Applied Theories of Economics, Vol. 11, No. 3, Atumn 2024

Nowcasting Iran's GDP Using Sentiment Analysis of Economic News

Morteza Beiranvand[1]
Seyed Saeed Malek Sadati[2]
Seyed Mohammad Javad Razmi[3]

Abstract

This study examines textual data's ability to nowcast Iran's gross domestic product (GDP). To this end, 301,498 economic news articles from March 2005 to December 2023 were extracted from the Fars news agency website using a web crawling technique. Following initial preprocessing, the news texts were sorted into various categories via the Dirichlet Latent Allocation (LDA) model, wherein each category corresponds to a distinct news topic. Subsequently, to ascertain whether an article conveys a positive or negative sentiment, we executed lexicon-based sentiment analysis utilizing SentiStrength. Ultimately, by aggregating the news sentiment scores seasonally under each topic, we constructed a seasonal sentiment time series. These time series were then assessed for their efficacy in nowcasting Iran's quarterly GDP, employing ridge regression, lasso regression, elastic net, and gradient boosting methods. The findings reveal that incorporating textual data can reduce prediction errors by 12 to 18 percent relative to a univariate time series model. Moreover, our results suggest that sentiment extracted from textual content, particularly news articles, is a viable approach. This strategy could potentially enable the provision of immediate GDP estimates following the end of each reference quarter.

Keywords: *Economic News, Nowcasting, GDP, Topic Modeling, Sentiment Analysis.*
JEL Classification: *E17, C53, C55.*

[1] Ph.D. Candidate in Economics, Ferdowsi University of Mashhad. beiranvand.mo@gmail.com
[2] Assistant Professor of Economics, Ferdowsi University of Mashhad. msadati@um.ac.ir
[3] Professor of Economics, Ferdowsi University of Mashhad. mjrazmi@um.ac.ir

Introduction

When monitoring and assessing the state of the economy in real-time, policymakers face the problem that Gross Domestic Product (GDP), the most comprehensive measure of the economy, is released with a lag. For instance, in Iran, the first (flash) GDP estimate for a reference quarter is released 60-90 days after the quarter's close. To address this lag, various forecasting and nowcasting methods have been developed. The toolkit for nowcasting and forecasting generally employs various models consisting of quantitative and qualitative data or so-called hard and soft data. The former refers to direct measures of economic activity such as industrial production, retail sales, and the unemployment rate. Soft data such as surveys provide qualitative assessments reflecting sentiment and expectations. Examples of such surveys include the Purchasing Managers Index (PMI)

In recent years, the rise of big data and technological advancements has led to an increasing interest in incorporating alternative data sources to enhance the performance of forecasting and nowcasting models. Alternative data refers to non-traditional or unconventional data sources used in analysis, decision-making, or research alongside or instead of traditional data sources such as official statistics or surveys. Alternative data can come from a wide range of sources, including but not limited to satellite imagery, social media posts, online search trends, credit card transactions, Transportation data, Energy consumption data, and more. These data sources offer unique and often real-time insights into various aspects of human behavior, economic activity, market trends, and other phenomena. The use of alternative data has gained prominence with the increasing availability of big data and advancements in data processing and analytic techniques, allowing for new and innovative approaches to gaining insights and making predictions in different fields.

News texts, as one of the alternative data sources and one of the largest types of big data, are a key source of information and a way to assess the sentiments and opinions of journalists, analysts, and experts regarding economic conditions. Recent advancements in statistical and computational methods have enabled economic researchers to access and process large volumes of this textual data. This research aims to expand the growing body of economic studies demonstrating how big data and unconventional methods can improve economic predictions, particularly in the short term. This goal has been achieved through the use of news texts in Persian and the application of new text mining and sentiment analysis methods.

Methodology

The corpus utilized in this study comprises 301,498 economic news articles from the Fars News Agency, covering the period from March 2005 to November 2023. This corpus was compiled using web crawling techniques to systematically extract data from the Fars News Agency's website. The news data are decomposed into time series representing newspaper topics using a Latent Dirichlet Allocation (LDA) model. This model is an unsupervised machine-learning algorithm that learns the underlying topics of a set of documents. It is based on a generative probabilistic approach to inferring the distribution of words that defines a topic while simultaneously annotating articles with a distribution of topics. To automatically detect whether the news article is conveying a positive or negative tone, we perform a lexicon-based sentiment analysis. SentisStrength is the lexicon used in this research. This tool utilizes a lexicon of emotion words and their Ultimately, by weights, which linguistics experts have manually selected. aggregating the news sentiment scores seasonally under each topic, we constructed a seasonal sentiment time series. These time series were then assessed for their efficacy in nowcasting Iran's quarterly GDP, employing ridge regression, lasso regression, elastic net, and gradient boosting methods.

Results and Discussion

The findings of this study demonstrate that incorporating textual data into Gross Domestic Product (GDP) nowcasting models can significantly reduce prediction errors by 12 to 18 percent compared to traditional univariate time series models. Using sentiment analysis in this way makes it possible to provide GDP estimates promptly after the end of each reference quarter. Unlike official statistics, news articles are published daily and do not experience publication delays. Furthermore, the online availability of news from various agencies greatly minimizes the challenges of accessing this type of textual data, unlike other potential big data sources for evaluating economic conditions, such as transactions from electronic payment systems. This approach allows policymakers and economic analysts to use real-time data from news articles to gain immediate insights into economic performance, facilitating more informed decision-making and timely policy interventions.

فصلنامه نظریه‌های کاربردی اقتصاد/ سال یازدهم/ شماره ۳/ پاییز ۱۴۰۳/ صفحات ۱۶۴-۱۳۵

پیش‌بینی به‌هنگام تولید ناخالص داخلی ایران با استفاده از تحلیل احساس اخبار اقتصادی[1]

مرتضی بیرانوند

دانشجوی دکتری اقتصاد دانشگاه فردوسی مشهد، beiranvand.mo@gmail.com

سید سعید ملک‌الساداتی*

استادیار اقتصاد دانشگاه فردوسی مشهد، msadati@um.ac.ir

سید محمدجواد رزمی

استاد اقتصاد دانشگاه فردوسی مشهد، mjrazmi@um.ac.ir

تاریخ دریافت: ۱۴۰۳/۰۴/۲۷ تاریخ پذیرش: ۱۴۰۳/۰۹/۰۴

چکیده

این پژوهش، روشی برای کمّی‌سازی اطلاعات بدون ساختار اخبار اقتصادی برای به‌کارگیری در ارزیابی به‌هنگام شرایط اقتصادی را پیشنهاد می‌دهد. به همین منظور، اخبار اقتصادی به صورت روزانه از ابتدای سال ۱۳۸۴ تا انتهای آذرماه سال ۱۴۰۲، از پایگاه اینترنتی خبرگزاری فارس استخراج شده است. متون خبری، پس از پیش‌پردازش اولیه، با استفاده از مدل تخصیص پنهان دیریکله (LDA) در دسته‌های مختلفی طبقه‌بندی شدند به نحوی که هر دسته، یک موضوع خبری را نشان می‌دهد. سپس با استفاده از رویکرد تحلیل احساس مبتنی بر واژه‌نامه، امتیاز یا نمره حسی هر خبر مشخص شده است. از تجمیع فصلی امتیازات حسی اخبار ذیل هر موضوع، سری‌های زمانی حسی ایجاد و توانایی این سری‌های زمانی در پیش‌بینی تولید ناخالص داخلی فصلی ایران با استفاده از روش‌های ریج، لسو، الستیک‌نت و تقویت گرادیان ارزیابی شده‌اند. نتایج نشان داده‌اند که به‌کارگیری داده‌های حسی می‌تواند خطای پیش‌بینی را بین ۱۲ تا ۱۸ درصد نسبت به الگوی سری زمانی تک‌متغیره کاهش دهد. علاوه بر این، با استفاده از رویکرد پیشنهادی این پژوهش می‌توان بلافاصله بعد از اتمام هر فصل مرجع و با استفاده از اخبار اقتصادی منتشر شده در همان فصل، برآوردی به‌هنگام از تولید ناخالص داخلی فصلی ارائه کرد.

واژه‌های کلیدی: اخبار اقتصادی، پیش‌بینی به‌هنگام، تولید ناخالص داخلی، مدل‌سازی موضوعی، تحلیل احساس.

طبقه‌بندی JEL: E17، C53، C55.

---

[1] این مقاله مستخرج از رساله دکترای نویسنده اول است.

* نویسنده مسئول مکاتبات

۱- مقدمه

یکی از مهمترین دغدغه‌های اقتصاددانان و سیاست‌گذاران، ریشه‌یابی نوسانات متغیرهای اقتصادی و به حداقل رساندن آن است. زیرا رشد اقتصادی در فضای باثبات اقتصادی امکان‌پذیر است (افشار و همکاران[1]، ۱۴۰۲). همچنین سیاست‌گذاران اقتصادی، به‌ویژه سیاست‌گذاران حوزه پولی، باید ارزیابی دقیقی از مدت‌زمان و نحوه اثرگذاری این سیاست‌ها بر بخش‌های مختلف اقتصادی داشته باشند؛ چرا که اعمال این سیاست‌ها می‌تواند آثار متفاوت و نامتوازنی در بخش‌های مختلف اقتصاد داشته باشد (آقانیا و همکاران[2]، ۱۴۰۱). شناسایی نوسانات اقتصادی و ارزیابی اثر سیاست‌ها در ایران، یک چالش مشخص دارد که عبارت است از: تأخیر در انتشار داده‌ها. اطلاعات و آمار مربوط به تولید ناخالص داخلی، به عنوان شاخصی کلیدی در اقتصاد کلان که سطح فعالیت اقتصادی کشور را منعکس می‌کند و همه تحلیل‌گران، تولیدکنندگان، سرمایه‌گذاران و انواع مختلف سازمان‌های دولتی و خصوصی، آن را پیگیری می‌کنند؛ با تأخیر قابل-توجهی منتشر می‌شود. این تأخیر، مانع از دستیابی به‌هنگام عوامل اقتصادی و سیاست-گذاران به اطلاعات مهم و ضروری است. اگر چه ابعاد این چالش در ایران گسترده‌تر است و مطابق با تقویم انتشارات و تجدیدنظرهای مرکز آمار ایران، برآورد اولیه تولید ناخالص داخلی فصلی با ۶۰ تا ۹۰ روز تأخیر منتشر می‌شود، اما این چالش تنها مختص به ایران نیست و ابعادی جهانی دارد. به عنوان مثال، باربا‌گلیا و همکاران[3] (۲۰۲۴) در تحقیق خود اشاره کرده‌اند که اداره آمار اروپا، برآورد اولیه تولید ناخالص داخلی فصلی را ۳۰ روز پس از پایان هر فصل منتشر می‌کند. برای رفع این مشکل و به دست آوردن برآورد اولیه از شاخص‌های کلیدی اقتصاد، اغلب از داده‌های به‌هنگام‌تری استفاده می-شود که با متغیر هدف ارتباط دارند (کینی و همکاران[4]، ۲۰۱۲).

طی سال‌های اخیر، پیشرفت فناوری، امکان به‌کارگیری منابع داده جدیدی را به منظور رصد و ارزیابی سطح فعالیت اقتصادی در زمان واقعی فراهم کرده است که به داده‌های جایگزین[5] معروف شده‌اند. داده‌های جایگزین، به منابع داده جدیدی اطلاق می‌شوند که

---

[1] Afshar et al. (2023)
[2] Aghania et al. (2023)
[3] Barbaglia et al.
[4] Keeney et al.
[5] Alternative Data

در مطالعات اقتصادی در کنار (یا به جای) منابع داده سنتی و متعارفی مانند آمار رسمی و پرسش‌نامه‌ها استفاده می‌شوند (منچادو[1]، ۲۰۲۳). داده‌های جایگزین مورد استفاده برای ارزیابی به‌هنگام وضعیت اقتصادی، طیف وسیعی از اطلاعات همچون داده‌های موتورهای جست‌وجوی اینترنتی (فرارا و سیمونی[2]، ۲۰۲۳)، داده‌های پرداخت الکترونیکی (گالبریث و تکاچ[3]، ۲۰۱۸)، داده‌های حمل‌ونقل (باربا‌گلیا و همکاران، ۲۰۲۳)، داده‌های مصرف انرژی (لورنس و روآ[4]، ۲۰۲۱)، تصاویر ماهواره‌ای (هیو و یائو[5]، ۲۰۲۲) و داده‌های متنی (ثورسرود[6]،۲۰۲۰) را در بر می‌گیرند. گرایش به استفاده از این نوع اطلاعات از مزایای آن نسبت به داده‌های آماری مرسوم ناشی می‌شود. از جمله این مزایا می‌توان به حجم بالای اطلاعات، سرعت جمع‌آوری بالا، تنوع زیاد و اعتبار بالا اشاره کرد.

متون خبری، به عنوان یکی از داده‌های جایگزین و از بزرگترین انواع کلان‌داده نه تنها یک منبع کلیدی اطلاعات هستند، بلکه راهی برای ارزیابی احساسات و عقاید روزنامه-نگاران، تحلیل‌گران و کارشناسان در مورد شرایط اقتصاد نیز می‌باشند. وقایع مهم اقتصادی معمولاً زمانی رخ می‌دهند که تفکر مشابهی در بین گروه‌های زیادی از مردم شکل بگیرد. متون خبری، ابزاری ضروری برای انتشار و گسترش این افکار مشابه هستند و مانند اکثر ارتباطات انسانی، اغلب محتوای روایی و حسی دارند. اخبار، توانایی انتشار تغییرات در روایت‌های در حال چرخش اقتصاد را دارند، بنابراین می‌توانند هم تغییر در احساسات را نشان دهند هم تغییرات در آنچه شیلر[7] (۲۰۲۰)، از آن به عنوان «تفکر مشابهی در میان گروه‌های زیادی از مردم» یاد می‌کند. دشواری در پردازش بلادرنگ این حجم عظیم از اطلاعات متنی، سبب شده است که محققان در زمینه‌های مختلف، در حال توسعه روش‌های آماری برای استخراج اطلاعات و تجزیه و تحلیل متون باشند. از جمله روش‌های پیشرو در تجزیه و تحلیل متون، روش تحلیل احساس[8] است

---

[1] Manchado
[2] Ferrara & Simoni
[3] Galbraith & Tkacz
[4] Lourenço & Rua
[5] Hu & Yao
[6] Thorsrud
[7] Shiller
[8] Sentiment Analysis

که با تبدیل اطلاعات حسی موجود در متون به داده‌های عددی، محققان را قادر می‌سازد تا اطلاعاتی را کشف کنند که در داده‌های کمّی مرسوم وجود ندارند.

این پژوهش با طراحی مدلی برای بررسی خودکار احساسات موجود در متون خبری اقتصادی، از روش‌های پردازش زبان طبیعی[1] و یادگیری ماشین بهره می‌گیرد تا ابزاری کارآمد برای رصد و پیش‌بینی به‌هنگام شرایط اقتصاد کلان ایجاد کند. این مدل، فرصت ارزیابی به‌هنگام شرایط اقتصادی را قبل از انتشار داده‌های رسمی برای سیاست‌گذاران و فعالان اقتصادی فراهم می‌کند. با توجه به کمبود پیکره متنی[2] فارسی در حوزه اخبار اقتصادی، در ابتدا یک پیکره متنی از اخبار اقتصادی در این پژوهش ایجاد شده است. در گام بعدی، به جای ساخت یک شاخص حسی منفرد از این اخبار، چندین سری زمانی در موضوعات مختلف اقتصادی با استفاده از روش مدل‌سازی موضوعی[3] ساخته می‌شود و برخلاف بسیاری از پژوهش‌های مشابه که تنها از عنوان خبر برای تعیین قطبیت یا نمره حسی استفاده کرده‌اند، نمره حسی هرخبر با روش تحلیل احساس واژه‌نامه‌محور[4] و با استفاده از تمامی اجزای آن خبر (عنوان، خلاصه، متن) محاسبه می‌شود. اگرچه تمرکز این پژوهش بر استفاده از متون خبری برای پیش‌بینی تولید ناخالص داخلی است، اما ایده‌های ارائه شده در به‌کارگیری کلان‌داده‌های متنی، قابلیت استفاده در پیش‌بینی به‌هنگام شاخص‌های مختلف اقتصادی را دارند.

۲- ادبیات موضوع

۲-۱- مبانی نظری

احساسات از نظر روان‌شناختی، یکی از خصوصیات ذهنی انسان است که نشان‌دهنده خواسته‌ها، اعتقادات و ادراکات حسی فرد است و از طریق ارتباط متنی، صوتی یا تصویری قابل درک است (مونزرو و همکاران[5]، ۲۰۱۴). به عنوان مثال، آهنگ و لحن صدای فرد، یا دفعات و مدت زمانی که حین سخن گفتن مکث می‌کند، دربردارنده

[1] Natural Language Processing (NLP)
[2] Text Corpus
[3] Topic Modeling
[4] Lexicon-based
[5] Munezero et al.

احساسات ضمنی فرد است. همین حالت برای حرکات صورت و بدن (زبان بدن)، کلمات نوشته شده در یک سند متنی و همچنین رنگ‌های موجود در یک تصویر وجود دارد.

حداقل از زمان کار کینز[1] (۱۹۳۶)، اقتصاددانان به تأثیر احساسات بر تصمیم‌گیری اقتصادی افراد توجه داشته‌اند. درک ارتباط بین احساسات و تصمیم‌گیری در سطوح خرد و کلان همچنان در نظریه‌های اقتصادی از اهمیت بالایی برخوردار است (آنجلیتوس و همکاران[2]، ۲۰۱۸). از نمونه‌های کلاسیک استفاده از تحلیل احساس می‌توان به آثار محققانی چون کوز[3] (۱۹۶۰) در زمینه اقتصاد حقوق و فریدمن و شوارتز[4] (۱۹۶۳) در زمینه اقتصاد سیاسی اشاره کرد (اش و هانسن[5]، ۲۰۲۳). یکی از دلایل اهمیت تحلیل احساس در اقتصاد این است که در بسیاری از موارد به کشف پیش‌گویی‌های خودمحقق‌کننده[6] کمک می‌کند. برای مثال، هنگامی که بدبینی مصرف‌کنندگان یا صاحبان کسب‌وکار نسبت به رشد اقتصادی افزایش می‌یابد، این احساسات می‌توانند به‌طور مستقیم به کاهش رشد اقتصادی منجر شوند. مثال مشخص‌تر، عملیات بانکی است. زمانی که سپرده‌گذاران نگرانی شدیدی درباره وضعیت سایر سپرده‌گذاران و ذخایر بانکی داشته باشند، احتمال وقوع فرار بانکی افزایش می‌یابد.

در اقتصاد سیاسی نیز، نقش پوشش خبری و لحن اخبار در تأثیرگذاری بر میزان سرمایه اجتماعی و اعتماد به دولت مورد توجه قرار گرفته است (اشبو[7]، ۲۰۱۰). رابرت شیلر به عنوان یکی از پیشگامان استفاده از داده‌های متنی و تحلیل آنها در زمینه اقتصاد با معرفی رویکرد «اقتصاد روایی[8]»، به اهمیت نقش روایت‌ها و داستان‌های عمدتاً متنی در احساسات و انتظارات عوامل اقتصادی پرداخته است. او بر این نکته تأکید دارد که روایت‌ها نقش تعیین‌کننده‌ای در شکل‌دهی سریع فرهنگ، جوّ فکری و رفتارهای اقتصادی ایفا می‌کنند و مطالعه آن‌ها می‌تواند توانایی ما را در پیش‌بینی رویدادهای اقتصادی و آماده‌سازی برای آن‌ها افزایش دهد. شیلر با بررسی روایت‌های عامه دوره‌های

---

[1] Keynes
[2] Angeletos et al.
[3] Coase
[4] Friedman & Schwartz
[5] Ash & Hansen
[6] Self-Fulfilling Prophecy
[7] Eshbaugh
[8] Narrative Economics

مختلف و به کمک علم همه‌گیرشناسی[1]، به بررسی ارتباط بین این روایت‌ها و رفتارهای اقتصادی و پیامدهای آن‌ها از جمله تورم، رکود، حباب‌های اقتصادی و حتی جنگ پرداخته است. برای مثال، شیلر نشان داده است که فراگیر شدن روایت‌هایی مانند منحنی لافر و دیگر داستان‌های طرف عرضه، به مطالبات شدید مردمی برای کاهش مالیات منجر شد. همچنین، تلفیق روایت‌های مرتبط با کاهش مالیات و کوچک‌تر شدن دولت، به پیشبرد جنبش اجتماعی موسوم به کارآفرینی کمک کرد. همان‌طور که اشاره شد استفاده از احساسات به عنوان متغیر یا پارامتر، قدمتی دیرینه در اقتصادسنجی دارد. اگر چه از لحاظ تاریخی، استفاده از پرسش‌نامه‌ها یا متغیرهای نماینده[2] برای تعیین کمیّت متغیرهای حسی غالب بوده است، اما در سال‌های اخیر، تحلیل احساسات موجود در داده‌های متنی، صوتی و تصویری به عنوان روشی جایگزین، رایج شده است. این رویکرد نوین به محققان اجازه می‌دهد تا بینش‌های جدیدی در مورد تأثیر احساسات بر رفتارهای اقتصادی به‌دست آورند.

دیجیتالی شدن رسانه‌های ارتباطی و افزایش حجم ذخیره‌سازی و قدرت پردازش رایانه‌ها، حجم عظیمی از داده‌ها که حاوی اطلاعات ارزشمندی برای تحلیل اقتصادی هستند را در دسترس قرار داده است. همین امر باعث ایجاد رشته جدیدی از تحقیقات اقتصادسنجی شده است که به استخراج احساس از کلان‌داده‌ها و تبدیل داده‌های احساسی کیفی به متغیرهای کمّی و سپس استفاده از آنها در تحلیل اقتصادسنجی و کشف روابط بین احساسات و سایر متغیرها می‌پردازد. این رشته نوظهور که به ترکیب احساسات و اقتصادسنجی می‌پردازد، به نام سنتومتریک[3] شناخته می‌شود (آلگبا و همکاران[4]، ۲۰۲۰).

تحلیل احساس، این توانایی را دارد که به حل یا درک بسیاری از مسائل مربوط به استفاده از اقتصادسنجی در زمینه‌های اقتصاد، مالی، حسابداری، بازاریابی و غیره کمک کند. شاخص‌های حسی مبتنی بر داده‌های کیفی می‌توانند ابزاری مستقیم برای ارزیابی انواع شوک‌های اقتصادی یا نماینده‌ای برای مفاهیمی مانند اطمینان یا انتظارات باشند.

---

[1] Epidemiology
[2] Proxies
[3] Sentometrics
[4] Algaba et al.

احساس، یک متغیر پنهان است، به این معنی که به راحتی قابل مشاهده نیست. اگر چه انتخاب روش مناسب تجزیه و تحلیل داده‌های حسی به هدف و کاربرد مطالعه بستگی دارد اما یک زمینه مشترک برای اقتصادسنجی به کار رفته در تحلیل احساس این است که ابتدا باید احساس را اندازه‌گیری کرد. در ادبیات اقتصادی، از سه رویکرد کلی برای کمّی‌سازی احساسات استفاده می‌شود. در رویکرد اول، احساسات با استفاده از شاخص‌های مبتنی بر نظرسنجی برآورد می‌شوند. در رویکرد دوم، با استفاده از روش‌های داده‌کاوی، احساسات به‌طور مستقیم از یک منبع داده کیفی استخراج می‌شوند و در رویکرد سوم، پژوهش‌گر، شاخص‌های حسی را که قبلا توسط یک توسعه-دهنده داده[1] و با محاسبات پیچیده محاسبه شده (برای مثال شاخص‌های روانشناختی بازار تامسون رویترز[2]) را استفاده می‌کند (آلگبا و همکاران، ۲۰۲۰). بعد از اندازه‌گیری احساس، این اطلاعات حسی در مدل‌های پیش‌بینی استفاده می‌شوند تا بهترین پیش‌بینی ممکن در هر زمان به دست آید. متغیرهای حسی را می‌توان به تنهایی به عنوان متغیرهای توضیحی در مدل‌ها استفاده کرد یا به مجموعه‌ای از سایر متغیرهای توضیحی استاندارد افزود. سپس، می‌توان بررسی کرد که آیا این ادغام، عملکرد پیش‌بینی را بهبود می‌بخشد یا خیر. فراتر از بهبود نتایج مدل‌های پیش‌بینی، استفاده از روش‌های تحلیل احساس در مقایسه با روش‌های سنتی استخراج احساسات مزایای دیگری نیز دارد. از جمله این مزایا، انعطاف‌پذیری بالاتر است. به این معنی که هرگونه تغییر و به‌روزرسانی در روش تحلیل احساس، به سادگی قابلیت انجام شدن و آزمایش با داده‌های موجود را دارد، درحالی‌که اصلاح ساختار یک نظرسنجی، نیاز به ارسال مجدد نظرسنجی برای به دست آوردن نتایج جدید دارد. از طرف دیگر، اطلاعات به دست آمده از داده‌های حسی مستخرج از تحلیل احساس، معمولاً وقفه انتشار ندارند. این ویژگی باعث می‌شود داده‌های حسی به متغیرهایی ایده‌آل برای بهبود عملکرد مدل‌های پیش‌بینی در زمان واقعی تبدیل شوند و در نتیجه، این امکان فراهم می‌شود که واکنش‌های سیاستی در زمان مناسب‌تر اعمال شوند. در مطالعات اقتصادی اگرچه صوت و تصویر نیز مبنایی برای کشف احساس بوده‌اند اما تحلیل احساس بیشتر بر روی اسناد متنی انجام شده است.

---

[1] Data Developer
[2] Thomson Reuters MarketPsych Indices

۲-۲- پیشینه پژوهش

باربا‌گلیا و همکاران (۲۰۲۴)، با استفاده از تحلیل احساس متون خبری، به پیش‌بینی متغیرهای اقتصاد کلان از جمله تولید ناخالص داخلی برای پنج اقتصاد بزرگ اروپا پرداخته‌اند. نتایج پژوهش آنها نشان می‌دهد که شاخص‌های به‌هنگام ساخته شده از تحلیل احساس اخبار اقتصادی، پیش‌بینی‌کننده‌های مهمی برای ارزیابی سطح فعالیت اقتصادی هستند. در پژوهش دیگری که توسط اپریگلیانو و همکاران[1] (۲۰۲۳) انجام شده است، نویسندگان در مطالعه خود، تحلیل احساس مبتنی بر اخبار اقتصادی را برای ردیابی تحولات کوتاه‌مدت شرایط اقتصادی پیشنهاد کرده‌اند. همچنین، نویسندگان برای پاسخ به این سوال که آیا می‌توان از اخبار روزنامه‌ها برای پیش‌بینی سطح فعالیت اقتصادی استفاده کرد یا خیر، یک واژه‌نامه اقتصادی برای زبان ایتالیایی ساختند و با به‌کارگیری آن روی حدود ۱/۵ میلیون خبر، شاخصی مبتنی بر احساس را برای کشور ایتالیا ایجاد کردند که با سیکل‌های تجاری و اتفاقات مهم شکل‌دهنده آن مطابقت دارد. شپیرو و همکاران[2] (۲۰۲۲)، با استفاده از رویکرد مبتنی بر واژه‌نامه، سری‌های زمانی حسی را با استفاده از تحلیل احساس اخبار اقتصادی در دوره زمانی ۱۹۸۰-۲۰۱۵ ایجاد کرده‌اند و با استفاده از روش‌های یادگیری ماشین نشان داده‌اند که این سری‌های زمانی می‌توانند برای پیش‌بینی به‌هنگام متغیرهای اقتصادی به کار گرفته شوند. در پژوهشی دیگر، کالامارا و همکاران[3] (۲۰۲۲)، با تحلیل احساس اخبار اقتصادی سه روزنامه بریتانیا نشان داده‌اند که استفاده از تحلیل احساس اخبار اقتصادی، پیش‌بینی‌های متغیرهای کلان اقتصادی از جمله تولید ناخالص داخلی را به‌طور قابل توجهی بهبود می‌بخشد. بورتولی و همکاران[4] (۲۰۱۸)، با تحلیل احساس بیش از یک میلیون مقاله خبری منتشر شده در روزنامه لوموند در دوره زمانی ۱۹۹۰-۲۰۱۷ و ساخت یک شاخص حسی، نشان داده‌اند که احساسات نهفته در رسانه‌ها در مورد وضعیت اقتصاد، خطای پیش‌بینی تولید ناخالص داخلی فرانسه را کاهش می‌دهد.

---

[1] Aprigliano et al.
[2] Shapiro et al.
[3] Kalamara et al.
[4] Bortoli et al.

مطالعات دیگری نشان داده‌اند که در مواقعی که اقتصاد با اتفاقات پیش‌بینی نشده‌ای مواجه می‌شود، عملکرد مدل‌هایی که از شاخص‌های حسی استفاده می‌کنند در تشخیص به‌هنگام سیکل‌های تجاری از مدل‌های مرسوم بهتر است. برای مثال، اگیلار و همکاران[1] (۲۰۲۱)، با استفاده از روش‌های متن‌کاوی، یک شاخص مبتنی بر احساس ایجاد کرده‌اند که جهت‌گیری اخبار اقتصادی را به تصویر می‌کشد. این شاخص، منعکس‌کننده تعداد مقالات خبری حاوی کلمات کلیدی مرتبط با رونق و رکود در چرخه تجاری اسپانیا است و نسبت به شاخص‌های مرسوم در پیش‌بینی به‌هنگام تولید ناخالص داخلی اسپانیا در زمان همه‌گیری کرونا، عملکرد بهتری دارد. آشوین و همکاران[2] (۲۰۲۱) نیز با استفاده از اخبار اقتصادی ۱۵روزنامه اروپایی، که به صورت ماشینی به انگلیسی ترجمه شده‌اند، شاخص‌های حسی ایجاد کرده‌اند که می‌توانند به‌طور قابل‌توجهی پیش‌بینی‌های به‌هنگام از رشد تولید ناخالص داخلی حقیقی برای اتحادیه اروپا (منطقه یورو) را بهبود دهند.

در مطالعات پیش‌بینی اقتصادی که از تحلیل احساس متون خبری استفاده می‌کنند، علاوه بر مطالعاتی که حس اخبار را تنها در یک شاخص حسی منفرد خلاصه کرده‌اند، مطالعاتی نیز سری‌های زمانی حسی را بر مبنای موضوع اخبار و با استفاده از روش‌های متن‌کاوی ساخته‌اند. برای مثال، بایبی و همکاران[3] (۲۰۲۰)، محتوای متن کامل ۸۰۰،۰۰۰ مقاله مربوط به وال‌استریت ژورنال[4] برای دوره زمانی ۱۹۸۴-۲۰۱۷ را به عنوان مضامین موضوعی که به راحتی قابل تفسیر هستند خلاصه‌سازی و سپس نسبت پوشش خبری هر موضوع در هر مقطع زمانی را کمّی‌سازی کرده‌اند. آن‌ها نشان داده‌اند که به‌کارگیری موضوعات خبری در مقایسه با شاخص‌های استاندارد، حاوی اطلاعات مهم‌تری از نظر قدرت پیش‌بینی در مورد وضعیت آتی اقتصاد است. ثورسرود (۲۰۲۰) نیز اخبار اقتصادی نروژ را در دوره ۱۹۸۸-۲۰۱۴ از یک روزنامه اقتصادی استخراج و با روش LDA این داده‌های متنی را به سری‌های زمانی موضوعی تجزیه کرده است. در گام بعدی، با تحلیل احساس این سری‌های زمانی و افزودن نمره حسی به اخبار، یک

[1] Aguilar et al.
[2] Ashwin et al.
[3] Bybee et al.
[4] The Wall Street Journal (WSJ)

شاخص چرخه تجاری روزانه ایجاد شده است که استفاده از آن در یک مدل عامل پویا، چرخه تجاری حقیقی را در هر دوره با دقت بالایی پیش‌بینی می‌کند.

۳- روش‌شناسی

۳-۱- مدل‌سازی موضوعی با روش تخصیص پنهان دیریکله[1] (LDA)

روش‌های مدل‌سازی موضوعی، به منظور کشف مضامین پنهان و ساختارهای موضوعی در مجموعه‌های بزرگ اسناد مبتنی بر متن (پیکره‌های متنی) با استفاده از الگوریتم‌های ریاضی طراحی شده‌اند. این روش‌ها بر این مفهوم استوار هستند که در پس‌زمینه انتخاب هر کلمه (از میان ذخیره 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 که در نگاه اول آشکار نیستند (والدز و همکاران[2]، ۲۰۱۸). گسترش این الگوریتم‌ها تأثیر زیادی در مدل‌سازی و استفاده از پیکره‌های متنی در مطالعات اقتصادی داشته است. (به منابع بایبی و همکاران (۲۰۲۰)، ثورسرود (۲۰۲۰)، ازکیتا[3] (۲۰۲۰) و آردیا و همکاران[4] (۲۰۱۹) مراجعه نمایید.)

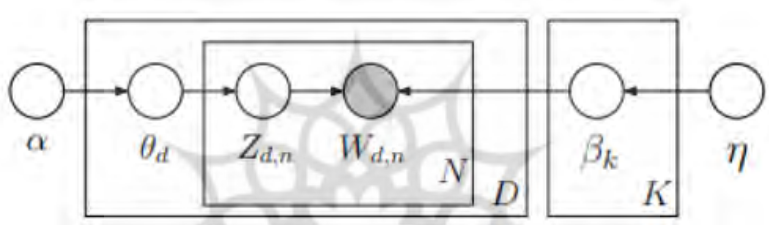
روش‌های متعددی برای مدل‌سازی موضوعی ایجاد شده است که انواع روابط و محدودیت‌ها را در مجموعه داده‌ها در نظر می‌گیرند. روش تخصیص پنهان دیریکله (LDA)، رایج‌ترین این روش‌ها است. LDA ، یک روش احتمالی مولد[5] بدون نظارت است (وانگ و همکاران[6]، ۲۰۱۹). در این روش، هر پیکره متنی شامل تعدادی کلمه است که می‌توان به هر کلمه آن، تعدادی موضوع با احتمال مشخص نسبت داد، به‌طوری که در نهایت 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

---

[1] Latent Dirichlet Allocation (LDA)
[2] Valdez et al.
[3] Azqueta
[4] Ardia et al.
[5] Generative Probabilistic
[6] Wang et al.

(موضوعات) با احتمالات مختلف تعلق دارد و به‌طور مشابه، هر موضوع نیز تا حدی به تمام اسناد با احتمالات متفاوت متعلق است (رایزنبیکلر و ریاترر[1]، ۲۰۱۹).

نمودار ۱، نموداری از مدل گرافیکی LDA را نشان می‌دهد که نحوه ارتباط بین متغیرهای تصادفی مختلف را نشان می‌دهد. تنها متغیری که مشاهده می‌شود (نتیجه آن مشخص است)، هاشور خورده است. اگر نتیجه متغیر دوم به مقدار متغیر اول بستگی داشته باشد، یک فلش از یک متغیر تصادفی دیگر به آن کشیده شده است. همچنین، لازم به توضیح است که صفحه مستطیل شکل در اطراف مجموعه‌ای از متغیرها به این معناست که این مجموعه چندین بار تکرار می‌شود.



نمودار (۱): مدل گرافیکی LDA

تعداد کل اسناد	D
تعداد کلمات هر سند	N
تعداد کل موضوعات	k
پارامتر توزیع پیشین دیریکله برای موضوعات به ازای هر سند	α
پارامتر توزیع پیشین دیریکله برای کلمات به ازای هر موضوع	η
توزیع کلمات برای موضوع k- ام	β_k
توزیع موضوعات برای سند d- ام	θ_d
موضوع کلمه n- ام در سند d- ام	$Z_{d,n}$
یک کلمه مشخص	$w_{d,n}$

فرآیند تولیدی برای LDA با توزیع مشترک متغیرهای پنهان و مشاهده شده، مطابق با رابطه (۱) قابل بیان است (بِلی[2]، ۲۰۱۲):

[1] Reisenbichler & Reutterer
[2] Blei

$$p(\beta_{1:K}, \theta_{1:D}, Z_{1:D}, W_{1:D}) = \prod_{i=1}^{K} p(\beta_i) \prod_{d=1}^{D} p(\theta_d)\left(\prod_{n=1}^{N} p(z_{d,n}|\theta_d)p(w_{d,n}|\beta_{1:k}, z_{d,n})\right) \quad (۱)$$

خروجی مدل‌سازی پایه در LDA مجموعه‌ای از گروه‌های کلمات (یعنی موضوعات) است که هر یک با احتمال عضویت معین در هر سند شناسایی می‌شوند. به طور معمول، این خروجی به تعداد محدودی از کلمات و موضوعاتی که بالاترین احتمال را دارند، محدود می‌شود. در نهایت، برچسب‌گذاری موضوعات به‌صورت دستی توسط محقق صورت می‌گیرد.

۲-۳-تحلیل احساس با رویکرد واژه‌نامه‌محور

تحلیل احساس، یک حوزه میان‌رشته‌ای است که ریشه 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 اغلب پژوهش‌ها صرف کشف احساس در متون طبیعی شده است (کرامت‌فر[1]،۱۴۰۰). تشخیص قطبیت[2]، 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[3]، ۱۴۰۱). رویکرد واژه‌نامه‌محور یکی از مهمترین رویکردهای تحلیل احساس است. در این روش، سیستم‌های توسعه‌یافته با استفاده از واژه‌نامه‌های از پیش‌ساخته که متشکل از واژگان و عبارات حاوی احساس و نمره عددی مرتبط با آن‌ها هستند، یک نمره کلی برای عبارت، جمله، یا سند مورد نظر محاسبه می‌کنند و این عدد را به عنوان شاخص احساس آن سند در نظر می‌گیرند (کرامت‌فر،۱۴۰۰). به عنوان مثال، کلماتی مانند "خوب"، "زیبا" و "شگفت‌انگیز" به انسان احساس مثبت می‌دهند و کلماتی مانند "بد"، "زشت" و

[1] Keramatfar (2021)
[2] Polarity
[3] Rajabi et al. (2022)

"ترسناک" کلماتی با قطبیت منفی هستند. لازم به ذکر است که اغلب کلماتی که بار حسی دارند صفت و قید هستند اما برخی از نام‌ها مانند زباله و افعالی مانند نفرت داشتن و عشق‌ورزیدن نیز بار حسی دارند. این کلمات ممکن است دارای وزن یا بدون وزن باشند. منظور از وزن، یک عدد یا احتمالی است که برای هر کلمه در نظر گرفته می‌شود تا سطح مثبت یا منفی بودن آن را نشان دهد. اغلب، از این رویکرد برای محاسبه جهت‌گیری اسناد با توجه به جهت‌گیری معنایی کلمات و عبارات درون اسناد استفاده می‌شود (همتیان و سهرابی[1]،۲۰۱۹).

واژه‌نامه‌های زیادی برای تعیین قطبیت کلمه‌ها وجود دارند. هر واژه‌نامه، سازوکار خاصی را برای تعیین قطبیت استفاده می‌کند و روش خاصی را برای نشان دادن قطبیت کلمه-ها به کار می‌برد. یکی از معروف‌ترین واژه‌نامه‌های حسی سنتی‌استرنگت[2] است که توسط تلوال و همکاران[3] (۲۰۱۰)، برای زبان انگلیسی و در ادامه برای سایر زبان‌ها از جمله زبان فارسی، توسعه پیدا کرده است و برای تحلیل احساس و محاسبه نمره حسی متون، از لیست‌های زیر استفاده می‌کند:

- **لیست کلمات حسی:** مجموعه‌ای از اصطلاحات مثبت و منفی است که هر کدام دارای مقداری از ۵- تا ۵ هستند.
- **لیست کلمات پرسشی:** حاوی چند کلمه پرسشی پرتکرار است.
- **لیست کلمات نفی:** حاوی چند کلمه نفی پرتکرار برای معکوس کردن کلمات حسی بعد از خود است.
- **لیست نمادهای حسی:** مکمل لیست اول و ترکیب علامت‌هایی است که برای بیان احساسات استفاده شده است.

سنتی‌استرنگت، احساس کلی یک متن طولانی را با احساسات مثبت و منفی کلی هر جمله محاسبه می‌کند. استفاده از این واژه‌نامه مانند سایر الگوریتم‌هایی که تحلیل احساس بر مبنای واژه‌نامه را به کار می‌برند بر این فرض استوار است که نویسنده‌ای که احساس منفی نسبت به یک موضوع دارد، از واژگانی با بار احساسی منفی بیشتر استفاده می‌کند و بالعکس. در مطالعات اقتصادی این واژه‌نامه برای کشف احساس متون

[1] Hemmatian & Sohrabi
[2] SentiStrength
[3] Thelwall et al.

در زمینه‌های مختلفی به کار رفته است. (به منابع استریچارز و همکاران[۱] (۲۰۱۸)، رافمن و یاکار[۲] (۲۰۱۹) و همچنین بروسیوس و همکاران[۳] (۲۰۲۰) مراجعه نمایید.) استفاده از نسخه فارسی این واژه‌نامه در مطالعات اقتصادی به زبان فارسی نیز سابقه دارد. (به منابع آهنگری و همکاران[۴] (۱۴۰۲) و زارعی و همکاران[۵] (۱۳۹۹) مراجعه نمایید.)

۳-۳- رگرسیون‌های انقباضی[۶] ریج[۷]، لسو[۸] و الستیک‌نت[۹]

رگرسیون‌های ریج، لسو و الستیک‌نت از تکنیک‌های میزان‌سازی[۱۰] برای کاهش پیچیدگی مدل استفاده می‌کنند و ضمن جلوگیری از بیش‌برازش در رگرسیون، باعث افزایش دقت پیش‌بینی و تفسیرپذیری مدل آماری نیز می‌شوند.

این روش‌ها که تحت عنوان الگوریتم‌های رگرسیون انقباضی نیز شناخته می‌شوند، اغلب به منظور ایجاد مدل ساده‌تر به کار می‌روند، به‌ویژه زمانی که تعداد متغیرهای پیش‌بینی در مجموعه نسبت به تعداد مشاهدات بالا باشد یا زمانی که احتمال هم‌خطی چندگانه در مجموعه داده‌های مربوط به متغیرهای پیش‌بینی وجود داشته باشد (ویلکوکس[۱۱]، (۲۰۱۹) ؛پارک و کنیشی[۱۲]، (۲۰۱۶)).

ضرایب تخمین زده شده در رگرسیون ریج، هرگز دقیقاً به صفر کاهش پیدا نمی‌کنند، اما امکان دارد برخی از عوامل با سهم اندک در متغیر پاسخ، به صفر نزدیک شوند. معادله رگرسیون ریج به شکل رابطه (۲) است:

$$\beta = argmin\left[\sum_{i=1}^{I}\left(y_i - \beta_0 - \sum_{j=1}^{P} x_{ij}\,\beta_j\right)^2 + \lambda \sum_{j=1}^{P} \beta_j^2\right] \qquad (۲)$$

---

[1] Strycharz et al.
[2] Rothman & Yakar
[3] Brosius et al.
[4] Ahangari et al. (2023)
[5] Zarei et al. (2020)
[6] Shrinkage Regression
[7] Ridge
[8] Lasso
[9] Elastic Net
[10] Regularization
[11] Wilcox
[12] Park & Konishi

رگرسیون لسو، بسیار شبیه به رگرسیون ریج است، با این تفاوت که با استفاده از یک تابع جریمه روی جمع قدرمطلق ضرایب، تعداد پارامترها کنترل می‌شود و ضریب پارامترها را می‌توان در طول فرآیند تنظیم مدل به صفر رساند. رگرسیون لسو، به شکل زیر در رابطه (۳) مشخص می‌شود:

$$\beta = argmin\left[\sum_{i=1}^{I}\left(y_i - \beta_0 - \sum_{j=1}^{P} x_{ij}\,\beta_j\right)^2 + \lambda \sum_{j=1}^{P}|\beta_j|\right] \qquad (۳)$$

رگرسیون الستیک‌نت، ترکیبی از جریمه‌های ریج و لسو را شامل می‌شود و معادله رگرسیون آن به شکل رابطه (۴) است:

$$\beta = argmin\left[\sum_{i=1}^{I}\left(y_i - \beta_0 - \sum_{j=1}^{P} x_{ij}\,\beta_j\right)^2 + \lambda \sum_{j=1}^{P}(1-\alpha)\beta_j^2 + (\alpha)|\beta_j|\right] \qquad (۴)$$

وزن نسبی جریمه‌ها توسط متغیر تنظیم $\boldsymbol{\alpha}$ تعیین می‌شود (ریچاردسون و همکاران،[1] ۲۰۲۱).

۳-۴- روش یادگیری ماشین تقویت گرادیان[2]

مدل تقویت گرادیان، یک مدل یادگیری جمعی[3] است. مدل‌های یادگیری جمعی با ترکیب چندین مدل پیش‌بینی یا یادگیرنده‌های ضعیف که عملکردی اندکی بهتر از حدس تصادفی دارند، به کاهش احتمال وقوع بیش‌برازش، کاهش واریانس و افزایش صحت پیش‌بینی مدل نهایی کمک می‌کنند. با توجه به داده‌های ورودی $\{(x_i.y_i)\}_{i=1}^{n}$ و تابع زیان $L(y_i.F(X))$ ، مراحل الگوریتم این مدل به شرح زیر است:

گام اول: شروع مدل با یک مقدار ثابت:

$$F_0(x) = argmin_\gamma \sum_{i=1}^{n} L\,(y_i.\gamma) \qquad (۵)$$

که در آن y_i مقدار مشاهده شده و γ مقدار پیش بینی شده است. $F_0(x)$ نیز میانگین مقادیر مشاهده شده است.

گام دوم: برای مقادیر m از ۱ تا M:

الف) محاسبه مقدار γ_{im} برای مقادیر i از ۱ تا n

$$\gamma_{im} = -\left[\frac{\partial L(y_i.F(x_i))}{\partial F(x_i)}\right]_{F(x)=F_{m-1}(x)} \qquad (۶)$$

ب) برازش یک درخت رگرسیون[4] به مقادیر γ_{im} و ایجاد R_{jm} برای مقادیر j از ۱ تا J_m

---

[1] Richardson et al.
[2] Gradient Boosting
[3] Ensemble Learning Model
[4] Regression Tree

پ) محاسبه γ_{jm} برای مقادیر j از ۱ تا J_m

$$\gamma_{jm} = argmin_{\gamma} \sum_{x_i \in R_{ij}} L\left(y_i, F_{m-1}(x_i) + \gamma\right) \quad (۷)$$

ث) به‌روزرسانی $F_m(x)$

$$F_m(x) = F_{m-1}(x) + \nu \sum_{j=1}^{J_m} \gamma_m I\left(x \epsilon R_{jm}\right) \quad (۸)$$

که در آن ν نرخ آموزش[1] است.

گام سوم:

$$\hat{F}(x) = F_M(x) \quad (۹)$$

پس از انجام تمامی تکرارهای M و به روزرسانی تابع $F_m(x)$، مدل نهایی $\hat{F}(x)$ ، رابطه بین متغیرهای مستقل و متغیر وابسته را تقریب می‌زند (یون[2]، ۲۰۲۱).

۳-۵- مدل معیار[3] و معیارهای ارزیابی

در این پژوهش، الگوی سری زمانی تک‌متغیره به عنوان مدل معیار انتخاب شده است و عملکرد مدل‌های ساخته شده توسط سری‌های زمانی مبتنی بر احساس متون خبری، با این الگو مقایسه می‌شوند.

هرچند الگوهای سری‌زمانی مبانی نظری اقتصادی ندارند، اما پیش‌بینی‌های نسبتاً دقیقی را از متغیر مدنظر ارائه می‌دهند. در الگوی سری‌زمانی تک‌متغیره، مقادیر حال یک متغیر اقتصادی به صورت تابعی از مقادیر گذشته آن در نظر گرفته می‌شود. بدین ترتیب، تنها اطلاعاتی که برای پیش‌بینی مقادیر آتی متغیر لازم است مقادیر گذشته همان متغیر است و به اطلاعات متغیرهای دیگر نیازی نیست (نوفرستی[4]، ۱۴۰۰). الگوهای سری‌زمانی تک‌متغیره مربوط به خانواده مدل‌های ARIMA[5] هستند که اغلب بر مبنای روش باکس و جنکینز (۱۹۷۶) مدل‌سازی می‌شوند. رابطه (۱۰)، شکل کلی این مدل‌ها را نشان می‌دهد (سوری[6]،۱۴۰۰):

$$\emptyset(L)\Delta^d Y_t = \mu + \theta(L)u_t \quad (۱۰)$$

---

[1] Learning Rate
[2] Yoon (2021)
[3] Benchmark
[4] Noferesti (2021)
[5] Autoregressive Integrated Moving Average
[6] Souri (2021)

همچنین برای اندازه‌گیری دقت پیش‌بینی مدل‌ها از معیارهای ارزیابی مرسوم شامل میانگین قدر مطلق خطا[1](MAE)، جذر میانگین مربعات خطا[2]$(RMSE)$ و میانگین درصد قدرمطلق خطا[3]$(MAPE)$ استفاده خواهد شد. هر سه معیار، مطابق با روابط (۱۱-۱۳)، مقدار خطا را از مقایسه مقادیر واقعی(Y_t) و پیش‌بینی(Y_t^f) محاسبه می‌کنند (سوری،۱۴۰۰).

$$MAE = \frac{\sum_{t=T+1}^{T+m}\left|Y_t^f - Y_t\right|}{m} \quad (۱۱)$$

$$RMSE = \sqrt{\frac{\sum_{t=T+1}^{T+m}\left(Y_t^f - Y_t\right)^2}{m}} \quad (۱۲)$$

$$MAPE = \frac{\sum_{t=T+1}^{T+m}\left|\frac{Y_t^f - Y_t}{Y_t}\right|}{m} * 100 \quad (۱۳)$$

m طول دوره پیش بینی است که از T+1 تا T+m می باشد.

۴- مجموعه داده‌ها

در این مطالعه، از داده‌های تولید ناخالص داخلی فصلی ایران مستخرج از سایت مرکز آمار ایران به قیمت‌های ثابت سال ۱۳۹۰ و در بازه زمانی فصل اول سال ۱۳۸۴ تا فصل سوم سال ۱۴۰۲، به عنوان متغیر هدف (متغیری که تخمین مقدار آن در پایان هر فصل مرجع مد نظر است) استفاده شده است. انتخاب این دوره زمانی به‌دلیل امکان دسترسی آنلاین به آرشیو خبرگزاری فارس جهت ساخت متغیرهای پیش‌بینی مبتنی بر اخبار اقتصادی صورت گرفته است. لازم به توضیح است که خبرگزاری فارس با توجه به سه معیار زیر از میان خبرگزاری‌های فارسی زبان انتخاب شده است:

۱) قابلیت دسترسی برخط[4] به آرشیو اخبار: از لحاظ فنی امکان استخراج برخط اخبار از پایگاه داده خبرگزاری وجود داشته باشد.

۲) آرشیو جامع از اخبار: متون خبری به صورت روزانه در دوره مورد بررسی موجود باشد.

---

[1] Mean Absolute Error
[2] Root Mean Square Error
[3] Mean Absolute Percentage Error
[4] Online

۳) مرجع بودن خبرگزاری: برای تعیین مرجع بودن یک سایت، از معیارهای تحلیل دامنه استفاده می‌شود.

جدول (۱): معیارهای تحلیل دامنه خبرگزاری فارس

امتیاز (رتبه)	کاربرد	معیار
۶ از ۱۰	معیار مورد استفاده موتور جست‌وجوی گوگل برای اندازه	Google PageRank
۷ از ۱۰	معیاری برای سنجش اثربخشی محتوای سایت در جذب	CPR Score
۷۹ از ۱۰۰	معیاری برای سنجش اعتبار دامنه سایت	Domain Authority:
۶۳ از ۱۰۰	معیاری برای سنجش اعتبار صفحات سایت	Page Authority
۴۲ در ایران	معیاری برای سنجش محبوبیت سایت بر اساس ترافیک	Alexa Rank
۲ از ۴ (۱۳۹۹)	معیار مقایسه با سه خبرگزاری فارسی زبان (مهر، ایرنا، ایسنا) پربازدید	Alexa Rank (Comparative)

منبع: checkpagerank.net

۵- یافته‌ها

در ابتدا با روش خزش[1]، ۳۰۱،۴۹۸ خبر اقتصادی در دوره زمانی مورد بررسی از پایگاه اینترنتی خبرگزاری فارس استخراج شده است. در ادامه، با استفاده از روش‌های متن‌کاوی، پیش‌پردازش متن[2] به منظور تمیزکردن داده‌های متنی و تبدیل آن‌ها به قالبی کاربردی برای پردازش‌های بعدی روی خبرهای استخراج شده انجام شد. در مرحله پیش‌پردازش، از کتابخانه پارسی‌ور[3] و سه ابزار نرمال‌سازی متن (برای استانداردسازی و یکپارچه‌سازی متون)، واحدساز[4] (تبدیل و تفکیک جملات به کلمات (توکن)) و حذف ایست‌واژه‌ها[5] (حذف واژه‌هایی که مفهومی را منتقل نمی‌کنند (به منظور افزیش دقت و سرعت مدل‌ها)) استفاده شد. در گام بعدی، با استفاده از الگوریتم LDA موضوعات اصلی متون خبری شناسایی شد. جدول ۲، 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.

---

[1] Web Crawling
[2] Text Preprocessing
[3] Parsivar
[4] Word Tokenization
[5] Stop Words

جدول (۲): نتایج مدل‌سازی موضوعی با LDA روی اخبار اقتصادی

ردیف	برچسب موضوع	کلمه ۱	کلمه۲	کلمه۳	کلمه۴	کلمه۵	کلمه۶	کلمه۷	کلمه۸	کلمه۹	کلمه ۱۰
۱	بورس	سهام	بورس	سهم	سرمایه	بازار	اوراق	شاخص	معاملات	بهادار	فرابورس
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	حمل‌ونقل	حمل‌ونقل	قطار	ناوگان	بنادر	جاده	سوخت	مسافر	تردد	محور	ترافیک
۷	بخش عمومی	بودجه	مجلس	دولت	کمیسیون	قانون	یارانه	سازمان	مالیات	لایحه	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

منبع: یافته‌های تحقیق

خروجی دیگر الگوریتم LDA احتمال عضویت هر خبر به یکی از ۱۰ موضوع جدول فوق است. در ادامه به هر خبر، موضوع با بیشترین احتمال، به عنوان موضوع آن خبر تخصیص داده شده است. جدول ۳، توزیع فراوانی موضوعات روی کل اخبار را نشان می‌دهد. بعد از تخصیص یک موضوع مشخص به هر خبر، نمره حسی نیز با استفاده از نسخه جاوای واژه‌نامه سنتی‌استرنگت به هر خبر تخصیص داده شد. همان‌طور که پیش‌تر توضیح داده شد، این ابزار از یک واژه‌نامه از کلمات حسی و وزن آنها استفاده می‌کند که به صورت دستی توسط متخصصان زبان‌شناسی انتخاب شده است. نکته قابل‌توجه دیگر این است که 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 برای هر دوره سه ماهه، ۱۰ سری زمانی فصلی مبتنی بر احساس تولید شده است. این سری‌های زمانی را می‌توان در مدل‌های پیش‌بینی مختلف به کار برد.

جدول (۳): توزیع فراوانی اخبار روی موضوعات

موضوع	تعداد خبر	درصد از کل اخبار
بورس	۲۲،۸۱۰	۷/۵۷
پول و بانک	۳۷،۰۰۳	۱۲/۲۷
انرژی	۴۴،۲۴۲	۱۴/۶۷
اقتصاد بین‌الملل	۳۷،۳۲۵	۱۲/۳۸
صنعت و کشاورزی	۳۸،۹۵۵	۱۲/۹۲
حمل‌ونقل	۱۱،۳۱۶	۳/۷۵
بخش عمومی	۴۲،۵۲۸	۱۴/۱۱
توسعه	۱۹،۲۸۸	۶/۴۰
فناوری	۳۵،۶۳۵	۱۱/۸۲
بازار	۱۲،۳۹۶	۴/۱۱
مجموع	۳۰۱،۴۹۸	۱۰۰

منبع: یافته‌های تحقیق

قبل از تخمین مدل‌های پژوهش، لازم است آزمون نرمال بودن داده‌ها به‌عنوان فرض زیربنایی تخمین مدل‌ها انجام شود. مطابق با نتایج جدول (۴)، نرمال بودن داده‌های سری‌های زمانی حسی و همچنین GDP تایید می‌شود.

جدول (۴):نتایج آزمون نرمال بودن داده‌ها

نوع آزمون	متغیر	مقدار آماره	مقدار احتمال	متغیر	مقدار آماره	مقدار احتمال
آزمون جارک-برا[1] $H_0: X_t \sim N(0.\sigma^2)$	GDP	۳/۰۱	۰/۲۲	X6	۲/۰۷	۰/۳۵
	X1	۳/۱۲	۰/۲۰	X7	۲/۶۹	۰/۲۶
	X2	۴/۵۰	۰/۱۰	X8	۱/۷۴	۰/۴۱
	X3	۰/۸۳	۰/۶۵	X9	۴/۵۴	۰/۱۰
	X4	۰/۶۸	۰/۷۱	X10	۲/۹۹	۰/۲۲
	X5	۱/۰۵	۰/۵۸			

منبع: یافته‌های تحقیق

در گام بعد، مدل‌های لسو، ریج و الستیک‌نت و همچنین مدل تقویت گرادیان برای پیش‌بینی تولید ناخالص داخلی به عنوان متغیر هدف و بر پایه استفاده از این ۱۰ سری

[1] Jarque-Bara

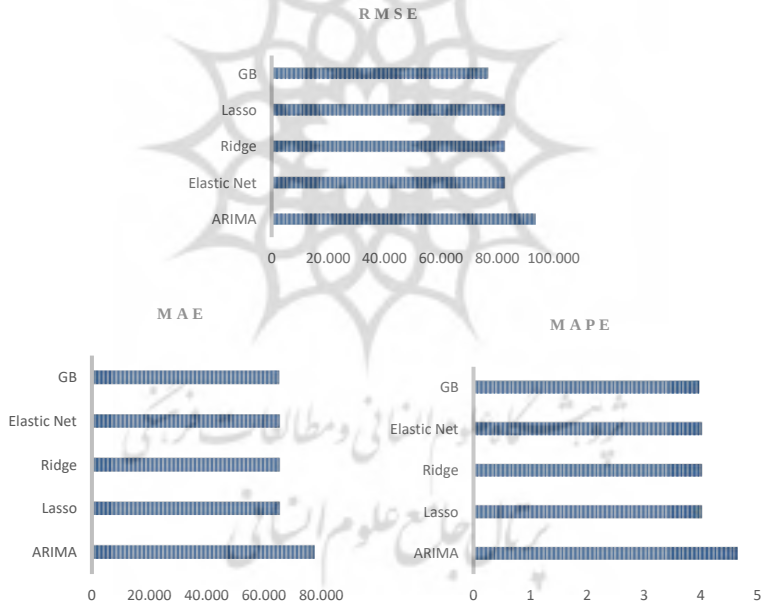
زمانی حسی به عنوان متغیرهای توضیحی برآورد شد. این مدل‌ها با استفاده از ۹۰ درصد داده‌ها آموزش داده شد و ۱۰ درصد از داده‌ها نیز به صورت تصادفی به عنوان داده‌های آزمون یا داده‌های آزمایشی (test size=0.01, random state=42) در نظر گرفته شدند. به منظور مقایسه عملکرد این مدل‌ها، الگوی سری زمانی تک‌متغیره برآورد شد. مرتبه این الگو بر مبنای روش باکس-جنکینز به صورت ARIMA(1,1,0) تعیین شد. الگوی آریما، به دلیل اتکا بر مقادیر گذشته متغیر و عدم نیاز به اطلاعات اضافی برای برآورد، همواره پس از هر فصل مرجع، قابلیت به‌روزرسانی دارد و مبنای مناسبی برای ارزیابی سایر مدل‌های پیش‌بینی محسوب می‌شود.

جدول (۵): مقایسه نتایج مدل‌های پیش‌بینی تولید ناخالص داخلی فصلی ایران

دوره‌های زمانی آزمون (سال:فصل)	مقدار واقعی تولید ناخالص داخلی (میلیارد ریال)	مقادیر پیش‌بینی‌شده				
		مدل معیار	رگرسیون‌های انقباضی			مدل یادگیری ماشین
		ARIMA(1,1,0)	Lasso	Ridge	Elastic Net	Gradient Boosting
۱۳۸۵:۱	۱،۴۵۳،۳۱۸	۱،۴۱۶،۹۲۴	۱،۳۸۵،۷۰۱	۱،۳۸۵،۸۷۹	۱،۳۸۶،۷۶۴	۱،۵۵۹،۶۶۵
۱۳۹۹:۴	۱،۹۲۹،۷۳۷	۱،۸۹۶،۵۲۱	۱،۸۴۴،۸۳۶	۱،۸۴۴،۷۸۶	۱،۸۴۴،۵۴۲	۲،۰۲۷،۳۵۳
۱۳۸۶:۳	۱،۷۰۹،۸۹۷	۱،۶۹۴،۵۶۲	۱،۷۴۳،۰۶۷	۱،۷۴۳،۰۸۰	۱،۷۴۳،۱۴۱	۱،۶۷۶،۷۷۸
۱۳۸۶:۲	۱،۷۱۹،۲۵۶	۱،۵۹۰،۱۳۱	۱،۷۰۴،۱۹۸	۱،۷۰۴،۱۳۹	۱،۷۰۳،۸۸۳	۱،۶۱۱،۱۹۹
۱۳۹۲:۴	۱،۶۰۰،۴۸۶	۱،۶۴۸،۹۸۳	۱،۷۰۳،۹۶۷	۱،۷۰۴،۰۷۹	۱،۷۰۴،۶۳۸	۱،۷۰۴،۱۱۶
۱۳۹۹:۲	۱،۸۶۰،۱۷۰	۱،۷۴۸،۷۷۵	۱،۸۹۸،۶۸۰	۱،۸۹۸،۵۸۴	۱،۸۹۸،۱۰۵	۱،۸۵۷،۲۶۵
۱۳۹۱:۱	۱،۵۳۶،۵۴۸	۱،۷۱۱،۵۱۹	۱،۷۰۷،۱۴۰	۱،۷۰۷،۱۵۶	۱،۷۰۷،۲۳۳	۱،۵۸۰،۴۱۸
۱۳۸۷:۱	۱،۵۸۷،۳۱۵	۱،۶۵۹،۲۹۶	۱،۵۷۷،۳۷۹	۱،۵۷۷،۳۳۵	۱،۵۷۷،۱۱۵	۱،۵۶۰،۳۲۷
RMSE		۹۳،۴۷۰	۸۲،۵۳۷	۸۲،۵۴۴	۸۲،۵۷۸	۷۶،۶۵۹
MAE		۷۷،۶۱۴	۶۵،۴۰۸	۶۵،۴۱۱	۶۵،۴۱۷	۶۵،۳۱۷
MAPE		۴/۶۵	۴/۰۲	۴/۰۲	۴/۰۲	۳/۹۷

منبع: یافته‌های تحقیق

مطابق نتایج بدست آمده از جدول ۵ و همچنین مقایسه معیارهای ارزیابی در نمودار ۲، استفاده از داده‌های حسی در مدل‌های رگرسیون انقباضی و یادگیری ماشین در مقایسه با الگوی سری زمانی تک‌متغیره که تنها مبتنی بر مقادیر گذشته متغیر هدف است، باعث کاهش خطای پیش‌بینی می‌شود. به بیان دقیق‌تر، خطای پیش‌بینی بر اساس معیار RMSE به ترتیب در مدل‌های رگرسیون انقباضی در حدود ۱۲ درصد و در روش تقویت گرادیان در حدود ۱۸ درصد کمتر از مدل معیار است. با توجه به معیار MAE، خطا در هر دو دسته مدل، حدود ۱۶ درصد کاهش داشته است. معیار MAPE نیز نشان می‌دهد که استفاده از مدل‌های مبتنی بر تحلیل احساس اخبار اقتصادی، خطای پیش‌بینی را در مدل‌های رگرسیون انقباضی ۱۴ درصد و در روش تقویت گرادیان ۱۵ درصد نسبت به مدل معیار کاهش می‌دهد.



نمودار (۲): مقایسه معیارهای ارزیابی مدل‌ها

منبع: یافته‌های تحقیق

۶- نتیجه‌گیری

این پژوهش نشان می‌دهد که متون خبری حاوی اطلاعات مفیدی راجع به وضعیت اقتصاد هستند و می‌توان از این اطلاعات برای پیش‌بینی مهم‌ترین شاخص عملکرد

اقتصاد کلان، یعنی تولید ناخالص داخلی استفاده کرد. برخلاف آمار رسمی، اخبار به صورت روزانه منتشر می‌شوند و وقفه انتشار ندارند. از طرفی، امکان دسترسی برخط به اخبار منتشر شده در خبرگزاری‌ها باعث می‌شود که محدودیت دسترسی به این داده‌های متنی به مراتب نسبت به سایر کلان داده‌های بالقوه مفید در ارزیابی شرایط اقتصادی، برای مثال تراکنش‌های سیستم پرداخت الکترونیکی، کمتر باشد. به همین منظور، ابتدا با انتخاب سایت فارس به عنوان یکی از خبرگزاری‌های مرجع فارسی‌زبان، ۳۰۱،۴۹۸ خبر اقتصادی به صورت روزانه از ابتدای سال ۱۳۸۴ تا انتهای آذر ماه سال ۱۴۰۲ از پایگاه اینترنتی این خبرگزاری استخراج و پیش‌پردازش‌های اولیه شامل نرمال‌سازی، واحدساز و حذف ایست‌واژه‌ها، روی آنها انجام شده است. سپس با استفاده از روش LDA به عنوان رایج‌ترین روش مدل‌سازی موضوعی، اخبار در موضوعات خبری به‌طور جداگانه دسته‌بندی شدند. در ادامه، به منظور ایجاد سری‌های زمانی فصلی از این موضوعات، نمره حسی هر خبر با استفاده از نسخه فارسی واژه‌نامه سنتی‌استرنگت مشخص و تجمیع سه‌ماهه انجام شده است. نتایج نشان داده است که استفاده از این سری‌های زمانی حسی در مدل‌های پیش‌بینی، خطای پیش‌بینی را نسبت به مدل معیار که تنها از مقادیر گذشته متغیر تولید ناخالص داخلی فصلی برای پیش‌بینی استفاده می‌کند، تا ۱۸ درصد کاهش می‌دهد. این پژوهش از نخستین مطالعاتی است که در ایران از داده‌های متنی خبری به زبان فارسی به‌عنوان منبعی برای پیش‌بینی اقتصادی بهره گرفته است. به‌عنوان پیشنهادهایی برای تحقیقات آتی، می‌توان این پژوهش را از چند جنبه گسترش داد. اول: از حیث غنی نمودن منبع داده، یعنی می‌توان اخبار را از خبرگزاری‌های دیگر و یا حتی از منابع خبری دیگری همچون شبکه‌های اجتماعی، استخراج و تحلیل کرد. دوم: از لحاظ روش تجزیه و تحلیل، می‌توان عملکرد سایر واژه‌نامه‌های حسی و دیگر روش‌های تحلیل احساس همچون روش‌های مبتنی بر یادگیری ماشین را برای کشف احساس مورد بحث و بررسی قرار داد. سوم: از نظر متغیر هدف، با اتخاذ رویکرد مناسب از حیث منبع داده و روش تحلیل احساس، می‌توان پیش‌بینی به‌هنگام از سایر متغیرهای اقتصاد کلان همچون تورم، بیکاری، مصرف و سرمایه‌گذاری را ارائه کرد.

تضاد منافع

نویسندگان نبود تضاد منافع را اعلام می‌دارند.

فهرست منابع

۱. افشار، پروین، منوچهری، صلاح‌الدین و امانی، رامین (۱۴۰۲). نااطمینانی اقتصاد کلان، ریسک سیاسی و نوسانات بازار ارز در ایران. *فصل‌نامه نظریه‌های کاربردی اقتصاد*، ۱۰(۳)، ۱۰۲-۶۷.

۲. آقانیا، پریسا، حیدری، حسن و جهانگیری، شهاب (۱۴۰۱). بررسی تأثیر شوک‌های سیاست پولی بر رشد اقتصادی و تورم در اقتصاد ایران: شواهد تجربی بر اساس مدل TVP-SFAVAR-SV. *فصل‌نامه نظریه‌های کاربردی اقتصاد*، ۹(۴)، ۶۱-۹۶.

۳. آهنگری آهنگرکلائی، مرتضی، سبطی، علی و یعقوبی، مهدی (۱۴۰۲). ساخت واژگان به صورت خودکار برای تحلیل نظرات در حوزه بورس. *فصل‌نامه پردازش علائم و داده‌ها*، ۲۰(۲)، ۳-۲۰.

۴. رجبی، زینب، ولوی، محمدرضا و حورعلی، مریم (۱۴۰۱). مروری بر روش‌های تحلیل احساس در متون فارسی. *فصل‌نامه پردازش علائم و داده‌ها*، ۱۹(۲)، ۱۰۷-۱۳۲.

۵. زارعی، عظیم، فیض، داود و طاهری، غزاله (۱۳۹۹). ارائه چارچوب هوشمندی بازار اجتماعی مبتنی بر وب ۲/۰ با استفاده از تکنیک متن‌کاوی در وب‌سایت‌های رسانه‌های اجتماعی (مورد مطالعه: تحلیل رقابتی در بین برندهای سامسونگ و امرسان). *فصل‌نامه پژوهش‌های مدیریت در ایران*، ۲۴(۴)، ۹۸-۱۲۵.

۶. سوری، علی (۱۴۰۰). *اقتصاد سنجی پیشرفته:* جلد دوم. *انتشارات نور علم*، تهران.

۷. کرامت‌فر، عبدالصمد (۱۴۰۰). *مدل‌سازی چند جریانی زمینه نظرات برای تحلیل احساس*. رساله دکتری. دانشگاه قم.

۸. نوفرستی، محمد (۱۴۰۰). *اقتصاد سنجی کاربردی داده‌های سری زمانی*. انتشارات دانشگاه شهید بهشتی، تهران.

1. Afshar, P. A., Manochehri, S., & Amani, R. (2023). Macroeconomic Uncertainty, Political Risk and Exchange Rate Market Fluctuations in Iran. *Quarterly Journal of Applied Theories of Economics*, 10(3), 67-102 (In Persian).
2. Aghania, P., Heidari, H., & Jahangiri, Sh. (2023). Investigating the Impact of Monetary Policy Shocks on Economic Growth and Inflation in the Iranian Economy: Empirical Evidence Based on the TVP-TVP-

SFAVAR-SV Model. *Quarterly Journal of Applied Theories of Economics*, 9(4), 61-96 (In Persian).

3. Aguilar, P., Ghirelli, C., Pacce, M., & Urtasun, A. (2021). Can news help economic sentiment?. An application in COVID-19 times. *Economics Letters*, 199, 109730.
4. Ahangari, M., Sebti, A., & Yaghoubi, M. (2023). Automatically generate lexicon for the Persian stock market. *Signal and Data Processing*, 20(2), 3-20 (In Persian).
5. 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.
6. Angeletos, G. M., Collard, F., & Dellas, H. (2018). Quantifying confidence. *Econometrica*, 86(5), 1689-1726.
7. Aprigliano, V., Emiliozzi, S., Guaitoli, G., Luciani, A., Marcucci, J., & Monteforte, L. (2023). The power of text-based indicators in forecasting Italian economic activity. *International Journal of Forecasting*, 39(2), 791-808.
8. Ardia, D., Bluteau, K., & Boudt, K. (2019). Questioning the news about economic growth: Sparse forecasting using thousands of news-based sentiment values. *International Journal of Forecasting*, 35(4), 1370-1386.
9. Ash, E., & Hansen, S. (2023). Text algorithms in economics. *Annual Review of Economics*, 15(1), 659-688.
10. Ashwin, J., Kalamara, E., & Saiz, L. (2021). Nowcasting euro area GDP with news sentiment: a tale of two crises. *Journal of Applied Econometrics*, 1–19.
11. Azqueta Gavaldon, A.(2020). Text-mining in macroeconomics: the wealth of words . *Doctoral dissertation*, University of Glasgow.
12. Barbaglia, L., Consoli, S., & Manzan, S. (2024). Forecasting GDP in Europe with textual data. *Journal of Applied Econometrics*, 39(2), 338-355.
13. Barbaglia, L., Frattarolo, L., Onorante, L., Pericoli, F. M., Ratto, M., & Pezzoli, L. T. (2023). Testing big data in a big crisis: Nowcasting under COVID-19. *International Journal of Forecasting*, 39(4), 1548-1563.
14. Blei, D. M. (2012). Probabilistic topic models. *Communications of the ACM*, 55(4), 77-84.
15. Bortoli, C., Combes, S., & Renault, T. (2018). Nowcasting GDP growth by reading newspapers. *Economie et Statistique*, 505(1), 17-33.
16. Brosius, A., van Elsas, E. J., & de Vreese, C. H. (2020). Bad news, declining trust? Effects of exposure to economic news on trust in the European Union. *International Journal of Public Opinion Research*, 32 (2): 223–242.

17. Bybee, L., Kelly, B. T., Manela, A., & Xiu, D. (2020). The structure of economic news. Tech. rep. *NBER Working paper*, 26648.
18. Coase, R. H. (1960). The problem of social cost. *J. Law Econ*, 3,1–44
19. Eshbaugh-Soha, M. (2010). The tone of local presidential news coverage. *Political Communication*, 27(2), 121-140.
20. Ferrara, L., & Simoni, A. (2023). When are Google data useful to nowcast GDP? An approach via preselection and shrinkage. *Journal of Business & Economic Statistics*, 41(4), 1188-1202.
21. Friedman, M., & Schwartz, A. J. (1963). A Monetary History of the United States: 1867–1960. *Princeton, NJ*: Princeton Univ Press.
22. Galbraith, J. W., & Tkacz, G. (2018). Nowcasting with payments system data. *International Journal of Forecasting*, 34(2), 366-376
23. Hemmatian, F., & Sohrabi, M.(2019).A survey on classification techniques for opinion mining and sentiment analysis. *Artificial intelligence review*, 52(3), 1495-1545.
24. Hu, Y., & Yao, J. (2022). Illuminating economic growth. *Journal of Applied Econometrics*, 37(5), 896-919.
25. Kalamara, E., Turrell, A., Redl, C., Kapetanios, G., & Kapadia, S. (2022). Making text count: economic forecasting using newspaper text. *Journal of Applied Econometrics*, 37(5), 896-919.
26. Keeney, M., Kennedy, B., & Liebermann, J. (2012). The value of hard and soft data for short-term forecasting of GDP. *Economic Letters Series*, 11/EL/12, Central Bank of Ireland.
27. Keramatfar, A. (2021). *Multi-stream modeling of comments' contexts for sentiment analysis*. Ph.D. Thesis, University of Qom (In Persian).
28. Lourenço, N., & Rua, A. (2021). The Daily Economic Indicator: tracking economic activity daily during the lockdown. *Economic Modelling*, 100, 105500.
29. Manchado Marcos, L. (2023). *Nowcasting with Alternative Data* (Bachelor's thesis, Universitat Politècnica de Catalunya).
30. Munezero, M., Montero, C. S., Sutinen, E., & Pajunen, J. (2014). Are they different? Affect, feeling, emotion, sentiment, and opinion detection in text. *IEEE transactions on affective computing*, 5(2), 101-111.
31. Noferesti, M. (2021). Applied Econometric Time Series. *Shahid Beheshti University Press*, Tehran (In Persian).
32. Park, H., & Konishi, S. (2016). Robust logistic regression modelling via the elastic net-type regularization and tuning parameter selection. *Journal of Statistical Computation and Simulation*, 86(7), 1450-1461.
33. 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).

34. Reisenbichler, M., & Reutterer, T. (2019). Topic modeling in marketing: recent advances and research opportunities. *Journal of Business Economics*, 89(3), 327-356.
35. Richardson, A., van Florenstein Mulder, T., & Vehbi, T. (2021). Nowcasting GDP using machine-learning algorithms: A real-time assessment. *International Journal of Forecasting*, 37(2), 941-948.
36. Rothman, T., & Yakar, C. (2019). Empirical AnalysiscTowardsrtherEffectr of Social Media on Cryptocurrency Price and Volume. *European Scientific Journal*, ESJ, 15, 31-52.
37. Shapiro, A. H., Sudhof, M., & Wilson, D. J. (2022). Measuring news sentiment. *Journal of econometrics*, 228(2), 221-243.
38. Shiller, R. J. (2020). *Narrative economics: How stories go viral and drive major economic events*. Princeton University Press.
39. Souri, A. (2021). *Advanced Econometrics: Volume Two*. Noor Elm Press, Tehran (In Persian).
40. Strycharz, J., Strauss, N., & Trilling, D. (2018). The role of media coverage in explaining stock market fluctuations: Insights for strategic financial communication. *International Journal of Strategic Communication*, 12(1), 67-85.
41. Thelwall, M., Buckley, K., Paltoglou, G., Cai, D., & Kappas, A. (2010). Sentiment strength detection in short informal text. *Journal of the American society for information science and technology*, 61(12), 2544-2558.
42. Thorsrud, L. A. (2020). Words are the new numbers: A newsy coincident index of the business cycle. *Journal of Business & Economic Statistics*, 38(2), 393-409.
43. Valdez, D., Pickett, A. C., & Goodson, P. (2018). Topic modeling: latent semantic analysis for the social sciences. *Social Science Quarterly*, 99(5), 1665-1679.
44. 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.
45. Wilcox, R. R. (2019). Multicolinearity and ridge regression: results on type I errors, power and heteroscedasticity. *Journal of applied statistics*, 46(5), 946-957.
46. Yoon, J. (2021). Forecasting of real GDP growth using machine learning models: Gradient boosting and random forest approach. *Computational Economics*, 57(1), 247-265
47. Zarei, A., Feiz, D., & Taheri, Gh. (2021). Providing Social Market Intelligence Framework based on web 2.0 Using Text-Mining Technique on Social Media Websites (Case Study: Competitive Analysis between Samsung and Emersun Brands). *Management Research in Iran*, 24(4), 98-125 (In Persian).