده از چالش‌ها و نقطه‌نظرات عاملین در سال ۱۴۰۱: با توجه به اینکه تکرار زیاد هم رخدادی بین برخی کلمات ازجمله "بسته‌بندی، ضعیف، ظرف، چلیک، تنوع، ضعیف، کارتن، رنگ‌بندی، اجبار، پکیج، جنس، خرید، رقیب، تبلیغ، لباس تبلیغاتی، تابلو تبلیغاتی" در **شکل الف- ۱** نشان‌دهنده عباراتی است که تکرار بیشتری داشته‌اند، چهار عبارت که در **جدول الف- ۹** آمده‌اند، استنتاج شدند.

همان‌طور که در **جدول الف- ۹** مشاهده می‌شود، عبارت "بسته‌بندی با ظرف چلیک ضعیف است" بیشترین درصد رخداد را دارد. درواقع، بیشترین چالش مطرح‌شده توسط عاملین پالایشگاه در سال ۱۴۰۱، مربوط به ضعف ظرف چلیک استفاده شده در بسته‌بندی محصولات پالایشگاه است. در این راستا، پالایشگاه می‌تواند برای افزایش سطح رضایت عاملین و تلاش در راستای افزایش تعداد عاملین فروش خود به‌وسیله تبلیغات دهان‌به‌دهان،

این موضوع را بررسی کرده و از راهکاری نظیر تغییر مظروف مورداستفاده در بسته‌بندی یا تقویت جنس بسته‌بندی چلیک استفاده نماید. درواقع، استفاده از استراتژی توسعه محصول در این مورد توصیه می‌شود.

عبارات استخراج‌شده از نظرات کاربران سایت شرکت در مورد روغن‌های خودرویی در بازه زمانی ۱۳۹۸ تا ۱۳۹۹: با توجه به اینکه تکرار زیاد هم رخدادی بین برخی کلمات ازجمله "پیشنهاد، *10W40*، *20W50*، روغن‌موتور شماره ۴، روغن‌موتور شماره ۵، فلومکس، کیفیت، شماره ۴، تفاوت" در **شکل الف- ۳** نشان‌دهنده عباراتی است که تکرار بیشتری داشته‌اند، دو عبارت که در **جدول الف- ۱۰** آمده‌اند، استنتاج شدند.

با توجه به درصد رخداد محاسبه‌شده، عبارت "تقاضای ارایه محصول پیشنهادی مناسب خصوصا ازنظر گرانروی برای انواع خودروها" از بیشترین درصد رخداد برخوردار است. لذا شرکت می‌تواند برای پیشرفت عملکرد خود در بخش فروش و در راستای کمک به انتخاب صحیح محصول از سوی خریداران، نسبت به راه‌اندازی یک سیستم توصیه گر بر مبنای نوع و سال تولید خودرو یا اضافه کردن چنین مواردی در بخش فیلترها در سایت، برای ارایه راهنمایی بهتر به مشتریان در سایت پالایشگاه اقدام نماید.

راهکار دیگری که در این مورد توصیه می‌شود این است که در رابطه با اینکه هرکدام از محصولات خودرویی این پالایشگاه برای چه خودروهایی مناسب هستند، اطلاعاتی در بخش مقالات سایت اضافه شود؛ به‌عبارت‌دیگر، ارایه راهنمایی و اطلاعات در این حوزه یا ایجاد سیستم توصیه محصولات بر اساس ویژگی‌های انتخاب‌شده برای خودرو توسط مشتریان، موجب افزایش رضایت مشتریان از خرید شده و از دو طریق تبلیغات دهان‌به‌دهان و بازاریابی دیجیتال، موجب افزایش حجم فروش خواهد شد. به عبارتی استراتژی‌های تکنولوژی اطلاعات یا نفوذ در بازار به دلیل افزایش تلاش‌های فروش برای این مورد مناسب است.

عبارات استخراج‌شده از نظرات کاربران سایت شرکت در مورد روغن‌های خودرویی در بازه زمانی ۱۳۹۹ تا ۱۴۰۰: با توجه به اینکه تکرار زیاد هم رخدادی بین برخی کلمات ازجمله " اصالت، بازار، تقلبی، فلومکس، روغن‌موتور شماره ۴، کیفیت، روغن‌موتور شماره ۲، *euro 4*، پراید، راهنمایی، *TU5*" در **شکل الف- ۴** نشان‌دهنده عباراتی است که تکرار بیشتری داشته‌اند، دو عبارت که در **جدول الف- ۱۱** آمده‌اند، 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 می‌تواند از طریق ایجاد رمزینه[1] یا رمزینه پاسخ سریع[2] بر روی کالاهای خود و اطلاع‌رسانی این موضوع از طریق کانال‌های تبلیغاتی شرکت به مصرف‌کنندگان، به آن‌ها اطمینان دهد که می‌توانند از اصالت محصولی که خریداری کرده‌اند مطمئن شوند.

عبارات استخراج‌شده از نظرات کاربران سایت شرکت در مورد روغن‌های خودرویی در بازه زمانی ۱۴۰۰ تا ۱۴۰۲: با توجه به اینکه تکرار زیاد هم رخدادی بین برخی کلمات ازجمله "روغن‌موتور شماره ۲، قوطی، تعویض، فلزی، پلاستیکی، گالن، کیفیت، تیبا، دوگانه‌سوز، پیشنهاد، پراید، راهنمایی، اصالت، محصول" در **شکل الف- ۵** نشان‌دهنده عباراتی است که تکرار بیشتری داشته‌اند، سه عبارت که در **جدول الف- ۱۲** آمده‌اند، استنتاج شدند.

عبارت "پلاستیکی یا فلزی بودن مظروف محصول روغن‌موتور شماره ۲" بیشترین درصد رخداد را به خود اختصاص داده است که با توجه به اینکه این عبارت از داده‌های نظرات سایت پالایشگاه استخراج شده است، نشان‌دهنده ابهام در مورد مظروف این محصول است. درصورتی‌که اخیرا تغییری در بسته‌بندی این محصول صورت گرفته است، شرکت می‌تواند از طریق کانال‌های ارتباطی و تبلیغاتی خود این موضوع را به اطلاع مشتریان برساند. وجود چنین ابهام‌هایی ممکن است موجب شود مشتریان به اصالت کالا مطمئن نباشند و به همین دلیل از خرید خود منصرف شوند.

[1] Barcode

[2] Quick Response Code (QR Code)

عبارات استخراج‌شده در مورد روغن‌موتور شماره ۱: با توجه به اینکه تکرار زیاد هم رخدادی بین برخی کلمات ازجمله " قیمت، به‌صرفه، پایین، ارزان، اقتصادی، نشتی، قوطی، بسته، ضربه، ریخت، خودرو، قدیمی، پراید، کاربراتور، دیجی‌کالا، بازار، منصفانه" در **شکل الف- ۶** نشان‌دهنده عباراتی است که تکرار بیشتری داشته‌اند، چهار عبارت که در **جدول الف- ۱۳** آمده‌اند، استنتاج شدند.

بین عبارات استخراج‌شده از نظرات کاربران سایت دیجی‌کالا در مورد روغن‌موتور 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 دیجی‌کالا در خرید انجام‌شده توسط این دسته از مشتریان، این موضوع باید علت‌یابی و رفع شود. چراکه این موضوع هم موجب نارضایتی مشتریان شده، هم ممکن است بر تصمیم خرید مشتریان بالقوه تاثیر منفی بگذارد.

عبارات استخراج‌شده در مورد روغن‌موتور شماره ۴: با توجه به اینکه تکرار زیاد هم‌رخدادی بین برخی کلمات ازجمله " قیمت، بازار، دیجی‌کالا، دولتی، به‌صرفه، کیفیت، مناسب، روغن" در **شکل الف- ۷** نشان‌دهنده عباراتی است که تکرار بیشتری داشته‌اند، دو عبارت که در **جدول الف- ۱۴** آمده‌اند، استنتاج شدند.

در مورد محصول روغن‌موتور پارس روغن‌موتور شماره ۴، بیشترین درصد رخداد مربوط به عبارت "تفاوت قیمت محصول در بازار و دیجی‌کالا و قیمت به‌صرفه محصول در دیجی‌کالا" می‌باشد. لذا شرکت با تامین موجودی این محصول در دیجی‌کالا می‌تواند این کانال فروش موثر را نگه دارد که خریداران بیشتری بتوانند با قیمت مناسب از این محصول بهره‌مند شوند و درنتیجه میزان فروش محصول افزایش یابد.

عبارات استخراج‌شده در مورد ضد یخ شماره ۱: با توجه به اینکه تکرار زیاد هم‌رخدادی بین برخی کلمات ازجمله " کیفیت، ضدیخ، قیمت، مناسب، روغنی، محصول، قوطی، تعویض، فلزی، بازار، دیجی کالا" در **شکل الف- ۹** نشان‌دهنده عباراتی است که تکرار بیشتری داشته‌اند، سه عبارت که در **جدول الف- ۱۵** آمده‌اند، استنتاج شدند.

بیشترین عبارت تکرار شده بین نظرات کاربران سایت دیجی‌کالا در مورد ضد یخ شماره ۱، "کیفیت و قیمت مناسب ضد یخ" می‌باشد که نشان‌دهنده رضایت مصرف‌کنندگان از این محصول است.

عبارات استخراج‌شده در مورد ضد یخ شماره ۲: با توجه به اینکه تکرار زیاد هم‌رخدادی بین برخی کلمات ازجمله " کیفیت، ضدیخ، قیمت، روغنی، محصول، قوطی، فلزی" در **شکل الف-۱۰** نشان‌دهنده عباراتی است که تکرار بیشتری داشته‌اند، دو عبارت که در **جدول الف- ۱۶** آمده‌اند، استنتاج شدند.

درصد رخداد نزدیک به ۷۰% برای عبارت "کیفیت خوب، روغنی بودن و قیمت ضد یخ" به‌دست آمده است که نشان‌دهنده مزایای این محصول در نگاه مصرف‌کنندگان آن است.

۵- بحث و نتیجه‌گیری

در این پژوهش، مدلی برای استخراج موضوعات مطرح‌شده در داده‌های متنی نظرات ذی‌نفعان سازمان ارایه شد. اجزای اصلی این مدل، مدلسازی موضوع و شبکه هم رخدادی واژگان هستند که با استفاده از نتایج به دست آمده از آن‌ها عبارات مطرح‌شده در مجموعه داده‌ها مشخص می‌شوند و درصد رخداد آن‌ها با توجه به کلمات و یال‌های مربوط به این عبارات در داده‌ها محاسبه می‌شود. این مدل روی نظرات گروهی از ذی‌نفعان پالایشگاه شامل عاملین فروش پالایشگاه، کاربران سایت پالایشگاه و خریدارانی که از طریق فروشگاه‌های اینترنتی (فروشگاه دیجی‌کالا) اقدام به خرید محصولات پالایشگاه نموده‌اند پیاده‌سازی شده است. خلاصه نتایج به‌دست آمده در **جدول ۷** آمده است.

جدول ۷- خلاصه نتایج به‌دست آمده از پژوهش.
Table 7- Summary of results obtained from the research.

عنوان داده	عبارت پرتکرار	درصد رخداد عبارت	پیشنهاد مدیریتی ارایه‌شده
چالش‌ها و نقطه‌نظرت عاملین فروش در سال ۱۴۰۱	بسته‌بندی با ظرف چلیک ضعیف است.	31.01%	بهبود جنس ظرف بسته‌بندی یا تغییر نوع ظرف / استراتژی توسعه محصول
نظرات کاربران سایت پالایشگاه در بازه زمانی ۱۳۹۹-۱۳۹۸	تقاضای ارایه محصول پیشنهادی مناسب خصوصا ازنظر گرانروی برای انواع خودروها	32.26%	ارایه راهنمایی و اطلاعات در مورد اینکه هر محصول برای چه نوع و مدل خودروهایی مناسب است یا ایجاد سیستم توصیه محصول مناسب بر اساس ویژگی‌های خودروی موردنظر خریدار/استراتژی‌های تکنولوژی اطلاعات یا استراتژی نفوذ در بازار
نظرات کاربران سایت پالایشگاه در بازه زمانی ۱۴۰۰-۱۳۹۹	نگرانی در مورد اصالت کالا و وجود کالای تقلبی در بازار	35.57%	ایجاد رمزینه یا رمزینه پاسخ سریع روی کالاها و اطلاع‌رسانی به مشتریان از طریق کانال‌های ارتباطی و تبلیغاتی شرکت
نظرات کاربران سایت پالایشگاه در بازه زمانی ۱۴۰۲-۱۴۰۰	پلاستیکی یا فلزی بودن مظروف روغن‌موتور شماره ۲	36.67%	اطلاع‌رسانی تغییرات بسته‌بندی از طریق کانال‌های ارتباطی و تبلیغاتی شرکت
نظرات کاربران سایت دیجی‌کالا در مورد روغن‌موتور شماره ۱	قیمت مناسب و اقتصادی روغن	21.99%	کنترل هزینه‌های تولید این محصول جهت ثبات قیمت
	ضربه خورده بودن قوطی و نشت مواد از قوطی	17.44%	پیگیری برای علت‌یابی و رفع این مساله
نظرات کاربران سایت دیجی‌کالا در مورد روغن‌موتور پارس روغن‌موتور شماره ۴	تفاوت قیمت محصول در بازار و دیجی‌کالا	44.48%	فعال نگه‌داشتن کانال فروش اینترنتی از طریق دیجی‌کالا
نظرات کاربران سایت دیجی‌کالا در مورد ضد یخ شماره ۱	کیفیت و قیمت مناسب ضد یخ	34.39%	کنترل کیفیت و قیمت ضدیخ جهت تداوم رضایت مشتریان
نظرات کاربران سایت دیجی‌کالا در مورد ضد یخ شماره ۲	کیفیت خوب، روغنی بودن و قیمت ضد یخ	60.94%	کنترل کیفیت و قیمت ضدیخ جهت تداوم رضایت مشتریان

برای روشن‌تر شدن **جدول ۷**، به‌عنوان مثال ردیف اول جدول که در مورد داده‌های چالش‌ها و نقطه نظرات عاملین در سال ۱۴۰۱ است، به‌طور مفصل توضیح داده می‌شود.

با توجه به کلمات کلیدی که به‌وسیله اجرای الگوریتم تخصیص دیریکله پنهان روی سری داده‌های مذکور مشخص شد (**جدول الف- ۱**) و روابطی که بین این کلمات در شبکه هم رخدادی قابل مشاهده است (**شکل الف- ۱**)، چند عبارت از این سری داده‌ها استخراج شد (**جدول الف- ۹**). همچنین بر اساس تکرار هم رخدادی کلمات کلیدی در سراسر سری داده‌ها، درصد رخداد هر عبارت مشخص شده است. عبارتی که بیشترین درصد رخداد را داشت، عبارت " بسته‌بندی با ظرف چلیک ضعیف است" می‌باشد که می‌توان با بهبود ظرف بسته‌بندی یا تغییر جنس ظرف که در دسته استراتژی‌های توسعه محصول قرار می‌گیرد، این عامل نارضایتی را از بین برد. جدول فوق شامل خلاصه نتایج به‌دست آمده است لذا برای هر سری داده‌ها عبارت‌هایی که بیشترین درصد رخداد را داشته‌اند، در این جدول قرار دارند.

تشکر و قدردانی

محققان از اساتید محترم دانشکده مهندسی صنایع و سیستم‌ها و همچنین پرسنل واحد مهندسی فروش پالایشگاه روان‌کار که در نگارش این مقاله همراهی نموده‌اند، کمال تشکر و قدردانی را اعلام می‌دارند.

منابع مالی

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

تعارض با منافع

هیچ تضادی در منافع در مورد انتشار این نسخه از پژوهش وجود ندارد. همه نویسندگان، نسخه نهایی ارسال‌شده را مطالعه و تایید کرده‌اند. نویسندگان تضمین می‌کنند که این پژوهش، اثر اصلی آن‌ها بوده، قبلا چاپ نشده و در حال حاضر تحت انتشار نمی‌باشد.

منابع

[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. https://doi.org/10.33168/JSMS.2021.0301

[3] Konchady, M. (2006). *Text mining application programming*. Charles River Media. https://dl.acm.org/doi/abs/10.5555/1137800

[4] Tanwar, G., Tripathi, S., Gaur, D., & Shakya, D. (2022). An approach to improve analysis of product reviews using sentiment analysis. *2022 4th international conference on advances in computing, communication control and networking (ICAC3N)* (pp. 1526–1530). IEEE. https://doi.org/10.1109/ICAC3N56670.2022.10074049

[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/s10479-023-05213-9

[6] Sunil, N., & Shirazi, F. (2023). Customer review classification using machine learning and deep learning techniques. *International conference on human-computer interaction* (pp. 581–597). 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., & Bagherzadeh-Mohasefi, J. (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. *IFIP international conference on artificial intelligence applications and innovations* (pp. 659–672). Springer, Cham. https://doi.org/10.1007/978-3-030-79150-6_52

[10] Farzadnia, S., & Raeesi Vanani, I. (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/10.1007/978-3-031-35915-6_42

[12] Ratniasih, N. L., & Jayanti, N. W. N. (2022). Sentiment analysis of stakeholder satisfaction measurement. *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*, *16*, 217–226. https://doi.org/10.22146/ijccs.72245

[13] Belyaeva, Z., Petrosyan, A., & Shams, S. M. 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. *International conference on innovative computing and communications: Proceedings of ICICC 2022* (pp. 397–406). Singapore: Springer Nature Singapore. 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* (Vol. 2457, p. 40006). AIP Publishing. https://doi.org/10.1063/5.0118679

[17] Soni, A., Rasool, A., Dubey, A., & Khare, N. (2022). Visualizing commenters opinions through topic analysis. *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), 4020078. https://doi.org/10.1061/(ASCE)ME.1943-5479.0000845

[19] Han, S., Li, X., & Hwang, H. (2022). Analysis of news data on 'super app'using topic modeling. *International conference on human-computer interaction* (pp. 33–39). Springer, Cham. 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*. 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. https://doi.org/10.1016/j.ipm.2022.103215

[22] Austin, E., Trabelsi, A., Largeron, C., & Zaïane, O. R. (2022). Hierarchical topic model inference by community discovery on word co-occurrence networks. *Australasian conference on data mining* (pp. 148–162). Springer, Singapore. 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. (1984). *Strategic management: A stakeholder approach*. Pitman. https://www.scirp.org/reference/referencespapers?referenceid=1209107

[25] Freeman, R. E., Wicks, A. C., & Parmar, B. (2004). Stakeholder theory and "the corporate objective revisited". *Organization Science*, *15*(3), 364–369. https://doi.org/10.1287/orsc.1040.0066

[26] Freeman, R. E., Harrison, J. S., Wicks, A. C., Parmar, B. L., & De Colle, S. (2010). *Stakeholder theory: The state of the art*. Cambridge University Press. http://www.cambridge.org/9780521190817

[27] Rothaermel, F. T. (2018). *Strategic management*. McGraw-Hill. https://www.amazon.com/dp/1259927628?lv=shuf&channelId=500&plpRedirect=mhFallback

[28] Hanson, D., Hitt, M. A., Ireland, R. D., & Hoskisson, R. E. (2016). *Strategic management: Competitiveness and globalisation*. Cengage AU. https://www.scirp.org/reference/referencespapers?referenceid=3559080

[29] Ansari Mahyari, Z. (2012). *Market development strategies of Iran water & power contracting company in the middle east* [Thesis]. (**In Persian**). https://elmnet.ir/doc/10559406-53113

[30] 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

[31] Bahaderestani, A., Ghalenovi, M., & Motahari- Farimani, N. (2018). Designing a model of key stakeholder expectations in metro construction projects: A case study of Mashhad urban train lines. *Amirkabir Journal of Civil Engineering*, *50*(1) 227-242. (**In Persian**). https://www.magiran.com/paper/1831363

[32] Dellarocas, C., Zhang, X., & Awad, N. F. (2007). Exploring the value of online product reviews in forecasting sales: The case of motion pictures. *Journal of Interactive Marketing*, *21*(4), 23–45. https://doi.org/10.1002/dir.20087

[33] Sotiriadis, M. D., & Van Zyl, C. (2013). Electronic word-of-mouth and online reviews in tourism services: The use of twitter by tourists. *Electronic Commerce Research*, *13*(1), 103–124. https://doi.org/10.1007/s10660-013-9108-1

[34] Siering, M., Deokar, A. V, & Janze, C. (2018). Disentangling consumer recommendations: Explaining and predicting airline recommendations based on online reviews. *Decision Support Systems*, *107*, 52–63. https://doi.org/10.1016/j.dss.2018.01.002

[35] Ansoff, H. I. (1957). Strategies for diversification. *Harvard Business Review*, *35*(5), 113–124. https://www.scirp.org/reference/referencespapers?referenceid=989821

[36] Salton, G ., & Mcgill, M. J. (1983). *Introduction to modern information retrieval*. McGraw-Hill Education, New York. https://www.scirp.org/reference/referencespapers?referenceid=2273012

[37] Papadimitriou, C. H., Tamaki, H., Raghavan, P., & Vempala, S. (1998). Latent semantic indexing: A probabilistic analysis. *Proceedings of the seventeenth ACM sigact-sigmod-sigart symposium on principles of database systems* (pp. 159–168). ACM Press. https://doi.org/10.1145/275487.275505

[38] Hofmann, T. (1999). Probabilistic latent semantic indexing. *Proceedings of the 22nd annual international acm sigir conference on research and development in information retrieval* (pp. 50–57). ACM. https://doi.org/10.1145/312624.312649

[39] 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

[40] Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of Machine Learning Research*, *3*, 993–1022. https://dl.acm.org/doi/10.5555/944919.944937

[41] Gheyoumi, M. (2021). Computational analysis of the evolution of Persian-language and linguistics articles. *Language and Linguistics*, *17*(33), 99–130. (**In Persian**). https://doi.org/10.30465/lsi.2021.8089

[42] 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

[43] Edwards, D. (1995). Graphical modelling. In *recent advances in descriptive multivariate analysis* (pp. 135–156). Clarendon Press. https://doi.org/10.1093/oso/9780198522850.003.0007

[44] He, Q. (1999). Knowledge discovery through co-word analysis. *Library Trends*, *48*(1), 133–159. https://hdl.handle.net/2142/8267

[45] Ding, Y., Chowdhury, G. G., & Foo, S. (2001). Bibliometric cartography of information retrieval research by using co-word analysis. *Information Processing & Management*, *37*(6), 817–842. https://doi.org/10.1016/S0306-4573(00)00051-0

[46] Katbi, F. (2014). *Co-occurrence analysis of knowledge management vocabulary in web of science database from 1993 to 2012 for visualization of vocabulary infrastructure* [Thesis]. (**In Persian**). https://elmnet.ir/doc/10759819-13552

[47] Zipf, G. K. (1949). *Uman behavior and the principle of Least Effort: An Introduction to Human Ecology*. Addison-Wesley Press. https://psycnet.apa.org/record/1950-00412-000

[48] Ahmadi, H. & Osareh, F. (2017). Co-word analysis concept, definition and application. *Librarianship and Information Organization Studies*, *28*(1), 125-145 (**In Persian**). https://nastinfo.nlai.ir/article_1132.html?lang=en

[49] Callon, M., Courtial, J. P., Turner, W. A., & Bauin, S. (1983). From translations to problematic networks: An introduction to co-word analysis. *Social Science Information*, *22*(2), 191–235. https://doi.org/10.1177/053901883022002003

[50] Callon, M., Law, J., & Rip, A. (1986). *Mapping the dynamics of science and technology*. Macmillan. https://www.amazon.com/dp/0333372239?lv=shuf&channelId=500&plpRedirect=mhFallback

[51] Ahmadi, H. & Osareh, F. (2017). Co-word analysis concept, definition and application. *Librarianship and Information Organization Studies*, *28*(1), 125-145 (**In Persian**). https://nastinfo.nlai.ir/article_1132.html?lang=en

[52] Courtial, J. P., & Law, J. (1989). A co-word study of artificial intelligence. *Social Studies of Science*, *19*(2), 301–311. https://doi.org/10.1177/030631289019002005

[53] Whittaker, J. (1989). Creativity and conformity in science: Titles, keywords and co-word analysis. *Social Studies of Science*, *19*(3), 473–496. https://doi.org/10.1177/030631289019003004

[54] Callon, M., Courtial, J. P., & Laville, F. (1991). Co-word analysis as a tool for describing the network of interactions between basic and technological research: The case of polymer chemsitry. *Scientometrics*, *22*(1), 155–205. https://doi.org/10.1007/BF02019280

[55] Turner, W., & Rojouan, F. (1991). Evaluating input/output relationships in a regional research network using co-word analysis. *Scientometrics*, *22*(1), 139–154. https://doi.org/10.1007/bf02019279

[56] Salemi, N., & Koosha, K. (2014). Co-citation analysis and co-word analysis in bibliometrics mapping: A methodological evaluation. *Iranian Journal of Information Processing and Management*, *29*(1), 253–266. (**In Persian**). https://doi.org/10.35050/JIPM010.2013.011

[57] Simonetti, A., Albano, A., Plaia, A., & Tumminello, M. (2025). Ranking coherence in topic models using statistically validated networks. *Journal of Information Science*, *51*(3), 744–765. https://doi.org/10.1177/01655515221148369

[58] Murakami, R., & Chakraborty, B. (2021). Neural topic models for short text using pretrained word embeddings and its application to real data. *2021 IEEE 4th international conference on knowledge innovation and invention (ICKII)* (pp. 146–150). IEEE. https://doi.org/10.1109/ICKII51822.2021.9574752

پیوست الف

جدول کلمات کلیدی استخراج‌شده و تعداد تکرار آن‌ها برای سری‌های مختلف داده‌ها به شرح زیر است.

جدول الف- ۱-کلمات کلیدی استخراج‌شده و تعداد تکرار آن‌ها برای داده‌های چالش‌ها و نقطه‌نظرات 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
بسته‌بندی	17	صادر	6	بشکه	2
حواله	16	افزایش	6	قوی	2
تبلیغ	16	ضدیخ	6	وجه	2
کارتن	16	سهمیه	6	کشاورز	2
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
رقیب	12	مصرف‌کننده	5	موجود	2
کیفیت	11	حذف	5	مظروف	2
اجبار	11	تقاضا	5	راننده	2
عرضه	11	طول	4	صلاح‌دید	2
مصرف	10	قیمت	4	عاملین	1
خرید	10	لباس تبلیغاتی	4	تجارت	1
رنگ‌بندی	10	تعداد	4	معتبر	1
بازار	10	محصولات	4	سررسید	1
تولید	9	انبار	4	راضی‌کننده	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	راه‌اندازی	1
پرداخت	6	بنزین	3	نظرسنجی	1
شیوه	6	خارج	3	تعداد کلمات کلیدی	83

جدول الف- ۲- کلمات کلیدی استخراج‌شده و تعداد تکرار آن‌ها برای داده‌های نظرات کاربران سایت پالایشگاه در مورد روغن‌های 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	دوگانه‌سوز	11	پژو پارس	7
روغن‌موتور شماره ۴	38	پراید	10	گرید	7
10W40	22	گیربکس	10	تابستان	6
20W50	22	تعویض	10	سمند XU7	5
روغن‌موتور شماره ۵	19	تیبا	9	زمستان	5
راهنمایی	16	ال نود	9	منطقه	5
کیفیت	12	بنزین	8	یورو ۴	4
روغن‌موتور شماره ۲	12	۲۰۶ تیپ ۲	7	پژو پارس XU7	4
تفاوت	12	تک‌سوز	7	تعمیر	3
فلومکس	11	بازار	7	تعداد کلمات کلیدی	29

جدول الف-۳- کلمات کلیدی استخراج‌شده و تعداد تکرار آن‌ها برای داده‌های نظرات کاربران سایت پالایشگاه در مورد روغن‌های 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
روغن‌موتور 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
روغن‌موتور شماره ۲	11	تیوفایو	7	توربو شارژ	3
اصالت	10	قوطی	6	تعداد کلمات کلیدی	20

جدول الف- ۴- کلمات کلیدی استخراج‌شده و تعداد تکرار آن‌ها برای داده‌های نظرات کاربران سایت پالایشگاه در مورد روغن‌های 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
روغن‌موتور شماره ۲	22	گیربکس	9	یورو	4
پراید	19	قوطی	8	بازار	4
پیشنهاد	18	کیفیت	8	گرید	4
روغن‌موتور شماره ۴	14	بهران	8	نیمه سنتتیک	3
دوگانه‌سوز	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
فلومکس	11	صدا	5	بادامک	1
تیبا	10	تهیه	5	تعداد کلمات کلیدی	35

جدول الف- ۵-کلمات کلیدی استخراج‌شده و تعداد تکرار آن‌ها برای داده‌های نظرات کاربران دیجی‌کالا در مورد روغن‌موتور 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
دیجی‌کالا	84	سالم	17	پژو	8
بازار	63	نشتی	17	کارکرد	7
موتور	62	باکیفیت	17	استاندارد	6
توصیه	61	اصل	16	حمل	6
به‌صرفه	60	برند	19	تحویل	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
قوطی	27	مشکل	13	منصفانه	4
پیکان	27	دقت	12	مصوب	4
راضی	25	کاربراتور	12	رنگ	4
ریخت	24	سفارش	12	کارامد	4
موتورسیکلت	21	متوسط	12	دولتی	4
برند	19	روا	11	تعداد کلمات کلیدی	55
قدیمی	19	هوندا	10		

جدول الف-۶- کلمات کلیدی استخراج‌شده و تعداد تکرار آن‌ها برای داده‌های نظرات کاربران دیجی‌کالا در مورد روغن‌موتور 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	ایکس یوسون (XU7)	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	قابل‌قبول	5
روغن‌موتور	29	درجه	9	سیاه	5
اصل	29	پایین	8	تحویل	5
ایران	26	حد	8	کارخانه	4
بازار	26	سال	8	تخفیف	3
به‌صرفه	21	تقلبی	8	اطمینان	3
مدل	19	دویست و شش	8	اقتصادی	2
محصول	18	ایرانول	8	راضی‌کننده	2
پژو	18	فلزی	7	شدت	2
صدا	18	ارزان	7	بی‌انصاف	1
روغن‌موتور شماره	17	فوق‌العاده	7	استعلام	1
برند	15	فلومکس	7	تعداد کلمات کلیدی	71

جدول الف- ۷-کلمات کلیدی استخراج‌شده و تعداد تکرار آن‌ها برای داده‌های نظرات کاربران دیجی‌کالا در مورد ضد یخ شماره ۱.

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

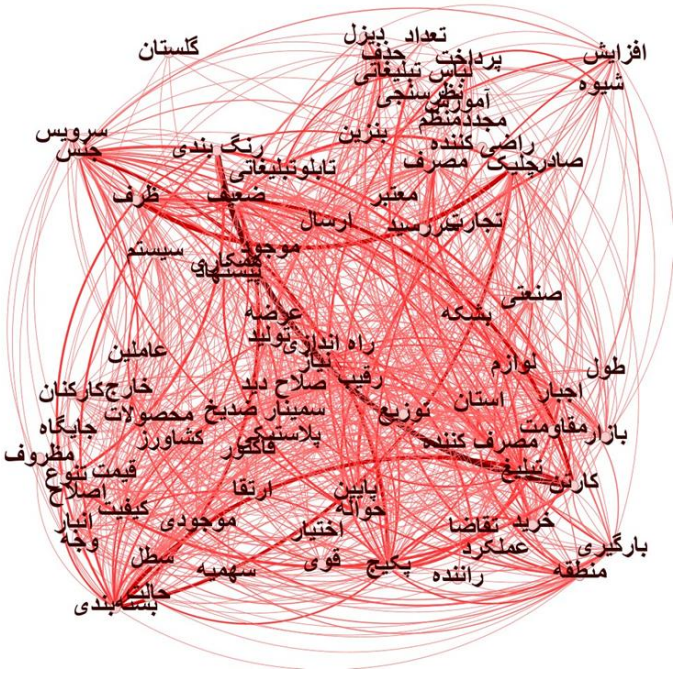
کلمه	تعداد تکرار	کلمه	تعداد تکرار	کلمه	تعداد تکرار
ضدیخ	62	خودرو	10	شک	3
قیمت	54	غلظت	9	باکیفیت	3
کیفیت	47	تقلبی	9	به‌صرفه	3
قوطی	25	مکانیک	9	تعریف	3
توصیه	24	اطمینان	8	درب	3
محصول	21	مطمئن	8	رقیق	3
بازار	18	کاسپین	8	راضی‌کننده	2
برند	17	تعویض	8	سیلندر	2
فلزی	16	جوش	6	کارامد	2
مناسب	16	پالایشگاه	6	ترجیح	1
رادیاتور	14	زمستان	5	تسمه	1
تاریخ	14	جواب	5	دانگیل	1
دیجی‌کالا	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 استخراج‌شده و تعداد تکرار آن‌ها برای داده‌های نظرات کاربران دیجی‌کالا 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
دیجی‌کالا	6	تقلبی	3	روشن	1
سفارش	5	پالایشگاه	3	امپر	1
برند	5	اسیدی	3	تعداد کلمات کلیدی	29

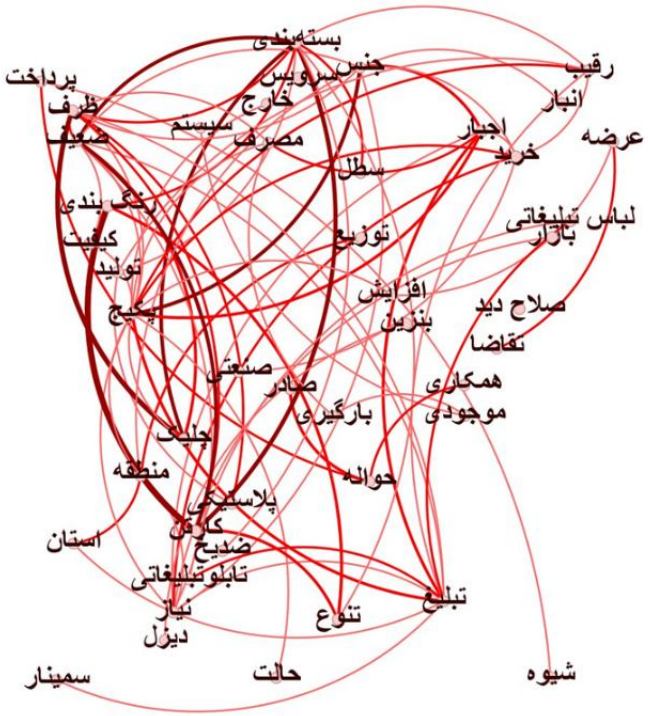
شبکه هم رخدادی کلمات کلیدی برای سری‌های مختلف داده‌ها به‌صورت زیر است.



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

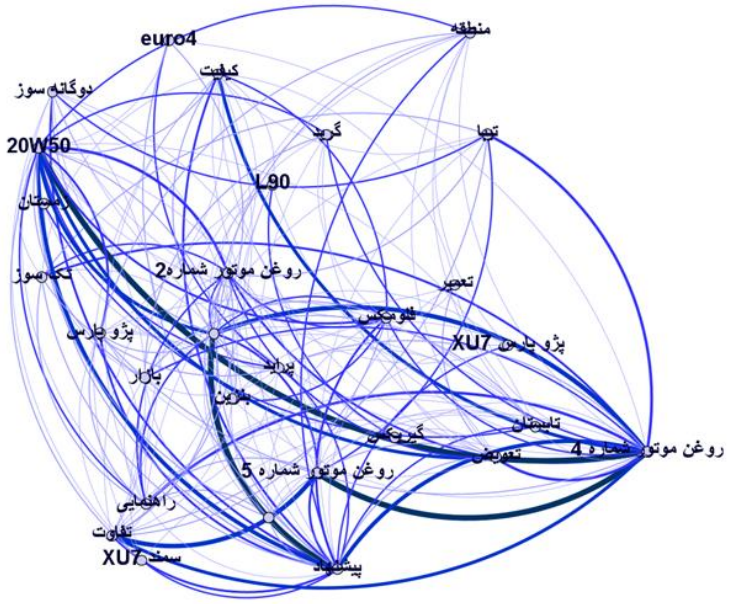
Figure A-1- Word co-occurrence network for data on stakeholders' challenges and viewpoints in the year 1401.

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



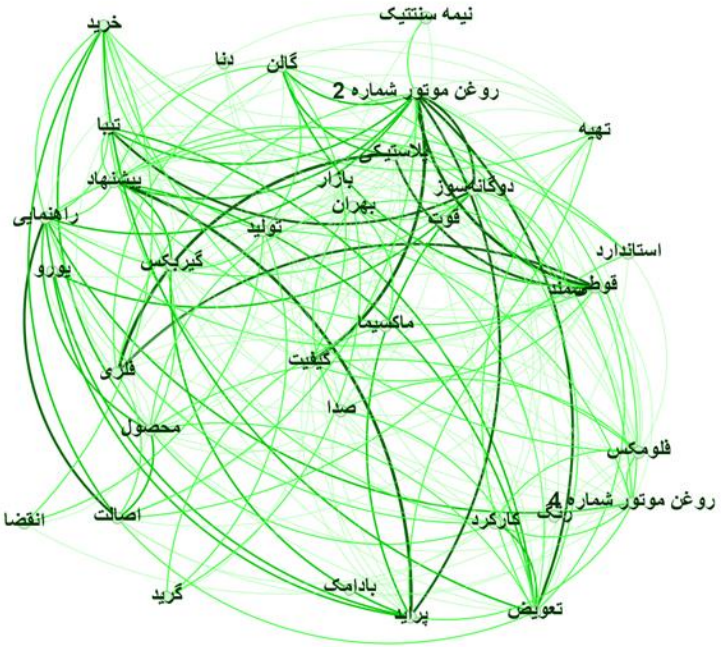
شکل الف- ۲- شکل کاهش‌یافته شبکه هم‌رخدادی واژگان برای داده‌های چالش‌ها و نقطه‌نظرات عاملین در سال ۱۴۰۱.

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



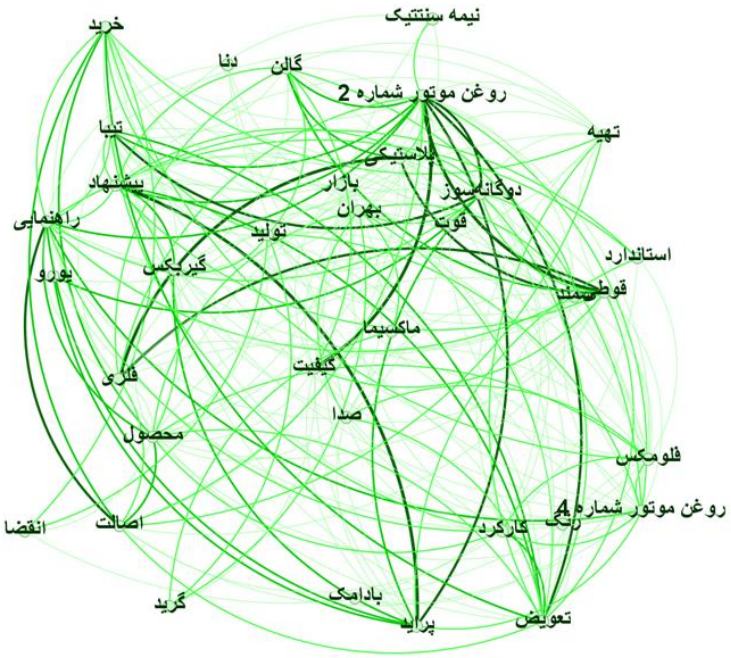
شکل الف- ۳- شبکه هم‌رخدادی واژگان برای داده‌های نظرات کاربران سایت پالایشگاه در مورد روغن‌های خودرویی در بازه زمانی ۱۳۹۸ تا ۱۳۹۹.

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



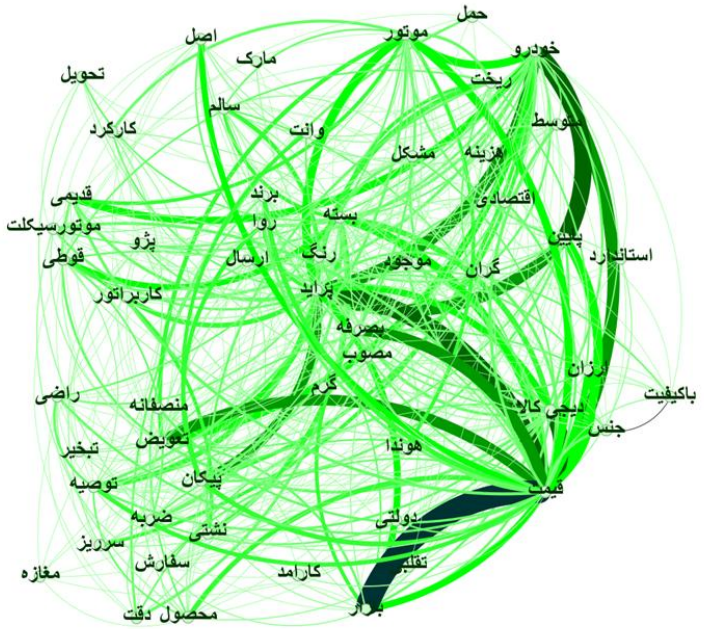
شکل الف-۴- شبکه هم‌رخدادی واژگان برای داده‌های نظرات کاربران سایت پالایشگاه در مورد روغن‌های خودرویی در بازه زمانی ۱۳۹۹ تا ۱۴۰۰.

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



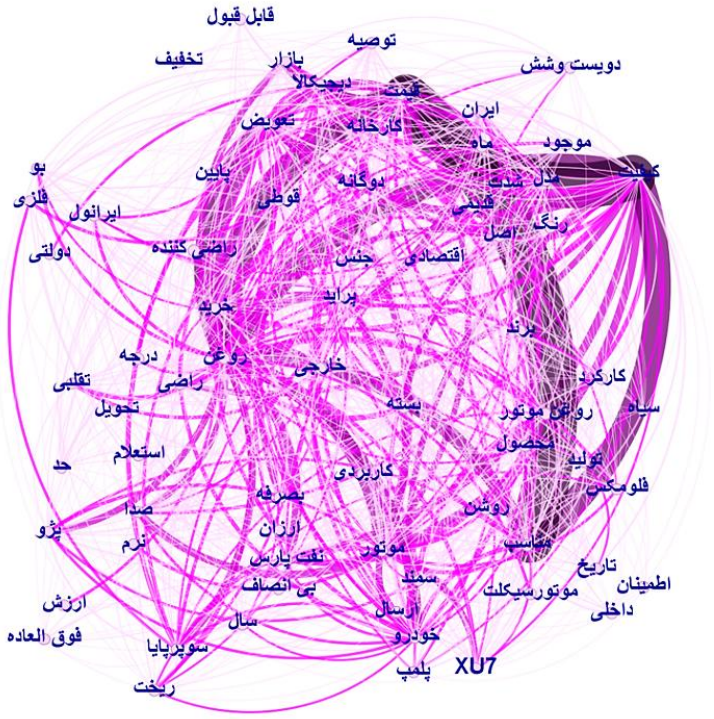
شکل الف- ۵- شبکه هم‌رخدادی واژگان برای داده‌های نظرات کاربران سایت پالایشگاه در مورد روغن‌های خودرویی در بازه زمانی ۱۴۰۰ تا ۱۴۰۲.

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



شکل الف- ۶- شبکه هم‌رخدادی واژگان برای داده‌های نظرات کاربران دیجی‌کالا در مورد روغن‌موتور شماره ۱.

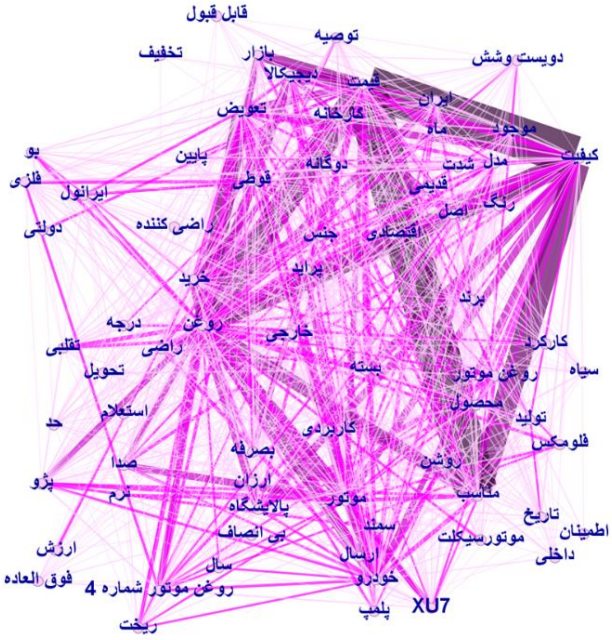
Figure A-6- Word co-occurrence network for Digikala user reviews regarding engine oil No. 1.



شکل الف- ۷- شبکه هم‌رخدادی واژگان برای داده‌های نظرات کاربران دیجی‌کالا در مورد روغن‌موتور شماره ۴.

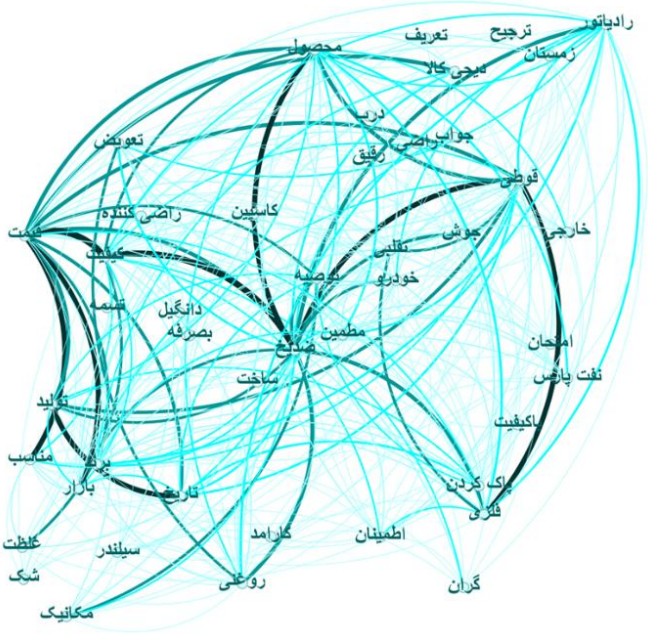
Figure A-7- Word co-occurrence network for Digikala user reviews regarding engine oil No. 4.

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.



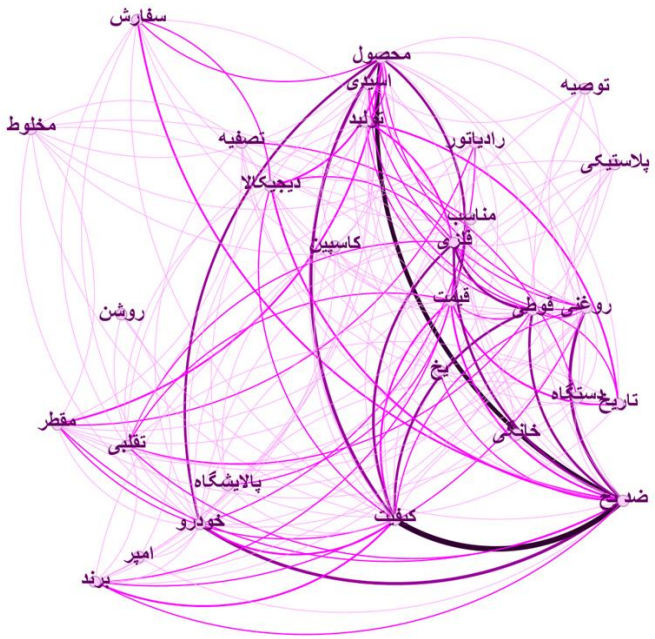
شکل الف- ۸- شکل شبکه هم‌رخدادی واژگان با یال‌های غیر منحنی برای داده‌های نظرات کاربران دیجی‌کالا در مورد روغن‌موتور پارس روغن‌موتور شماره ۴.

Figure A-8- Word co-occurrence network with straight edges for Digikala user reviews regarding Pars Engine Oil No. 4.



شکل الف- ۹- شبکه هم‌رخدادی واژگان برای داده‌های نظرات کاربران دیجی‌کالا در مورد ضد یخ شماره ۱.

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 دیجی‌کالا در مورد ضد یخ شماره ۲.

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

عبارت‌های استخراج‌شده از شبکه هم رخدادی واژگان برای سری‌های مختلف داده‌ها به‌صورت زیر می‌باشد.

جدول الف- ۹- عبارات استخراج‌شده از شبکه هم‌رخدادی واژگان داده‌های چالش‌ها و نقطه‌نظرات 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 بسته‌بندی، ضعیف، ظرف و چلیک	بسته‌بندی با ظرف چلیک ضعیف است	31.01%
کلمات تنوع، ضعیف، کارتن و رنگ‌بندی	ضعیف بودن تنوع رنگ‌بندی کارتن	15.5%
کلمات اجبار، پکیج، جنس، خرید و رقیب	فروش محصولات به‌صورت پکیج اجباری	13.95%
کلمات تبلیغ، لباس تبلیغاتی و تابلو تبلیغاتی	عاملین خواستار تبلیغ به‌وسیله لباس و تابلو تبلیغاتی هستند.	3.88%

جدول الف- ۱۰- عبارات استخراج‌شده از شبکه هم‌رخدادی واژگان داده‌های نظرات کاربران سایت شرکت در مورد روغن‌های 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، روغن‌موتور شماره ۴، روغن‌موتور شماره ۵، فلومکس، کیفیت و شماره ۴	تقاضای ارایه محصول پیشنهادی مناسب خصوصا ازنظر گرانروی برای انواع خودروها	32.26%
کلمات تفاوت، روغن‌موتور شماره ۴ و روغن‌موتور شماره ۵	تفاوت بین روغن‌موتور شماره ۴ و شماره ۵ برای خریداران مبهم است	18.06%

جدول الف- ۱۱- عبارات استخراج‌شده از شبکه هم‌رخدادی واژگان داده‌های نظرات کاربران سایت شرکت در مورد روغن‌های 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، روغن‌موتور شماره ۴، کیفیت و روغن‌موتور شماره ۲	نگرانی در مورد اصالت کالا و وجود کالای تقلبی در بازار	35.57%
کلمات euro 4، پراید، روغن‌موتور شماره ۲، روغن‌موتور 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 استخراج‌شده 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 روغن‌موتور شماره ۲، قوطی، تعویض، فلزی، پلاستیکی، گالن	پلاستیکی یا فلزی بودن مظروف محصول روغن‌موتور شماره ۲	36.67%
کلمات روغن‌موتور شماره ۲، کیفیت، تیبا، دوگانه‌سوز، پیشنهاد و پراید	تقاضای راهنمایی برای خرید محصول مناسب با توجه به استاندارد و نوع خودرو	21.67%
کلمات راهنمایی، اصالت و محصول	تقاضای راهنمایی در مورد نحوه حصول اطمینان از اصالت کالا	8.33%

جدول الف- ۱۳- عبارات استخراج‌شده از شبکه هم‌رخدادی واژگان داده‌های نظرات کاربران دیجی‌کالا در مورد روغن‌موتور 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، به‌صرفه، 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، دیجی‌کالا، بازار، منصفانه، ارزان، به‌صرفه، اقتصادی	تفاوت قیمت محصول در دیجی‌کالا و بازار و قیمت منصفانه دیجی‌کالا	15.26%

جدول الف- ۱۴- عبارات استخراج‌شده از شبکه هم‌رخدادی واژگان داده‌های نظرات کاربران دیجی‌کالا در مورد روغن‌موتور 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، دیجی‌کالا، دولتی، به‌صرفه	تفاوت قیمت محصول در بازار و دیجی‌کالا و قیمت به‌صرفه محصول در دیجی‌کالا	44.48%
کلمات قیمت، کیفیت، مناسب و روغن	قیمت و کیفیت مناسب روغن	7.84%

جدول الف- ۱۵- عبارات استخراج‌شده از شبکه هم‌رخدادی واژگان داده‌های نظرات کاربران دیجی‌کالا 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، ضدیخ، قیمت، مناسب، روغنی و محصول	کیفیت و قیمت مناسب ضدیخ	34.39%
کلمات قوطی، تعویض، فلزی، ضدیخ	رضایت از فلزی بودن قوطی ضدیخ	14.03%
کلمات قیمت، بازار، مناسب، دیجی‌کالا	تفاوت قیمت بازار و قیمت دیجی‌کالای ضدیخ	11.76%

جدول الف- ۱۶- عبارات استخراج‌شده از شبکه هم‌رخدادی واژگان داده‌های نظرات کاربران دیجی‌کالا 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، ضدیخ، قیمت، روغنی و محصول	کیفیت خوب، روغنی بودن و قیمت ضدیخ	60.94%
کلمات قوطی، فلزی، ضدیخ، محصول	رضایت از فلزی بودن قوطی ضدیخ	20.31%