



## پیش‌بینی تعداد گردشگران براساس رکوردهای اطلاعاتی گوگل ترندز با روش یادگیری ماشینی (مورد مطالعه: گردشگران شهر یزد)

حامد فلاح تفتی<sup>۱</sup>

### چکیده

با توسعه صنعت گردشگری و رشد کسب‌وکارهای مرتبط، کسب اطلاعات به‌روز در برنامه‌ریزی صحیح و برآورد دقیق تعداد گردشگران، به‌منظور به‌کارگیری کارآمد منابع با هدف توسعه زیرساخت‌ها و افزایش درآمد، ضروری است. وجود برنامه‌های دقیق در نهایت به ارتقای سطح رضایت گردشگران ورودی منجر می‌شود. با توسعه فرهنگ جست‌وجوی اطلاعات، گردشگران معمولاً پیش از آغاز سفر، به جست‌وجوی اطلاعات مربوط به اقامتگاه‌ها و خدمات گردشگری موجود در مقصد، از طریق منابع اینترنتی، اقدام می‌کنند. در پژوهش حاضر، با استفاده از داده‌های منتخب مربوط به پرس‌وجوهای کاربران سراسر جهان در موتور جست‌وجوی گوگل در مورد امکانات و توانمندی‌های گردشگری شهر یزد، تعداد گردشگران آتی این شهر پیش‌بینی شده است. داده‌های پژوهش را آمارهای جست‌وجوی کاربران تشکیل می‌دهد که از سامانه گوگل ترندز پایین‌گذاری شده و با استفاده از روش یادگیری ماشینی مدل پیش‌بینی، طراحی و اعتبارسنجی شده است. پس از آماده‌سازی و تحلیل داده‌ها، مشخص شد که پرس‌وجوهای ثبت‌شده در گوگل ترندز، قدرت فراوانی (بیش از ۹۵ درصد) در پیش‌بینی تعداد گردشگران شهر یزد در بازه زمانی سال ۲۰۱۴ تا ۲۰۱۹ در مقاطع ماهانه دارد.

تاریخ دریافت: ۱۳۹۸/۱۱/۰۶

تاریخ پذیرش: ۱۳۹۹/۰۸/۱۵

### واژه‌های کلیدی:

پیش‌بینی تعداد گردشگران، تقاضای گردشگری، گوگل ترندز، یادگیری ماشینی، شهر یزد

### مقدمه

موردنیاز در توسعه این صنعت برآمده‌اند؛ از این‌رو دسترسی به پیش‌بینی دقیق از تعداد گردشگران ورودی به هر منطقه، با در نظر گرفتن روند توزیع فصلی آن‌ها، اهمیت بالایی در برنامه‌ریزی صحیح متولیان دارد و موجب توسعه کارآفرینی در حوزه گردشگری و به دنبال آن، اشتغال‌زایی مستقیم و غیرمستقیم متناسب با حجم صنعت گردشگری منطقه می‌شود.

لازمه پیش‌بینی صحیح هر متغیر، دسترسی به داده‌های به‌روز است؛ البته گردآوری داده‌های آماری مربوط به حوزه گردشگری به علت تعدد بازیگران دولتی و خصوصی، وجود مراکز ارائه خدمات متعدد، زمان‌بر و پرهزینه است. بررسی‌ها نشان می‌دهد که برای دستیابی به داده‌های گردشگری، میان‌رخداد تا دستیابی به آمار، تأخیر حداقل دوماهه وجود دارد که بیشتر به فرایندهای گردآوری، تحلیل و آماده‌سازی داده‌ها برمی‌گردد (Onder, 2017).

پیش‌بینی تعداد گردشگران ورودی به مقصدهای گردشگری، مسئله مهمی در حوزه‌های گوناگون مدیریت گردشگری است. آژانس‌های مسافرتی، اقامتگاه‌ها و هتل‌ها و در مجموع نهادهای متولی دولتی و خصوصی در صورتی که برآورد صحیحی از تعداد گردشگران ورودی به مقصد در بازه‌های زمانی پیش‌رو داشته باشند، قادر خواهند بود برنامه‌ریزی مناسب، و خدمات مطلوب اقامتی و تفریحی را در اختیار بازدیدکنندگان قرار داده، سطح درآمدزایی خود را افزایش دهند. علاوه بر این، داشتن برآورد صحیح از تقاضای گردشگری موجب برنامه‌ریزی صحیح‌تر و در نهایت ارتقای رضایت گردشگران خواهد شد (Zhang & Song, 2018). در سال‌های اخیر، با توجه به فرصت‌های اقتصادی پدیدآمده در صنعت گردشگری، بسیاری از نواحی شهری و روستایی درصد توجه به جاذبه‌های گردشگری خود و سرمایه‌گذاری در خلق و گسترش زیرساخت‌های

۱. استاد پار گروه مدیریت، دانشکده علوم انسانی، دانشگاه علم و هنر، یزد، ایران؛ h.fallah@sau.ac.ir



به‌کارگیری روش‌های تحلیل داده‌های حجیم<sup>۱</sup> برای داده‌هایی که روند زمانی طولانی‌تری با الگوهای فصلی و رفتاری پیش‌بینی‌پذیر دارند، کارآمد است (Xu et al., 2016). می‌توان با استفاده از داده‌های تاریخی، به طراحی مدل‌های پیش‌بینی تقاضا پرداخت و پس از اطمینان از کارکرد مطلوب آن برای دوره‌های گذشته، با هدف عملیاتی‌کردن آن‌ها به‌منظور پیش‌بینی روند تقاضا برای دوره‌های آتی اقدام کرد. روش تحلیل داده‌های حجیم شامل گردآوری، آماده‌سازی و پردازش مجموعه داده‌های بسیار حجیم (بر مبنای گیگابایت یا ترابایت) است که به دنبال آن تحلیل‌های پیچیده‌ای انجام می‌شود که نیازمند فناوری‌های نوآورانه و پیشرفته برای ذخیره‌سازی، مدیریت، تحلیل و نمایش اطلاعات در قالب نمودارهاست (Chen et al., 2012). برای مثال، تحلیل اطلاعات مشتریان بانک‌ها و خدمات الکترونیکی پولی، داده‌هایی که وبسایت‌های اینترنتی با کاربران زیاد مانند گوگل<sup>۲</sup>، ویکی‌پدیا<sup>۳</sup> و تریپ‌ادوایزر<sup>۴</sup> ثبت کرده‌اند نمونه‌هایی از تحلیل داده‌های حجیم‌اند که پردازش آن‌ها نیازمند روش‌ها و فناوری‌های نوین و قدرتمند است.

اینترنت به‌منزله مأخذی برای دستیابی به اطلاعات در مورد مقصدهای گردشگری، اقامتگاه‌ها و وسایل حمل‌ونقل به جزئی جدایی‌ناپذیر از فرایندهای برنامه‌ریزی سفر تبدیل شده است. رفتار جست‌وجوی اطلاعات<sup>۵</sup> در اینترنت به‌دست گردشگران، موجب شده است موتورهای جست‌وجو منابع ارزشمندی برای دستیابی به اطلاعات مرتبط با نیاز کاربران در خصوص مقصدهای گردشگری باشد. موتور جست‌وجوی گوگل، در حکم مشهورترین جست‌وجوگر تحت وب، را می‌توان به‌مثابه اصلی‌ترین مأخذ دستیابی به اطلاعات لازم در مورد سفر دانست؛ به‌نحوی که این شرکت حدود ۶۷ درصد سهم بازار را از میان رقبای خود دارد (Dergiades et al., 2018). شرکت گوگل با علم به این که افراد نیازهای خود را در سامانه موتور جست‌وجوی این شرکت دنبال می‌کنند، تلاش کرده است که داده‌های کاربران را با اهداف تبلیغاتی و شخصی‌سازی نتایج و همچنین افزایش سطح کارایی خود ذخیره کند. سال‌هاست که این داده‌ها در سامانه گوگل ترندز<sup>۶</sup> به‌صورت خلاصه‌شده و به‌صورت رایگان در اختیار پژوهشگران قرار گرفته است تا با استفاده از آن‌ها

مطالعات خود را پیش‌برند. از طریق گوگل ترندز می‌توان به تفکیک مناطق جغرافیایی، به بررسی داده‌ها و روند جست‌وجوی کاربران در مورد کلیدواژه‌های مدنظر آن‌ها اقدام کرد. این داده‌ها، که به تفکیک بازه‌های زمانی (از سال ۲۰۰۴)، تفکیک موضوعی و تفکیک جغرافیایی کاربر جست‌وجوگر در اختیارمان است، با دقت بالا خواسته‌ها و نیازهای اطلاعاتی کاربران را منعکس می‌کند. این داده‌ها با منطقه جغرافیایی انتخاب‌شده، به‌صورت ماهانه یا هفتگی به شکل رایگان قابلیت پایین‌گذاری از طریق وبسایت گوگل ترندز را دارند.

فرض ابتدایی در اجرای این پژوهش این است که مسافران بالقوه در مورد ویژگی‌ها و امکانات مقصد مدنظر، پیش از سفر، در گوگل جست‌وجو می‌کنند؛ بنابراین شاخص گوگل ترندز، به‌منزله متغیری در پیش‌بینی تقاضای گردشگری آتی، در هر مقصدی می‌تواند استفاده شود. هرچند معمولاً میان جست‌وجوی کاربران و بازدید از مقصد مدنظر، به‌منظور آماده‌سازی سفر فاصله می‌افتد، اما توان پیش‌بینی دقیق تعداد گردشگران مقصد با استفاده از آمار جست‌وجوی واژگان، قدرت این متغیرها را در دقت برآوردها نشان می‌دهد.

نوآوری و هدف مطالعه حاضر، پیش‌بینی تعداد گردشگران بازدیدکننده از مقصدهای گردشگری شهری در قالب آمار شب اقامت است. به این منظور، می‌توان از داده‌های مستخرج از سامانه گوگل ترندز، به‌منزله منبع قدرتمند، که نیازهای اطلاعاتی کاربران را در دوره‌های گذشته نشان می‌دهد، استفاده کرد.

به این منظور و با هدف پیش‌بینی تقاضای گردشگری مقصد، علاوه بر استفاده از روش‌های غیرعلی برای پیش‌بینی تعداد گردشگران - که مبتنی بر داده‌های آماری گذشته است - از روش‌های علی براساس داده‌های آماری دسته‌بندی‌شده مستخرج از گوگل ترندز، که علاقه‌مندی و جست‌وجوی واژگان خاص در مورد مقصد گردشگری را نشان می‌دهد، استفاده شده است؛ بنابراین براساس پرس‌وجوهای کاربران در خصوص مواردی مانند جاذبه‌های گردشگری مقصد، امکانات و مراکز اقامتی و تفریحی، وسایل ارتباطی و حمل‌ونقل، آژانس‌های مسافرتی، و وجود تورهای گردشگری اختصاصی انجام داده‌اند، مبنای پیش‌بینی تعداد گردشگران خارجی به مقصد تحلیل شده‌اند. همچنین با توجه به روند فصلی گردشگری در مقصدها (که عمدتاً به دلایل آب‌وهوایی و یا تعطیلات اتفاق می‌افتد)، برای حذف داده‌های پرت و ناکارآمد در فرایند پیش‌بینی هموارسازی داده‌ها انجام شده است.

1. Big Data Analysis
2. Google
3. Wiki Pedia
4. Trip Advisor
5. Information Seeking Behavior
6. Google Trends



در تحلیل سری‌های زمانی، معمولاً از سری‌های زمانی سالانه در مطالعات پیش‌بینی تقاضای گردشگری استفاده می‌شود. براساس پژوهشی که سانگ و لی (۲۰۰۸) انجام داده‌اند، از ۱۲۱ مطالعه منتشر شده درباره پیش‌بینی تقاضای گردشگری، فقط ۳۰ مورد از آن‌ها از سری‌های زمانی ماهانه استفاده کرده‌اند. مزایای استفاده از داده‌های ماهانه به جای داده‌های سالانه، دسترسی به مشاهدات بیشتر برای دوره مشابه است، که برای برآورد و ارزیابی پیش‌بینی با دقت بالاتر کاربرد دارند (Gunter & Onder, 2015). علاوه بر این، به کارگیری داده‌های ماهانه برای بررسی تقاضای گردشگری به منظور شناسایی فصلی بودن بازدید از مقصدها حائز اهمیت است.

همچنین در زمینه پیش‌بینی تقاضای گردشگری در سطح شهرها مطالعاتی انجام شده است؛ برای مثال، میان سال‌های ۲۰۰۰ تا ۲۰۰۷، وو و ترنر (۲۰۰۶)، سانگ و لی (۲۰۰۸)، اسمرال (۲۰۱۴) و گانتر و اوندر (۲۰۱۵)، با استفاده از داده‌های ماهانه، مطالعه‌ای در سطح مقصدهای گردشگری شهری انجام داده‌اند.

در گذشته در برخی پژوهش‌ها با در نظر گرفتن متغیرهای مستقل در زمان گذشته، در حکم پیش‌بینی کننده و تأثیر آن‌ها در مقدار متغیر وابسته در آینده، به علت تأخیر موجود در تأثیرگذاری آن‌ها به روش مدل سازی اشاره شده و مفروضات جدیدی در خصوص مسئله پیش‌بینی دقیق تعداد گردشگران آینده، ارائه شده است (Onder, 2017). این رویکرد شامل خلق مدل‌های معمول پیش‌بینی تقاضاست که با در نظر گرفتن تعدادی متغیر پیش‌بینی کننده (مستقل) - که مقدار آن‌ها در دوره‌های  $t-1, t-2, \dots, t-n$  اندازه‌گیری و تحلیل شده و به منزله مفسر مقدار متغیر وابسته در دوره  $t+1, t+2, \dots, t+n$  استفاده شده‌اند (Croes & Venegas, 2005).

### داده‌های گوگل ترندز

گوگل ترندز بایگانی واژه‌های جست‌وجوشده به دست کاربران سراسر جهان است که آن‌ها را در خوشه‌های متفاوتی از جمله حوزه‌های کلانی مانند سفر، هنر و سرگرمی، بازی‌های رایانه‌ای، سلامت، علائق و لذت‌ها، و مواردی مشابه قرار می‌دهد. خوشه مربوط به سفر، شامل واژه‌های جست‌وجوشده در مورد مقصدهای گردشگری و جاذبه‌های آن مکان‌هاست؛ بنابراین استخراج آمارهای جست‌وجو در خوشه مربوط به گردشگری شامل عبارت‌هایی است که در دسته‌بندی‌های جاذبه‌های گردشگری مقصد گردشگری، پرس‌وجوهای مربوط به کمیّت و کیفیت مراکز اقامتی، شرایط و وسایل در دسترس

علت انتخاب شهر یزد برای پیش‌بینی تقاضای گردشگری در این پژوهش، قرارگیری این شهر در حلقه گردشگری فرهنگی - تاریخی اصلی کشور ایران است که سالانه بخش زیادی از گردشگران خارجی ورودی کشور از این شهر و جاذبه‌های آن بازدید می‌کنند (Bagheri et al., 2018)، همچنین ثبت شهر یزد در فهرست میراث جهانی سازمان یونسکو در سال ۲۰۱۷ (Farashah et al., 2018)، باعث شده است که توجه گردشگران بسیاری از سراسر جهان به این شهر جلب شود. همچنین به علت فراهم بودن داده‌های رسمی آمار شب اقامت گردشگران خارجی در بازه زمانی ابتدای سال ۱۳۸۹ تا پایان سال ۱۳۹۷، از نظر تحلیل داده‌ها، پیش‌بینی تقاضای گردشگری شهر یزد میسر بوده است.

با توجه به این که گردشگران بالقوه در همان زمانی که اقدام به جست‌وجوی اطلاعات گردشگری در مورد یک مقصد می‌کنند نمی‌توانند سفر خود را آغاز کنند؛ تأخیری چندماهه میان جست‌وجو تا اقامت در مقصد گردشگری متصور است. به این منظور، با فرایند پنجره‌سازی داده‌ها<sup>۱</sup> در آماده‌سازی و پالایش داده‌ها، تلاش شده است تا مجموع جست‌وجوهای انجام شده در چند دوره گذشته (t-n)، به منزله پیش‌بینی کننده‌های مناسب برای تقاضای گردشگری در زمان حاضر (t)، استفاده شود.

### مبانی نظری و پیشینه تحقیق

#### پیش‌بینی تقاضای گردشگری

اکثر مطالعات انجام شده در پیش‌بینی تعداد گردشگران از روش‌های کمی استفاده می‌کنند، که عمدتاً شامل روش‌های غیرعلی (پیش‌بینی تک‌متغیره) مانند تحلیل سری‌های زمانی هستند. همچنین بررسی ادبیات پژوهش نشان می‌دهد تلاش‌هایی برای استفاده از روش‌های علی (پیش‌بینی چندمتغیره) در پیش‌بینی تقاضای گردشگران انجام شده است (مانند مدل‌های اقتصادسنجی<sup>۲</sup>).

در پژوهش‌های متعدد، توجه به روند خط‌های فصلی و روش‌های هموارسازی نمایی به خوبی در پیش‌بینی تقاضای گردشگری به کار گرفته شده و دقت آن‌ها تأیید شده‌اند (Athanasopoulos et al., 2011). اما روش‌های اتورگرسیو<sup>۳</sup> به روند مقدار یک متغیر در گذشته توجه دارد و همان را مبنای پیش‌بینی قرار می‌دهد و تأکید زیادی بر رویدادهای گذشته دارد. از آنجاکه همیشه گذشته نمی‌تواند آینه‌ای از اتفاقات آینده باشد، نیاز به رویکردهایی است که بتوان با استفاده از سایر متغیرها و شاخص‌ها، تغییرات متغیر وابسته را پیش‌بینی کرد.

1. Data Windowing  
2. Econometric  
3. Autoregressive



برای حمل و نقل (شامل پرواز، اتوبوس و قطار)، غذاهای محلی و سرگرمی‌های ویژه مقصد مدنظر و دیگر موضوعات مربوط به حوزه گردشگری است (google, 2014). داده‌های گوگل ترندز در پژوهش‌های متعددی در گذشته استفاده شده‌اند که برای مثال می‌توان به پیش‌بینی نرخ بی‌کاری در ایالات متحده (Ettredge et al., 2005) و شناسایی میزان همه‌گیری بیماری آنفولانزا (Doomik, 2009) براساس پرس‌وجوهای کاربران گوگل اشاره کرد.

در حوزه گردشگری و مهمان‌نوازی، با استفاده از داده‌های جست‌وجو در وب (همانند گوگل ترندز) مطالعات متعددی انجام شده است که قدرت این داده‌ها را در پیش‌بینی میزان تقاضای گردشگری نشان می‌دهد. برای مثال گاولیک و همکاران (۲۰۱۱) از داده‌های گردآوری شده از موتورهای جست‌وجو برای پیش‌بینی تعداد بازدیدکنندگان از مقصدهای گردشگری استفاده کردند. آن‌ها متوجه شدند که داده‌های مربوط به جست‌وجوهای گذشته کاربران، می‌تواند در پیش‌بینی تقاضای آتی گردشگری نقش کاربردی و اثربخشی ایفا کند.

یانگ و همکاران (۲۰۱۴) متوجه شدند که داده‌های مربوط به ترافیک صفحات وب به‌منظور پیش‌بینی تقاضای هتل‌ها در آینده، دقت و قدرت مناسبی برای استفاده دارند. در پژوهشی دیگر، داده‌های گوگل ترندز درباره هتل‌های جزایر کارائیب و پرواز به این مقصد گردشگری، با هدف پیش‌بینی تعداد گردشگران آن محل، استفاده شد و محققان به مفیدبودن این داده‌ها در پیش‌بینی‌ها اشاره کردند (Bangwayo-Skeete & Skeete, 2015).

یانگ و همکاران (۲۰۱۵) قدرت پیش‌بینی داده‌های وبگاه گوگل ترندز و موتور جست‌وجوی بایدو (موتور جست‌وجوی چینی که بیشتر کاربران این کشور استفاده می‌کنند) را برای پیش‌بینی تعداد گردشگران کشور چین مقایسه کردند و متوجه شدند رکوردهای اطلاعاتی موتور جست‌وجوی بایدو توان بیشتری در پیش‌بینی تعداد گردشگران مورد انتظار در این کشور دارد که به علت دسترسی و شهرت بیشتر این موتور جست‌وجو در کشور چین بوده است.

گانتر و اوندرا (۲۰۱۶) در پژوهش خود برای پیش‌بینی تعداد گردشگران در کشورها و شهرهای اروپایی با استفاده از داده‌های گوگل ترندز به این نتیجه رسیدند که این داده‌ها دقت بالاتری به نسبت الگوهای مرسوم غیرعلی

(پیش‌بینی متغیری صرفاً مبتنی بر داده‌های گذشته همان متغیر) برای پیش‌بینی دارند. آن‌ها به این منظور، با استفاده از نتایج پرسش‌جوی کاربران از سال ۲۰۱۴، شاخص‌های قرار گرفته در دسته گردشگری گوگل ترندز را استخراج کرده، برای یادگیری ماشینی آمارهای اقامت گردشگران از اداره آمار اتحادیه اروپا استفاده کردند. پس از طراحی مدل یادگیری، براساس داده‌های محفوظ نگاه‌داشته شده برای آزمون مدل به پیش‌بینی اقدام کردند. همچنین پژوهش‌های دیگری نیز در خصوص پیش‌بینی تعداد گردشگران ورودی و خروجی با استفاده از داده‌های جست‌وجوی اینترنتی براساس گوگل ترندز صورت گرفته است که می‌توان به پژوهش‌های بانگواو -اسکیت و اسکیت (۲۰۱۵) و گانتر و اوندرا (۲۰۱۶) اشاره کرد.

حال‌بینی<sup>۲</sup> (پیش‌بینی مقادیر نزدیک به رویدادهای کنونی) به اندازه پیش‌بینی مقدار یک متغیر در آینده اهمیت دارد؛ زیرا پتانسیل ارائه علائم هشدار اولیه در خصوص تغییرات را دارد و امکان ارزیابی فوری وضعیت فعلی را فراهم می‌آورد (Castle et al., 2019). چوی و واریان (۲۰۱۲) نشان دادند که با داده‌های گوگل ترندز می‌توان حال‌بینی را برای میزان فروش خودرو، برنامه‌های مسافرتی، نرخ بی‌کاری و اعتماد مصرف‌کنندگان به برند امکان‌پذیر ساخت. نتایج مطالعه آن‌ها نشان می‌دهد که پرس‌وجوهای ثبت شده در موتورهای جست‌وجو می‌تواند به‌منظور پیش‌بینی میزان خریدهای مصرفی برای محصولات یا خدماتی که برای خرید آن‌ها، مصرف‌کنندگان قبل از تصمیم‌نهایی جست‌وجو و برنامه‌ریزی می‌کنند، استفاده کرد.

مطالعه‌ای که جونز و همکاران (۲۰۰۸) بر روی پرس‌وجوهای اینترنتی مرتبط با اماکن جغرافیایی انجام دادند نشان داد که ۸۴ درصد از جست‌وجوها شامل نام شهرهاست؛ در حالی که ۱۴ درصد نام کشور را شامل می‌شوند و بقیه شامل نام‌های استانی و منطقه‌ای می‌شوند. همچنین پرس‌وجوهای مربوط به سفرها (نزدیک به ۹۱ درصد) کمتر از پنج کلیدواژه را شامل می‌شوند (Xiang & Pan, 2005). همچنین جست‌وجوگران وب اغلب جست‌وجوهای خود را با افزودن، حذف کردن، ویرایش یا بازنویسی عبارت مدنظر انجام می‌دهند؛ بنابراین دستاورد پژوهشی آن‌ها مبنای مناسبی برای تعیین ماهیت متغیرها و دسته‌بندی آن‌ها به شکل کاربردی در اختیار سایرین قرار دارد؛ برای مثال، مسافری بالقوه، که با کشور ایران ناآشناست و



این سامانه به صورت خودکار اهمیت عبارت‌های جست‌وجوشده درمورد موضوع مدنظر را در بازه درصدی نشان می‌دهد. این مقادیر، که شاخص گوگل ۱ نام دارند، در بازه ۰ تا ۱۰۰ درصد تعیین شده و قابلیت پایین‌گذاری به دست پژوهشگران را دارد. به علت گستردگی بازه انتخاب‌شده، که از آغاز سال ۲۰۱۴ تا پایان فوریه ۲۰۱۹ را شامل می‌شود (در حدود شش سال میلادی)، داده‌ها به صورت بازه‌های ماهانه استخراج شدند. هرچه شاخص گوگل متغیر مدنظر درصد بالاتری داشته باشد، بیانگر اهمیت بیشتر آن کلیدواژه در آن ماه برای کاربران است. همچنین به علت این‌که پژوهش حاضر به دنبال پیش‌بینی تقاضای سفر گردشگران خارجی است، جست‌وجوهای انجام‌شده در سراسر جهان در هر کلیدواژه پایین‌گذاری شدند.

جدول ۱ کلیدواژه‌های منتخب از سامانه گوگل ترندز را برای پیش‌بینی تعداد گردشگران، که کاربران سراسر جهان در بازه‌های ماهانه انجام داده‌اند، نشان می‌دهد. بررسی کیفیت این داده‌ها نشان می‌دهد که به علت ناموجود بودن یا مرسوم نبودن برخی از خدمات گردشگری در شهر یزد، مانند به اشتراک گذاشتن خودرو یا سفر اشتراکی، غذاهای محلی و سرگرمی‌های محلی جست‌وجوی این کلیدواژه‌ها نتایج مناسبی را به کاربران باز نگردانده و بنابراین گوگل ترندز اهمیت کمی (در حد صفر درصد) برای برخی از بازه‌های زمانی برای آن‌ها در نظر گرفته است. همچنین به علت تعداد زیاد داده‌های از دست‌رفته در برخی از کلیدواژه‌ها، در پالایش ابتدایی داده‌ها این دسته از کلیدواژه‌ها کنار گذاشته شده و برای تحلیل استفاده نشدند.

گفتنی است مبنای انتخاب متغیرهای ذکر شده در جدول ۱، میزان کافی داده‌ها در دوره‌های مبنای یادگیری مدل بوده است؛ به نحوی که امکان استفاده از متغیر پیش‌بینی‌کننده در محاسبات وجود داشته باشد. برای مثال جست‌وجوهای کاربران درمورد غذاهای محلی یک منطقه نیز می‌تواند عاملی برای انتخاب مقصد، به خصوص در گردشگری غذا باشد، که لازمه آن دسترسی به تعداد مناسب رکوردهای اطلاعاتی است.

می‌خواهد از یک شهر این کشور بازدید کند، ممکن است جست‌وجوی خود را با نام کشور از قبیل «سفر به ایران» آغاز کند. سپس می‌تواند براساس هدف و قصد سفر خود، عبارت جست‌وجوشده را دقیق‌تر کند و عبارتی مانند «هتل‌های یزد» را برای اجرای دوباره جست‌وجو اصلاح کند. بر این اساس، یکی از نتایج مدنظر این مطالعه این است که الگوهای پیش‌بینی با شاخص‌های جست‌وجوی نام شهرها دقت بالاتری در پیش‌بینی تعداد گردشگران دارد.

جنسن و همکاران (۲۰۰۰) نشان دادند ۲۲ درصد از کاربران وب به طور متوسط ۲/۸۴ پرس‌وجو در هر دور انجام می‌دهند. این بدان معنی است که کاربران پرس‌وجوی خود را براساس نتایج جست‌وجوی پیشین اصلاح می‌کنند و همین باعث می‌شود دقت این داده‌ها در دستیابی به اطلاعات مدنظر به دست کاربر افزایش یابد؛ از این رو می‌توان از کاربردی بودن داده‌های پرس‌وجوی کاربران موتورهای جست‌وجو برای بررسی شرایط بازار و پیش‌بینی متغیرهای مدنظر اطمینان کافی به دست آورد.

### روش تحقیق

با توجه به هدف پژوهش، که پیش‌بینی تقاضای سفر گردشگران خارجی به شهر یزد است، پژوهش حاضر از نوع کاربردی است. ماهیت پژوهش نیز به علت استفاده از داده‌های واقعی (متغیرهای مستقل و وابسته) از نوع توصیفی-پیمایشی و روش تحلیل داده‌ها از نوع کمی است. همچنین شیوه گردآوری داده‌ها با استفاده از آمارهای واقعی ثبت‌شده در سامانه گوگل ترندز برای متغیرهای مستقل، و معاونت آمار سازمان مدیریت و برنامه‌ریزی استان یزد است.

با توجه به تنوع بالای دسته‌بندی داده‌ها در گوگل ترندز، به منظور دستیابی به پیش‌بینی دقیق متغیر مدنظر، داده‌های مرتبط با حوزه گردشگری شهر یزد از این سامانه استخراج شد. کلیدواژه‌هایی که احتمال دارد با تصمیم نهایی در خصوص سفر مرتبط باشند، هریک در قالب متغیرهای کمی استخراج شدند. با توجه به این‌که داده‌های گوگل ترندز براساس شاخص‌هایی مانند اهمیت جست‌وجوی عبارت به دست کلیه کاربران و براساس زمان جست‌وجو و همچنین موقعیت جغرافیایی آن‌ها انجام می‌شود،



جدول ۱: کلیدواژه‌های استخراج‌شده از سامانه گوگل ترندز براساس جست‌وجوی کاربران از ژانویه ۲۰۱۴ تا پایان فوریه ۲۰۱۹

کلیدواژه لاتین	زمینه موضوعی انتخاب‌شده	انتخاب برای تحلیل
Yazd Travel	سفر به یزد	انتخاب
Iran Travel	سفر به ایران	انتخاب
Air Travel	سفر هوایی	انتخاب
Bus & Rail	قطار و اتوبوس	انتخاب
Hotel & Accommodations	هتل‌ها و مراکز اقامتی	انتخاب
Specialty Travel	سفرهای اختصاصی	انتخاب
Tourist Destinations	مقصد‌های گردشگری	انتخاب
Travel Guides & Travelogues	راهنمایان تور و گردشگری	عدم انتخاب
Car Rental & Taxi Services	کرایه خودرو و خدمات تاکسی	عدم انتخاب
Luggage & Travel Accessories	وسایل و لوازم سفر	عدم انتخاب
Carpooling & Ridesharing	اشتراک‌گذاری خودرو و راندگی	عدم انتخاب
Travel Agencies & Services	آژانس‌های مسافرتی و خدماتی	انتخاب

داده‌هاست (Kumar & Reinartz, 2018). به‌این‌منظور، ابتدا با در نظر گرفتن داده‌های از دست‌رفته (ناموجود) بازه زمانی مناسب برای نگه‌داشت داده‌ها تعیین شد. از این طریق، ۵۸ رکورد اطلاعاتی در بازه می ۲۰۱۴ تا فوریه ۲۰۱۹ برای اجرای روش یادگیری ماشین نگه داشته شد. همچنین تعداد متغیرهای مستقل برای پیش‌بینی تقاضای گردشگری شهر یزد در قالب هشت کلیدواژه منتخب از عبارات‌های جست‌وجوشده در این بازه زمانی از سایت گوگل ترندز بود. علاوه بر این، برای اطمینان از دقت مناسب این متغیرها در پیش‌بینی دقیق تعداد گردشگران، آمار واقعی شب اقامت گردشگران در هر ماه نیز به ماتریس داده‌ها اضافه و به‌منزله متغیری مستقل استفاده شد. این کار با هدف بررسی معناداری این متغیر مستقل در فرایند پیش‌بینی انجام شد.

برای اطمینان از وجود رابطه خطی میان متغیرها، تمامی داده‌های استخراج‌شده از گوگل ترندز و همچنین آمارهای شب اقامت گردشگران خارجی استان یزد، با استفاده از روش میانگین‌گیری متحرک و استفاده از لگاریتم طبیعی، تعدیل یافتند و نقاط حدی موجود در متغیرها اصلاح شدند. میانگین

در مجموع هشت کلیدواژه متفاوت - که بنابر نظر پژوهشگران بیشترین ارتباط را با فرایند تصمیم‌گیری گردشگران در انتخاب مقصد (شهر یزد) داشته و از طرفی حجم داده مناسبی از آن‌ها در اختیار بود - انتخاب شدند. برای اجرای روش یادگیری ماشین، به آمار دقیق گردشگران خارجی بر حسب شب اقامت در هتل‌ها و مراکز اقامتی شهر یزد نیاز است. در این پژوهش، داده‌های رسمی که سازمان میراث فرهنگی در بازه فروردین ۱۳۸۴ تا پایان اسفند ۱۳۹۷ ارائه کردند جمع‌آوری شد. با توجه به اینکه این آمار باید با داده‌های گوگل ترندز از نظر ماه مطابقت داشته باشد و همچنین با توجه به تعداد کمتر داده‌های در دسترس آمار گردشگران، بازه‌های زمانی بر حسب ماه‌های میلادی در نظر گرفته شد. با بررسی اولیه، با توجه به تغییر الگوریتم محاسبه در سامانه گوگل ترندز و نادقیق بودن آمارهای گردشگران خارجی شهر یزد در سال‌های ابتدایی انتشار، بازه نهایی برای اجرای پژوهش از ماه می ۲۰۱۴ (معادل بار دیهشت ۱۳۹۳) تا پایان فوریه ۲۰۱۹ (معادل با اسفند ۱۳۹۷) تعیین شد.

### تحلیل داده‌ها

#### آماده‌سازی داده‌ها

اولین گام در اجرای روش‌های داده‌کاوی و یادگیری ماشین، آماده‌سازی و پالایش





زمان  $t$  عینی‌تر سازد. به این منظور بازه زمانی متغیر برای هموارسازی داده‌ها انتخاب شد و مجموع آمار جست‌وجوی کاربران در این بازه‌های زمانی مبنای اجرای روش یادگیری ماشینی قرار گرفت.

در فرایند پنجره‌سازی داده‌ها در دوره‌های زمانی معین، ماتریس داده‌ها خرد شده و به زیرمجموعه‌هایی تقسیم می‌شود که ابعاد هر یک برابر با تعداد متغیرها ضرب در عرض دوره پنجره‌سازی است. بنابراین در پایان برای یکپارچه‌سازی مجدد داده‌ها و دستیابی به یک ماتریس داده واحد، مقدار مجموع متغیرها در بازه زمانی پنجره‌سازی محاسبه شده و یک ماتریس داده جدید، که شامل متغیرهای مستقل تجمیع شده‌اند، ارائه می‌شود. این ماتریس در نهایت برای اجرای روش پیش‌بینی یادگیری ماشینی استفاده شد.

برای آماده‌سازی و تحلیل داده‌های پژوهش حاضر و ارائه الگوی پیش‌بینی با استفاده از روش یادگیری ماشینی، از نرم‌افزار کد باز ریپیدماینر<sup>۳</sup> نسخه ۹ استفاده شده است.

### پیش‌بینی تقاضای گردشگری به روش یادگیری ماشینی نظارت شده

پس از آماده‌سازی داده‌ها به منظور اجرای یادگیری ماشینی نظارت شده<sup>۴</sup>، روش رگرسیون خطی برای پیش‌بینی تقاضای گردشگری شهر یزد انتخاب شد. متغیر وابسته، تعداد گردشگران خارجی ورودی به شهر یزد است که مقدار واقعی آن در ۶۷ ماه گذشته، پس از میانگین‌گیری متحرک، در دسترس است. همچنین نه متغیر مستقل، که داده‌های مستخرج از سامانه گوگل ترندز در حوزه موضوعی سفر به یزد، و یک متغیر مجزا، که آمار گردشگران ورودی به شهر یزد در دوره‌های گذشته را دربر دارد، برای پیش‌بینی مقدار متغیر وابسته در هر ماه انتخاب شدند.

در یادگیری ماشینی به روش نظارت شده، نخست ماتریس داده‌ها براساس یک متغیر وابسته، که مقدار واقعی آن در دوره‌های گذشته در دسترس است، برچسب‌گذاری<sup>۵</sup> می‌شود و از همین روی، ورودی‌ها و خروجی‌های مطلوب متناظر با آن‌ها به صورت جفت‌های ورودی/خروجی در اختیار قرار می‌گیرند. هدف از ارائه این جفت‌های ورودی/خروجی این است که الگوریتم با مقایسه پاسخ درست، که از پیش تعیین شده است و پاسخی که خود پیش‌بینی کرده،

متحرک<sup>۱</sup>، یکی از شاخص‌های مهمی که بسیار استفاده می‌شود، در تحلیل‌های مربوط به پیش‌بینی یک متغیر است که با حذف نوسانات موجود در داده‌ها کمک می‌کند تا تصمیم‌گیرنده بتواند تصویر بهتری از روند تغییرات داده‌ها را طی زمان ترسیم کند. میانگین متحرک یکی از شاخص‌های دنباله‌رو است؛ زیرا متوسط مقدار داده‌ها در بازه زمانی انتخابی گذشته را نشان می‌دهد (برای مثال دو، ده یا چند دوره گذشته). روش تحلیل پنجره (پنجره‌سازی)<sup>۲</sup> با امکان پذیر کردن ترکیب مشاهدات در سری‌های زمانی و مقطعی تاحدودی مشکل کافی نبودن مشاهدات را در پیش‌بینی‌های سری زمانی برطرف می‌کند (Mueller & Massaron, 2016). این روش براساس میانگین متحرک عمل می‌کند و برای یافتن روندهای عملکرد یک واحد در طول زمان مفید است. به این صورت که با هر واحد داده (رکورد) در یک دوره، مانند واحدی مستقل رفتار می‌شود. در این صورت، عملکرد یک واحد در دوره‌ای خاص در مقابل عملکرد خود آن واحد در سایر دوره‌ها قرار گرفته و در کنار عملکرد سایر واحدها ارزیابی می‌شود. این وضعیت باعث افزایش تعداد دوره‌های بررسی شده در تحلیل می‌شود که در هنگام بررسی نمونه‌هایی در اندازه کوچک مفید است. تغییر عرض پنجره، یعنی کم‌وز یادکردن تعداد دوره‌های زمانی مطالعه شده، طیف گسترده‌ای از تحلیل‌های هم‌زمان را میسر می‌سازد که فقط شامل مشاهدات دوره زمانی مشخصی نبوده و تغییرات روند را به خوبی منعکس می‌کند؛ به ویژه به علت تأخیر میان زمان جست‌وجوی کاربر در اینترنت و آغاز سفر، پنجره‌سازی داده‌ها می‌تواند اثر این تأخیر را تا حد زیادی خنثی کند.

در تحلیل پنجره‌ای، داده‌ها برای دوره‌ای مشخص جمع می‌شوند؛ بدین معنا که علاوه بر در نظر گرفتن مقدار متغیر در دوره مدنظر برای پیش‌بینی، دوره‌های قبلی نیز به تعداد مشخص شده به‌منزله متغیری جدید در ماتریس داده‌ها اضافه می‌شوند. برای مثال اگر عرض پنجره ۳ در نظر گرفته شود، پیش‌بینی متغیر وابسته، براساس مقدار متغیرهای مستقل در دوره  $t-1$  و  $t-2$  در نظر گرفته خواهد شد. این کار علاوه بر افزایش حجم داده‌های کم در پیش‌بینی هر متغیر، به علت تأخیر زمانی موجود میان پرس‌وجوی کاربر تا شروع سفر به صورت عملی می‌تواند تأثیر این پرس‌وجوهای اینترنتی در گذشته را بر آمار شب اقامت گردشگران در

3. Rapid Miner

4. Supervised Machine Learning

5. Labeled

1. Moving Average

2. Windowing Data



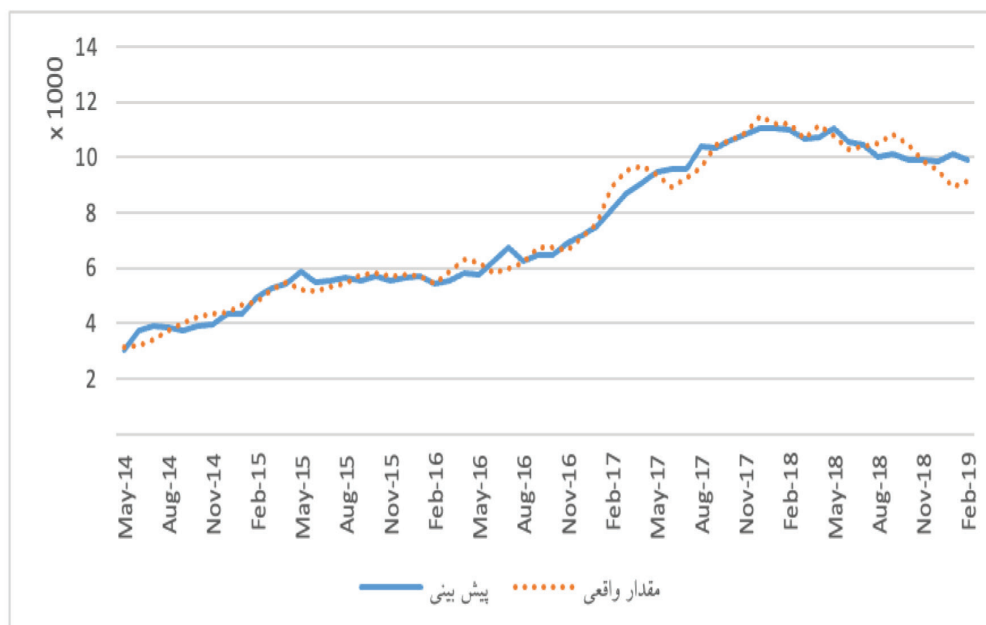
در هر نوبت ۹۰ درصد از داده‌ها به صورت تصادفی، به منزله داده‌های آموزشی برای الگوسازی، و ۱۰ درصد باقی مانده به منزله داده‌های آزمایشی برای ارزیابی مدل طراحی شده استفاده شدند. با میانگین‌گیری از عملکرد به دست آمده از نتایج این ده نوبت، نتایج نهایی پیش‌بینی و ارزیابی دقت مدل‌سازی به دست آمد.

نمودار زیر مقدار واقعی متغیر وابسته (تعداد گردشگران خارجی بر حسب شب اقامت در دوره‌های گذشته بر حسب هزار نفر) و مقدار پیش‌بینی شده آن با استفاده از روش رگرسیون خطی برای مقایسه دقت پیش‌بینی را نشان می‌دهد. در نمودار ۱ مشاهده می‌شود که از نظر ظاهری میان مقادیر پیش‌بینی شده و واقعی تفاوت اندکی وجود دارد، خطای پیش‌بینی (متوسط اختلاف میان این دو مقدار) در خور توجه نیست (نمودار ۱).

خطاهای موجود را شناسایی کند و به این ترتیب، برای ایجاد خروجی‌های بعدی آموزش دیده و الگویی که خود استفاده می‌کند را براساس این آموزش‌ها اصلاح کند. فرایند آموزش تا زمانی که الگوی پیش‌بینی الگوریتم به دقت کافی برسد و نتایج پیش‌بینی با نتیجه از قبل تعیین شده یکسان باشند ادامه خواهد یافت و از همین روی در یادگیری نظارت‌شده سیستم پس از آموزش دیدن و با استفاده از الگوهای که طی آموزش خلق کرده، قادر خواهد بود در مواجهه با داده‌های به اصطلاح بدون برچسب<sup>۱</sup>، خروجی مناسب را پیش‌بینی کند (Mueller & Massaron, 2016).

به منظور تقسیم‌بندی داده‌ها به دو دسته آموزشی و آزمایشی، در پژوهش حاضر از روش نمونه‌گیری مکرر<sup>۲</sup> استفاده شده است؛ به این ترتیب که داده‌ها در ده نوبت مجزا در فرایند الگوسازی وارد شدند.

نمودار ۱: مقایسه مقدار پیش‌بینی شده واقعی تعداد گردشگران خارجی شهر یزد در بازه زمانی می ۲۰۱۴ تا فوریه ۲۰۱۹



حاصل می‌شود؛ به نحوی که در سطح اطمینان ۹۵ درصد ضریب معناداری (p-value) باید کمتر از ۰/۰۵ باشد (مقدار آماره t بیش از ۱/۹۶) باشد. جدول ۲ اهمیت هر یک از متغیرهای مستقل، و عرض از مبدأ را در پیش‌بینی متغیر وابسته نشان می‌دهد.

از آنجاکه پیش‌بینی در روش‌های رگرسیون خطی با استفاده از یک یا چند متغیر مستقل انجام می‌شود، ضریب رگرسیونی نشان‌دهنده قدرت هر یک از این متغیرها در تبیین متغیر وابسته است. البته با بررسی ضریب معناداری هر یک از متغیرهای مستقل ابتدا از تصادفی نبودن ضریب رگرسیون آن متغیر اطمینان

1. Unlabeled  
2. Bootstrapping





ستون پایانی جدول ۲ وضعیت پذیرش یا رد متغیر پیش‌بینی‌کننده را در تأثیرگذاری در مقدار متغیر هدف (تعداد گردشگران آتی) نشان می‌دهد. نکته درخور توجه این است که برخلاف انتظار، متغیر مستقل گردشگران خارجی شهر یزد در ماه‌های گذشته - که به الگو اضافه شده بود - تأثیر معناداری در پیش‌بینی تعداد گردشگران آتی ندارد. به عبارت دیگر، آمارهای گذشته گردشگران نتوانسته است تصویر مناسبی از وضعیت آینده گردشگری مقصد ارائه کند.

جدول ۲: ضریب رگرسیون و شاخص‌های معناداری هر یک از متغیرهای مستقل الگوی رگرسیون خطی در سطح اطمینان ۹۵ درصد

متغیر (پرس و جوی ثبت شده)	ضریب رگرسیون	انحراف معیار	ضریب رگرسیون استاندارد	انحراف	t مقدار آماره	معناداری (p-value)	پذیرش / رد
یزد (سفر)	-۲۸۸۴/۳۳	۶۱۳/۷۳۸	-۴/۶۲۴	۰/۰۹	-۴/۷	۰	پذیرش
اتوبوس و قطار	۱۵۸۴/۸۰۴	۳۸۵/۵۳۳	۳/۶۴۶	۰/۰۶۹	۴/۱۱۱	۰	پذیرش
سفرهای اختصاصی	۲۰۷/۴۰۱	۵۸/۷۱۳	۰/۲۸۵	۰/۳۱۳	۳/۵۳۲	۰/۰۰۱	پذیرش
سفر هوایی	۴۵۱/۶۱۸	۱۴۳/۶۲۱	۰/۳۳۱	۰/۳۰۴	۳/۱۴۵	۰/۰۰۳	پذیرش
مقصدهای گردشگری	۷۹۹/۸۸۶	۳۰۸/۶۸۴	۱/۰۹	۰/۰۹۹	۲/۵۹۱	۰/۰۱۳	پذیرش
آمار واقعی گردشگری در دوره‌های پیشین	۰/۰۲۵	۰/۰۱۴	۰/۳۲۱	۰/۰۹۹	۱/۱۷۵۹	۰/۰۸۵	رد
ایران (سفر)	-۲۹۲/۴۵۶	۲۲۳/۹۱۹	-۰/۱۱۲	۰/۸۹۶	-۱/۳۰۶	۰/۱۹۸	رد
هتل‌ها و مراکز اقامتی	۳۱/۵۲۳	۹۲/۴۲۲	۰/۰۳۹	۰/۰۲۵۴	۰/۳۴۱	۰/۷۳۵	رد
عرض از مبدأ*	۷۴۰/۱۲۷	۵۳۲۱/۰۵۳	-	-	۱/۳۹۱	۰/۱۷۱	رد

بنابراین رابطه رگرسیون خطی زیر، میزان تأثیرگذاری هر یک از متغیرهای مستقل را در پیش‌بینی متغیر وابسته نشان می‌دهد.

$$\text{quantity}_{\text{tourism}} = -2884.33 * \text{Yazd}_{\text{travel}} + 1584.804 * \text{Bus\&Train} + 207.401 * \text{Special Travel} + 451.618 * \text{Air Travel} + 799.886 * \text{Tourist Destinations}$$

در رابطه فوق، متغیر هدف تعداد تقاضای گردشگران خارجی شهر یزد در هر ماه است. همچنین ضرایب نشان داده شده برای هر یک از متغیرهای مستقل، ضریب رگرسیونی آن‌هاست که اهمیت هر یک را در پیش‌بینی تعداد گردشگران نشان می‌دهد. مشخص است که با توجه به یکسان بودن مقیاس سنجش این متغیرها، که از گوگل ترندز پایین‌گذاری شده‌اند، بیشترین اهمیت متعلق به متغیر سفر به یزد و جست‌وجوهای انجام شده در خصوص ویژگی‌های عمومی گردشگری این شهر است.

در اجرای روش‌های مدل‌سازی پیش‌بینی، به شاخص‌هایی برای اطمینان از کارکرد صحیح مدل در پیش‌بینی وضعیت حال حاضر نیاز است (حال‌بینی). در صورتی که مدل‌های یادگیری ماشینی بتوانند با استفاده از داده‌های برچسب‌گذاری شده وضعیت فعلی

را دقیق نشان دهند، می‌توان این مدل‌ها را برای پیش‌بینی وضعیت‌های آینده نیز به‌کار برد. جدول ۳ برخی از مهم‌ترین شاخص‌های اعتبارسنجی مدل پیش‌بینی را نشان می‌دهد. خطای نسبی در این پژوهش، برابر با ۵ درصد است که نشان‌دهنده خطای اندک میان وضعیت واقعی و مقادیر پیش‌بینی است. همچنین مقدار خطای مطلق، که به معنای خطای برآورد تعداد دقیق گردشگران به صورت متوسط در هر ماه است، برابر با ۳۲۰ نفر است. شاخص ضریب تعیین قدرت پیش‌بینی‌کنندگی متغیرهای مستقل را در برآورد متغیر وابسته نشان می‌دهد که مقدار آن نزدیک به ۹۹ درصد است. این مسئله به معنای قدرت بالای پنج متغیر مستقل شناسایی شده در برآورد تعداد گردشگران شهر یزد است.

استفاده از داده‌های برچسب‌گذاری شده وضعیت فعلی



جدول ۳: شاخص‌های ارزیابی مدل رگرسیون در پیش‌بینی تعداد گردشگران

شاخص ارزیابی	مقدار به دست آمده	انحراف معیار
خطای مطلق	۳۱۹/۷۶	۱۳۸/۳۳
خطای نسبی	۰/۰۵	۰/۰۲
ریشه دوم میانگین مربعات خطا	۰/۱۸	۰/۰۹
مربع خطاها	۱۷۰۱۴۱/۷۳	۱۰۹۰۵۵/۲۸
ضریب تعیین	۰/۹۹	۰/۰۲
میانگین پیش‌بینی	۷۴۱۵/۴۱	۸۰۲/۸۳
ضریب هم‌بستگی اسپیرمن	۰/۹۹	۰/۰۰۲

هنگامی حاصل شد که هر دو پارامتر بر روی ۱۲ ماهه تنظیم شده بودند. به عبارت دیگر، میانگین ۱۲ ماهه تعداد گردشگران ورودی شهر یزد به بهترین صورت روند تغییرات استقبال گردشگران خارجی از این شهر را نشان می‌دهد. همچنین در نظر گرفتن مجموع جست‌وجوهای اینترنتی ۱۲ ماه گذشته کاربران گوگل در مورد ویژگی‌های گردشگری مقصد بهترین تصویر را برای پیش‌بینی دقیق متغیر وابسته ایجاد می‌کند. جدول ۴ پنج مورد از بهترین نتایج اجرای روش بهینه‌سازی پارامترها را نشان می‌دهد که مشاهده می‌شود در تکرار ۳۶ و ۴۰ام، مقدار خطای پیش‌بینی به ۵ درصد رسیده است که کمترین مقدار از ۱۴۴ آزمایش مستقلی است که انجام شده؛ بنابراین مقادیر ۱۲ ماهه برای پارامترهای مدل مناسب تشخیص داده شد.

جدول ۴: تنظیم پارامترهای دوره زمانی میانگین متحرک و پنجره‌سازی در ماتریس داده‌ها

تکرار	بازه زمانی در نظر گرفتن متغیرهای مستقل	بازه زمانی سنجش میانگین متحرک متغیر وابسته	مقدار خطای نسبی پیش‌بینی
۴۰	۱۲	۱۲	۰/۰۵
۳۶	۱۲	۱۲	۰/۰۵
۳۳	۳	۱۲	۰/۰۶
۳۷	۳	۱۲	۰/۰۶
۲۸	۱۲	۹	۰/۰۸

### بحث و نتیجه‌گیری

روش رگرسیون خطی با استفاده از منطق یادگیری ماشینی برای پیش‌بینی تعداد گردشگران خارجی ورودی شهر یزد در بازه ماه می ۲۰۱۴ تا فوریه ۲۰۱۹ (جمعاً ۵۸ ماه) اجرا شد. به این منظور، آماده‌سازی داده‌ها شامل انتخاب متغیرهای دارای مقادیر و واریانس مطلوب، محاسبه میانگین متحرک تعداد گردشگران ورودی در دوره‌های ۱۲ ماهه، و هموارسازی داده‌های مربوط به پرس‌وجوهای انجام‌شده در خصوص سفر به یزد در بازه ۱۲ ماهه انجام شد. برای اجرای روش رگرسیون ۱۰ نوبت متمایز فرایند مدل‌سازی اجرا شد که هر بار ۹۰ درصد از داده‌ها به صورت تصادفی برای یادگیری ماشینی و ۱۰ درصد باقی‌مانده برای آزمون عملکرد مدل در پیش‌بینی تعداد گردشگران در دوره‌های گذشته، که داده‌های آن در دسترس بود، استفاده شد. حاصل میانگین این ده نوبت نشان می‌دهد که دقت پیش‌بینی در برآورد تعداد گردشگران ورودی به شهر یزد در بازه‌های ماهانه

### بهینه‌سازی پارامترهای مدل

یکی از مهم‌ترین مسائل در اجرای روش‌های یادگیری ماشین و مدل‌سازی‌های پیش‌بینی متغیرها، تعیین نوع و مقدار پارامترهای هر یک از مراحل اجرای تحلیل است. به عبارت دیگر، با تنظیمات متفاوت، کیفیت نتایج نیز دستخوش تغییر می‌شوند. به همین علت در روش‌های بهینه‌سازی پارامترها<sup>۱</sup>، که به منزله طراحی آزمایش‌ها هم شناخته می‌شوند، تلاش می‌شود با دست‌کاری نوع و مقدار پارامترهای هر یک از اجزای تحلیل بهترین جواب در فرایند مدل‌سازی حاصل شود.

در این پژوهش، دو شاخص بااهمیت در فرایند آماده‌سازی داده‌ها از طریق فرایند بهینه‌سازی پارامترها تعیین شد. ابتدا با توجه به هموارسازی متغیر وابسته در دوره‌های گذشته و حذف نقاط حدی از طریق روش میانگین متحرک، لازم است دوره‌های مطلوب برای اخذ این میانگین مشخص شود. به این منظور، بازه زمانی ۱ تا ۱۲ ماه (به علت تغییرات فصلی) در تحلیل مجاز دانسته شد. همچنین در فرایند پنجره‌سازی متغیرهای مستقل مدل، که با هدف کاهش اثر تأخیر زمانی میان جست‌وجوی اینترنتی کاربران تا آغاز سفر و اقامت آن‌ها در مقصد انجام شده است، نیز بازه زمانی ۱ تا ۱۲ ماه مجاز دانسته شد.

با اجرای روند بهینه‌سازی پارامترها، که مشتمل بر اجرای ۱۴۴ مدل متمایز (با توجه به ترکیب  $12 * 12$ ) حالت ناشی از پارامترهای مجاز برای دو عملگر (مدل) بود، کمترین مقدار خطای نسبی مدل رگرسیون خطی

1. Parameter Optimization



گردشگری این شهر. نکته جالب توجه این است که با توجه قرارگیری یزد در حلقه طلایی گردشگری ایران، شاخص پرس و جوی انجام شده در خصوص سفر به ایران (به‌منزله شاخص کلان‌علاقه به سفر به ایران) نمی‌تواند پیش‌بینی‌کننده‌ای قوی برای تقاضای گردشگری در یک مقصد گردشگری شهری باشد (ضریب معناداری بیش از ۰/۰۵). همچنین به‌رغم انتظار، جست‌وجوهای انجام شده در خصوص هتل‌ها و مراکز اقامتی شهر مقصد نیز ضریب معناداری بیش از ۰/۰۵ کسب کرده است و نمی‌تواند مبنایی برای برآورد تعداد گردشگران متقاضی سفر به شهر یزد باشد. بررسی مطالعات پیشین نشان می‌دهد یافته‌های پژوهش حاضر با مطالعات گانتز و اوندر (۲۰۱۶)، یانگ و همکاران (۲۰۱۴) و گاولیک و همکاران (۲۰۱۱)، که با استفاده از داده‌های گوگل ترندز اقدام به پیش‌بینی تقاضای گردشگری در شهرها و کشورهای متفاوت در بازه‌های زمانی متفاوت کرده‌اند، مطابقت دارد.

از یافته‌های تحقیق مشخص است که از میان شاخص‌هایی مانند امکانات حمل‌ونقل و دستیابی به مقصد گردشگری شهری (اتوبوس و قطار، و هواپیما) و همچنین امکانات و جذابیت‌های گردشگری مقصد بیشترین تأثیر مثبت را در دقیق بودن پیش‌بینی داشته‌اند. به‌عبارت دیگر، این موضوعات مهم‌ترین دغدغه‌های گردشگران بالقوه شهر یزد است که در تصمیم‌گیری نهایی آن‌ها تأثیر بسزایی دارد. پیشنهاد می‌شود تصمیم‌گیرندگان و متولیان حوزه گردشگری به‌جای توجه صرف به امکانات اقامتی و توسعه هتل‌ها، که پیش‌بینی‌کننده خوبی برای تقاضای گردشگری نیستند، بر روی معرفی جاذبه‌ها و امکانات گردشگری (غیر اقامتی)، و گسترش وسایل حمل‌ونقل و توسعه راه‌های ارتباطی به شهر یزد سرمایه‌گذاری کنند. همچنین به علت گسترش روزافزون رفتار جست‌وجوگری اطلاعات میان مردم سراسر جهان، که طبیعتاً متقاضیان گردشگری نیز در میان آن‌ها هستند، طراحی و استقرار و به‌روزرسانی مستمر پایگاه‌های اطلاع‌رسانی در خصوص معرفی امکانات و جاذبه‌های تفریحی، سفر و اقامت در مقصدها، که نیاز اطلاعاتی افراد را پاسخ‌گو باشد، می‌تواند نقش مستقیمی در جذب گردشگران بالقوه به آن مقصد ایفا کند. در صورتی که افراد بتوانند اطلاعات لازم را به ساده‌ترین شکل ممکن همراه با تصاویر و اطلاعات چندرسانه‌ای کسب کنند، تصمیم‌گیری آن‌ها به‌منظور شکل‌گیری تصویر مناسب گردشگری تحت تأثیر قرار خواهد گرفت.

با استفاده از متغیرهای مستقل مستخرج از سامانه گوگل ترندز فقط شامل ۵ درصد خطای نسبی است که از نظر آماری درخور توجه نیست.

نتایج به‌دست آمده نشان می‌دهد فقط با استفاده از ۵ شاخص مربوط به پرس‌وجوهای انجام شده در خصوص موضوعات مربوط به گردشگری شهر یزد، می‌توان با ۹۵ درصد اطمینان به پیش‌بینی تعداد گردشگران خارجی ورودی به این شهر در ماه‌های آتی اقدام کرد. همان‌گونه که در نمودار ۱ مشخص شده است، خط روند تغییرات مقدار واقعی گردشگران با تعداد پیش‌بینی شده آن‌ها مشابهت بسیاری دارد و از نظر تغییرات دوره‌ای مثبت یا منفی اتفاق افتاده در این روند، مقدار پیش‌بینی شده تعداد گردشگران با واقعیت اتفاق افتاده در دوره‌های گذشته مطابقت بسیاری دارد؛ به‌نحوی که براساس شاخص‌های ارزیابی مدل، به‌طور متوسط در بازه زمانی ۵۸ ماهه میزان خطای مطلق (گردشده به بالا) فقط ۳۲۰ نفر است؛ بدین معنا که در صورت در نظر گرفتن متغیرهای مستقل پنج‌گانه منتخب برای پیش‌بینی تقاضا، به‌طور متوسط اختلاف ۳۲۰ نفر گردشگران خارجی برای تعداد شب اقامت با مقدار واقعی، مشاهده می‌شود که با توجه به ۳۰ روزه بودن دوره‌های پیش‌بینی مجموع ۳۲۰ نفر در هر ماه تقریباً برابر با برآورد اشتباه ده نفر گردشگر خارجی ورودی به شهر یزد، در یک روز (با احتساب ماه‌های ۳۰ روزه) است.

در این پژوهش، برای اطمینان از کارکرد صحیح مدل رگرسیون خطی و اعتبار متغیرهای مستخرج از گوگل ترندز، متغیر تعداد شب اقامت گردشگران خارجی در ماه‌های گذشته نیز به‌منزله پیش‌بینی‌کننده موقعیت فعلی به مجموعه داده‌ها اضافه شد. با توجه به این که معناداری این متغیر مستقل در پیش‌بینی تعداد گردشگران رد شد، نتیجه گرفته می‌شود آمارهای گذشته نمی‌تواند مبنای مناسبی برای پیش‌بینی تغییرات آینده قرار گیرد.

براساس یافته‌ها، مهم‌ترین شاخص‌های گوگل ترندز در موضوع گردشگری شهر یزد - که می‌تواند از میان دوازده شاخص منتخب پژوهشگران برای پیش‌بینی دقیق تقاضای گردشگری این شهر در مقصدهای مدنظر قرار گیرد - به ترتیب اهمیت عبارت‌اند از: اطلاعات عمومی در مورد سفر به یزد، وضعیت و شرایط حمل‌ونقل با اتوبوس و قطار به مقصد، تورها و سفرهای اختصاصی موجود در مقصد که گردشگر می‌تواند تهیه کند، امکانات و شرایط سفر هوایی به یزد، و در نهایت ویژگی‌ها و امکانات



performance of tourist arrivals? Mixed-data sampling approach". *Tourism Management*, 46, 454-464.

Castle, J. L., Fawcett, N. W., & Hendry, D. F. (2009). "Nowcasting is not just contemporaneous forecasting". *National Institute Economic Review*, 210(1), 71-89.

Chen, H., Chiang, R. H., & Storey, V. C. (2012). Business intelligence and analytics: From big data to big impact. *MIS quarterly*, 1165-1188.

Choi, H., & Varian, H. (2012). "Predicting the present with Google Trends". *Economic record*, 88(2012), 2-9.

Croes, R. R., & Vanegas Sr, M. (2005). "An econometric study of tourist arrivals in Aruba and its implications". *Tourism management*, 26(6), 879-890.

Dergiades, T., Mavragani, E., & Pan, B. (2018). "Google Trends and tourists' arrivals: Emerging biases and proposed corrections". *Tourism Management*, 66, 108-120.

Doomnik, J. A. (2009). Improving the timeliness of data on influenza-like illnesses using Google search data. Working paper, 1-21.

Ettredge, M., Gerdes, J., & Karuga, G. (2005). "Using web-based search data to predict macroeconomic statistics". *Communications of the ACM*, 48(11), 87-92.

Farashah, M. D. P., Aslani, E., & Khademzade, M. (2018). "Strategic planning of cultural tourism development in historic city of Yazd (case study: Fahadan neighbourhood)". *Almatourism: Journal of Tourism, Culture and Territorial Development*, 9(18), 23-44.

Gawlik, E., Kabaria, H., & Kaur, S. (2011). "Predicting tourism trends with Google Insights". Accessed December, 1, 2012.

Gunter, U., & Önder, I. (2015). "Forecasting international city tourism demand for Paris: Accuracy of uni-and multivariate models employing monthly data". *Tourism management*, 46, 123-135.

Jansen, B. J., Spink, A., & Saracevic, T. (2000). "Real life, real users, and real needs: a study and analysis of user queries on the web". *Information processing & management*, 36(2), 207-227.

Jones, R., Zhang, W. V., Rey, B., Jhala, P., & Stipp,

در اجرای این پژوهش، تلاش شد با گردآوری داده‌هایی که حداقل تورش<sup>۱</sup> و بالاترین دقت را داشتند، برای تقاضای آینده مدل مناسب پیش‌بینی ارائه شود؛ اما با توجه به این که آمارهای مربوط به اقامت گردشگران با تأخیر ارائه شده و این که تمامی گردشگران در مراکز اقامتی رسمی اقامت نمی‌کنند، داده‌های مربوط به تعداد گردشگران در هر ماه می‌تواند با واقعیت تفاوت اندکی داشته باشد. پدیده‌هایی مانند کاناپه‌گردی<sup>۲</sup>، اقامت در کمپ‌ها و مراکز غیررسمی (مانند منازلی که برای اقامت کرایه داده می‌شوند) در مقصد گردشگری، موجب شده است این اختلافات در آمارهای واقعی و منتشر شده دیده شود؛ بنابراین روش‌های آمارگیری مستقیم می‌تواند راهکاری برای غلبه بر این مسئله باشد. همچنین پیشنهاد می‌شود در پژوهش‌های بعدی، به صورت مقایسه‌ای خطای پیش‌بینی تعداد گردشگران با استفاده از داده‌های موتورهای جست‌وجو در چندین مقصد گردشگری مقایسه و از این طریق تأثیر تصویر مقصد گردشگری بر پرس‌وجوهای گردشگران بالقوه شناسایی شود.

با توجه به گستردگی انواع اهداف گردشگری، می‌توان متغیرهای دیگری را به منزله شاخص‌های پیش‌بینی‌کننده تعداد گردشگران ورودی به یک مقصد در نظر گرفت که پیشنهاد می‌شود در پژوهش‌های آتی به شرط دسترسی به داده‌های کافی - برای مثال میزان جست‌وجوی غذاهای محلی در یک مقصد یا دسترسی به امکانات درمانی و سلامت به منزله شاخص‌های پیش‌بینی‌کننده - در نظر گرفته شوند.

## منابع

Athanasopoulos, G., Hyndman, R. J., Song, H., & Wu, D. C. (2011). "The tourism forecasting competition". *International Journal of Forecasting*, 27(3), 822-844.

Bagheri, M., Shojaei, P., & Khorami, M. (2018). "A comparative survey of the condition of tourism infrastructure in Iranian provinces using VIKOR and TOPSIS". *Decision Science Letters*, 7(1), 87-102.

Bangwayo-Skeete, P. F., & Skeete, R. W. (2015). "Can Google data improve the forecasting

### 1. Bias

۲. وب‌سایت گردشگری است که به صورت شبکه اجتماعی و یکی از معروف‌ترین پرتال‌های گردشگری در اینترنت در زمینه «مهمان‌نوازی و اقامت رایگان» فعالیت می‌کند و چند سالی است که بین ایرانیان به محبوبیت درخور توجهی دست یافته است.

Xiang, Z., & Pan, B. (2011). "Travel queries on cities in the United States: Implications for search engine marketing for tourist destinations". *Tourism Management*, 32(1), 88-97.

Xu, F., Lin, Y., Huang, J., Wu, D., Shi, H., Song, J., & Li, Y. (2016). "Big data driven mobile traffic understanding and forecasting: A time series approach". *IEEE transactions on services computing*, 9(5), 796-805.

Yang, X., Pan, B., Evans, J. A., & Lv, B. (2015). "Forecasting Chinese tourist volume with search engine data". *Tourism Management*, 46, 386-397.

Yang, Y., Pan, B., & Song, H. (2014). "Predicting hotel demand using destination marketing organization's web traffic data". *Journal of Travel Research*, 53(4), 433-447.

Zhang, X., & Song, H. (2018). "An integrative framework for collaborative forecasting in tourism supply chains". *International Journal of Tourism Research*, 20(2), 158-171.

E. (2008). "Geographic intention and modification in web search". *International Journal of Geographical Information Science*, 22(3), 229-246.

Kumar, V., & Reinartz, W. (2018). *Data mining. In Customer Relationship Management* (pp. 135-155). Springer, Berlin, Heidelberg.

Mueller, J. P., & Massaron, L. (2016). *Machine learning for dummies*. John Wiley & Sons.

Önder, I. (2017). "Forecasting tourism demand with Google trends: Accuracy comparison of countries versus cities". *International Journal of Tourism Research*, 19(6), 648-660.

Smeral, E. (2014). "Forecasting the city hotel market". *Tourism Analysis*, 19(3), 339-349.

Song, H., & Li, G. (2008). "Tourism demand modelling and forecasting—A review of recent research". *Tourism management*, 29(2), 203-220.

Vu, J. C., & Turner, L. W. (2006). "Regional data forecasting accuracy: The case of Thailand". *Journal of Travel Research*, 45(2), 186-193.