

## Evaluating the Use of Media-Based Data in Improving Economic Forecasting

**Morteza Beiranvand:** PhD student in Development Economics, Faculty of Administrative Sciences and Economics, Ferdowsi University of Mashhad, Mashhad, Iran. **email:** beiranvand.mo@gmail.com

**Saeed Malek Sadati:** Assistant Professor, Department of Economics, Faculty of Administrative Sciences and Economics, Ferdowsi University of Mashhad (Corresponding Author), Mashhad, Iran. **email:** msadati@um.ac.ir

**Seyed Mohammad Javad Razmi:** Professor, Department of Economics, Faculty of Administrative Sciences and Economics, Ferdowsi University of Mashhad, Mashhad, Iran. **email:** mjrazmi@um.ac.ir

The expansion of media accessibility and continuous development of information extraction techniques have introduced a new source of data in economic research—media-based data. The growing use of such data in economic analysis signals a significant shift in information collection and assessment approaches within the field. Using news texts as a major media-based data resource and applying artificial intelligence and machine learning methods, this study offers an innovative framework for analyzing and monitoring Iran's economic conditions. First, a corpus of 50,216 daily Persian economic news reports was constructed. Next, to convert qualitative news information into quantitative data, topic modeling was applied for categorization, followed by sentiment analysis (negative, neutral, positive) using a lexicon-based approach. Finally, the accuracy and reliability of extracted data in forecasting the Purchasing Managers' Index (PMI/SHAMAK)—a leading measure of economic activity—were evaluated. Results show that the use of media-based data reduces forecasting error compared to conventional economic models. Furthermore, the proposed approach enables assessment of economic conditions at the end of each month using media news, prior to the publication of official statistics.

**Keywords:** Media-based Data, Economic News, Purchasing Managers' Index, Machine learning, Text Mining.

---

**How to cite this paper:** Beiranvand, M., Malek Sadati, S., & Razmi, S. M. J. (2026). Evaluating the Use of Media-Based Data in Improving Economic Forecasting. *Rasaneh*, 36(4), 115-134. **[In persian]**

## ارزیابی استفاده از داده‌های رسانه‌بنیان در بهبود پیش‌بینی‌های اقتصادی

تاریخ دریافت: ۱۴۰۳/۰۵/۱۳

تاریخ پذیرش: ۱۴۰۳/۰۹/۱۱

نوشته

مرتضی بیرانوند\*

سعید ملک‌الساداتی\*\*

سید محمدجواد رزمی\*\*\*

### چکیده

امروز، افزایش فراگیری و سهولت دسترسی به رسانه‌ها، همراه با تکامل مستمر روش‌های استخراج اطلاعات، منجر به ظهور منبع جدیدی از داده‌ها در تحقیقات اقتصادی شده است که به داده‌های رسانه‌بنیان شهرت یافته‌اند. روند فزاینده استفاده از این نوع داده‌ها در مطالعات و تحلیل‌های اقتصادی، نشان‌دهنده تحولی مهم در شیوه‌های جمع‌آوری و تحلیل اطلاعات در حوزه اقتصاد است. این پژوهش با استفاده از متون خبری به‌عنوان یکی از اصلی‌ترین منابع داده رسانه‌بنیان و بهره‌گیری از روش‌های هوش مصنوعی و یادگیری ماشین، یک رویکرد نوآورانه برای تحلیل و پیش‌بینی وضعیت اقتصادی ایران ارائه می‌دهد. در ابتدا، پیکره متنی شامل ۵۰/۲۱۶ خبر اقتصادی روزانه به زبان فارسی ایجاد شد. در گام بعد، برای تبدیل اطلاعات کیفی متون خبری به داده‌های کمی، ابتدا با استفاده از روش مدل‌سازی موضوعی، این اخبار به دسته‌های موضوعی مشخصی طبقه‌بندی و سپس جهت‌گیری حسی (منفی، خنثی، مثبت) اخبار ذیل هر موضوع با استفاده از رویکرد تحلیل احساس مبتنی بر واژه‌نامه مشخص شد. در پایان، میزان دقت و قابلیت اتکای داده‌های استخراج‌شده از اخبار در پیش‌بینی شاخص مدیران خرید (شامخ) - که یکی از معتبرترین و پرکاربردترین شاخص‌های ارزیابی وضعیت اقتصاد کشور محسوب می‌شود - مورد سنجش و ارزیابی قرار گرفت. نتایج نشان‌دهنده کارایی داده‌های رسانه‌بنیان می‌تواند خطای پیش‌بینی را نسبت به الگوهای مرسوم اقتصادی کاهش دهد. افزون بر این، روش پیشنهادی این پژوهش این امکان را فراهم می‌کند که بلافاصله پس از پایان هر ماه، با استفاده از اخبار منتشرشده در رسانه‌ها، وضعیت اقتصادی را پیش از انتشار آمار رسمی تحلیل و تبیین کرد. کلیدواژه: داده‌های رسانه‌بنیان، اخبار اقتصادی، شاخص مدیران خرید، یادگیری ماشین، متن‌کاوی.

\* دانشجوی دکتری اقتصاد توسعه، دانشکده علوم اداری و اقتصاد، دانشگاه فردوسی مشهد، مشهد، ایران  
beiranvand.mo@gmail.com

\*\* استادیار گروه اقتصاد، دانشکده علوم اداری و اقتصاد، دانشگاه فردوسی مشهد (نویسنده مسئول)، مشهد، ایران  
msadati@um.ac.ir

\*\*\* استاد گروه اقتصاد، دانشکده علوم اداری و اقتصاد، دانشگاه فردوسی مشهد، مشهد، ایران  
mjrazmi@um.ac.ir

نحوه استناد به این مقاله: بیرانوند، مرتضی؛ ملک‌الساداتی، سعید و رزمی، سید محمدجواد (۱۴۰۴). ارزیابی استفاده از داده‌های رسانه‌بنیان در بهبود پیش‌بینی‌های اقتصادی. رسانه، (۴)۳۶، ۱۱۵-۱۳۴.

## مقدمه

حضور رسانه‌ها چنان با تجربه‌های عموم افراد و اعضای جامعه پیوند خورده است که نمی‌توان از آن تنها به عنوان یک ابزار یاد کرد، بلکه باید آن را عضو فعالی از زندگی روزمره دید که توان اثربخشی و بازسازی ماندگاری اعمال، افکار و تعامل اعضای جامعه را دارد (نادریله‌رود و قاسمی، ۱۴۰۲). به همین سبب امروزه همگان این عبارت را مورد توجه قرار داده‌اند که اهمیت جهان امروز، یافتن و دسترسی به اطلاعات و ارتباطات است و بسیاری از صاحب‌نظران علوم اجتماعی و ارتباطات بر جامعه اطلاعاتی و جامعه شبکه‌ای تأکید داشته‌اند (کبیری و همکاران، ۱۴۰۲). این فراگیری و گسترش رسانه‌ها در جامعه باعث شده است که استفاده از داده‌های رسانه‌بنیان<sup>۱</sup> به عنوان یک رویکرد نوآورانه در مطالعات اقتصادی مطرح شود. این رویکرد با بهره‌گیری از حجم فراوان اطلاعات موجود در منابع مختلف رسانه‌ای، امکان تحلیل دقیق‌تر و به‌روزتری از شرایط اقتصادی در سطوح خرد و کلان فراهم می‌کند. به‌هنگام بودن، تفصیلی بودن و حجم زیاد سه ویژگی کلیدی این نوع داده‌ها هستند. به‌هنگام بودن این داده‌ها، بینش‌های جدیدی درباره روندها و رویدادهای اقتصادی در زمان مناسب فراهم می‌کند. تفصیلی بودن این داده‌ها، دسترسی به اطلاعات کاملی درباره صنایع، شرکت‌ها و نواحی جغرافیایی خاص را ممکن می‌سازد و حجم زیاد این داده‌ها نیز درک جامع‌تری از روندها و الگوهای اقتصادی به همراه دارد. مقالات و گزارش‌های خبری، شبکه‌های اجتماعی، وب‌نوشت‌ها و انجمن‌های برخط از جمله مهم‌ترین منابع داده‌های رسانه‌بنیان هستند که مطالعه آن‌ها بینش‌های ارزشمندی در خصوص روندها و رویدادهای اقتصادی، تغییرات سیاسی، انتظارات بازار و نگرانی‌های اقتصادی ارائه می‌دهد.

رسانه‌های خبری به‌عنوان واسطه اصلی اطلاعات در جامعه، به‌طور مداوم رویدادهای مختلف از جمله رویدادهای اقتصادی را به توصیف کلامی تبدیل و در قالب اخبار منتشر می‌کنند. این فرایند به بررسی جامع و عمیق رویدادها و پدیده‌های مختلف می‌پردازد و تلاش می‌کند تا با توصیف دقیق، تفسیر معنا، شناسایی تأثیر و استنباط علل، به درک کامل‌تری از این رویدادها دست یابد (بابی<sup>۲</sup> و همکاران، ۲۰۲۰). اخبار با گسترش و نشر روایت‌ها یا همان داستان‌ها و افکار رایج در جامعه، باعث ایجاد تغییر سریع در فرهنگ، جو فکری و رفتار اقتصادی می‌شوند و مطالعه آن‌ها می‌تواند توانایی ما را در پیش‌بینی رخدادها و اقتصادی و مهیا شدن برای آن‌ها بالا ببرد. مطالعاتی که به بررسی اخبار منتشر شده در رسانه‌ها از جنبه اقتصادی پرداخته‌اند، بیشتر اوقات بر تحلیل داده‌های متنی متمرکز بوده‌اند. دلیل این تمرکز این است که متون می‌توانند انواع مختلفی از داده‌های غیرساختاریافته را در خود جای دهند، از جمله داده‌های صوتی، تصویری و ویدیویی که می‌توان آن‌ها را به متن تبدیل و سپس تحلیل کرد (آلگابا<sup>۳</sup> و همکاران، ۲۰۲۰).

1. Media-based Data  
2. Bybee  
3. Algaba

استفاده از تحلیل اسناد متنی برای بررسی پدیده‌های اقتصادی سابقه طولانی دارد. از نمونه‌های کلاسیک در این زمینه می‌توان به آثار محققانی چون کوز<sup>۱</sup> (۱۹۶۰) در زمینه اقتصاد حقوق و فریدمن و شوارتز<sup>۲</sup> (۱۹۶۳) در زمینه اقتصاد سیاسی اشاره کرد. در گذشته تجزیه و تحلیل متن از طریق خوانش دقیق انسانی انجام می‌شد. این رویکرد به دلیل محدودیت‌های انسانی، نمی‌تواند با سرعت و مقیاس مورد نیاز برای بررسی داده‌های متنی عظیمی که اکنون با گسترش رسانه‌ها در دسترس پژوهشگران قرار دارند، هم‌گام شود (اش و هانسن<sup>۳</sup>، ۲۰۲۳). به‌منظور غلبه بر این محدودیت و استخراج اطلاعات از داده‌های متنی، استفاده از روش‌های پردازش زبان طبیعی<sup>۴</sup> (NLP) رایج شده است. پردازش زبان طبیعی، یکی از شاخه‌های هوش مصنوعی است که به بررسی و پردازش متون و گفتار انسانی توسط رایانه‌ها و سیستم‌های هوشمند می‌پردازد (کنسولی<sup>۵</sup> و همکاران، ۲۰۲۱).

این پژوهش با بهره‌گیری از داده‌های خبری و بهره‌مندی از روش‌های پیشرفته متن‌کاوی و یادگیری ماشین، نقش و تأثیر رسانه‌ها در تحلیل‌های اقتصادی را بررسی می‌کند و نشان می‌دهد که پژوهشگران اقتصادی برای شناسایی الگوها و روندهای اقتصاد ایران، با توجه به اهمیت فزاینده‌ای که رسانه‌ها در جمع‌آوری و انتشار اطلاعات دارند، باید در کنار استفاده از داده‌های اقتصادی مرسوم، از اطلاعات موجود در رسانه‌ها نیز استفاده کنند. در این پژوهش با توجه به کمبود وجود پیکره متنی خبری اقتصادی برای زبان فارسی، این پیکره با استخراج اخبار اقتصادی از اینترنت و پیش‌پردازش آن‌ها تهیه شده است. در گام بعد، از مدل‌سازی موضوعی<sup>۶</sup> که در پردازش زبان طبیعی و مدیریت اطلاعات برای شناسایی و استخراج موضوعات مختلف موجود در یک مجموعه متنی به‌کار می‌رود استفاده شد و با تعیین توزیع احتمالی کلمه‌ها برای هر موضوع، اخبار بر اساس این توزیع‌ها به موضوعات مختلف تخصیص داده می‌شود. سپس با استفاده از دیگر ابزار مهم پردازش زبان طبیعی یعنی تحلیل احساس<sup>۷</sup>، که هدف اصلی آن تعیین نگرش و احساس نویسنده نسبت به یک موضوع خاص است، جهت‌گیری حسی اخبار مشخص می‌شود. در پایان، داده‌های حاصل از تبدیل اطلاعات کیفی متنی به اطلاعات کمی، برای پیش‌بینی شاخص مدیران خرید<sup>۸</sup> (شامخ) به‌عنوان یکی از مهم‌ترین نماگرهای وضعیت اقتصاد به کار می‌رود. این پیش‌بینی با استفاده از روش یادگیری ماشین<sup>۹</sup> تقویت‌گرادیان<sup>۱۰</sup> انجام می‌شود.

1. Coase
2. Friedman & Schwartz
3. Ash & Hansen
4. Natural Language Processing (NLP)
5. Consoli
6. Topic Modeling
7. Sentiment Analysis
8. Purchasing Managers' Index (PMI)
9. Machine Learning
10. Gradient Boosting

## پیشینه پژوهش

### پیشینه نظری

#### رسانه و اقتصاد روایی<sup>۱</sup>

رسانه‌ها منبعی غنی و پویا از اطلاعات را ارائه می‌دهند که می‌توانند پیش‌بینی اقتصادی را به‌طور قابل توجهی بهبود دهند. در ادامه انواع اصلی این منابع اطلاعاتی به همراه کاربرد اقتصادی آن‌ها معرفی شده است.

**الف. مقالات و گزارش‌های خبری:** تحلیلگران با تجزیه و تحلیل احساسات مقالات خبری مرتبط با اقتصاد، حوزه‌های مختلف یا شرکت‌های خاص می‌توانند انتظارات کلی بازار را بسنجند. همچنین رویدادهای مهمی همچون تغییرات سیاستی، بلایای طبیعی، رویدادهای ژئوپلیتیکی که بر اقتصاد مؤثرند را از طریق تحلیل اخبار می‌توان شناسایی کرد و با رصد اخبار مالی می‌توان واکنش‌های بازار به شاخص‌های مختلف اقتصادی و اطلاعیه‌های شرکت‌های بورسی را درک کرد.

**ب. شبکه‌های اجتماعی:** تحلیل احساسات عمومی در شبکه‌های اجتماعی، می‌تواند بینش‌هایی در زمان مناسب در مورد احساسات عمومی و اعتماد مصرف‌کننده ارائه دهد. ردیابی موضوعات پرطرفدار نیز می‌تواند به شناسایی مسائل اقتصادی نوظهور و تغییرات در رفتار مصرف‌کننده کمک کند.

**پ. وب‌نوشت‌ها و انجمن‌ها:** وب‌نوشت‌ها و انجمن‌ها بیشتر اوقات دارای بینش و اظهارات آینده‌نگر تحلیلگران و کارشناسان هستند که می‌تواند برای تکمیل منابع داده‌سستی استفاده شود. انجمن‌ها و بخش‌های نظرات نیز می‌توانند بازخوردی بدون واسطه از نظرات و عقاید مصرف‌کننده را ارائه دهند که برای ارزیابی تقاضای بازار و احساسات اقتصادی مفید است.

تأکید به استفاده از منابع داده‌های جدید برای بررسی‌های اقتصادی و نقد رویکردهای سنتی به پژوهش‌های اقتصاد خرد و کلان موضوع جدیدی نیست. جالینگ کوپمانس<sup>۲</sup> (۱۹۴۷)، در مقاله‌ای مشهور با عنوان "سنجش بدون نظریه"، از رویکردی انتقاد می‌کند که برای پیش‌بینی اقتصادی، تنها به ویژگی‌های آماری داده‌های شاخص‌های کلیدی اقتصادی مانند نرخ تولید ناخالص ملی یا نرخ بهره توجه دارد. انقلاب اقتصاد رفتاری<sup>۳</sup> در چند دهه اخیر بر چگونگی تأثیر عواطف، عادات و قوانین اجتماعی بر رفتار اقتصادی تأکید دارد و به همین دلیل می‌تواند

1. Narrative Economics  
2. Tjalling Koopmans  
3. Behavioral Economics

به تکمیل درک ما از داده‌های آماری کمک کند. در ادامه مسیر اقتصاد رفتاری، شیلر<sup>۱</sup> (۲۰۱۹) با ایجاد مفهومی به نام "اقتصاد روایی" به اهمیت تأثیر رسانه‌ها و رویدادهای اجتماعی بر پدیده‌های اقتصادی تأکید کرده است.

اقتصاد روایی مطالعه نحوه انتشار و همه‌گیری روایت‌هایی است که می‌تواند به اقتصاددانان در تحلیل و پیش‌بینی پدیده‌های اقتصادی کمک کند. روایت، به‌عنوان شرحی از رخداد‌های واقعی یا تخیلی، پدیده‌ای مخصوص انسان است. روایت‌ها عامل ایجاد تغییر سریع در فرهنگ، جو فکری و رفتار اقتصادی‌اند و مطالعه آن‌ها می‌تواند توانایی ما را در پیش‌بینی رخداد‌های اقتصادی و مهیا شدن برای آن‌ها بالا ببرد (شیلر، ۲۰۱۹). شیلر، با بررسی روایت‌های عامه دوره‌های مختلف و به کمک علم همه‌گیرشناسی<sup>۲</sup> به ارتباط بین این روایت‌ها و رفتارهای اقتصادی و پیامدهای آن مانند تورم، رکود، حباب‌های اقتصادی و حتی جنگ می‌پردازد. برای نمونه وی نشان می‌دهد که فراگیر شدن روایت‌هایی مانند منحنی لافر<sup>۳</sup> و سایر داستان‌های طرف عرضه به مطالبه شدید مردمی، برای کاهش مالیات منجر شد. همچنین به هم پیوستن روایت‌هایی درباره کاهش مالیات و کوچک‌تر شدن دولت، به پیشبرد جنبشی اجتماعی به نام "کارآفرینی" کمک کرد.

در گذشته ایده‌ها و روایت‌ها به‌طور تصادفی و غیرسازمان‌یافته شایع می‌شدند اما امروزه رسانه‌های اجتماعی، پایگاه‌های خبری و موتورهای جست‌وجو، کاری می‌کنند که آن‌ها با سرعت بسیار بیشتری اشاعه یابند. روایت‌ها به‌صورت توالی واژگانی شکل می‌گیرند که بر اساس اصول زبان‌شناسی پشت سر هم قرار گرفته‌اند. واژه‌ها هم معانی ساده و سرراست و هم معانی ضمنی و کاربردهای استعاره‌ای دارند. شرط اصلی استفاده از روایت‌ها در مطالعات اقتصادی کشف الگوهای قابل فهم و کاربردی از آن‌ها است. بیشتر تلاش‌ها برای فهم این الگوها، به اسناد متنی معطوف بوده است و با پیشرفت روش‌های پژوهش و گردآوری داده‌ها از رسانه‌ها، تحلیل متنی به عامل تأثیرگذاری در شناسایی الگوهای روایی و کاربرد آن‌ها در شناسایی و پیش‌بینی نوسانات اقتصادی تبدیل شده است.

در نخستین گام می‌توان با شمردن واژگان و عبارات موجود در متون دیجیتالی در دسترس، چیزهایی درباره روایت‌های اقتصادی عامه‌پسند آموخت ولی برای سنجش قدرت روایت‌های ضدونقیضی که در گذر زمان باز ترکیب می‌شوند و رخداد‌های اقتصادی بزرگ را رقم می‌زنند، باید از هوش مصنوعی در داده‌های غیرساخت‌یافته متنی یا آنچه امروزه متن‌کاوی نامیده می‌شود، استفاده کرد.

1. Shiller  
2. Epidemiology  
3. Laffer curve

## محدودیت‌های استفاده از داده‌های رسانه‌بنیان

اگرچه استفاده از داده‌های رسانه‌بنیان در مطالعات اقتصادی و اجتماعی می‌تواند مزایای زیادی داشته باشد، اما این امر با چالش‌ها و محدودیت‌هایی نیز همراه است. یک دسته از این محدودیت‌ها، محدودیت‌های فنی است که به حجم و ماهیت این داده‌ها مرتبط است و دسته دیگر مربوط به سوگیری‌هایی است که بر کیفیت این داده‌ها تأثیر می‌گذارد. نخستین محدودیت فنی در زمینه جمع‌آوری و استفاده از داده‌ها، مسائل و مقررات قانونی است. مقررات مختلف ملی و بین‌المللی ممکن است محدودیت‌هایی را بر جمع‌آوری و نحوه استفاده از داده‌ها ایجاد کنند. افزون‌بر این، سیاست‌های سخت‌گیرانه‌ای که بیشتر رسانه‌ها در رابطه با امکان دسترسی و استخراج اطلاعات منتشر شده در بستر آن‌ها دارند، داده‌های مورد نیاز برای اهداف تحقیقاتی را محدود می‌کند.

محدودیت فنی دیگر، به حجم بالای این داده‌ها مرتبط است که جمع‌آوری، ذخیره‌سازی، پردازش و تحلیل آن‌ها را نیازمند کسب دانش در زمینه‌های مختلفی مانند داده‌کاوی، آمار و یادگیری ماشین می‌کند. از طرف دیگر این حجم بالا از داده‌ها، شامل مقادیر قابل توجهی از اطلاعات نامربوط و اضافه نیز می‌شود که برای حذف آن‌ها و تمرکز بر داده‌های اصلی و مرتبط، انجام پیش‌پردازش‌های مختلف ضروری می‌شود. پویایی‌های زمانی نیز می‌تواند به سرعت داده‌های رسانه‌بنیان را منسوخ کند، به طوری که ممکن است یافته‌های ناشی از اطلاعات جمع‌آوری شده در گذشته اخیر، واقعیت‌های فعلی را منعکس نکند و در نتیجه به روزرسانی مداوم این داده‌ها ضروری شود (مارتی<sup>۱</sup> و همکاران، ۲۰۱۹).

سوگیری‌هایی را که بر داده‌های رسانه‌بنیان تأثیر می‌گذارد می‌توان به دو دسته سوگیری‌های دسترسی<sup>۲</sup> و سوگیری‌های تحریریه<sup>۳</sup> تقسیم کرد (پارکینسون<sup>۴</sup>، ۲۰۲۴). در سوگیری دسترسی، خبرنگاران با محدودیت‌هایی مواجه هستند که توانایی آن‌ها را در پوشش کامل و دقیق رویدادها کاهش می‌دهد. این محدودیت‌ها می‌تواند شامل سانسور دولتی، مسائل امنیتی و محدودیت‌های مالی باشد. اگر گروه‌ها یا افراد خاصی بیشتر از دیگران در دسترس خبرنگاران باشند، نظرات آن‌ها به طور بیشتری در رسانه‌ها منعکس می‌شود و اگر تحت فشارهای سیاسی و دولتی پرداختن به برخی موضوعات محدودیت داشته باشد، آن موضوعات در رسانه‌ها به حاشیه رانده می‌شوند. در رسانه‌های اجتماعی نیز همه گروه‌های جمعیتی به یک اندازه فعال نیستند و توزیع جغرافیایی کاربران نیز متفاوت است. همه این موارد بر تعمیم‌پذیری یافته‌های پژوهشی که از داده‌هایی که در معرض سوگیری دسترسی بوده‌اند مؤثر است و به درک نادرست از رویدادها و موضوعات منجر می‌شوند.

1. Martí  
2. Access Biases  
3. Editorial Biases  
4. Parkinson

سوگیری تحریریه به تمایل ارائه اطلاعات به روشی اشاره دارد که با علایق سیاسی، ایدئولوژیکی یا تجاری خاصی همخوانی دارد. این سوگیری به دو صورت ظاهر می‌شود. نخست، به‌طور خودخواسته، که در آن هیئت تحریریه آگاهانه یک دیدگاه خاص را ترویج می‌کند و دوم، به‌طور ناخواسته، که ناشی از ترجیحات و پیشینه‌های شخصی خبرنگاران و سردبیران است. این سوگیری باعث می‌شود برخی موضوعات به‌طور مداوم در چرخه‌های خبری مورد توجه قرار گیرند، در حالی که موضوعات دیگر کمتر گزارش می‌شوند. گاهی رسانه با انتخاب کلمات، تصاویر، و زمینه‌ای خاص یکی از جنبه‌های موضوع یا رویدادی را برجسته و روایت را به گونه‌ای چارچوب‌بندی می‌کند که با دیدگاه‌های ایدئولوژیک خاصی هم‌سو باشد. همچنین رسانه‌ها ممکن است برای جلب توجه و برانگیختن احساسات خاصی از مخاطبان، به زبان اغراق‌آمیز و استفاده از کلمات با بار حسی قوی متوسل شوند و بر نحوه روایت داستان‌های خبری و تعامل مخاطبان تأثیر بگذارند (رودریگو<sup>۱</sup> و همکاران، ۲۰۲۴).

افزون بر موارد فوق، پیامدهای اخلاقی استفاده از داده‌های رسانه‌های اجتماعی می‌تواند فرایندهای تحقیقاتی را پیچیده کند. یکی از دغدغه‌های اخلاقی مهم، موضوع رضایت آگاهانه<sup>۲</sup> است. کاربران رسانه‌های اجتماعی ممکن است آگاه نباشند که داده‌های آن‌ها برای اهداف تحقیقاتی جمع‌آوری می‌شود. محققان باید در نظر بگیرند که آیا استفاده از این داده‌ها بدون رضایت صریح کاربران، اخلاقی است یا خیر. حفظ حریم خصوصی افراد نیز یک ملاحظه اخلاقی مهم است. حتی زمانی که داده‌ها برای عموم قابل دسترسی هستند، نگرانی‌هایی در مورد احتمال استفاده از این داده‌ها برای شناسایی افراد یا افشای اطلاعات حساس وجود دارد (مارتی و همکاران، ۲۰۱۹).

## شاخص مدیران خرید در ایران

شاخص مدیران خرید در ایران توسط اتاق بازرگانی، صنایع، معادن و کشاورزی ایران منتشر می‌شود. این شاخص از مهرماه سال ۱۳۹۷ برای بخش صنعت و از مهرماه سال ۱۳۹۸ برای کل اقتصاد با عنوان اختصاری "شامخ" انتشار یافته است. اطلاعات مربوط به شامخ کل اقتصاد ایران، از طریق نظرسنجی و نمونه‌گیری از واحدهای مهم اقتصادی (شامل صنعت، خدمات و کشاورزی) در سطح کشور جمع‌آوری می‌شود. در محاسبه شامخ کل اقتصاد، سهم ارزش افزوده هر بخش در کل اقتصاد نیز در نظر گرفته می‌شود. این روش جمع‌آوری اطلاعات و محاسبه شاخص با هدف ایجاد یک تصویر جامع از وضعیت اقتصادی کشور طراحی شده است.

1. Rodrigo  
2. Informed Consent

پرسشنامه الکترونیکی شاخص مدیران خرید ایران در دو بخش مؤلفه‌های اصلی و کمکی با الهام از رویکرد مؤسسه بین‌المللی "IHS MARKIT" طراحی شده است.

### ۱. مؤلفه‌های اصلی (وزن دار):

- مقدار تولید محصولات (پرسش اصلی با ضریب ۲۵ درصد)
- میزان سفارش‌های جدید (پرسش اصلی با ضریب ۳۰ درصد)
- سرعت انجام و تحویل سفارش (پرسش اصلی با ضریب ۱۵ درصد)
- موجودی مواد اولیه (پرسش اصلی با ضریب ۱۰ درصد)
- میزان استخدام و به‌کارگیری نیروی انسانی (پرسش اصلی با ضریب ۲۰ درصد)

### ۲. مؤلفه‌های کمکی (بدون وزن):

- قیمت خرید مواد اولیه
- موجودی محصول نهایی در انبار
- میزان صادرات کالا
- قیمت محصولات تولیدشده
- مصرف حامل‌های انرژی
- میزان فروش محصولات
- انتظارات تولید در ماه آینده

در نظرسنجی، از پاسخ‌دهندگان خواسته می‌شود که با پاسخ به ۱۲ پرسش مطرح‌شده، با انتخاب یکی از گزینه‌های زیر، وضعیت فعلی را نسبت به ماه قبل ارزیابی کنند.

۱. وضعیت نسبت به ماه گذشته بهتر شده است. (امتیاز: ۱)
۲. وضعیت نسبت به ماه گذشته بدتر شده است. (امتیاز: ۰)
۳. وضعیت نسبت به ماه گذشته تغییری نکرده است. (امتیاز: ۰/۵)

در این نظرسنجی، به هر پاسخ امتیازی اختصاص داده می‌شود. مجموع حاصل ضرب درصد پاسخ‌دهندگان به هر یک از گزینه‌ها در امتیاز همان گزینه، مبین عدد شامخ در آن پرسش است. در پایان، معیار سنجش این شاخص عددی بین صفر تا ۱۰۰ است. عدد صفر به معنای آن است که ۱۰۰ درصد پاسخ‌دهندگان گزینه "بدتر شده است" را انتخاب کرده‌اند، و عدد ۱۰۰ به معنای آن است که ۱۰۰ درصد پاسخ‌دهندگان گزینه "بهتر شده است" را انتخاب کرده‌اند. عدد ۵۰ نشان‌دهنده عدم تغییر وضعیت نسبت به ماه قبل است. نمره‌های بالای ۵۰ نشانه

بهبود وضعیت نسبت به ماه قبل و نمره‌های زیر ۵۰ به معنای بدتر شدن وضعیت نسبت به ماه قبل است.

## پیشینه تجربی

برخی از مطالعاتی که هدفشان استخراج اطلاعات از متون اقتصادی و مالی است، این کار را با جست‌وجوی واژه‌ها و عباراتی خاص در متن انجام می‌دهند. این رویکرد به آن‌ها امکان می‌دهد تا از طریق شناسایی چنین کلمات و عباراتی، احساسات موجود در متن را آشکار کنند. دا<sup>۱</sup> و همکاران (۲۰۱۵)، در پژوهش خود از کلیدواژه‌های "رکود اقتصادی"، "بیکاری" و "ورشکستگی" در جست‌وجوهای برخط برای شناسایی و آشکارسازی احساسات موجود در بازار استفاده کردند. آن‌ها از این رویکرد برای ارائه یک شاخص جدید به نام "شاخص نگرش‌های مالی و اقتصادی آشکارشده توسط جست‌وجو (FEARS)"<sup>۲</sup> برای ارزیابی احساسات سرمایه‌گذاران بهره‌گرفتند. بیکر<sup>۳</sup> و همکاران (۲۰۱۶)، در پژوهش خود با جست‌وجو و شمارش کلیدواژه‌هایی همچون "اقتصادی"، "سیاست" و "نااطمینانی" در مجموعه‌ای از مقالات خبری اقتصادی آمریکا، یکی از مؤلفه‌های اصلی شاخص "نااطمینانی سیاست اقتصادی"<sup>۴</sup> (EPU) را طراحی کردند. این شاخص با سایر شاخص‌های نااطمینانی همبستگی بالایی دارد و می‌تواند در پیش‌بینی متغیرهایی همچون اشتغال و سرمایه‌گذاری استفاده شود. لوداتی و پسران<sup>۵</sup> (۲۰۲۳)، نیز در پژوهش خود با جست‌وجو و شمارش کلیدواژه‌هایی مانند "تحریم"، "ایران" و "اقتصاد" در مقالات خبری شش روزنامه مهم جهانی، شاخصی را برای محاسبه شدت تحریم‌ها طراحی کردند. نتایج استفاده از این شاخص نشان داد که تحریم‌ها به‌طور قابل توجهی درآمد صادرات نفت ایران را کاهش داده و به دنبال آن افزایش تورم و کاهش رشد تولید را به همراه داشته است.

در کنار رویکرد شمارش کلمات، استفاده از واژه‌نامه‌های حسی<sup>۶</sup> نیز برای تحلیل احساسات متون در مطالعات اقتصادی رایج بوده است. این واژه‌نامه‌ها شامل مجموعه‌ای از کلمات هستند که به هر کلمه یک برچسب حسی (مثبت، منفی یا خنثی) اختصاص داده شده است. با استفاده از این واژه‌نامه‌ها، محققان می‌توانند میزان احساسات مثبت، منفی و خنثی در متون را محاسبه کرده و شاخص‌های حسی مرتبط با پدیده‌های اقتصادی را ایجاد کنند. این رویکرد

1. Da
2. Financial and Economi Attitudes Revealed by Search
3. Baker
4. Economic Policy Uncertainty
5. Laudati & Pesaran
6. Sentiment Lexicons

مکمل روش شمارش کلمات بوده و امکان تحلیل عمیق‌تر محتوای متون را می‌دهد. تتلاک<sup>۱</sup> و همکاران (۲۰۰۸)، با استفاده از واژه‌نامه Harvard-IV نشان دادند که بین مقدار استفاده از کلمات با بار حسی منفی در اخبار مرتبط با هر شرکت و قیمت سهام آن شرکت، رابطه معنادار وجود دارد. لاگرن و مک‌دونالد<sup>۲</sup> (۲۰۱۱)، واژه‌نامه‌ای را برای تحلیل احساس متون از منظر تجاری و مالی ایجاد کردند. این واژه‌نامه اطلاعات ارزشمندی را در مورد محتوای گزارش‌های مالی شرکت‌های عضو بورس اوراق بهادار آمریکا، فراهم می‌کند. هستن و سینها<sup>۳</sup> (۲۰۱۵)، با استفاده از واژه‌نامه‌های LM<sup>۴</sup> و GI<sup>۵</sup> برای تحلیل احساس اخبار مالی و اقتصادی، نشان دادند که محتوای خبری می‌تواند برای پیش‌بینی بازده سهام به کار رود. شاپیرو<sup>۶</sup> و همکاران (۲۰۲۲)، با ترکیب واژه‌نامه‌های LM، GI، و HL<sup>۷</sup> نشان دادند که سری‌های زمانی حسی مبتنی بر اخبار اقتصادی می‌تواند برای پیش‌بینی به‌هنگام متغیرهای اقتصاد کلان در آمریکا به‌کار گرفته شوند. رویکرد سوم در مطالعات اقتصادی برای کشف اطلاعات از متون، به‌کارگیری روش‌های یادگیری ماشین با نظارت است. در این روش، نخست مجموعه‌ای از اسناد توسط ارزیابان متخصص، برچسب‌گذاری می‌شوند. سپس، با استفاده از این مجموعه داده‌های برچسب‌دار، الگوریتم‌های یادگیری ماشین به نحوی آموزش داده می‌شوند که بتوانند الگوهای موجود در اسناد را شناسایی کرده و برای اسناد جدید، برچسب‌های مربوطه را پیش‌بینی کنند. برای نمونه حبیبی نیکجو و همکاران (۱۴۰۲)، برای ساخت شاخص نااطمینانی برای اقتصاد ایران نخست تعدادی از اخبار که از شبکه‌های اجتماعی استخراج کرده بودند را توسط ارزیابان انسانی بر حسب اثرگذاری بر نااطمینانی، برچسب‌گذاری کردند. سپس این برچسب‌گذاری را با الگوریتم‌های یادگیری ماشین مختلف روی تمامی اخبار تعمیم دادند. نتایج نشان داد که شاخص آن‌ها با روند شواهد تاریخی و اتفاقات مهم سیاسی و اقتصادی همخوانی دارد.

## روش پژوهش

### مدل‌سازی موضوعی با روش تخصیص پنهان دیریکله<sup>۸</sup> (LDA)

روش‌های مدل‌سازی موضوعی، به‌منظور کشف مضامین پنهان و ساختارهای موضوعی در مجموعه بزرگی از اسناد مبتنی بر متن (پیکره متنی) با استفاده از الگوریتم‌های ریاضی طراحی شده‌اند. این روش‌ها بر این مفهوم استوارند که در پس‌زمینه انتخاب هر کلمه از میان ذخیره

1. Tetlock
2. Loughran & McDonald
3. Heston & Sinha
4. General Inquirer Dictionary Harvard
5. Loughran-McDonald
6. Shapiro
7. Hu and Liu lexicon
8. Latent Dirichlet Allocation (LDA)

واژگان و محل قرارگیری آن کلمه در متن، هدف و قصدی نهفته است و مدل‌سازی ریاضی ارتباط کلمات در یک متن، امکان آشکارسازی مفاهیم، مفروضات یا مقاصد زیربنایی را فراهم می‌آورد که بلافاصله قابل مشاهده نیستند (والدز<sup>۱</sup> و همکاران، ۲۰۱۸). روش تخصیص پنهان دیریکله (LDA) رایج‌ترین این روش‌ها است (وانگ<sup>۲</sup> و همکاران، ۲۰۱۹). در این روش، هر پیکره متنی شامل تعدادی کلمه است که می‌توان به هر کلمه آن، تعدادی موضوع با احتمال مشخص نسبت داد، به طوری که در پایان این کلمات از طریق ترکیب شدن با یکدیگر، یک متن و موضوع آن را تشکیل می‌دهند. برخلاف روش‌های کلاسیک خوشه‌بندی، که در آن‌ها عضویت یک متغیر به صورت دودویی است، هر واحد (کلمه) تا حدی به همه خوشه‌ها (موضوعات) با احتمالات مختلف تعلق دارد و به‌طور مشابه، هر موضوع نیز تا حدی به تمام اسناد با احتمالات متفاوت تعلق دارد (رایزنیکلر و ریاترر<sup>۳</sup>، ۲۰۱۹). خروجی مدل‌سازی پایه در LDA گروه‌هایی از کلمات (یعنی موضوعات) با احتمال عضویت و نسبت این موضوعات در هر سند است و به‌طور معمول، خروجی به تعداد کمی از کلمات و موضوعات با بیشترین احتمال محدود می‌شود و یک برجسب موضوع به صورت دستی توسط محقق به هر موضوع اختصاص می‌یابد.

#### تحلیل احساس با رویکرد واژه‌نامه محور<sup>۴</sup>

تحلیل احساس، یک حوزه میان‌رشته‌ای است که ریشه‌های خود را در پردازش زبان طبیعی، زبان‌شناسی، روان‌شناسی و هوش مصنوعی دارد. این حوزه، تلاش دارد تا ورای رسانه‌هایی که وظیفه انتقال مفاهیم را دارند، به کشف و درک احساس نهفته در آن‌ها مبادرت کند. اگرچه تحلیل احساس برای کشف احساس در محمل‌های اطلاعات همچون صوت، عکس و ویدیو نیز به کار می‌رود، اما بیشتر پژوهش‌ها صرف کشف احساس در متون طبیعی شده است (کرامت‌فر، ۱۴۰۰). تشخیص قطبیت، اساسی‌ترین و مشهورترین وظیفه در تحلیل احساس است تا جایی که در بسیاری از پژوهش‌ها، تشخیص قطبیت و تحلیل احساس به‌طور جایگزین استفاده می‌شوند. منظور از قطبیت سند، احساس و ارزیابی است که آن سند در ذهن ایجاد می‌کند. خروجی این وظیفه می‌تواند یک طبقه‌بندی باینری (مثبت یا منفی) یا طبقه‌بندی چندگانه (بسیار منفی، منفی، خنثی، مثبت یا بسیار مثبت) باشد (رجبی و همکاران، ۱۴۰۱).

واژه‌نامه‌های زیادی برای تعیین قطبیت کلمه‌ها وجود دارند. هر واژه‌نامه، سازوکار خاصی را برای تعیین قطبیت استفاده می‌کند و روش خاصی را برای نشان‌دادن قطبیت کلمه‌ها به

1. Valdez
2. Wang
3. Reisenbichler & Reutterer
4. Lexicon-based

کار می‌برد. برای نمونه، کلماتی مانند "خوب"، "زیبا" و "شگفت‌انگیز" احساسات مثبتی را به انسان منتقل می‌کنند، در حالی که کلماتی همچون "بد"، "زشت" و "ترسناک" دارای قطبیت منفی هستند. نکته مهم این است که بیشتر کلمات دارای بار حسی، در واقع صفت و قید هستند اما برخی نام‌ها مانند "زباله" و افعالی مثل "نفرت داشتن" و "عشق ورزیدن" نیز می‌توانند دارای بار حسی باشند. این کلمات ممکن است دارای وزن یا بدون وزن باشند. وزن در این جا به معنای یک عدد یا احتمال است که برای هر کلمه در نظر گرفته می‌شود تا سطح مثبت یا منفی بودن آن کلمه را نشان دهد. این رویکرد بیشتر اوقات برای محاسبه جهت‌گیری اسناد با توجه به جهت‌گیری معنایی کلمات و عبارات درون اسناد استفاده می‌شود (همتیان و سهرابی، ۲۰۱۹). سنتی استرنگت<sup>۱</sup>، از جمله این واژه‌نامه حسی است که توسط تلوال<sup>۲</sup> و همکاران (۲۰۱۰)، برای زبان انگلیسی و در ادامه برای سایر زبان‌ها از جمله زبان فارسی، توسعه پیدا کرده است. سنتی استرنگت، احساس کلی یک متن طولانی را با احساسات مثبت و منفی کلی هر جمله محاسبه می‌کند و مانند سایر الگوریتم‌هایی که تحلیل احساس بر مبنای واژه‌نامه را به کار می‌برند بر این منطق استوار است که نویسنده‌ای که احساس منفی نسبت به یک موضوع دارد برآیند کلماتی که استفاده می‌کند منفی‌تر است و برعکس.

### روش یادگیری ماشین<sup>۴</sup> تقویت گرادیان<sup>۵</sup>

یادگیری ماشین یکی از زیرمجموعه‌های هوش مصنوعی است که به رایانه‌ها این امکان را می‌دهد تا بدون این که به‌طور صریح برنامه‌ریزی شوند، از داده‌ها بیاموزند و به‌طور خودکار عملکردشان را بهبود دهند. در واقع، یادگیری ماشین به الگوریتم‌ها و تکنیک‌هایی اشاره دارد که به کمک آن‌ها سیستم‌ها می‌توانند الگوها را شناسایی کنند و پیش‌بینی‌هایی بر اساس داده‌های ورودی انجام دهند. روش تقویت گرادیان یکی از روش‌های محبوب در یادگیری ماشین برای پیش‌بینی است. این روش به‌صورت تکراری، مدل‌های ضعیف را با هم ترکیب می‌کند تا یک مدل قوی‌تر را ایجاد کند. اساس این روش بر این اصل استوار است که مدل‌های ضعیف می‌توانند به‌صورت تجمیعی، عملکرد مناسبی داشته باشند.

### مجموعه داده‌ها

در این مطالعه، از داده‌های شاخص مدیران خرید ماهانه، مستخرج از پایگاه اینترنتی اتاق بازرگانی، صنایع، معادن و کشاورزی ایران در بازه زمانی مهرماه ۱۳۹۸ تا اسفندماه سال ۱۴۰۲،

1. Hemmatian & Sohrabi
2. SentiStrength
3. TheWall
4. Machine Learning
5. Gradient Boosting

استفاده شده است. اخبار اقتصادی نیز در همان دوره زمانی و از پایگاه اینترنتی خبرگزاری فارس استخراج شده‌اند. انتخاب خبرگزاری فارس با توجه به سه معیار قابلیت دسترسی و امکان استخراج خبرها به صورت برخط، وجود آرشیو اخبار در دوره مورد بررسی و مرجع بودن خبرگزاری انجام شده است. مرجع بودن خبرگزاری با توجه به معیارهای تحلیل دامنه<sup>۱</sup> و رتبه‌بندی تارنمای آکسا<sup>۲</sup> تعیین شده است. بر اساس این رتبه‌بندی، خبرگزاری فارس با توجه به ترافیک ورودی تارنما در سال ۱۴۰۰ رتبه ۴۲ را در میان تمامی تارنماهای فارسی زبان داشته است و در سال ۱۳۹۹ در مقایسه با سه خبرگزاری پربازدید مهر، ایسنا و ایرنا در رتبه دوم قرار گرفته است.

### یافته‌های پژوهش

نخست با روش خزش<sup>۳</sup>، ۵۰/۲۱۶ خبر اقتصادی در دوره زمانی مورد بررسی از پایگاه اینترنتی خبرگزاری فارس استخراج شد. در ادامه، با استفاده از روش‌های متن‌کاوی، پیش‌پردازش متن<sup>۴</sup> به منظور یک‌دست کردن داده‌های متنی خبری و تبدیل آن‌ها به قالبی کاربردی برای پردازش‌های بعدی انجام شد. در مرحله پیش‌پردازش، از کتابخانه پارسی‌ور<sup>۵</sup> و سه ابزار نرمال‌سازی متن (برای استانداردسازی و یکپارچه‌سازی متون)، واحدساز<sup>۶</sup> (تبدیل و تفکیک جملات به کلمات (توکن)) و حذف ایست‌واژه‌ها<sup>۷</sup> (حذف واژه‌هایی که مفهومی را منتقل نمی‌کنند (به منظور افزایش دقت و سرعت مدل‌ها) استفاده شد. در گام بعدی با استفاده از الگوریتم LDA، موضوعات اصلی متون خبری شناسایی شد. جدول ۱، این موضوعات را به همراه ۱۰ کلمه اصلی هر موضوع نشان می‌دهد. برچسب هر موضوع با توجه به ارتباط بین کلمات به صورت دستی انتخاب شده است.

پژوهشگاه علوم انسانی و مطالعات فرهنگی  
پرتال جامع علوم انسانی

1. Domain Authority
2. Alexa
3. Web Crawling
4. Text Preprocessing
5. Parsivar
6. Word Tokenization
7. Stop Words

جدول ۱. نتایج مدل‌سازی موضوعی با LDA روی اخبار اقتصادی

ردیف	برجسب موضوع	کلمه ۱	کلمه ۲	کلمه ۳	کلمه ۴	کلمه ۵	کلمه ۶	کلمه ۷	کلمه ۸	کلمه ۹	کلمه ۱۰
۱	بورس	سهام	بورس	سهم	سرمایه	بازار	اوراق	شاخص	معاملات	بهادار	فرا بورس
۲	پول و بانک	بانک	ارز	بانکی	نرخ	تسهیلات	سامانه	مجلس	کمیسیون	ارزی	سپرده
۳	انرژی	نفت	گاز	نفتی	بشکه	انرژی	اوپک	جهانی	قیمت	بنزین	پتروشیمی
۴	اقتصاد بین‌الملل	آمریکا	ایران	تحریم	گاز	نفت	اروپا	چین	روسیه	جهان	تجاری
۵	صنعت و کشاورزی	صنعت	کشاورزی	تعاون	صنایع	محصولات	معدن	گندم	وزارت	بارش	هواشناسی
۶	حمل و نقل	حمل و نقل	قطار	ناوگان	بنادر	جاده	سوخت	مسافر	تردد	محور	ترافیک
۷	بخش عمومی	بودجه	مجلس	دولت	کمیسیون	قانون	یارانه	سازمان	مالیات	لاایحه	وزارت
۸	توسعه	توسعه	پروژه	صندوق	اعتبار	احداث	مسکن	برق	شهرسازی	فرودگاه	ساخت
۹	فناوری	تلفن	اطلاعات	ارتباطات	فناوری	سرمایه	پروژه	وزیر	شرکت	شبکه	مدیرعامل
۱۰	بازار	قیمت	بازار	نرخ	بازارهای	طلا	دلار	تومان	ارزش	معاملات	سکه

بعد از تخصیص یک موضوع مشخص به هر خبر، نمره حسی نیز با استفاده از نسخه جاوای واژه‌نامه سنتی استرنگت به هر خبر تخصیص داده شد. همان‌طور که پیش‌تر گفته شد این ابزار از یک واژه‌نامه از کلمات حسی و وزن آن‌ها که به صورت دستی توسط متخصصان زبان‌شناسی انتخاب شده است، استفاده می‌کند. در تحلیل احساس هر خبر، از تمامی اجزای آن خبر شامل عنوان، خلاصه و متن اصلی استفاده شده است. با تجمیع نمره حسی اخبار در هر موضوع، به صورت ماهانه و سپس تفاضل‌گیری امتیاز حسی تجمیع شده از مقدار تجمعی ماه قبل، ۱۰ سری زمانی مبتنی بر احساس اخبار ساخته شده است. بعد از ساخت سری‌های زمانی حسی، مدل تقویت‌گرادیان با استفاده از ۹۰ درصد داده‌ها (۴۸ ماه نخست)، شامل مقادیر واقعی شاخص مدیران خرید و سری‌های زمانی حسی خبری آموزش داده شد. برای سنجش عملکرد مدل، مدل با استفاده از داده‌های آزمون یا آزمایشی (۶ ماه آخر) برآورد شد. برای مقایسه عملکرد مدل، الگوهای سری زمانی تک‌متغیره نیز به عنوان مدل معیار برآورد شدند. این الگوها که تنها به مقادیر گذشته متغیره متکی هستند به صورت گسترده برای پیش‌بینی مقادیر متغیره‌های اقتصادی به کار می‌روند. جدول ۲، نتایج پیش‌بینی مدل با استفاده از داده‌های متون خبری و الگوهای سری زمانی تک‌متغیره را نشان می‌دهد.

**جدول ۲. مقایسه نتایج مدل‌های با داده‌های رسانه‌بنیان و مدل مرسوم اقتصادی در پیش‌بینی شاخص مدیران خرید از مهرماه سال ۱۴۰۲ تا اسفندماه ۱۴۰۲**

دوره‌های زمانی آزمون (سال / ماه)	میزان تأخیر در انتشار (روز)	مقدار واقعی شاخص مدیران خرید (درصد)	مدل مرسوم اقتصادی (سری زمانی تک‌متغیره)		مدل جدید با داده‌های رسانه‌بنیان (اخبار اقتصادی)
			الگو خود بازگشت مرتبه ۱	الگو خود بازگشت مرتبه ۴	روش الگوریتم یادگیری ماشین تقویت گرادیان
۱۴۰۲/۷	۲۰	۵۴/۱۵	۴۹/۹۸	۵۰/۴۵	۵۳/۴۵
۱۴۰۲/۸	۲۵	۵۰/۹۶	۴۹/۹۵	۵۰/۰۱	۵۳/۱۲
۱۴۰۲/۹	۱۹	۴۹/۸۸	۴۹/۳۴	۵۰/۶۸	۴۶/۳۶
۱۴۰۲/۱۰	۱۷	۵۰/۸۴	۴۹/۱۴	۵۰/۵۸	۴۶/۴۴
۱۴۰۲/۱۱	۱۶	۵۵/۰۵	۴۹/۳۲	۴۹/۸۹	۵۱/۵۵
۱۴۰۲/۱۲	۲۷	۵۵/۸۰	۵۰/۱۲	۵۰/۳۳	۵۱/۶۱
خطای پیش‌بینی (معیار RMSE)			۳/۷۹	۳/۴۶	۳/۳۳

مطابق نتایج به دست آمده، استفاده از داده‌های حسی اخبار در مقایسه با الگوی سری زمانی تک‌متغیره، باعث کاهش خطای پیش‌بینی می‌شود. از طرف دیگر، مدل با داده‌های رسانه‌بنیان در تشخیص رونق و رکود بهترین عملکرد را دارد به نحوی که کاهش و یا افزایش مقدار شاخص مدیران خرید را نسبت به ماه قبل در تمامی دوره‌ها به درستی تشخیص می‌دهد. روندهای مثبت احساسات که از اخبار اقتصادی به دست می‌آید به‌طور معمول با دوره‌های رونق اقتصادی هم‌راستا هستند، درحالی که احساسات منفی به رکود یا نااطمینانی مرتبط می‌شوند. به همین دلیل، تحلیل احساسات موجود در اخبار اقتصادی می‌تواند اطلاعات ارزشمندی درباره روندهای کوتاه‌مدت اقتصادی در ایران فراهم کند. همچنین با وجود ضرورت دسترسی به‌هنگام به آمار شاخص مدیران خرید برای فعالان اقتصادی، همان‌گونه که در ستون میزان تأخیر در انتشار رسمی شاخص مدیران خرید مشخص است، این آمار با تأخیر دو تا سه هفته‌ای بعد از پایان هر ماه منتشر می‌شود که این مسئله نیز با استفاده از تحلیل احساسات اخبار اقتصادی رفع می‌شود و به محض پایان هر ماه می‌توان برآوردی از این شاخص ارائه کرد.

## نتیجه گیری

امروزه، استفاد از داده‌های رسانه‌بنیان در مطالعات اقتصادی برای تحلیل و درک روندها، الگوها و پدیده‌های اقتصادی مورد توجه ویژه‌ای قرار گرفته است. این پژوهش نیز به منظور گسترش بدنه رو به رشد این مطالعات، اهمیت این نوع از داده‌ها را در رصد شرایط اقتصادی ایران بررسی کرده است. به همین منظور نخست ۵۰/۲۱۶ خبر روزانه اقتصادی از مهرماه سال ۱۳۹۸ تا اسفند سال ۱۴۰۲ از بستر اینترنت استخراج و پیش پردازش‌های اولیه شامل نرمال‌سازی، واحدسازی و حذف ایست‌واژه‌ها، روی آن‌ها انجام شده است. در ادامه با استفاده از دو ابزار مدل‌سازی موضوعی و تحلیل احساس و اژانه‌محور، شاخص‌های کمی مبتنی بر این اخبار ایجاد و در پایان توانایی این شاخص‌ها در پیش‌بینی شاخص مدیران خرید که وضعیت عمومی صنعت و بخش‌های مختلف اقتصادی را نشان می‌دهد، ارزیابی شد. نتایج نشان داد که استفاده از اخبار افزون بر این‌که منجر به کاهش خطای پیش‌بینی می‌شود، به حل مشکل تأخیر در انتشار اطلاعات شاخص مدیران خرید نیز کمک می‌کند و به محض پایان هر ماه و با استفاده از اخبار اقتصادی منتشر شده در همان ماه، می‌توان پیش‌بینی به‌هنگام و کارایی از این شاخص و شرایط فعلی اقتصاد و کسب‌وکارها داشت.

با وجود مزیت‌های استفاده از داده‌های رسانه‌بنیان در پژوهش‌های اقتصادی، به‌کارگیری این داده‌ها محدودیت‌هایی نیز به همراه دارد. نخستین چالش حجم زیاد و سرعت تولید بالای این نوع از داده‌هاست که پردازش و مدیریت آن‌ها را دشوار می‌کند. پژوهشگران با تحقیق و توسعه مداوم بر روی روش‌های هوش مصنوعی و ارائه ابزارهای جدید یادگیری ماشین، به دنبال حل این چالش بوده‌اند. چالش دوم، وجود سوگیری‌ها در اطلاعات و اخبار منتشر شده توسط رسانه‌ها است. از آن‌جا که اطلاعات نادرست یا گزارش‌های جانبدارانه نتایج تحلیل‌ها را دچار انحراف می‌کنند، تضمین صحت و قابل اعتماد بودن داده‌های رسانه‌بنیان اهمیت زیادی دارد. یک راه برای حل این چالش این است که خروجی مدل‌های تولید شده با داده‌های رسانه‌بنیان با سایر منابع داده مقایسه و ارزیابی شود؛ همان‌طور که در این پژوهش، صحت نتایج تحلیل احساس اخبار با پیش‌بینی اطلاعات واقعی شاخص مدیران خرید ارزیابی شد. به عنوان پیشنهادهایی برای مطالعات آینده، می‌توان از تحلیل احساس اخبار اقتصادی و گفتمان رسانه‌های اجتماعی به منظور پیش‌بینی نوسانات بازار سهام و سایر بازارهای مالی، شناسایی الگوهای مصرف و پس‌انداز در خانوارها، پیش‌بینی شاخص‌های اقتصاد کلان و ارزیابی تأثیر سیاست‌ها و مقررات اقتصادی استفاده کرد. همچنین، می‌توان اطلاعات منتشر شده در شبکه‌های اجتماعی و انجمن‌های تخصصی را کمی‌سازی کرد و آن‌ها را با داده‌های اقتصادی مرسوم ترکیب و در مدل‌های اقتصادی به کار گرفت.

## منابع

- حبیبی نیکجو، حبیب؛ چشمی، علی و سلیمی فر، مصطفی (۱۴۰۲). اندازه‌گیری شاخص ناطمینانی اقتصادی رسانه‌بنیان با الگوریتم‌های یادگیری ماشین در ایران و تأثیر آن بر نرخ ارز. اقتصاد پولی مالی، ۲۹(۲۳)، ۱-۴۶. <https://doi.org/10.22067/mfe.2022.75682.1175>
- رجبی، زینب؛ ولوی، محمدرضا و حورعلی، مریم (۱۴۰۱). مروری بر روش‌های تحلیل احساس در متون فارسی. پردازش علائم و داده‌ها، ۱۹(۲)، ۱۰۷-۱۳۲. <https://doi.org/10.52547/jsdp.19.2.107>
- کبیری، امید؛ تقی‌پور، فائزه و پیکری، حمیدرضا (۱۴۰۲). سنجش تنوع ارتباطات انسانی در شبکه اجتماعی اینستاگرام بین کاربران ایرانی. رسانه، ۳۴(۳)، ۵۵-۸۱. <https://doi.org/10.22034/bmsp.2022.332601.1697>
- کرامت‌فر، عبدالصمد (۱۴۰۰). مدل‌سازی چندجریانی زمینه‌نظرات برای تحلیل احساس. پایان‌نامه دکتری، دانشگاه قم.
- نادرپیلهرود، مقصود و قاسمی، محمد (۱۴۰۲). تحلیل پیام‌های اینستاگرامی سلب‌ریتی‌ها درباره کرونا. رسانه، ۳۴(۳)، ۱۸۵-۲۰۶. <https://doi.org/10.22034/bmsp.2022.324092.1673>
- Algaba, A., Ardia, D., Bluteau, K., Borms, S., & Boudt, K. (2020). Econometrics meets sentiment: An overview of methodology and applications. *World Bank Economic Review*, 8(3), 351-371. <https://doi.org/10.1111/joes.12370>
- Ash, E., & Hansen, S. (2023). Text algorithms in economics. *Annual Review of Economics*, 15, 659-688. <https://doi.org/10.1146/annurev-economics-082222-074352>
- Baker, S. R., Bloom, N., & Davis, S. J. (2016). Measuring economic policy uncertainty. *The quarterly journal of economics*, 131(4), 1593-1636. <https://doi.org/10.1093/qje/qjw024>
- Bybee, L., Kelly, B. T., Manela, A., & Xiu, D. (2020). *The structure of economic news* (No. w26648). National Bureau of Economic Research. (No. w26648). <https://doi.org/10.3386/w26648>
- Consoli, S., Pezzoli, L. T., & Tosetti, E. (2021). Emotions in macroeconomic news and their impact on the european bond market. *International Money and Finance*, 118, 102472. <https://doi.org/10.1016/j.jimonfin.2021.102472>
- Coase, R. H. (1960). The problem of social cost. *J. Law Econ.* 3, 1-44.
- Da, Z., Engelberg, J., & Gao, P. (2015). The sum of all FEARS investor sentiment and asset prices. *The Review of Financial Studies*, 28(1), 1-32. <https://doi.org/10.1093/rfs/hhu072>
- Friedman, M., & Schwartz, A. J. (1963). *A Monetary History of the United States: 1867-1960*. Princeton, NJ: Princeton Univ Press.
- Habibi, N, H., Cheshomi, A., Salimifard, M. (2023). Measuring the Media-based Economic Uncertainty index by Machine Learning Algorithms in Iran and its Effect on the Exchange Rate. *Monetary And Financial Economics*, 29(23), 1-46. [In persian] <https://doi.org/10.22067/mfe.2022.75682.1175>
- Hemmatian, F., & Sohrabi, M. (2019). A survey on classification techniques for opinion mining and sentiment analysis. *Artificial intelligence review*, 52(3), 1495-1545. <https://doi.org/10.1007/s10462-017-9599-6>
- Heston, S. L., & Sinha, N. R. (2015). News versus sentiment: Predicting stock returns from news stories. Robert H. Smith School Research Paper. <http://dx.doi.org/10.2139/ssrn.2311310>
- Kabiri, O., Taghipour, F., Peikari, H. R. (2023). Measuring the Diversity of Human Communication in the Instagram Social Network among Iranian Users. *Rasaneh*, 34(3), 55-81. [In persian] <https://doi.org/10.22034/bmsp.2022.332601.1697>
- Keramafar, A. (2021). Multi-stream modeling of comments' contexts for sentiment analysis. Ph.D. Thesis, University of Qom. [In persian]
- Koopmans, T. C. (1947). Measurement without theory. *The Review of Economics and Statistics*, 29(3), 161-172.
- Laudati, D., & Pesaran, M. H. (2023). Identifying the effects of sanctions on the Iranian economy using newspaper coverage. *Applied Econometrics*, 38(3), 271-294. <https://doi.org/10.1002/jae.2947>

- Loughran, T., & McDonald, B. (2011). When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks. *The Journal of finance*, 66(1), 35-6. <https://doi.org/10.1111/j.1540-6261.2010.01625.x>
- Martí, P., Serrano-Estrada, L., & Nolasco-Cirugeda, A. (2019). Social media data: Challenges, opportunities and limitations in urban studies. *Computers, Environment and Urban Systems*, 74, 161-174. <https://doi.org/10.1016/j.compenvurbsys.2018.11.001>
- Nader Pilehroud, M., & Ghasemi, M. (2023). Analysis of Celebrities' Instagram Messages about Corona. *Rasaneh*, 34(3), 185-206. [In persian] <https://doi.org/10.22034/bmsp.2022.324092.1673>
- Parkinson, S. E. (2024). Unreported Realities: The Political Economy of Media-Sourced Data. *American Political Science Review*, 118(3), 1527-1532. <https://doi.org/10.1017/S0003055423001181>
- Rajabi, Z., Valavi, M., & Hourali, M. (2022). Sentiment analysis methods in Persian text: A survey. *Signal and Data Processing*, 19(2), 107-132. [In persian] <https://doi.org/10.52547/jsdp.19.2.107>
- Reisenbichler, M., & Reutterer, T. T. (2019). Topic modeling in marketing: recent advances and research opportunities. *Business Economics*, 89(3), 327-356. <https://doi.org/10.1007/s11573-018-0915-7>
- Rodrigo-Ginés, F. J., Carrillo-de-Albornoz, J., & Plaza, L. (2024). A systematic review on media bias detection: What is media bias, how it is expressed, and how to detect it. <https://doi.org/10.1016/j.eswa.2023.121641>
- Shapiro, A. H., Sudhof, M., & Wilson, D. J. (2022). Measuring news sentiment. *Econometrics*, 228(2), 221-243. <https://doi.org/10.1016/j.jeconom.2020.07.053>
- Shiller, R. J. (2019). *Narrative economics: How stories go viral and drive major economic events*. Princeton University Press.
- Tetlock, P. C., Saar-Tsechansky, M., & Macskassy, S. (2008). More than words: Quantifying language to measure firms' fundamentals. *finance*, 63(3), 1437-1467. <https://doi.org/10.1111/j.1540-6261.2008.01362.x>
- Thelwall, M., Buckley, K., Paltoglou, G., Cai, D., & Kappas, A. (2010). Sentiment strength detection in short informal text. *American society for information science and technology*, 61(12), 2544-2558. <https://doi.org/10.1002/asi.21416>
- Valdez, D., Pickett, A. C., & Goodson, P. (2018). Topic modeling: latent semantic analysis for the social sciences. *Social Science Quarterly*, 99(5), 1665-1679. <https://doi.org/10.1111/ssqu.12528>
- Wang, H., Wang, J., Zhang, Y., Wang, M., & Mao, C. (2019). Optimization of Topic Recognition Model for News Texts Based on LDA. *J. Digit. Inf. Manag.*, 17(5), 257.

© Authors, Published by Department of Media Development Education. This is an open-access paper distributed under the CC BY (license <http://creativecommons.org/licenses/by/4.0/>).

