

Marketing Research in the Fourth Industrial Revolution, Using 'Big Data' Analytics and 'Machine Learning' to Provide Value to the Customer

Sayed Mohsen Mousavi¹, Seyed Fathollah Amiri Aghdaie^{2*}

1- PhD Candidate of Business Management, Department of Management, Faculty of Administrative Sciences and Economics, University of Isfahan, Isfahan, Iran

mohsen_mousavi@pds.ui.ac.ir

2- Associate Professor, Department of Management, Faculty of Administrative Sciences and Economics, University of Isfahan, Isfahan, Iran

s.aghdaie@ase.ui.ac.ir

Abstract

One of the most important consequences of the fourth industrial revolution is the digitalization of the business environment and the expansion of data-driven marketing. The consequences of these developments will have a profound effect on marketing research. The purpose of this study was to identify the components of value among a large volume of customers' feedback in social networks using 'Machine Learning' to analyze 'Big Data'. This research was applied in terms of purpose, descriptive-survey in terms of data collection method, and qualitative-quantitative (mixed methods) in terms of implementation. For this purpose, the tourism industry was selected as a case study. And 8290 customers' opinions were collected from the Internet in this regard and the components of value were extracted using two different methods of data clustering and extraction of association rules. The results of the research in the data clustering section included the identification of 20 value components in relation to the values desired by tourists. Also, using the second method (association rules), seven laws were extracted from the hidden knowledge in the relationships between the phrases used in the opinions of tourists. In general, the results showed that using 'Big Data' analytics and 'Machine Learning', the process of marketing research is possible with higher speed and accuracy and lower relative cost.

Keywords: Big Data, Marketing Research, the Fourth Industrial Revolution, Machine Learning, Value Creation.

تحقیقات بازاریابی در انقلاب صنعتی چهارم، استفاده از تحلیل کلان داده‌ها و «یادگیری ماشین» برای ارائه ارزش به مشتری

سید محسن موسوی^۱، سید فتح‌الله امیری عقدایی^{۲*}

۱- دانشجوی دکتری مدیریت بازرگانی، گروه مدیریت، دانشکده علوم اداری و اقتصاد، دانشگاه اصفهان، اصفهان، ایران

۲- دانشیار گروه مدیریت، دانشکده علوم اداری و اقتصاد، دانشگاه اصفهان، اصفهان، ایران

s.aghdaie@ase.ui.ac.ir

چکیده

از مهم‌ترین پیامدهای چهارمین انقلاب صنعتی، دیجیتالیزه شدن محیط کسب و کار و گسترش بازاریابی داده‌محور است که پیامدهای این تحولات، تأثیرات عمیقی بر تحقیقات بازاریابی خواهد داشت. هدف این پژوهش، شناسایی ویژگی یا مؤلفه‌های ارزش از میان حجم انبوه نظرات مشتریان در شبکه‌های مجازی با استفاده از یادگیری ماشین برای تحلیل کلان داده‌ها، با رویکرد بازاریابی پیش‌بینانه است که متناسب با فضای موجود انجام گرفته است. این پژوهش از لحاظ هدف کاربردی، از نظر روش گردآوری داده‌ها توصیفی-پیمایشی و از نظر اجرا کیفی-کمی (آمیخته) است. بدین منظور، صنعت گردشگری به عنوان مطالعه موردی انتخاب شد و تعداد ۸۲۹۰ نظر از مشتریان در این باره از بستر اینترنت جمع‌آوری شد و با استفاده از دو روش «خوشه‌بندی داده‌ها» و «استخراج قوانین انجمنی»، مؤلفه‌های ارزش استخراج شدند. نتایج پژوهش در قسمت خوشه‌بندی شامل شناسایی ۲۰ مؤلفه ارزش در رابطه با ارزش‌های مدنظر گردشگران است. همچنین، با استفاده از روش دوم (قوانین انجمنی)، هفت قانون از میان دانش پنهان در روابط بین عبارات به کار برده شده در نظرات گردشگران استخراج شد. به طور کلی، نتایج نشان می‌دهد با استفاده از تحلیل کلان داده‌ها و یادگیری ماشین، انجام فرایند «تحقیقات بازاریابی» با سرعت و دقت بیشتر و هزینه نسبی کمتر امکان پذیر است.

کلیدواژه‌ها: کلان داده، تحقیقات بازاریابی، انقلاب صنعتی چهارم، یادگیری ماشین، خلق ارزش.

۱. مقدمه

در حال حاضر، فضای کسب و کار شاهد تحولات عمیق و اساسی است و بسیاری از صاحب نظران در حوزه تجارت و اقتصاد بر این باورند که این شواهد حاکی از شروع دوره جدیدی یعنی انقلاب صنعتی چهارم است (ماآر^۱، ۲۰۱۶؛ گوپتا و همکاران^۲، ۲۰۱۷). امروزه در محیط رقابتی کسب و کار به وجود آمده، شرکت‌ها در مواجهه با مسئله داده‌های بزرگ یا کلان داده‌ها با چالش‌هایی مانند تصمیم‌گیری سریع برای بهبود بهره‌وری روبه‌رو هستند؛ زیرا بسیاری از سیستم‌های تولید آمادگی لازم برای مدیریت کلان داده‌ها را به دلیل نبود ابزارهای تحلیلی هوشمند ندارند (لی و همکاران^۳، ۲۰۱۴). چهارمین انقلاب صنعتی (اصطلاحی که کلاوس شواب^۴، بنیان‌گذار و رئیس اجرایی اجلاس جهانی اقتصاد معرفی کرده است) دنیایی را توصیف می‌کند که به افراد این توانایی را می‌دهد که برای زندگی کردن و مدیریت آن با استفاده از فناوری ارتباطات بین حوزه‌های دیجیتال و واقعیت آفلاین^۵ حرکت کنند (شو و همکاران^۶، ۲۰۱۸).

نخستین انقلاب صنعتی زندگی و اقتصاد را به طور کلی دگرگون کرد و اقتصاد کشاورزی را به اقتصادی تغییر داد که در آن صنعت و ماشین‌آلات تحت سلطه انسان‌اند. نفت و برق، تولید انبوه را در دومین انقلاب صنعتی تسهیل کرد و در سومین انقلاب صنعتی از فناوری اطلاعات برای خودکارسازی تولید استفاده شد و در نهایت در چهارمین انقلاب صنعتی، به افزایش قدرت شناختی تولیدات انسانی توجه شد و این قدرت تقویت شد (شو و همکاران، ۲۰۱۸). باتوجه به تعاریف

مختلف و مباحث علمی که برای توصیف سه انقلاب صنعتی نخست استفاده شده‌اند، می‌توان گفت از ابتدای قرن حاضر انقلاب چهارم شروع شده است که آن را با عنوان «انقلاب دیجیتال» نیز معرفی می‌کنند. ویژگی‌های این دوره فراگیر بودن، اینترنت موبایل، حسگرهای کوچک‌تر، قوی‌تر و درعین حال ارزان‌تر، هوش مصنوعی و «یادگیری ماشین» است (صنایعی، ۱۳۹۶). نکته مهمی که در این باره باید در نظر داشت، این است که سرعت و اندازه تغییرات ناشی از انقلاب صنعتی چهارم را نباید نادیده گرفت؛ زیرا این تغییرات باعث تغییر و انتقال قدرت، ثروت و دانش خواهد شد. فقط با آگاهی از این تغییرات و سرعتی که این اتفاق می‌افتد می‌توان اطمینان حاصل کرد که مزایای پیشرفت‌های دانش و فناوری به دست آمده است (شو و همکاران، ۲۰۱۸).

همچنین، گسترش بی‌سابقه و فزاینده اینترنت باعث رشد سریع تولید محتوا توسط کاربران^۷، در شبکه‌های اجتماعی گوناگون شده است و این حجم عظیم داده به یکی از مهم‌ترین منابع اطلاعاتی برای هردو طرف مصرف‌کنندگان و کسب و کارها تبدیل شده است (دوان و همکاران^۸، ۲۰۱۳). همچنین رسانه‌های اجتماعی^۹ در بستر اینترنت راه و روش ارتباط، همکاری و تعامل جامعه را تغییر داده‌اند. درحقیقت، رسانه‌های اجتماعی به مجموعه‌ای از ابزارهای آنلاین گفته می‌شود که ارتباطات میان افراد را ایجاد کرده‌اند و باعث تبدیل شدن ارتباط یک‌طرفه یا مونولوگ^{۱۱} به تعامل دوجانبه یا دیالوگ^{۱۲} شده‌اند (حلوانی و همکاران^{۱۳}، ۲۰۱۹).

7. Sensor

8. User Generated Content (UGC)

9. Duan et al.

10. Social Media

11. Monologue

12. Dialogue

13. Halawani et al.

1. Maar

2. Gupta et al.

3. Lee et al.

4. Klaus Schwab

5. Offline

6. Xu et al.

از کلان داده‌ها در صنعت ۴,۰ انجام نگرفته است. موارد گفته شده اهمیت و ضرورت مطالعه را آشکار می‌کند و این پژوهش می‌تواند در راستای این چالش‌ها مسیر را برای انجام پژوهش‌های آتی روشن‌تر کند. به بیان کلی، این پژوهش، بر استفاده از کلان داده‌ها با کمک یادگیری ماشین و هوش مصنوعی (AI) و تأثیر آنها بر تحقیقات بازار و بازاریابی تمرکز دارد.

بنابراین، پرسش‌های اصلی این پژوهش عبارت‌اند از:
 ۱) ارزش‌های موردانتظار از دید گردشگران چه مواردی هستند (چه عواملی باعث رضایت مشتریان است)؟

۲) هر گروه ارزش، شامل چه مؤلفه‌هایی هستند؟

۲. مبانی نظری پژوهش و پیشینه پژوهش

۲-۱- پیشینه نظری

۲-۱-۱- ارزش

هر کسب‌وکار با توجه به مدل کسب‌وکار خود به خلق و ارائه «ارزش» به مشتری می‌پردازد. در حقیقت، کسب‌وکار برای بقا و رشد باید بتواند مشتری را به پرداخت (پول) در برابر دریافت آن «ارزش» متقاعد کند. در سال‌های اخیر مطالعه مفهوم «ارزش» به یکی از زمینه‌های موردعلاقه مدیران و پژوهشگران حوزه مدیریت تبدیل شده است. از سوی دیگر، ظهور ابزارهای جدید ارتباط با مشتریان از جمله شبکه‌های اجتماعی، چالش‌ها و درعین حال فرصت‌های جدیدی را در زمینه چگونگی معرفی و ارائه ارزش به ذی‌نفعان سازمان به وجود آورده است (لابرک^۲، ۲۰۱۴). مفهوم ارزش در زمینه‌های مختلف با تعاریف متفاوتی استفاده شده است. از دیدگاه مشتری، ارزش زمانی ایجاد می‌شود که مزایای به‌دست آمده از مصرف یک

با در نظر گرفتن فضای جدید به وجود آمده در انقلاب صنعتی چهارم و اهمیت داده‌ها به عنوان شالوده این انقلاب و تأثیر آن بر محیط کسب‌وکار، هدف این پژوهش شناسایی و بررسی مؤلفه‌های ارزش از نظر مشتریان بر اساس تحلیل کلان داده‌های مربوط به نظرات آنهاست؛ بدین منظور، صنعت گردشگری به عنوان نمونه‌ای است که مطالعه می‌شود. نظرات آنلاین، توصیف کننده تجربیات مسافران از اقامت در هتل‌هاست و ارزیابی گردشگران از هتل‌ها را منعکس می‌کند. این نظرات بر اساس سبک زندگی، نحوه تفکر و میزان لذت مسافران در مقصد هستند و نشان‌دهنده میزان رضایت مشتری از تجربه ماندن در مقصد یا هتل مربوطه‌اند (سانچز و همکاران^۱، ۲۰۱۹).

برای انجام این پژوهش، در گام اول عملیات پیش‌پردازش داده‌ها برای آماده‌سازی آنها انجام می‌شود. سپس در مرحله دوم به منظور استخراج مؤلفه‌های ارزش، نظرات با روش «خوشه‌بندی داده‌ها» تجزیه و تحلیل می‌شوند. در نهایت در مرحله سوم با استفاده از روش استخراج «قوانین انجمنی» به کشف دانش پنهان از مجموعه عبارات و لغات به کاررفته در نظرات گردشگران پرداخته می‌شود.

بررسی مطالعات داخلی و خارجی انجام شده نشان می‌دهد به دلیل جدید بودن موضوع در زمینه فرصت‌ها و تهدیدات به وجود آمده در حوزه تحقیقات بازاریابی که تأثیر گرفته از پیامدهای «انقلاب صنعتی چهارم» است، پژوهش‌های انجام شده در ابتدای راه هستند و زمینه گسترده‌ای برای پژوهش در این حوزه وجود دارد. از این مسئله به خصوص در داخل کشور غفلت بیشتری شده است و پژوهش مستقلی به‌ویژه با رویکرد استفاده

². Labrecque

¹. Sánchez et al.

(کالاهای فیزیکی و خدمات) «ارزش» دریافت شده مشتری را افزایش می‌دهد؛ سوم، نوآوری مشارکتی به بررسی تجربیات مشتری، خدمات مبتنی بر داده و امکانات مناسب نیاز دارد و چهارم، با فراهم شدن بسترهای جهانی برای مدل‌های تجاری جدید استعداد، فرهنگ و شکل‌های جدید ساختار سازمان باید دوباره بررسی شود و در آن تجدید نظر شود (شواب^۹، ۲۰۱۵).

۲-۱-۳. کلان داده‌ها

امروزه سازمان‌ها و کسب و کارها مقادیر عظیمی از داده‌ها را جمع‌آوری و ذخیره می‌کنند؛ به امید آنکه در آینده مفید واقع شوند؛ در این حالت، چالش‌هایی از قبیل بارگذاری^{۱۰} داده‌ها و استخراج دانش مناسب به منظور پشتیبانی از فرایند تصمیم‌گیری مدیران به وجود آمده است (آمادو و همکاران^{۱۱}، ۲۰۱۸). اصطلاح «کلان داده‌ها» به مجموعه داده‌هایی اطلاق می‌شود که از نظر سرعت، حجم و تنوع در سطح بالایی هستند و با تکنیک‌ها و ابزارهای سنتی پردازش نمی‌شوند (الجنیدی و الراگال^{۱۲}، ۲۰۱۴). در حال حاضر، کلان داده‌ها همه جا هستند؛ چه به شکل داده‌های ساخت یافته^{۱۳}، مانند پایگاه داده‌های سنتی سازمان (مثلاً سیستم مدیریت ارتباط با مشتری) یا داده‌های بدون ساختار^{۱۴} که فناوری‌های جدید ارتباطی و بستری^{۱۵} هستند که کاربر می‌تواند آنها را توسعه دهد یا ویرایش کند (مثلاً متون، تصاویر و فیلم‌ها) (لنسلی و لونگلی^{۱۶}، ۲۰۱۶). دو منبع مهم داده‌های امروزی رسانه‌های اجتماعی و برنامه‌های کاربردی موبایل هستند که

محصول یا خدمات از هزینه‌هایی که بابت آن متحمل شده است، بیشتر شوند (ژانگ و همکاران^۱، ۲۰۱۹). در این بازار پویا مشتری از سازمان یا کسب و کار انتظار دارد بیشترین ارزش‌ها را با مناسب‌ترین قیمت عرضه کند و سازمان‌ها نیز پیوسته به دنبال روش‌های جدید در خلق و ارائه «ارزش» هستند (لا و کاندامپولی^۲، ۲۰۰۴). همچنین کسب و کارها برای دستیابی به مزیت رقابتی باید رویکرد و تمرکز خود را از فروش کالا یا خدمات، به خلق ارزش برای مشتری تغییر دهند (کانز و اینگوالد^۳، ۲۰۱۶). با بررسی پژوهش‌های انجام شده می‌توان گفت مفهوم ارزش پیشنهادی به مشتری^۴ (CVP) نقش مهمی در ارتباط با نحوه فراهم ساختن ارزش برای مشتری دارد؛ ولی این مفهوم همچنان به طور ضعیفی فهمیده می‌شود (پین و همکاران^۵، ۲۰۱۷).

۲-۱-۲. انقلاب صنعتی چهارم

چهارمین انقلاب صنعتی با استفاده از دیجیتالیزه شدن، فناوری اطلاعات و ارتباطات، یادگیری ماشین، رباتیک و هوش مصنوعی باعث افزایش قدرت تصمیم‌گیری و انتقال بیشتر این فرایند از انسان به ماشین‌ها خواهد شد و این تحولات تأثیر عمیقی بر تحقیقات بازاریابی و مدیریت فروش می‌گذارد (سیام و شارما^۶، ۲۰۱۸). مفهوم کسب و کار از چهار جنبه تحت تأثیر صنعت ۴،۰^۸ قرار می‌گیرد: نخست، مشتری انتظار دریافت خدمات متفاوتی دارد؛ دوم، اضافه کردن قابلیت‌های دیجیتال به محصولات

⁹. Schwab

¹⁰. Upload

¹¹. Amado et al.

¹². Elgendy & Elragal

¹³. Structured Data

¹⁴. Unstructured Data

¹⁵. Platform

¹⁶. Lansley & Longley

¹. Zhang et al.

². La & Kandampully

³. Kans & Ingwald

⁴. Customer Value Proposition (CVP)

⁵. Payne et al.

⁶. Digitization

⁷. Syam & Sharma

⁸. Industry 4.0

(سیام و شارما^۶، ۲۰۱۸). به بیان دیگر، یادگیری ماشین^۷ به تنظیم و اکتشاف شیوه‌ها و الگوریتم‌هایی می‌پردازد که براساس آنها رایانه توانایی یادگیری پیدا می‌کند.

یادگیری ماشین به‌طور کلی به دو دسته اصلی یادگیری با نظارت^۸ و «یادگیری بدون نظارت»^۹ تقسیم می‌شود. مفهوم یادگیری نظارت‌شده یا «یادگیری آماری نظارت‌شده» به معنای ایجاد مدل آماری برای پیش‌بینی یا برآورد یک خروجی براساس یک یا چند ورودی است؛ در حالی که در «یادگیری بدون نظارت» ورودی وجود دارد؛ اما بر روی خروجی نظارت وجود ندارد و از پیش مشخص و محدود شده نیست (جیمز و همکاران^{۱۰}، ۲۰۱۳). در یادگیری با نظارت، برچسب داده‌های آموزشی معلوم است. یعنی می‌دانیم هر داده متعلق به چه کلاسی یا چه طبقه‌ای است. درحقیقت، برچسب داده‌ها نظارت را انجام می‌دهد؛ ولی در یادگیری بدون نظارت، برچسب داده‌ها معلوم نیست و این یعنی نظارتی بر داده‌ها وجود ندارد (راسچکا و میرجلالی^{۱۱}، ۲۰۱۹).

در مقایسه با روش‌های آماری سنتی، روش یادگیری ماشین^{۱۲} در فرایند پیش‌بینی عمکرد بهتری دارد؛ زیرا توانایی بیشتری برای به کارگیری روابط غیرخطی و پیچیده به‌منظور استخراج مقادیر متغیرهای خروجی براساس متغیرهای ورودی دارد؛ با این حال، نقطه ضعف روش یادگیری ماشین، سختی بیشتر تقسیم نتایج به دست آمده نسبت به مدل‌های سنتی است (هاستی و همکاران^{۱۱}، ۲۰۱۷). یکی از زیرشاخه‌های یادگیری ماشین، «متن کاوی»^{۱۲} است که توسعه‌ای بر

مطالعات نشان می‌دهد تأثیر زیادی بر تصمیمات مشتریان می‌گذارند و به‌طور مستقیم بر «ساخت برند»^۱ مؤثر هستند (مورو و همکاران^۲، ۲۰۱۶).

با استفاده از تحلیل کلان داده‌ها می‌توان فرایند تصمیم‌گیری را که یکی از مسائل و چالش‌های اصلی بازاریابان است، پشتیبانی و تسهیل کرد. این کار با ارائه بینش مناسب از بازار با پاسخ به سؤالات مهمی از قبیل موارد زیر انجام می‌گیرد: مناسب‌ترین محصول برای یک بازار خاص چیست؟ چگونه در یک بازار برای چنین محصولی تبلیغ شود؟ از طریق چه کانال‌های ارتباطی، در چه مقطع زمانی و با چه قیمتی؟ و کمپین بازاریابی با چه نوع اقدامات «پیش‌برد فروش» و تبلیغاتی پشتیبانی شود؟ با توجه به این موارد، جای تعجب ندارد که حوزه بازاریابی از نخستین حوزه‌هایی باشد که کلان داده‌ها را آزمایش می‌کند و از آنها استفاده می‌کند (بندل و وانگ^۳، ۲۰۱۶).

۲-۱-۴. یادگیری ماشین^۴

یادگیری ماشین در حوزه داده کاوی و به‌منظور ساخت مدل‌های پیش‌بینی و تجزیه و تحلیل کلان داده‌ها استفاده می‌شود. اصطلاح «یادگیری ماشین» را ابتدا آرتور لی ساموئل^۵ (۱۹۹۵) یکی از پیشگامان در زمینه هوش مصنوعی ابداع کرد و منظور از این اصطلاح، استفاده از رایانه است برای واکنش در برابر شرایط، بدون اینکه به‌طور صریح برای آن برنامه‌ریزی شده باشد. پژوهشگران هوش مصنوعی، یادگیری ماشین را بهترین و امیدوارکننده‌ترین رویکرد برای نزدیک شدن هوش مصنوعی به سطح هوش انسان عنوان کرده‌اند

⁶. Syam & Sharma

⁷. Supervised Learning

⁸. Unsupervised Learning

⁹. James et al.

¹⁰. Raschka & Mirjalili

¹¹. Hastie et al.

¹². Text Mining

¹. Brand Building

². Moro et al.

³. Bendle & Wang

⁴. Machine Learning

⁵. Arthur Lee Samuel

داده کاوی، با هدف استخراج الگوهای معنادار از اسناد متنی است. متن کاوی صفحات شبکه‌های اجتماعی کمک می‌کند درک بهتری از مشتریان به دست آید. کسب و کارها از متن کاوی نظرات مشتریان برای پیش‌بینی روند فروش آینده، مدیریت ارتباط با مشتری، به‌دست آوردن دانش رقبا، عملکرد برند، تحلیل احساسات و نظرات مشتریان و تصمیم‌گیری‌های مهم و دانش محور استفاده می‌کنند (یی‌لیاوو و پی‌تان^۱، ۲۰۱۴).

۲-۲. پیشینه تجربی

در حال حاضر، بسیاری از کشورها بر توسعه صنعت گردشگری به‌عنوان یکی از مهم‌ترین صنایع مؤثر در رشد اقتصادی کشورها تأکید دارند. صنعت گردشگری، چشم‌اندازهای رشد اقتصادی در ایران را بهبود می‌بخشد و موجب افزایش سرانه تولید ناخالص داخلی می‌شود و همچنین توسعه صنایع مرتبط با گردشگری موجب شدت یافتن تأثیرگذاری مثبت توسعه گردشگری بر رشد اقتصادی کشور است (رضاقلی‌زاده، ۱۳۹۹). پژوهش رحیم‌نیا و همکاران (۱۳۹۲) درباره ابعاد ویژه برند در صنعت هتل‌داری نشان می‌دهد کیفیت ادراک‌شده، بُعد اصلی در ایجاد ارزش ویژه برند است و تأثیر زیادی بر سایر ابعاد ارزش برند دارد و تعیین‌کننده‌ای قوی در ارزش ویژه برند محسوب می‌شود. شفیع‌ی و همکاران (۱۳۹۶) در پژوهش دیگری به بررسی تأثیر فناوری اطلاعات بر توسعه صنعت گردشگری پرداخته‌اند. این پژوهش ضرورت استفاده از مدل‌های داده‌محور کسب و کار در این صنعت را که به‌دلیل تحولات ایجادشده ناشی از توسعه فناوری اطلاعات است، بیان کرده و با استفاده از رویکرد تفسیر گرانه و روش فراترکیب به تجزیه و تحلیل

داده‌ها پرداخته است. یافته‌های این پژوهش ابعاد اقتصادی، اجتماعی و محیطی را برای توسعه گردشگری پایدار تعیین کرده است و شاخص‌های هر بُعد، الزامات قانونی و تأثیرات فناوری اطلاعات برای آنها را استخراج کرده و علاوه بر این اهداف، توانایی‌ها و مؤلفه‌های موردنیاز برای پیاده‌سازی مقاصد گردشگری هوشمند را شناسایی کرده است. خلیل‌نژاد و همکاران (۱۳۹۸) در پژوهش خود، تأثیر تجربه برند هتل، در ارزش ادراک‌شده میهمانان و قوت برند هتل را بررسی کرده‌اند. یافته‌های این پژوهش نشان می‌دهد تجربه برند هتل در ارزش ادراک‌شده مالی، ارزش ادراک‌شده کارکردی و ارزش ادراک‌شده اجتماعی تأثیر می‌گذارد؛ اما این تجربه بر ارزش ادراک‌شده فردی تأثیر ندارد؛ بلکه ارزش‌های ادراک‌شده مالی، کارکردی و اجتماعی هستند که بر ارزش ادراک‌شده فردی مؤثرند و در نهایت، ارزش ادراک‌شده فردی در قوت برند هتل تأثیر می‌گذارد. به‌دلیل نوظهور بودن نسبی حوزه استفاده از کلان‌داده‌ها در صنعت گردشگری، مطالعات محدودی در داخل کشور انجام گرفته است و رو به گسترش است. در زمینه تأثیر کاربرد کلان‌داده‌ها و یادگیری ماشین در فرایند بازاریابی، محمدیان و نائلی (۱۳۹۶) در پژوهشی به بررسی تحلیلی مفهوم کلان‌داده در حوزه بازاریابی نوین پرداخته‌اند. نتایج این پژوهش نشان می‌دهد شرکت‌ها و سازمان‌هایی که کلان‌داده‌ها را به‌چشم‌دارایی نگاه می‌کنند، در بازار رقابتی موفق‌تر عمل می‌کنند و می‌توانند با محوریت قراردادن مشتریان، میزان وفاداری آنها را افزایش دهند. همچنین این پژوهش به کمبود سرمایه‌گذاری و پژوهش‌های داخلی در این حوزه اشاره کرده است و شناخت و بهره‌مندی از این علم را جزء ملزومات نوین مدیریتی معرفی کرده و به مدیران

¹. Yee Liao & Pei Tan

حفظ بقا و جلوگیری از حذف شدن آن توسط رقبای دیگر در شرایط بد اقتصادی می‌شود. امیرخانی و متقی (۱۳۹۷) در پژوهش دیگری به معرفی استفاده یادگیری ماشین از کلان داده‌ها برای طراحی مدل، جهت پیش‌بینی پرداخته‌اند و با رویکردی جدید در طراحی مدل یادگیری ماشین، میزان محصول تولیدشده یک شرکت دارویی را پیش‌بینی کرده‌اند. نتایج این پژوهش نشان می‌دهد روش پیشنهادی آنها می‌تواند به پیش‌بینی تولید محصول با دقت مطلوب و درصد خطای بسیار پایین پردازد؛ همچنین نتایج به دست آمده نشان می‌دهد استفاده از یادگیری ماشین می‌تواند مشکلات تجزیه و تحلیل داده‌های پیچیده و بزرگ را در سیستم‌های مدیریت تولید محصول حل کند. ناصر و همکاران (۱۳۹۷) در پژوهش دیگری به مطالعه رفتار مشتریان در رزرواسیون آنلاین با استفاده از تکنیک داده کاوی و مدل کانو پرداخته‌اند. در این پژوهش هتل‌های چهار و پنج ستاره شهر تهران به عنوان مطالعه موردی بررسی شده‌اند و الگویی تلفیقی از مدل کانو و قوانین انجمنی برای طبقه‌بندی نیازهای مشتریان و تجزیه و تحلیل رفتار آنها ارائه شده است. نتایج پژوهش، نیازهای مشتریان را در سه دسته نیازهای اساسی، نیازهای عملکردی و نیازهای مهبج طبقه‌بندی کرده است. علاوه بر این، نتایج نشان می‌دهد داشتن امکان رزرو آنلاین در وبسایت شرکت ارائه‌دهنده خدمات رزرو بلیط، برای مردان اهمیت بیشتری نسبت به زنان دارد. سلطانی زنوزی (۱۳۹۸) در پژوهش دیگری نقش کلان داده‌ها در بازاریابی و کسب و کار را بررسی کرده است و دسترسی به کلان داده‌ها را دارای اهمیت زیادی برای شناخت رفتار مشتری، برنامه‌ریزی کمپین‌های تبلیغاتی و تصمیم‌گیری درباره آمیخته بازاریابی بیان کرده است. همچنین، به کاربردهای دیگر کلان داده‌ها مانند ایجاد

شرکت‌ها و سازمان‌ها پیشنهاد سرمایه‌گذاری در این حوزه به منظور بهره‌برداری از مزایای آن را داده است. آقایی و اسماعیلی (۱۳۹۶) در پژوهشی دیگر، تأثیر استفاده از کلان داده‌ها در مدیریت بازاریابی و تجزیه و تحلیل رفتار مصرف‌کننده را بررسی کرده‌اند و همچنین با اشاره به گسترش نفوذ اینترنت اشیا، به اهمیت و پیشرفت حوزه تجزیه و تحلیل کلان داده‌ها پرداخته‌اند. این پژوهش استفاده از کلان داده‌ها را روشی جدید و کارا برای شناخت بهتر مصرف‌کنندگان و تجزیه و تحلیل رفتار آنها معرفی کرده است. فقیه و اسدی (۱۳۹۷) در پژوهش خود با اشاره به افزایش روزافزون نقش کلان داده‌ها در بازاریابی، به بررسی مزایای کلان داده در حوزه بازاریابی پرداخته‌اند و توانایی‌های بالقوه استفاده از کلان داده‌ها برای کسب بینش لازم برای تصمیم‌گیری مدیران بازاریابی را یکی از مزایای راهبردی آن بیان کرده‌اند. نتایج بررسی‌های پژوهش نشان‌دهنده آن است که استفاده از کلان داده‌ها به شرط جمع‌آوری داده‌های صحیح و مرتبط و تحلیل درست آنها، تأثیر مستقیم و مثبتی در افزایش فروش و سودآوری دارد. مؤذن رضامحله (۱۳۹۷) در پژوهش خود تأثیر استفاده از مدیریت دانش و کلان داده‌ها را در مدیریت بازاریابی شرکت‌های بزرگ بررسی کرده است و با اشاره به تغییر شرایط کسب و کار و تحولات به وجود آمده در فرایند بازاریابی و فروش محصولات و اهمیت مدل بازاریابی یک شرکت برای پیروزی بر رقبای در بازار رقابتی، استفاده از کلان داده‌ها را راهکاری مناسب برای این کسب و کارها معرفی کرده است. نتایج این پژوهش نشان می‌دهد استفاده از کلان داده‌ها در مدیریت دانش باعث افزایش قدرت کسب و کار برای

¹. Internet of Things (IoT)

تغییر در محصول و انتخاب کانال‌های توزیع، پیاده‌سازی راهبردهای بازاریابی الکترونیکی و تولید محتوا اشاره کرده است. یافته‌های پژوهش نشان‌دهنده آن است که استفاده از کلان‌داده‌ها فرایند بازاریابی و فروش را با بهینه‌سازی و شخصی‌سازی راهبردهای بازاریابی بهبود می‌بخشد و باعث افزایش رضایت و وفاداری مشتریان می‌شود و در نتیجه با ایجاد توازن میان محصولات یک سازمان با نیازهای مشتری هم برای کسب و کار و هم برای مشتری سود و ارزش بیشتری همراه خواهد داشت.

استفاده از قابلیت‌های بالقوه سیستم‌های اطلاعاتی، یکی از عوامل مهم موفقیت برای رقابت در صنعت میهمان‌نوازی^۱ (گردشگری) است (اسپارک و همکاران^۲، ۲۰۱۶؛ راگیسو و همکاران^۳، ۲۰۱۷). همچنین به علت آنکه سرمایه‌گذاری در صنعت گردشگری مستلزم هزینه‌های زیادی است، بررسی و تجزیه و تحلیل ویژگی‌هایی که مسافران از خدمات دریافت‌شده خود توصیف می‌کنند یا در ذهن خود مجسم و بازسازی می‌کنند و در نظرات آنلاین خود به اشتراک می‌گذارند، امری معقول و ضروری است (سانچز و همکاران، ۲۰۱۹). ظهور رسانه‌های اجتماعی، وب ۲.۰ و کانال‌های دیجیتال باعث به وجود آمدن توصیه‌ها، بررسی‌ها و بیان نظرات مشتریان به صورت آنلاین شده است و به همین دلیل، نفوذ بازاریابی دهان‌به‌دهان (WoM^۴) به طور روزافزونی گسترش یافته است و در نهایت به پیدایش بازاریابی دهان‌به‌دهان الکترونیکی (eWoM) انجامیده است. «بازاریابی

دهان‌به‌دهان الکترونیکی» شکل جدید بازاریابی دهان‌به‌دهان است که از طریق اینترنت، وسایل الکترونیکی و شبکه‌های اجتماعی انجام می‌شود (لئونگ و همکاران^۵، ۲۰۱۹). پژوهش چونگ و همکاران^۶ (۲۰۱۸) به اهمیت نظرات و کامنت‌های بحث‌شده در اینترنت اشاره کرده است و نتایج این مطالعه نشان می‌دهد اغلب مسافران، نظرات مسافران قبلی را منبع اطلاعاتی برای تصمیم‌گیری در نظر می‌گیرند. همچنین، یافته‌های این پژوهش تأییدکننده اهمیت و سودمندی تحلیل این داده‌ها در بازاریابی پیش‌بینانه است. گیل‌ستو و همکاران^۷ (۲۰۱۹) در مطالعه خود با استفاده از بررسی نظرات آنلاین مشتریان به تجزیه و تحلیل برداشت مشتریان (میهمانان) از شیوه‌های حفاظت هتل‌ها از محیط‌زیست پرداخته‌اند. نتایج پژوهش آنها نشان می‌دهد مشتریان اقداماتی را که هتل انجام داده است، در شش مورد^۱ (انرژی، ۲ آب، ۳ خرید، ۴ زباله، ۵ سایت و آموزش و ۶ نوآوری درک می‌کنند؛ ولی سطح تعهدات زیست‌محیطی هتل‌ها را تشخیص نمی‌دهند. پژوهش سانچز و همکاران (۲۰۱۹) با استفاده از یادگیری ماشین (روش یادگیری ماشین نظارت‌شده) به تحلیل نظرات مشتریان درباره هتل پرداخته و راهکارهایی را برای افزایش سطح کیفیت سرویس‌دهی به میهمانان برای جلب رضایت آنها معرفی کرده است؛ بدین منظور، برای استخراج نظرات مشتریان ۳۳ هتل از سایت یلپ^۸ (یکی از شبکه‌های اجتماعی محبوب در ایالات متحده) استفاده شده است و در نهایت مدلی ارائه کرده که این مدل می‌تواند به مدیران هتل در افزایش سطح کیفیت سرویس‌دهی به مشتریان کمک کند.

^۵. Leong et al.

^۶. Chong et al.

^۷. Gil-Soto et al.

^۸. Yelp

^۱. Hospitality

^۲. Sparks et al.

^۳. Raguseo et al.

^۴. Word of Mouth (WoM)

۳. روش‌شناسی پژوهش

پژوهش حاضر از لحاظ هدف کاربردی، از نظر روش گردآوری داده‌ها توصیفی-پیمایشی و از نظر اجرا کیفی-کمی (آمیخته) و با رویکرد استقرایی است و برای تحلیل محتوای نظرات مشتریان، از روش‌ها و الگوریتم‌های تحلیل کلان داده‌های متنی استفاده شده است. داده‌های استفاده‌شده در این پژوهش، نظرات ثبت‌شده مشتریان در بستر اینترنت (در زمینه هتل) است. مجموعه داده‌ها (دیتاست^۱) حاوی ۸۲۹۰ نظر (کامنت^۲) به زبان فارسی است. برای انجام فرایندهای داده‌کاوی، متن کاوی و تحلیل کلان داده‌ها از نرم‌افزار رپیدماینر^۳ و زبان برنامه‌نویسی پایتون^۴ استفاده شده است. مراحل اجرایی پژوهش شامل سه مرحله اصلی است که در زیر توضیح داده شده است.

۳-۱- پیش پردازش داده‌ها^۵

داده‌های ورودی در متن کاوی، اغلب اسناد متنی بدون ساختار است؛ به همین دلیل، ابتدا باید داده‌های ورودی پیش پردازش شوند. در این فرایند ابتدا متن به کلمات، نمادها و یا دیگر عناصر معنادار که «نشانه^۶» نامیده می‌شوند، تقسیم می‌شوند (ویز و همکاران^۷، ۲۰۱۵). برای استخراج کلمات یک متن، ابتدا با حذف علائم نقطه‌گذاری و دیگر کاراکترهای غیرمتنی، داده‌ها تبدیل به سطری از کلمات می‌شوند. در شکل ۱ «توزیع ابری کلمات^۸» نظرات گردشگران درباره هتل بعد از انجام فرایند پیش پردازش داده‌ها نمایش داده شده است.

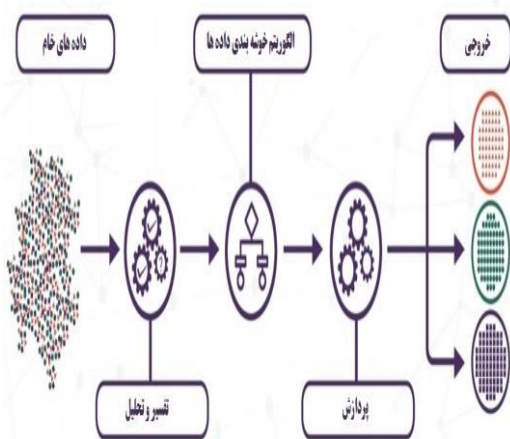


شکل ۱. نمایش ابری کلمات نظرات در

داده‌های متنی

۳-۲- خوشه‌بندی^۹ داده‌ها

در گام بعدی برای استخراج مؤلفه‌های «ارزش» از میان حجم انبوه نظرات مشتریان، از روش «خوشه‌بندی» استفاده می‌شود. خوشه‌بندی، فرایندی است که در آن، داده‌ها به گروه‌هایی که اعضای آن مشابه یکدیگر باشند تقسیم می‌شوند. یک خوشه مجموعه‌ای از اشیای شبیه یکدیگر است که با اشیای موجود در خوشه‌های دیگر غیرمشابه باشند (راسچکا و میرجلالی، ۲۰۱۹). در شکل ۲ فرایند خوشه‌بندی داده‌ها نشان داده شده است.



شکل ۲. نمایش شماتیک خوشه‌بندی (جداسازی) موضوعات اسناد متنی

1. Dataset
2. Comment
3. RapidMiner
4. Python
5. Data Pre-Processing
6. Token
7. Weiss et al.
8. Word Cloud

9. Clustering

اعضای درون خوشه با یکدیگر) و تفکیک پذیری^۶ (فاصله خوشه‌ها از همدیگر) ارائه شده است و با دو معیار «فاصله^۷» و «قطر^۸»، میزان فشردگی و تفکیک پذیری را محاسبه می‌کند. این معیار را رابطه^۱ تعریف می‌کند که در آن $d(C_i, C_j)$ مقدار فاصله بین خوشه^۱ i و j است و $diam(C_i)$ نماینده قطر خوشه^۱ است (فائزی راد و پویا، ۱۳۹۵).

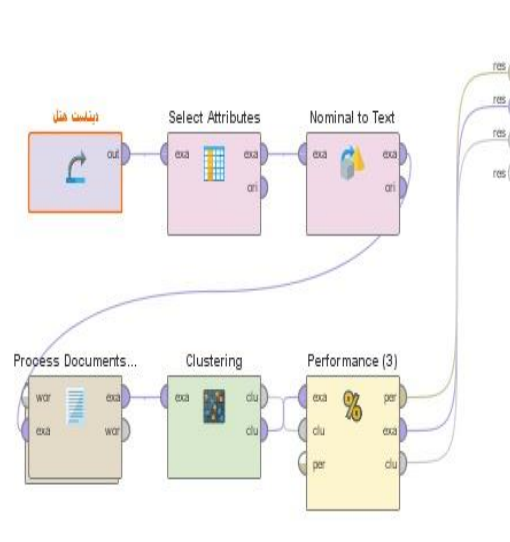
$$DI = \min_{1 \leq i \leq K} \left\{ \min_{i+1 \leq j \leq K} \left\{ \frac{d(C_i, C_j)}{\max_{1 \leq l \leq K} diam(C_l)} \right\} \right\} \quad \text{رابطه ۱.}$$

۳-۳- قوانین انجمنی^۹

قوانین انجمنی، رویکردی در داده کاوی است که برای کشف الگوهای مکرر، همبستگی‌ها، ارتباطات و ساختارهای علی در انواع پایگاه‌های داده و برای تحلیل و پیش‌بینی رفتار کاربران به کار می‌رود (شالینی و لال^{۱۰}، ۲۰۱۶). این قوانین با کشف ارتباطات بین مجموعه داده‌ها شرایطی را نشان می‌دهد که در یک مجموعه داده به‌طور مکرر با هم اتفاق می‌افتند و حضور برخی اقلام (موجودیت‌ها) را براساس سایر اقلام (موجودیت‌ها) شرح می‌دهند (انصاری و همکاران، ۱۳۹۸). هدف قوانین انجمنی در متن کاوی، کشف روابط میان کلمات و شناسایی قوانینی است که در متن، حضور یک مجموعه از کلمات وجود کلمات دیگر را ایجاب می‌کند.

در یک پایگاه داده اگر $I = \{I_1, I_2, \dots, I_m\}$ مجموعه‌ای از اقلام باشد و D مجموعه‌ای از پایگاه داده‌های تراکنش‌هاست که هر تراکنش دارای یک شناسه و همچنین مجموعه اقلامی نظیر $T \subseteq I$ است.

در متن کاوی هم، خوشه‌بندی تکنیکی است که برای گروه‌بندی اسناد مشابه (در یک موضوع) استفاده می‌شود. خوشه‌بندی متون زمانی نیاز می‌شود که حجم اسناد متنی بسیار زیاد باشد و امکان جداسازی آنها براساس موضوعی که هر گروه از اسناد به آن اشاره می‌کند برای انسان غیرممکن یا بسیار مشکل و زمان‌بر باشد. یکی از مقبول‌ترین روش‌ها، خوشه‌بندی داده‌ها به روش «k میانگین^۱» است که تشابه اسناد متنی و رابطه بین کلمات کلیدی را در نظر می‌گیرد (ویز و همکاران، ۲۰۱۵). در شکل ۳ نحوه انجام خوشه‌بندی داده‌ها توسط ریپدماینر در این پژوهش نشان داده شده است.



شکل ۳. خوشه‌بندی نظرات توسط ریپدماینر

در این پژوهش برای تعیین تعداد خوشه‌ها و اعتبارسنجی خوشه‌بندی^۲ از روش «شاخص دآن^۳» به‌دلیل کارایی بهتر نسبت به روش‌های دیگر در کلان‌داده‌ها استفاده شده است (کاسامبارا^۴، ۲۰۱۷). شاخص «دآن» با هدف تعیین مقدار فشردگی^۵ (فاصله

^۶. Separation

^۷. Cluster Distance

^۸. Diameter

^۹. Association Rule

^{۱۰}. Shalini & Lal

^۱. k-means Clustering

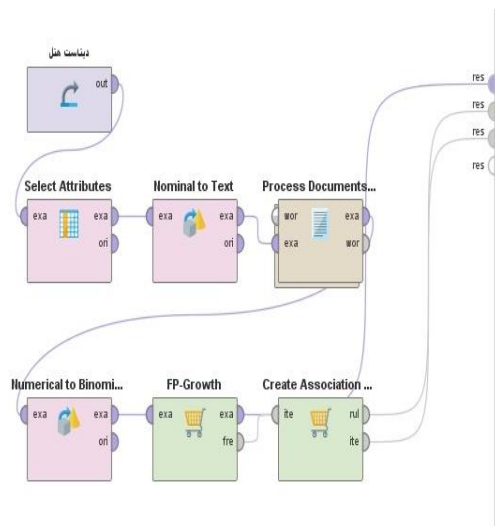
^۲. Clustering Validation

^۳. Dunn Index

^۴. Kassambara

^۵. Compactness

از بررسی این دو روش و مقایسه کارایی و سرعت آنها، از روش FP-Growth برای استخراج قوانین انجمنی استفاده شد (اسماعیلی پور، ۱۳۹۵). در شکل شماره ۴ نحوه انجام فرایند استخراج قوانین انجمنی در این پژوهش توسط رپیدماینر نشان داده شده است.



شکل ۴. استخراج قوانین انجمنی توسط رپیدماینر

۴. یافته‌های پژوهش

یافته‌های حاصل از متن کاوی نظرات مشتریان درباره صنعت گردشگری، در قالب دو بخش «خوشه‌بندی نظرات» و «استخراج قوانین انجمنی» ارائه می‌شود.

۴-۱. یافته‌های فرایند «خوشه‌بندی»

در این قسمت برای استخراج مؤلفه‌های ارزش‌موردانتظار مشتریان، نظرات جمع‌آوری شده با استفاده از الگوریتم «k-means» خوشه‌بندی می‌شوند و برای شناسایی موضوع یا ویژگی اشاره شده در هر خوشه، پرتکرارترین لغات در آن خوشه مشخص می‌شوند. نام‌گذاری هر خوشه با توجه به رابطه منطقی بین کلمات موجود در هر دسته از موضوعات انجام می‌گیرد که بیشترین تکرار (احتمال) را داشته‌اند (گوو و همکاران، ۲۰۱۷).

اگر A مجموعه‌ای از اقلام باشد، تراکشی مانند T حاوی A است، اگر داشته باشیم: $A \subseteq T$. یک قانون انجمنی به شکل $A \Rightarrow B$ بیان می‌شود که در آن $A \subset I$ و $A \cap B = \emptyset$ ، $B \neq \emptyset$ ، $A \neq \emptyset$ ، $B \subset I$ ، (همکاران، ۲۰۱۱). یک قانون انجمنی در شکل کلی خود بیان می‌کند که اگر رویداد A واقع شود، آنگاه رویداد B نیز رخ خواهد داد. از نماد $A \Rightarrow B$ برای نمایش یک قانون انجمنی استفاده می‌شود؛ A را مقدمه قانون و B را نتیجه قانون می‌نامند. مصدر تولید قوانین انجمنی هم‌زمانی‌هایی هستند که به اندازه کافی تکرار شده باشند که آنها را «مجموعه اقلام مکرر» می‌نامند (ناصر و همکاران، ۱۳۹۷). هر قانون $A \Rightarrow B$ شاخصی با نام «پشتیبان» دارد که مشخص می‌کند این هم‌زمانی به چه نسبتی تکرار شده است. مقدار پشتیبان قانون $A \Rightarrow B$ به کمک رابطه ۲ محاسبه می‌شود. صورت کسر برابر با تعداد هم‌زمانی‌های A و B است و مخرج کسر به تعداد کل رخدادها اشاره دارد.

$$\text{رابطه ۲. } Support(A \Rightarrow B) = N(A \& B) / N$$

پس از آنکه قوانینی با حداقلی از مقدار پشتیبان تولید شدند، می‌توان احتمال درست‌بودن قوانین را نیز محاسبه کرد. این شاخص که «مقدار اطمینان» نام دارد، مشخص می‌کند در چند درصد از زمان‌هایی که رویداد A واقع شده است، رویداد B نیز رخ داده است. مقدار اطمینان قانون $A \Rightarrow B$ به کمک رابطه ۳ محاسبه می‌شود (لی و همکاران، ۲۰۱۷).

$$\text{رابطه ۳. } Confidence(A \Rightarrow B) = N(A \& B) / N(A)$$

از مهم‌ترین و کاربردی‌ترین روش‌ها یا الگوریتم‌های استخراج قوانین انجمنی دو روش Apriori^۳ و FP-Growth^۳ هستند. در این پژوهش پس

^۱. Han

^۲. Lee

^۳. Frequent Pattern Growth Algorithm (FP-Growth)

^۴. Guo et al.

زیاد در آن خوشه داشته‌اند، موضوع بحث در آن خوشه مشخص می‌شود. از ۲۰ موضوع یا ویژگی شناسایی شده در این مجموعه پس از مرتب کردن، ۱۱ موضوع اول به مؤلفه‌های مثبت ارزش و در ادامه ۹ موضوع بعدی به مؤلفه‌های منفی ارزش از نگاه گردشگران اشاره دارد.

نتایج حاصل از خوشه‌بندی نظرات گردشگران به همراه هشت «کلمه» با بیشترین تکرار (احتمال) در هر خوشه برای شناخت موضوع یا ویژگی مدنظر مشتری در هر خوشه در جدول ۱ آمده است. با اعمال الگوریتم خوشه‌بندی بر روی نظرات گردشگران، ۲۰ خوشه مهم شناسایی شد. با بررسی هر خوشه و کلماتی که تکرار

جدول ۱. موضوعات استخراج شده از خوشه‌بندی نظرات و هشت کلمه با احتمال بالاتر در هر خوشه

موضوعات حاصل از خوشه‌بندی نظرات								«کلمات» با بیشترین تکرار (احتمال) در هر خوشه											
خوشه ۱:	نظرات دیگران درباره هتل	نظرات	دوستان	سایت	هتل	وقتی	انتخاب	خوندم	ستاره	خوشه ۱:	نظرات دیگران درباره هتل	نظرات	دوستان	سایت	هتل	وقتی	انتخاب	خوندم	ستاره
خوشه ۲:	تمیزی و بهداشت کلی هتل	اتاق	هتل	سرویس	ولی	صبحانه	تمیز	بهداشتی	رستوران	خوشه ۲:	تمیزی و بهداشت کلی هتل	اتاق	هتل	سرویس	ولی	صبحانه	تمیز	بهداشتی	رستوران
خوشه ۳:	برخورد پرسنل	هتل	تشکر	آقای	خانم	پرسنل	برخورد	مدیریت	خوب	خوشه ۳:	برخورد پرسنل	هتل	تشکر	آقای	خانم	پرسنل	برخورد	مدیریت	خوب
خوشه ۴:	نزدیکی به مراکز خرید و تفریح	هتل	پیاده	بازار	ساحل	مرکز	دقیقه	نزدیک	خرید	خوشه ۴:	نزدیکی به مراکز خرید و تفریح	هتل	پیاده	بازار	ساحل	مرکز	دقیقه	نزدیک	خرید
خوشه ۵:	سرویس بهداشتی و حمام مناسب	حمام	سرویس	حوله	تمیز	صدای	بهداشتی	دستشویی	خوب	خوشه ۵:	سرویس بهداشتی و حمام مناسب	حمام	سرویس	حوله	تمیز	صدای	بهداشتی	دستشویی	خوب
خوشه ۶:	تمیزی اتاق (ملافه، رومیزی و...)	اتاق	خوب	تمیز	برای	تخت	سفید	ملافه	میز	خوشه ۶:	تمیزی اتاق (ملافه، رومیزی و...)	اتاق	خوب	تمیز	برای	تخت	سفید	ملافه	میز
خوشه ۷:	زیبایی فضا و چیدمان داخلی	هتل	عالی	زیبا	اتاق	رستوران	فضای	بسیار	شیک	خوشه ۷:	زیبایی فضا و چیدمان داخلی	هتل	عالی	زیبا	اتاق	رستوران	فضای	بسیار	شیک
خوشه ۸:	امکانات منطقه‌ای برای تفریح	پارک	رفتیم	تومان	رستوران	کشتی	برید	داره	کیش	خوشه ۸:	امکانات منطقه‌ای برای تفریح	پارک	رفتیم	تومان	رستوران	کشتی	برید	داره	کیش
خوشه ۹:	کیفیت و تنوع صبحانه	پنیر	خوب	تخم	مربا	کره	صبحانه	عدسی	سوسیس	خوشه ۹:	کیفیت و تنوع صبحانه	پنیر	خوب	تخم	مربا	کره	صبحانه	عدسی	سوسیس
خوشه ۱۰:	خدمات ترانسفر و تحویل اتاق	اتاق	ساعت	هتل	برای	تحویل	ترانسفر	شدیم	فرودگاه	خوشه ۱۰:	خدمات ترانسفر و تحویل اتاق	اتاق	ساعت	هتل	برای	تحویل	ترانسفر	شدیم	فرودگاه
خوشه ۱۱:	امکانات جانبی	اتاق	تلفن	یخچال	چای	کافی	اینترنت	سرعت	قهوه	خوشه ۱۱:	امکانات جانبی	اتاق	تلفن	یخچال	چای	کافی	اینترنت	سرعت	قهوه
خوشه ۱۲:	هزینه زیاد تفریحات بیرون از هتل	تومان	پارک	قیمت	کشتی	رستوران	بلیط	کیش	گران	خوشه ۱۲:	هزینه زیاد تفریحات بیرون از هتل	تومان	پارک	قیمت	کشتی	رستوران	بلیط	کیش	گران
خوشه ۱۳:	قطعی برق	اتاق	برق	اتاق‌ها	تاریکی	قطع	باسلام	منطقه	برق‌ها	خوشه ۱۳:	قطعی برق	اتاق	برق	اتاق‌ها	تاریکی	قطع	باسلام	منطقه	برق‌ها
خوشه ۱۴:	تناسب نداشتن ستاره هتل با کیفیت آن	هتل	اتاق	ستاره	برخورد	کیفیت	سرویس	کثیف	اقامت	خوشه ۱۴:	تناسب نداشتن ستاره هتل با کیفیت آن	هتل	اتاق	ستاره	برخورد	کیفیت	سرویس	کثیف	اقامت
خوشه ۱۵:	تأخیر در پرواز	ساعت	دیر	تومان	تأخیر	شدیم	فرودگاه	تحویل	موقع	خوشه ۱۵:	تأخیر در پرواز	ساعت	دیر	تومان	تأخیر	شدیم	فرودگاه	تحویل	موقع
خوشه ۱۶:	وجود حشرات مزاحم در اتاق	ساس	آرامش	صبح	حشره	مزاحم	مهم‌تر	ازهمه	جوش	خوشه ۱۶:	وجود حشرات مزاحم در اتاق	ساس	آرامش	صبح	حشره	مزاحم	مهم‌تر	ازهمه	جوش
خوشه ۱۷:	کیفیت پایین لوازم بهداشتی	حوله	کیفیت	صابون	مسواک	شامپو	بد	زنگ	نفر	خوشه ۱۷:	کیفیت پایین لوازم بهداشتی	حوله	کیفیت	صابون	مسواک	شامپو	بد	زنگ	نفر
خوشه ۱۸:	ورود سروصدای مزاحم از بیرون	اتاق	صدای	طبقه	بیرون	بلند	پنجره	صدا	مزاحم	خوشه ۱۸:	ورود سروصدای مزاحم از بیرون	اتاق	صدای	طبقه	بیرون	بلند	پنجره	صدا	مزاحم
خوشه ۱۹:	نظافت نشدن اتاق و...	اتاق	کثیف	سرویس	حمام	دستشویی	خراب	توالت	حوله	خوشه ۱۹:	نظافت نشدن اتاق و...	اتاق	کثیف	سرویس	حمام	دستشویی	خراب	توالت	حوله
خوشه ۲۰:	قیمت بالا	تومان	هزینه	آژانس	اتاق	پرواز	تور	پول	گران	خوشه ۲۰:	قیمت بالا	تومان	هزینه	آژانس	اتاق	پرواز	تور	پول	گران

۴-۲- یافته‌های فرایند «قوانین انجمنی»

نظرات خود به کار برده‌اند، استخراج می‌شود. در جدول ۲ قوانین استخراج شده و دو شاخص سطح پشتیبان و سطح اطمینان آورده شده است.

هدف از شناسایی قوانین انجمنی، کشف روابط میان کلمات در مجموعه بزرگی از متون (نظرات مشتریان) است. در این روش، دانش پنهان از تحلیل کلمات یا عباراتی که مشتریان با یکدیگر و به‌وفور در

جدول ۲. نتایج استخراج قوانین انجمنی و میزان دو شاخص سطح پشتیبان و سطح اطمینان

شماره	مقدمه (قانون)	نتیجه (قانون)	سطح اطمینان	مقدار پشتیبان	قانون
۱	برخورد	اقامت، هتل	0.9650	0.3901	[اقامت هتل] → [برخورد]
۲	سرویس	اقامت، هتل	0.9679	0.2221	[اقامت هتل] → [سرویس]
۳	اتاق	اقامت، هتل	0.9686	0.5174	[اقامت هتل] → [اتاق]
۴	رستوران	اقامت، هتل	0.9689	0.3008	[اقامت هتل] → [رستوران]
۵	صبحانه	اقامت، هتل	0.9734	0.4023	[اقامت هتل] → [صبحانه]
۶	تمیز	اقامت، هتل	0.9746	0.2317	[اقامت هتل] → [تمیز]
۷	ستاره	اقامت، هتل	0.9847	0.2726	[اقامت هتل] → [ستاره]

ارزش پیشنهادی خود به مشتری، محصولی ارائه کند که به بهترین وجه، نیاز مشتریان آن بخش را پوشش دهد (شیهان و برونی بوسیو^۱، ۲۰۱۵). به‌طور کلی می‌توان گفت شناخت و ارائه «ارزش» مناسب به مشتری یکی از چالش‌های مهم برای موفقیت و سودآوری هر کسب‌وکار است.

این پژوهش با توجه به فرصت‌های به‌وجود آمده از پیامدهای انقلاب صنعتی چهارم که یکی از آنها دیجیتالیزه شدن محیط کسب‌وکار است، با استفاده از تحلیل کلان داده‌های شبکه‌های اجتماعی، چهارچوبی برای استخراج ویژگی‌ها یا مؤلفه‌های ارزش برای ارائه به مشتری معرفی می‌کند تا براساس آن، کسب‌وکارها بتوانند در فضای رقابتی بازار ارزش پیشنهادی مناسب‌تری به مشتریان خود عرضه کنند. نتایج این پژوهش با استفاده از دو روش متفاوت از یادگیری ماشین (خوشه‌بندی نظرات و استخراج قوانین انجمنی)

قوانین انجمنی به‌شکل «اگر ← آن‌گاه» تعریف می‌شوند. به‌عبارت دیگر، با استفاده از این روش روابط و وابستگی‌های پنهان در کلان داده‌ها آشکار می‌شود و کشف قوانین مهم و سودمند با ارائه اطلاعات مناسب، فرایند تصمیم‌گیری مدیران را تسهیل می‌کند.

۵. بحث، نتیجه‌گیری و پیشنهادها

هر کسب‌وکاری براساس یک «مدل کسب‌وکار» به فعالیت می‌پردازد. درحقیقت، مدل کسب‌وکار بیان‌کننده نحوه ایجاد ارزش و ارائه آن به مشتری است (مبینی‌دهکردی و همکاران، ۱۳۹۳). یکی از مهم‌ترین مؤلفه‌های مدل کسب‌وکار، «ارزش پیشنهادی به مشتری» است که بیان‌گر کالا یا خدماتی است که شرکت تعهد می‌دهد به مشتریان تحویل دهد (غلامی و همکاران، ۱۳۹۸). هنگامی که کسب‌وکار، بخش جذابی از بازار را به‌عنوان بازار هدف انتخاب می‌کند، باید با انتخاب مؤلفه‌ها یا ویژگی‌های مناسب برای

¹. Sheehan & Bruni-Bossio

به اکتشاف ویژگی‌ها یا مؤلفه‌های ارزش متناسب با انتظار مشتریان (گردشگران)، با تحلیل نظرات آنها در این زمینه پرداخته است.

نتایج حاصل شده با استفاده از روش اول (خوشه‌بندی داده‌ها) نشان‌دهنده ۲۰ مؤلفه (موضوع یا ویژگی) مهم در ارزش‌های درک شده گردشگران است که شامل ۱۱ مؤلفه مثبت ارزش و ۹ مؤلفه منفی ارزش است. ویژگی‌ها یا مؤلفه‌های مثبت ارزش ادراک شده گردشگران عبارت‌اند از: ۱) نظرات دیگران درباره هتل، ۲) تمیزی و بهداشت کلی هتل، ۳) برخورد خوب پرسنل هتل، ۴) نزدیکی هتل به مراکز خرید و تفریح، ۵) سرویس بهداشتی و حمام مناسب و تمیز، ۶) تمیزی اتاق‌های هتل (ملافه، حوله، رومیزی)، ۷) زیبایی فضای هتل و چیدمان داخلی، ۸) امکانات منطقه‌ای هتل برای تفریح، ۹) کیفیت و تنوع صبحانه، ۱۰) خدمات ترانسفر (جابه‌جایی) و تحویل اتاق و ۱۱) امکانات جانبی مانند تلفن، یخچال، اینترنت پرسرعت و ...

ویژگی‌ها یا مؤلفه‌های منفی ارزش ادراک شده گردشگران عبارت‌اند از: ۱) هزینه بالای تفریحات بیرون از هتل، ۲) قطعی برق اتاق‌ها، ۳) تناسب‌نداشتن ستاره هتل با کیفیت آن، ۴) تأخیر در پرواز، ۵) وجود حشرات مزاحم در اتاق، ۶) کیفیت پایین وسایل و لوازم بهداشت شخصی مانند حوله، شامپو و ...، ۷) ورود سروصداهای مزاحم از بیرون به داخل اتاق، ۸) نظافت‌نشدن اتاق و ... و ۹) قیمت بالا و نامتناسب.

نتایج به دست آمده با استفاده از روش دوم (قوانین انجمنی) شامل ۷ قانون مهم است که با استخراج دانش پنهان از روابط بین عبارات به کاربرده شده در نظرات گردشگران به دست آمده‌اند. یافته‌های این قسمت نشان می‌دهد ۷ موضوع از نظر مشتریان درباره اقامت در یک

هتل اهمیت بیشتری دارد که عبارت‌اند از: ۱) برخورد پرسنل هتل، ۲) نحوه و کیفیت سرویس‌دهی هتل به گردشگران، ۳) کیفیت و شرایط اتاق‌های هتل، ۴) کیفیت و شرایط رستوران هتل، ۵) کیفیت و تنوع صبحانه، ۶) تمیزی و بهداشت کلی هتل و ۷) تعداد ستاره‌های هتل. نتایج استخراج قوانین انجمنی از میان نظرات گردشگران، برای ارائه‌دهندگان خدمات گردشگری و هتل‌داران این موقعیت را فراهم می‌کند که بتوانند از نظرات مشتریان خود استفاده کنند و رضایت گردشگران را بیشتر جلب کنند. از سوی دیگر، مدیران با درک این قوانین، شناختی کلی از فضای کسب و کار و شرایط موجود هتل‌ها از دید گردشگران به دست می‌آورند و می‌توانند از این دانش در راستای ایجاد مزیت رقابتی استفاده کنند. همچنین، با استفاده از این قوانین ویژگی و مزیت‌های مدنظر گردشگران شناسایی می‌شود و مدیران می‌توانند از آنها برای تدوین برنامه‌های راهبردی به منظور ارائه ارزش به مشتری استفاده کنند.

نتایج پژوهش درباره «اهمیت کیفیت سرویس‌دهی به مشتریان» با پژوهش سانچز و همکاران (۲۰۱۹) تطابق دارد. نتایج پژوهش چونگ و همکاران (۲۰۱۸) نشان‌دهنده اهمیت نظرات بحث‌شده در اینترنت درباره هتل است و بیان‌گر این مطلب است که اغلب مسافران، نظرات مسافران قبلی را به‌عنوان منبع اطلاعاتی برای تصمیم‌گیری در نظر می‌گیرند. نتایج پژوهش حاضر نیز این مطلب را تأیید می‌کند و یکی از عناصر مثبت به دست آمده در بخش خوشه‌بندی، بررسی نظرات مسافران قبلی درباره هتل است. در باب اهمیت استفاده از نظرات گردشگران قبلی در فضای مجازی و استفاده از توانایی‌های بالقوه تحلیل داده‌ها در این زمینه باید گفت پژوهش‌های اسپارک و همکاران (۲۰۱۶) و راگسیو و

مصرف کننده با روش‌های جدید علوم داده‌ها^۱ مانند یادگیری ماشین و یادگیری عمیق^۲، از توانایی‌های بالقوه تحلیل کلان داده‌ها در زمینه تحقیقات بازار و دیگر شاخه‌های مرتبط با مدیریت کسب و کار بهره ببرند. درباره محدودیت‌های پژوهش می‌توان به استفاده از روش غیرنظارتی برای خوشه‌بندی نظرات در روش پژوهش اشاره کرد. روش خوشه‌بندی دارای مزایایی از قبیل سرعت بالا و بی‌نیازی به مداخله انسانی است؛ ولی همین مزیت ممکن است به کم شدن کیفیت نتایج منجر شود. برای حل این موضوع، به پژوهشگران پیشنهاد می‌شود از طریق تلفیق دو روش بانظارت و بی‌نظارت به پردازش متن پردازند و نتایج را مقایسه کنند. همچنین، محدودیت جامعه آماری به دلیل کمبود نسبی نظرات فارسی در اینترنت احتمال خطا در نمونه‌گیری را به وجود می‌آورد. برای حل این مشکل پیشنهاد می‌شود از نظرات خارجی هم‌زمان با نظرات داخلی به عنوان نمونه آماری استفاده شود و با مقایسه نتایج از سوگیری احتمالی نمونه پیش‌گیری شود. علاوه بر موارد ذکر شده، نبود منابع کافی و پژوهش‌های داخلی درباره روش و پیشینه پژوهش، از عوامل محدودکننده این پژوهش بوده‌اند.

منابع

۱. اسماعیلی پور، مهدی (۱۳۹۵)، آموزش گام به گام داده‌کاوی با ریپد ماینر، تهران: آتی‌نگر.
۲. آقایی، سپیده و اسماعیلی، سبحان (۱۳۹۶)، تأثیر فناوری کلان داده در سیستم‌های بازاریابی و تجزیه و تحلیل رفتار مصرف کننده، سومین کنفرانس بین‌المللی مدیریت و مهندسی صنایع، تهران، ایران.

همکاران (۲۰۱۷) و سانچز و همکاران (۲۰۱۹) با رویکرد پژوهش حاضر در یک راستا قرار دارند. پژوهش گیل ستو و همکاران (۲۰۱۹) نشان می‌دهد مشتریان درک کاملی از سطح تعهدات زیست محیطی هتل ندارند که این مورد در نتایج این پژوهش و تحلیل نظرات گردشگران ایرانی مشاهده نشد.

باتوجه به نتایج پژوهش، کارآفرینان حوزه گردشگری می‌توانند با استفاده از کلان داده‌ها اهمیت و میزان اعتبار ایده‌های جدید خود را بررسی کنند و در شناسایی و پیش‌بینی بازخورد احتمالی مشتریان از آن بهره‌گیرند. همچنین به کارآفرینان در حوزه گردشگری پیشنهاد می‌شود برای شناسایی نقاط ضعف موجود و ارائه خدمات جدید، از داده‌های شبکه‌های اجتماعی برای تأیید ایده‌ها، نگرش و تحلیل‌های خود استفاده کنند.

به مدیران کسب و کار در حوزه گردشگری پیشنهاد می‌شود در صورتی که توان و سرمایه لازم برای محقق کردن همه ارزش‌های موردانتظار گردشگران را ندارند، با بررسی نقاط قوت و ضعف خود، بر مهم‌ترین ابعاد ارزشی‌ای که توانایی توسعه آن وجود دارد، تمرکز کنند و برای خود در آن زمینه مزیت رقابتی ایجاد کنند. به عبارت دیگر، گوشه‌ای از بازار گردشگری را با تمرکز بر آن به دست آورند. باتوجه به اینکه در این پژوهش صنعت گردشگری تجزیه و تحلیل شد، به پژوهشگران آینده پیشنهاد می‌شود در شاخه‌های دیگر کسب و کار، مفهوم «ارزش» را مطالعه کنند. همچنین، باتوجه به پیشرفت سریع فناوری اطلاعات در حوزه هوش مصنوعی، پیشنهاد می‌شود با تلفیق روش‌های کلاسیک تحلیل و پیش‌بینی رفتار

¹. Data Science

². Deep learning

۳. امیرخانی، امیرحسین و متقی، وحید (۱۳۹۷). ارائه سیستم پیش‌بینی تولید محصول بر روی کلان‌داده با استفاده از یادگیری ماشین، *کنفرانس ملی تحقیقات نوین در مهندسی کامپیوتر برق فناوری اطلاعات*، مبارکه، اصفهان، ایران.
۴. انصاری، نسیم؛ وکیلی مفرد، حسین؛ منصوری‌زاده، محرم و امیری، محمدرضا (۱۳۹۸)، کشف الگوی دسترسی کاربران و تحلیل تراکنش‌های امانت منابع اطلاعاتی با استفاده از تکنیک قوانین انجمنی در داده‌کاوی (مطالعه موردی: کتابخانه‌ها و مراکز اطلاع‌رسانی دانشگاه علوم پزشکی همدان). *پژوهشنامه پردازش و مدیریت اطلاعات*، ۳۴(۳)، ۱۱۵۵-۱۱۸۶.
۵. خلیل‌نژاد، شهرام؛ سلیمان‌زاده، امید و کراقلی، مرضیه (۱۳۹۸)، تأثیر تجربه برند هتل بر ارزش ادراک‌شده گردشگران و قوت برند هتل، *گردشگری و توسعه*، ۱(۱)، ۱۰۲-۱۱۹.
۶. رحیم‌نیا، فریبرز؛ فاطمی، زهرا و هرنندی، عطاله (۱۳۹۲)، بررسی آثار متقابل ابعاد ارزش ویژه برند مبتنی بر مشتری (مورد مطالعه: هتل‌های پنج‌ستاره کلان‌شهر مشهد)، *مدیریت بازرگانی*، ۵(۴)، ۱-۲۰.
۷. رضاقلی‌زاده، مهدیه (۱۳۹۹)، نقش حمل و نقل هوایی، هتل‌ها و رستوران‌ها در رابطه بین گردشگری و رشد اقتصادی، *برنامه‌ریزی و توسعه گردشگری*، ۹(۳۲)، ۸۹-۱۰۶.
۸. سلطانی‌زنوزی، سیدمحمد (۱۳۹۸)، بررسی نقش کلان‌داده‌ها در بازاریابی و کسب‌وکار، دومین *کنفرانس ملی اندیشه‌های نوین در مدیریت کسب‌وکار*، تهران، ایران.
۹. شفیعی، ساناز؛ رجب‌زاده قطرمی، علی؛ حسن‌زاده، علیرضا و جهانیان، سعید (۱۳۹۶)، بررسی تأثیر فناوری اطلاعات بر توسعه پایدار مقاصد گردشگری به‌منظور توسعه مقاصد گردشگری هوشمند (با استفاده از رویکرد فراترکیب)، *تحقیقات بازاریابی نوین*، ۷(۴)، ۹۵-۱۱۶.
۱۰. صنایعی، علی (۱۳۹۶)، *انقلاب صنعتی چهارم*، اصفهان: جهاد دانشگاهی دانشگاه اصفهان.
۱۱. غلامی، ناصر؛ آقایی، نجف؛ محمدکاظمی، رضا و صفاری، مرجان (۱۳۹۸)، ارزش پیشنهادی به مشتری در مدل کسب‌وکارهای ورزشی، *مطالعات مدیریت ورزشی*، ۱۱(۵۳)، ۸۳-۹۸.
۱۲. فاضلی‌راد، محمدعلی و پویا، علیرضا (۱۳۹۵)، خوشه‌بندی فروشگاه‌های آنلاین از نگاه تأمین‌کننده با کمک بهینه‌یابی تعداد خوشه‌ها در الگوریتم دو مرحله‌ای SOM، *مطالعات مدیریت صنعتی*، ۱۴(۴۳)، ۱۰۹-۱۳۴.
۱۳. فقیه، شقایق و اسدی، سعید (۱۳۹۷)، بررسی مزایای کلان‌داده در حوزه بازاریابی، سومین *کنفرانس ملی فناوری در مهندسی برق و کامپیوتر*، سمنان، ایران.
۱۴. مبینی‌دهکردی، علی؛ رضوانی، مهران؛ داوری، علی؛ فروزان، فاطمه. (۱۳۹۳)، مدل کسب‌وکار نوآورانه B2C برای شرکت‌های پخش، *فصلنامه توسعه کارآفرینی*، ۷(۳)، ۵۶۹-۵۸۸.
۱۵. محمدیان، محمود و نائلی، مریم (۱۳۹۶)، بررسی تحلیلی مفهوم کلان‌داده در حوزه بازاریابی نوین، *کنفرانس بین‌المللی پژوهش‌های نوین در مدیریت، اقتصاد و توانمندی صنعت جهانگردی در توسعه*، مشهد، ایران.
۱۶. موذن رضامحله، محمدحسین (۱۳۹۷)، بازاریابی مؤثر شرکت‌های تجاری بزرگ با استفاده از مدیریت دانش و کلان‌داده‌ها از شبکه‌های

- Management*, 59, 467-483.
25. Gupta, S., Keen, M., Shah, A., Verdier, G., & Walutowy, M. F. (Eds.). (2017), *Digital revolutions in public finance*, Washington, DC: International Monetary Fund.
 26. Halawani, F. M., Soh, P. C., & Muthaiyah, S. (2019), The effect of social media on hotels' business performance in the Lebanese hotel sector: Effect of social media on hotels' business performance, *Journal of Electronic Commerce in Organizations (JECO)*, 17(3), 54-70.
 27. Han, J., Pei, J., & Kamber, M. (2011), *Data mining: concepts and techniques*, Elsevier.
 28. Hastie, T., Tibshirani, R., & Friedman, J. (2017), *The elements of statistical learning: data mining, inference, and prediction*, Springer Science & Business Media.
 29. James, G., Witten, D., Hastie, T., & Tibshirani, R. (2017), *An introduction to statistical learning*, New York: springer.
 30. Kans, M., & Ingwald, A. (2016), Business model development towards service management 4.0, *Procedia CIRP*, 47, 489-494.
 31. Kassambara, A. (2017), *Practical guide to cluster analysis in R: Unsupervised machine learning* (Vol. 1), STHDA.
 32. La, K. V., & Kandampully, J. (2004), Market oriented learning and customer value enhancement through service recovery management, *Managing Service Quality: An International Journal*, 14(5), 390-401.
 33. Labrecque, L. I. (2014), Fostering consumer-brand relationships in social media environments: The role of parasocial interaction, *Journal of Interactive Marketing*, 28(2), 134-148.
 34. Lansley, G., & Longley, P. (2016), Deriving age and gender from forenames for consumer analytics, *Journal of Retailing and Consumer Services*, 30, 271-278.
 35. Lee, E. B., Kim, J., & Lee, S. G. (2017), Predicting customer churn in mobile industry using data mining technology, *Industrial Management & Data Systems*. 117(1), 90-109.
 36. Lee, J., Kao, H. A., & Yang, S. (2014), اجتماعی، دومین کنفرانس ملی نقش مدیریت در چشم‌انداز ۱۴۰۴، رشت، ایران.
 ۱۷. ناصر صدرآبادی، علیرضا؛ الماسی سروستانی، راضیه و قبادی، الهام (۱۳۹۷)، مطالعه رفتار مشتریان در رزرواسیون آنلاین با استفاده از تکنیک داده کاوی و مدل کانو (مورد مطالعه: هتل‌های چهار و پنج ستاره شهر تهران)، مطالعات مدیریت گردشگری، ۱۳(۴۴)، ۱۹۲-۲۱۷.
 18. Amado, A., Cortez, P., Rita, P., & Moro, S. (2018), Research trends on Big Data in marketing: A text mining and topic modeling based literature analysis, *European Research on Management and Business Economics*, 24(1), 1-7.
 19. Bendle, N. T., & Wang, X. S. (2016), Uncovering the message from the mess of big data. *Business Horizons*, 59(1), 115-124.
 20. Chong, A. Y. L., Khong, K. W., Ma, T., McCabe, S., & Wang, Y. (2018), Analyzing key influences of tourists' acceptance of online reviews in travel decisions. *Internet Research*, 28(3), 564-585
 21. Duan, W., Cao, Q., Yu, Y., & Levy, S. (2013), Mining online user-generated content: using sentiment analysis technique to study hotel service quality. In *2013 46th Hawaii International Conference on System Sciences* (pp. 3119-3128). IEEE.
 22. Elgendy N., Elragal A. (2014), Big data analytics: A literature review paper. In: Perner P. (eds) *Advances in Data Mining. Applications and Theoretical Aspects*. ICDM 2014, Lecture Notes in Computer Science, vol 8557. Springer, Cham.
 23. Gil-Soto, E., Armas-Cruz, Y., Morini-Marrero, S., & Ramos-Henríquez, J. M. (2019), Hotel guests' perceptions of environmental friendly practices in social media, *International Journal of Hospitality Management*, 78, 59-67.
 24. Guo, Y., Barnes, S. J., & Jia, Q. (2017), Mining meaning from online ratings and reviews: Tourist satisfaction analysis using latent dirichlet allocation, *Tourism*

- (2015), Strategic value curve analysis: Diagnosing and improving customer value propositions, *Business Horizons*, 58(3), 317-24.
47. Sparks, B. A., So, K. K. F., & Bradley, G. L. (2016), Responding to negative online reviews: The effects of hotel responses on customer inferences of trust and concern, *Tourism Management*, 53, 74-85.
 48. Syam, N., & Sharma, A. (2018), Waiting for a sales renaissance in the fourth industrial revolution: Machine learning and artificial intelligence in sales research and practice, *Industrial Marketing Management*, 69, 135-146.
 49. Weiss, S. M., Indurkha, N., & Zhang, T. (2015), *Fundamentals of predictive text mining*, Springer.
 50. Xu, M., David, J. M., & Kim, S. H. (2018), The fourth industrial revolution: opportunities and challenges, *International Journal of Financial Research*, 9(2), 90-95.
 51. Yee Liao, B., & Pei Tan, P. (2014), Gaining customer knowledge in low cost airlines through text mining, *Industrial Management & Data Systems*, 114(9), 1344-1359.
 52. Zhang, T. C., Gu, H., & Jahromi, M. F. (2019), What makes the sharing economy successful? An empirical examination of competitive customer value propositions, *Computers in Human Behavior*, 95, 275-283.
 - Service innovation and smart analytics for industry 4.0 and big data environment, *Procedia Cirp*, 16(1), 3-8.
 37. Leong, L. Y., Hew, T. S., Ooi, K. B., & Lin, B. (2019), Do electronic word-of-mouth and elaboration likelihood model influence hotel booking? *Journal of Computer Information Systems*, 59(2), 146-160.
 38. Marr, B. (2016), Why everyone must get ready for the 4th industrial revolution, <https://www.forbes.com/sites/bernardmarr/2016/04/05/why-everyone-must-get-ready-for-4th-industrial-revolution/#26be9e2f3f90>.
 39. Moro, S., Rita, P., & Vala, B. (2016), Predicting social media performance metrics and evaluation of the impact on brand building: A data mining approach, *Journal of Business Research*, 69(9), 3341-3351.
 40. Payne, A., Frow, P., & Eggert, A. (2017), The customer value proposition: Evolution, development, and application in marketing, *Journal of the Academy of Marketing Science*, 45(4), 467-489.
 41. Raguseo, E., Neirotti, P., & Paolucci, E. (2017), How small hotels can drive value their way in infomediation, The case of 'Italian hotels vs. OTAs and Trip Advisor', *Information & Management*, 54(6), 745-756.
 42. Raschka, S., & Mirjalili, V. (2019), *Python Machine Learning: Machine Learning and Deep Learning with Python, scikit-learn, and TensorFlow 2*. Packt Publishing Ltd.
 43. Sánchez-Franco, M. J., Navarro-García, A., & Rondán-Cataluña, F. J. (2019), A naive Bayes strategy for classifying customer satisfaction: A study based on online reviews of hospitality services, *Journal of Business Research*, 101, 499-506.
 44. Schwab, K. (2015), The Fourth Industrial Revolution, What It Means and How to Respond? SNAPSHOT, December 12.
 45. Shalini, S., & Lal, K. (2016, April), Improved pseudo-association rules technique, In *2016 International Conference on Computing, Communication and Automation (ICCCA)* (pp. 890-895), IEEE.
 46. Sheehan, N. T., & Bruni-Bossio, V.