https://danesh24.um.ac.ir

CC BY

Research Article

Vol. 29, No. 1, 2022, p. 1-46

OPEN ACCESS

Measuring the Media-based Economic Uncertainty index by Machine Learning Algorithms in Iran and its Effect on the Exchange Rate

H. Habibi Nikjoo [1], A. Cheshomi [2]*, M. Salimifar[3]

1- Ph.D. Student of Economics Faculty of Economic and Administrative Sciences Ferdowsi University of Mashhad, Mashhad, Iran
2- Assistant Professor, Economic Faculty of Economic and Administrative Sciences Ferdowsi University of Mashhad, Mashhad, Iran
3- Professor, Economic Faculty of Economic and Administrative Sciences Ferdowsi University of Mashhad, Mashhad, Iran

(*- Corresponding Author Email: a.cheshomi@um.ac.ir)
https://doi.org/10.22067/mfe.2022.75682.1175

Received: 2021/04/04 Revised: 2022/07/04 Accepted: 2022/10/17	**How to cite this article:** Habibi nikjoo, H.; Cheshomi, A., & Salimifar, M. (2022). Measuring the Media-based Economic Uncertainty index by Machine Learning Algorithms in Iran and its Effect on the Exchange Rate. *Quarterly Monetary & Financial Economics*. *29*(1). 1-46. (in Persian with English abstract). https://doi.org/10.22067/mfe.2022.75682.1175
Available Online: 2022/10/17	

1- INTRODUCTION

Economic uncertainty is one of the important and influential factors on economic policies and their results, and in such a situation, rational decisions are replaced by other methods. Various studies has shown the effect of economic uncertainty on inflation, investment, economic growth, consumption and demand for money.

Uncertainty is difficult to measure due to its invisibility, and as the uncertainty measurement methods improve, the measurement of its effect on

various economic variables and markets and the prediction of their behavior in response to the actions of economic agents will be more accurate.

The main aim of this article is to measure the economic uncertainty index by using news published in social networks. This method of measurement has become very important with the widespread use of social networks.

2- THEORETICAL FRAMEWORK

Uncertainty is one of the most controversial concepts in the philosophy and methodology of economics. The history of the concept of economic uncertainty goes back to David Hume .There are three categories of theories about economic uncertainty .The first group believes that the future reality is unchangeable and predetermined and economic decision makers have perfect information. In this view, there is no such thing as uncertainty and the world is in complete certainty. 18th century the economists of were the first group to present this theory. The second group believes that the reality of the future is unchangeable and predetermined and the decision makers are able to know the future. These economists use objective conditional probability functions to solve the future uncertainty problem. The third class considers the future reality to be changeable and unknown. The starting point of these theories started from the study of the Chicago school economist Frank Knight titled "Risk, Uncertainty and Profit". He clearly distinguished between the two concepts of risk and uncertainty. Keynes also reached the same results as Knight. In general, in a situation where the economy has a high level of uncertainty, the theories of the first and second category have a good explanation. But in confronting with exogenous shocks such as the corona virus epidemic, war and financial crisis, the concept of uncertainty will be more appropriate in the theories of the third category. This study will measure this index based on fundamental uncertainty (the third category).

3- METHODOLOGY

In this article, the economic uncertainty index in Iran was measured from January 2017 to December 2020 by monitoring and analyzing 3,117,960 news from 28 popular and influential Iranian Telegram channels. To analyze these news, we used "supervised machine learning" methods. In the first step, 13,404 news items were labeled by human evaluators according to their impact on uncertainty. The labels had two modes "affecting uncertainty" and "neutral". Then by using four algorithms ("C4.5" from decision tree methods, "Multilayer Perceptron" from artificial neural network methods, "Logistics" from function-oriented methods and "Simple Bayes" from Bayesian methods) labeling of the whole news was done. The economic uncertainty index was calculated numerically and based on the number of news items that affect economic uncertainty, the measurement and value of this index was standardized, and then the quality of the index was evaluated with historical evidence, relabeling and comparison with the index based on Google data.

4- RESULTS & DISCUSSION

Among the 4 media-based uncertainty indicators, 3 indicators can better explain the historical events of this period. Among them, the best performance is determined by C4.5 algorithm from the decision tree methods. After this algorithm, multilayer perceptron, logistic has the best performance and the weakest performance belongs to the simple Bayes method. Media-based economic uncertainty index trend with C4.5 method is consistent with the important events of the study period, in such a way that the highest level of uncertainty occurred during the period when Trump announced his withdrawal from the JCPOA until the official withdrawal of the United States from the JCPOA. In general, it can be said that the fluctuations of the economic uncertainty index have been limited and have several jumps, which are due to the withdrawal of the United States from the JCPOA, the oil embargo and the assassination of Sardar Soleimani .

In the logistic algorithm, the highest level of uncertainty dates back to the end of 2020. The period that coincides with Trump's presidential

election. The level of economic uncertainty increases after Trump's official withdrawal from the JCPOA and reaches its peak with oil sanctions.

The output of the multilayer perceptron algorithm indicates that the average level of uncertainty has not changed significantly . In the simple Bayes algorithm, the highest level was also reached during the period of the withdrawal of the United States from the JCPOA and the increase in enrichment . The results of the regression showed that economic uncertainty has a positive and significant effect on the average logarithm of the exchange rate with multilayer perceptron, logistic and simple methods. This effect is larger in the multilayer perceptron model, which had better performance based on machine learning indicators.

5- CONCLUSIONS & SUGGESTIONS

The calculated economic uncertainty index is consistent with the important events of the study period, such as the US withdrawal from the JCPOA, Iran'soil sanctions, and the escalation of the US confrontation with Iran in the assassination of Sardar Soleimani. It is suggested that daily calculation of this index be used to reduce uncertainty in the managing future events. We employed GARCH model to test effect of Media-based Economic Uncertainty index on Iranian exchange rate. The results showed that Economic Uncertainty index has poisitve effect on exchange rate.

Keywords: Economic Uncertainty, Media, Machine Learning, Text Mining, Exchange Rate.

پژوهشگاه علوم انسانی و مطالعات فرهنگی
پرتال جامع علوم انسانی

اقتصاد پولی مالی، سال بیست و نهم، دوره جدید شمارهٔ ۱، پیاپی ۲۳، بهار و تابستان ۱۴۰۱

اندازه‌گیری شاخص نا اطمینانی اقتصادی رسانه‌بنیان با الگوریتم‌های یادگیری ماشین در ایران و تأثیر آن بر نرخ ارز

حبیب حبیبی نیکجو
دانشجوی دکتری علوم اقتصادی دانشگاه فردوسی مشهد
علی چشمی[۱]
دانشیار دانشگاه فردوسی مشهد
مصطفی سلیمی‌فر
استاد دانشگاه فردوسی مشهد

نوع مقاله: پژوهشی
https://doi.org/10.22067/mfe.2022.75682.1175

چکیده

هدف اصلی مقاله اندازه‌گیری شاخص نا اطمینانی اقتصادی با استفاده از اخبار منتشرشده در شبکه‌های اجتماعی است. این روش اندازه‌گیری با فراگیری استفاده از شبکه‌های اجتماعی اهمیت بالایی پیدا کرده است. در این مقاله، با پایش و تحلیل ۳٬۱۱۷٬۹۶۰ خبر از ۲۸ کانال تلگرامی پرمخاطب و اثرگذار ایران، شاخص نا اطمینانی اقتصادی در ایران را از ژانویه ۲۰۱۷ تا دسامبر ۲۰۲۰ اندازه‌گیری شد. برای تحلیل این اخبار از روش‌های «یادگیری ماشین با ناظر» بهره گرفته شد. در مرحله اول ۱۳۴۰۴ خبر توسط ارزیابان انسانی برحسب اثرگذاری بر نا اطمینانی برچسب‌گذاری شد. سپس با استفاده از چهار الگوریتم («**C4. 5**» از روش‌های درخت تصمیم، «پرسپترون چندلایه» از روش‌های شبکه عصبی مصنوعی، «لجستیک» از روش‌های تابع‌محور و «بیز ساده» از روش‌های بیزی) برچسب‌گذاری کل اخبار انجام شد. شاخص نا اطمینانی اقتصادی به‌صورت شمارشی و بر اساس تعداد اخباری که اثرگذار بر نا اطمینانی اقتصادی هستند، اندازه‌گیری و مقدار این شاخص استاندارد شده و سپس کیفیت شاخص با شواهد تاریخی، برچسب‌گذاری مجدد و مقایسه با شاخص مبتنی بر داده‌های گوگل ارزیابی شد. شاخص محاسبه شده با وقایع مهم دوره مطالعه مانند خروج آمریکا از برجام، تحریم نفتی و بالا گرفتن تقابل آمریکا با ایران در ترور سردار سلیمانی همخوانی دارد. برآورد تأثیر نا اطمینانی اقتصادی رسانه‌بنیان بر نرخ ارز با مدل گارچ، اثر مثبت و معنی‌دار این نا اطمینانی را نشان می‌دهد.

کلیدواژه‌ها: نا اطمینانی اقتصادی، رسانه، متن‌کاوی، یادگیری ماشین، نرخ ارز.
طبقه‌بندی JEL: D80, C10, E27,F31

---

[۱] نویسنده مسئول: a.cheshomi@um.ac.ir
تاریخ دریافت: ۱۴۰۰/۱۲/۱۵ تاریخ پذیرش: ۱۴۰۱/۰۷/۲۵
صفحات: ۴۶-۱

۱. مقدمه

یکی از عوامل مهم و اثرگذار بر سیاست‌های اقتصادی و نتایج آن، نا اطمینانی است. در حضور نا اطمینانی اقتصادی، بازیگران و سیاست‌گذاران با در نظرگیری اثر نا اطمینانی بر نتایج تصمیم‌گیری‌ها، روش‌های دیگری را جایگزین تصمیم‌گیری منطقی می‌کنند. در سال‌های اخیر سیاست‌گذاران در صدد شناخت اثر نا اطمینانی بر متغیرهای اقتصادی مهم برآمدند تا با ایجاد نوآوری در اتخاذ سیاست‌ها، در راستای بهبود وضعیت فعالیت‌های اقتصادی و بازارهای پولی در شرایط نا اطمینانی قدم بردارند (Sahinoz & Erdogan Cosar، 2013).

پژوهشگرانی که اثر نا اطمینانی بر متغیرهای مختلف اقتصادی را مطالعه می‌کنند معمولاً نشان می‌دهند نا اطمینانی بر تورم، سرمایه‌گذاری، رشد اقتصادی، مصرف و تقاضای پول اثرگذار است (Marcus, 1981؛ Bernanke، 1983؛ Aizenman & Marion، 1991؛ Rodrik، 1991 و Pourshahabi، 2009). بااین‌وجود، پس از وقوع بحران مالی سال ۲۰۰۷ و با توجه به پیامدهای زیان‌بار افزایش نا اطمینانی، توجه اقتصاددانان بر تأثیر نا اطمینانی روی متغیرهای مختلف اقتصادی افزایش یافت (Bloom، 2009؛ Baum et al.، 2010؛ Bachmann et al.، 2010؛ Jones and Olson، 2013 و Binde ، 2017). به‌طور مثال جونز و اولسن (2013) نشان می‌دهند اثر نا اطمینانی بر تورم چندان واضح نبوده و به شوک‌های بین‌المللی نظیر قیمت نفت بستگی دارد. ساهینز و اردوگان کوسار (۲۰۱۸) نیز نشان می‌دهند که نا اطمینانی بر رشد اقتصادی، مصرف و سرمایه‌گذاری ترکیه اثر نامطلوب داشته است و تأکید می‌کنند نا اطمینانی بالا اثر بیشتری بر کاهش سرمایه‌گذاری نسبت به کاهش تولید و مصرف دارد.

با توجه به غیرقابل مشاهده بودن نا اطمینانی، اندازه‌گیری آن دشوار است؛ بنابراین هرچه روش‌های اندازه‌گیری نا اطمینانی بهبود یابد، سنجش اثر آن بر متغیرهای و بازارهای اقتصادی گوناگون و پیش‌بینی رفتار آن‌ها در پاسخ به کنش عاملان اقتصادی دقیق‌تر خواهد بود. محققان به‌منظور پیش‌بینی معیار نا اطمینانی از روش‌های متفاوتی استفاده می‌کنند. بلوم (2007 و 2009)، بیکر و همکاران (2016)، بلوم و همکاران (2018)، به کمک روش‌های مبتنی بر داده‌کاوی رسانه‌ها به اندازه‌گیری شاخصی برای نا اطمینانی پرداخته‌اند. لودویگسون و همکاران (2021) از روش‌های مبتنی بر رگرسیون و نا اطمینانی خطا برای اندازه‌گیری شاخص نا اطمینانی اقتصاد کلان و نا اطمینانی مالی استفاده کرده‌اند. شاخص نا اطمینانی مصرف‌کننده لدوک و لیو (2016)، شاخص نا اطمینانی پیش‌بینی‌کنندگان حرفه‌ای روسی و سخپوسیان (2015) و شاخص نا اطمینانی کمیسیون اروپا (2020) نیز مبتنی بر روش نظرسنجی است.

پیشرفت تکنولوژی، افزایش پایگاه‌های دسترسی به داده و بهبود روش‌های سیستمی در سال‌های اخیر

اندازه‌گیری شاخص نا اطمینانی را بهبود بخشیده‌اند. به‌طور مثال باتوجه به آن‌که در سال‌های اخیر گوگل ۶۸۸ درصد از جست‌وجو در فضای اینترنت را به خود اختصاص داده، مانسکی (2013) به اندازه‌گیری شاخص نا اطمینانی با روش‌های جست‌وجو در منابع اطلاعاتی از طریق گوگل پرداخته است. بیکر (2016) و مور (2017) نیز از شاخص نا اطمینانی مبتنی بر اخبار استفاده کردند و معتقدند با این روش، اندازه‌گیری شاخص نا اطمینانی حتی به‌صورت روزانه ممکن است. لی‌لو و ژانگ (2015) از شاخص نا اطمینانی مبتنی بر بازارهای مالی استفاده می‌کنند و معتقدند مزیت این روش استخراج شاخص نا اطمینانی مرتبط با فعالیت‌های اقتصادی است.

مقاله حاضر با بهره‌گیری از متن ۲۸ کانال تلگرام خبری و با بهره‌گیری از روش‌های یادگیری ماشین، شاخص نا اطمینانی اقتصادی را اندازه‌گیری می‌کند. این مقاله از چهار جهت حائز اهمیت است: اول؛ پیکره مورداستفاده متن کامل اخبار کانال‌های تلگرام خبری فارسی است که برای اولین بار اتفاق می‌افتد. دوم؛ روش مورداستفاده، روش‌های مرسوم یادگیری ماشین است. این در حالی است که پژوهشگران در اکثر مطالعات گذشته، صرفاً بر جستجوی کلمات در متن (تیتر) خبر بسنده کرده‌اند. سوم، حجم اخبار مورداستفاده ۳ میلیون و ۱۱۷ هزار خبر است که حدود ۱۳۴۰۴ خبر توسط ارزیابان انسانی برچسب‌گذاری شد. این مقاله از منظر حجم بالای اخبار و ایجاد پیکره بزرگ برچسب‌گذاری شده، حائز اهمیت است. چهارم؛ اندازه‌گیری شاخص نا اطمینانی اقتصادی رسانه‌بنیان می‌تواند زمینه‌ساز انجام مطالعات تجربی در حوزه نا اطمینانی اقتصادی شود که پیش‌ازاین به دلیل نبود داده صورت نمی‌گرفت.

مقاله چنین تنظیم شده است که بعد از مقدمه در بخش دوم، مفهوم نا اطمینانی اقتصادی در مطالعات نظری مرور می‌شود. در بخش سوم، روش‌های اندازه‌گیری نا اطمینانی و مطالعات تجربی پیرامون آن موردبررسی قرار می‌گیرد. در بخش چهارم، روش اندازه‌گیری شاخص نا اطمینانی رسانه‌بنیان معرفی شده و شاخص محاسبه می‌شود. در بخش پنجم، شاخص نا اطمینانی اقتصادی رسانه‌بنیان با معیارهای مختلف ارزیابی می‌شود. در بخش ششم، اثر این شاخص بر نرخ دلار سنجیده شده و در پایان نتیجه‌گیری و پیشنهادها ارائه می‌شود.

پژوهشگاه علوم انسانی و مطالعات فرهنگی
پرتال جامع علوم انسانی

۲. مبانی نظری نا اطمینانی

نا اطمینانی یکی از مجادله‌آمیزترین مفاهیم در فلسفه و روش‌شناسی علم اقتصاد است. هیچ اجماعی

بین اقتصاددانان در مورد تعریف دقیق نا اطمینانی و ارتباط آن با نظریات اقتصادی وجود ندارد (لاوسون، 1988). تاریخچه مفهوم نا اطمینانی اقتصادی به دیوید هیوم[1] برمی‌گردد. اگرچه در طول تاریخ تعاریف متعددی از نا اطمینانی اقتصادی ارائه شده، اما تمامی این نظریات را می‌توان با توجه به قضاوت در مورد دو مؤلفه دسته‌بندی کرد:

مؤلفه هستی‌شناسی بیان می‌کند که واقعیت اقتصادی در آینده چگونه رقم می‌خورد. ازآنجایی‌که نا اطمینانی اقتصادی در طول زمان موضوعیت پیدا می‌کند، فرض در مورد چگونگی رخدادهای آتی بسیار کلیدی است. واقعیت اقتصادی در مورد آینده می‌تواند از پیش‌تعیین‌شده[2] و تغییرناپذیر[3] یا به‌صورت نامعلوم[4] و قابل‌تغییر[5] باشد.

مؤلفه معرفت‌شناختی بیان می‌کند که با فرض هر وضعیتی از آینده (قابل‌تغییر یا غیرقابل‌تغییر) آیا تصمیم‌گیران اقتصادی قادر هستند که آن را فهم‌کرده و نا اطمینانی نمایند. آیا انسان قدرت تحلیل آینده را دارد یا همواره با خطاهای شناختی و محاسباتی مواجه است. بر اساس دو مؤلفه فوق، سه دسته نظریه در مورد نا اطمینانی اقتصادی وجود دارد.

دسته اول: واقعیت آینده غیرقابل‌تغییر و از پیش‌تعیین شده است و تصمیم‌گیران اقتصادی دانش کامل دارند.

اقتصاددانان کلاسیک قرن ۱۹ با فرض دانش کامل و اطمینان کامل، اولین گروهی بودند که این نظریه را ارائه دادند. آن‌ها فرض می‌کردند که تصمیم‌گیران اقتصادی در دنیایی با اطمینان کامل زندگی می‌کنند و نسبت به واقعیت اقتصادی خارجی برنامه‌ریزی شده را در گذشته، حال و آینده دانش کامل دارند. ریکاردو (۱۸۱۷) به‌عنوان اقتصاددان کلاسیک قرن ۱۹ فرض می‌کند مسیر اقتصاد همانند مسیر سیاره‌ها در نظریه نیوتون است؛ به‌عبارت‌دیگر تصمیم‌گیران همه اطلاعات را در اختیار دارند و هیچ خطایی رخ نخواهد داد. در این دیدگاه اصلاً موضوعی به نام نا اطمینانی وجود ندارد و دنیا در قطعیت کامل قرار دارد؛ بنابراین، مدل‌های مطرح‌شده در این دوره مانند قانون‌های طبیعی غیرقابل‌تغییر، بدون زمان هستند. سیستم

پژوهشگاه علوم انسانی و مطالعات فرهنگی
پرتال جامع علوم انسانی

[1] David Hume
[2] Predetermined
[3] Immutable
[4] Unknowable
[5] Transmutable

معادلات والراس در قرن ۱۹ نمونه‌ای دیگر از فرض دانش‌کامل و آینده غیرقابل‌تغییر بود که با اطمینان کامل مواجه بود. اگر بازارهای فعلی و آتی وجود داشته باشد، با استفاده از دلال‌های والراسی همه تصمیم‌گیران از ابتدا تا ابد، از همه اطلاعات برخوردار خواهند بود.

دسته دوم: واقعیت آینده غیرقابل‌تغییر و از پیش‌تعیین شده است و تصمیم‌گیران قادر به شناخت آینده هستند.

اکثر اقتصاددانان جریان اصلی[1] علم اقتصاد ازجمله ساموئلسون و لوکاس معتقدند که آینده و تغییرات آن از قبل تعریف شده است اما در مورد دانش تصمیم‌گیران از کلاسیک‌های قدیمی محتاطانه‌تر رفتار می‌کنند. این اقتصاددانان برای حل مسئله نا اطمینانی آینده، از توابع احتمال شرطی عینی[2] استفاده می‌کنند. برای انجام نا اطمینانی قابل‌اتکا[3] از توابع احتمال شرطی عینی، باید از داده‌های آینده نمونه‌گیری کرده و آن داده‌ها را تحلیل کنیم اما چون این اتفاق نشدنی است، از فرض از پیش تعیین‌شده آینده با مفهوم ارگادیک‌بودن[4] استفاده می‌شود. ارگادیک‌بودن بر این موضوع دلالت می‌کند که رخدادهای آینده به‌صورت آماری سایه‌ای از سیگنال‌های بازاری حال و گذشته است. با اعمال فرض ارگادیک بودن، با استفاده از داده‌های گذشته می‌توان به نا اطمینانی‌های قابل‌اتکا آماری در مورد آینده دست‌یافت (Davidson، 1991). این نظریه که از اوایل قرن بیستم آغاز شده، مفهوم معادل اطمینان[5] و صرف ریسک احتمالی[6] جایگزین فرض دانش کامل در نظریه کلاسیک کرده است. صرف ریسک این امکان را فراهم می‌کند که نا اطمینانی ایجاد شود. به‌این‌ترتیب، ممکن است میان ارزش تخمین زده شده و رخداد آینده تفاوت وجود داشته باشد که می‌توان آن را نا اطمینانی تعریف کرد (Sargent & Lucas، 1981؛ Machina، 1987).

در دهه ۱۹۷۰، تحلیل صرف ریسک جای خود را به نظریه انتظارات‌عقلایی داد. در این نظریه، افراد تصمیم‌های خود را بر اساس تابع توزیع احتمال ذهنی[7] اتخاذ می‌کنند. این تابع ذهنی برمبنای «توزیع احتمال

---

[1] Mainstream economists
[2] Objective Conditional Probability Functions
[3] Reliable
[4] Ergodicity
[5] Certainty Equivalent
[6] Probabilistic Risk Premium
[7] Subjective Probability Dstribution

شرطی عینی غیرقابل‌تغییر» اتفاق می‌افتد (Lucas، 1971). اقتصاددانان ارتدوکس کنونی نا اطمینانی اقتصادی را معادل توزیع‌های احتمال عینی می‌دانند که اتفاقات آینده را تضمین می‌کند و برای همه افراد در زمان حال شناخته شده است (Sargent & Lucas، 1981). در این دیدگاه، انحراف معیار توزیع شاخص اندازه‌گیری نا اطمینانی اقتصادی است.

استفاده از انحراف معیار توزیع احتمال برای اندازه‌گیری نا اطمینانی اقتصادی، سبب شده است که اکثر تحلیل‌های اقتصاددانان کلاسیک پابرجا باقی بماند. برای مثال، هم چنان فرض غیرقابل‌تغییر بودن آینده پابرجاست؛ اما برخلاف مدل‌های دانش کامل، به تصمیم‌گیران اجازه می‌دهد که در کوتاه‌مدت دچار خطا محاسباتی شوند. در بلندمدت این فرض تضمین می‌کند که افراد با انتظارات عقلایی تابع توزیع احتمال عینی را می‌شناسند و بنابراین دچار خطا در بلندمدت نمی‌شوند.

دسته سوم: واقعیت آینده قابل‌تغییر و نامعلوم است.

دسته سوم از نظریات نا اطمینانی معتقدند که رفتارهای تصمیم‌گیران اقتصادی بخشی از آینده را شکل می‌دهد و ازاین‌رو آینده قابل‌تغییر است. نا اطمینانی بنیادین[1] واژه‌ای است که برای جدایی رویکرد خود از نا اطمینانی به‌منزله انحراف‌معیار توابع توزیع احتمال استفاده می‌شود. نقطه شروع این نظریات از مطالعه اقتصاددان مکتب شیکاگو، فرانک نایت (1921) با عنوان «ریسک، نا اطمینانی و سود» آغاز شد. او به‌صراحت بین دو مفهوم ریسک و نا اطمینانی تمایز قائل شد. تفسیر اولیه او از مطالعات هیوم، اسمیت و منگر الهام گرفته‌شده و نا اطمینانی را به‌عنوان نیروی اساسی که در همه‌ی اقدامات انسانی و اقتصادی مؤثر است مطرح می‌کند.

ایده اصلی نایت تمایز بین ریسک و نا اطمینانی است. نظریه وی بر این فرض استوار است که انتخاب اقتصادی و فعالیت کارآفرین و سرمایه‌گذار تنها درصورتی‌که دانش بشری ناقص باشد امکان‌پذیر است. نایت در مطالعه خود در مورد منشأ سود نشان می‌دهد که سود می‌تواند تنها در بازارهای ناقص ایجاد شود. بکرت (1996) به‌طور خلاصه نظریه نایت در مورد رد مدل بازار کامل و تئوری انتخاب عقلایی را به‌صورت زیر خلاصه کرد: «تحت فرض بازارهای کامل، تئوری‌های اقتصادی قدرت توضیح‌دهندگی سود را ندارند. زیرا بازار، تأمین‌کنندگان جدید را تا زمان برقراری تعادل میان قیمت کالا با هزینه نهایی تولید

[1] Fundamental Uncertainty

جذب می‌کند». نایت در تحلیل رفتار کارآفرینی، معتقد است که کارآفرینان ممکن است در مراحل دانشی متفاوتی قرار بگیرند. این مراحل عبارت‌اند از: قطعیت، ریسک[1] و نا اطمینانی.

کینز (1921) نیز به نتایج مشابهی با نایت دست یافت. بدین‌صورت که موقعیت‌های بسیار منحصربه‌فردی وجود دارد که احتمالات عددی[2] بتواند شکل بگیرد. وی در تحقیقات اقتصادی خود (۱۹۳۷-۱۹۳۶) به‌مانند منگر و نایت نتیجه گرفت که موقعیت‌های اقتصادی در بیشتر موارد به‌وسیله دو حالت ریسک و نا اطمینانی مشخص می‌شوند. درنتیجه، تجزیه‌وتحلیل آرمان‌گرایانه در تخصیص بهینه بازار که در سنت والراسی (1874) بیان شده است مطابق نظر او در بیشتر موارد بی‌معنی است. مهم‌ترین پدیده‌های اقتصادی مانند سود، رکود اقتصادی یا سرمایه‌گذاری فقط می‌توانند به فرض دانش ناقص و نه دانش کامل توضیح داده شوند (Bateman، 1996).

پساکینزین‌ها نیز دو رویکرد کاملاً متفاوت نسبت به نا اطمینانی اقتصادی دارند. رویکرد اول بیان می‌کند که نا اطمینانی مسئله‌ای هستی‌شناسی است و بر رد فرض ارگادیک بودن تمرکز می‌کنند؛ اما رویکرد دوم بر مسئله معرفت‌شناختی یا دانش ناکامل تصمیم‌گیران اقتصادی تأکید می‌کند. رویکرد اول از اوایل دهه ۱۹۸۰ شکل گرفته و برگرفته از نظریات کینز در رساله‌ای در باب احتمالات است. لاوسون (۱۹۸۵) یکی از کسانی است که این نظریه را بسط داده است. در مقابل رویکرد دوم با نظرات دیویدسون شکل گرفته است و در اواخر دهه ۱۹۹۰ بالغ شده است. این نظریات مبتنی بر مفاهیم نایت است و فرض ارگادیک بودن را زیر سؤال می‌برند.

در اقتصاد نهادی جدید، مسئله نا اطمینانی از منظر هزینه مبادله[3] تحلیل می‌شود. ویلیامسون در اقتصاد هزینه مبادله یکی از مهم‌ترین ابعاد مبادله را نا اطمینانی می‌داند و اهمیت آن را در تصمیمات فعالان اقتصادی نشان داده است. نا اطمینانی در اقتصاد هزینه مبادله به این دلیل مطرح می‌شود که هر یک از ساختارهای تدبیر مبادلات برای واکنش به اختلالات موجود از توانمندی‌های متفاوتی برخوردارند. دو نوع نا اطمینانی را می‌توان از یکدیگر تفکیک کرد که هرکدام اثر متفاوتی بر انتخاب ساختارهای تدبیر دارند: نا اطمینانی محیطی و نا اطمینانی رفتاری.

درمجموع، در شرایطی که اقتصاد از سطح بالای نا اطمینانی پذیری برخوردار است، نظریات دسته اول و

---

[1] Risk
[2] Numerical Probabilities
[3] Transaction Cost

دوم، توضیح‌دهندگی خوبی دارند. مفهوم نا اطمینانی در این دودسته نظریات، بیشتر به ریسک نزدیک است؛ اما در مواجهه با شوک‌های برون‌زا نظیر همه‌گیری ویروس کرونا، جنگ، بحران مالی مفهوم نا اطمینانی در نظریات دسته سوم مناسب‌تر خواهد بود. این مطالعه با مبنا قراردادن نا اطمینانی بنیادین (دسته سوم)، این شاخص را اندازه‌گیری خواهد کرد. علت انتخاب این مفهوم، نوسانات و شوک‌های زیاد برون‌زا در اقتصاد ایران در سال‌های گذشته بوده است. انتخاب دونالد ترامپ به‌عنوان رئیس‌جمهور آمریکا، خروج از برجام، بالا گرفتن تقابل آمریکا با ایران مانند ترور سردار سلیمانی و همه‌گیری کرونا، از مهم‌ترین شوک‌های برون‌زای این دوره بوده است.

۳. پیشینه تجربی تحقیق

نا اطمینانی اقتصادی مفهومی انتزاعی است. ازاین‌رو، تعیین متغیر جانشین[1] برای اندازه‌گیری آن در مطالعات تجربی با چالش‌های متعددی مواجه است. تقریباً ۴ متغیر جانشین برای اندازه‌گیری نا اطمینانی اقتصادی وجود دارد: شاخص مبتنی بر بازارهای مالی، شاخص عدم توافق نااطمینانی‌کنندگان، شاخص مبتنی بر داده‌های جست‌وجوی گوگل و شاخص مبتنی بر اخبار.

تلاطم بازار مالی یکی از شاخص‌های متداول برای اندازه‌گیری نا اطمینانی است. این شاخص مبتنی بر تعریف کلاسیک‌های جدید از نا اطمینانی است. این شاخص معمولاً با انحراف‌معیار توابع توزیع احتمال شرطی عینی یا ذهنی اندازه‌گیری می‌شود. این شاخص، محبوب‌ترین روش اندازه‌گیری نا اطمینانی اقتصادی است؛ چراکه به‌صورت لحظه‌ای قابل‌اندازه‌گیری بوده و به‌سادگی می‌توان آن را در کشورهای مختلف مقایسه کرد. مطالعات تجربی این حوزه عموماً با استفاده از خانواده مدل‌های آرچ و گارچ صورت می‌پذیرد. اشکال بزرگ شاخص‌های مبتنی بر نوسانات بازار این است که به‌صورت غیرمستقیم با فعالیت‌های اقتصادی مرتبط است. اگرچه درآمدهای شرکت‌ها به فعالیت‌های اقتصادی مرتبط است، اما تغییرات قیمت سهام در کوتاه‌مدت به عوامل دیگری ارتباط دارد (لی، لو و ژانگ، ۲۰۱۵). اگرچه این عوامل به نا اطمینانی اقتصادی مرتبط می‌شوند، اما ارتباط آن‌ها با فعالیت‌های اقتصادی به‌صورت صریح روشن نیست. بایر و همکاران (۲۰۱۹) و گارسیا و همکاران (۲۰۲۰) از این مفهوم در پژوهش خود استفاده کرده‌اند.

---

[1] Proxy

شاخص دیگر اندازه‌گیری نا اطمینانی که مبتنی بر بازارهای مالی است، پراکندگی در نا اطمینانی‌های تحلیل‌گران برای بازارهای آتی است. این شاخص برمبنای تعاریف نا اطمینانی بنیادین است. در این روش اندازه‌گیری نا اطمینانی فرض اصلی این است که تصمیم‌گیران اقتصادی در مورد آینده دانش کامل ندارند؛ بنابراین، تصمیم‌گیران مختلف نظرات متفاوتی نسبت به آینده دارند. هر چه اختلاف میان نا اطمینانی از آینده بیشتر باشد، نشان‌دهنده سطح بالاتری از نا اطمینانی اقتصادی است. مزیت بزرگ این شاخص ربط مستقیم آن به فعالیت‌های اقتصادی است. این مزیت در مقابل شاخص نوسانات بازار سهام تعریف می‌شود. یک نقد جدی به این روش وارد است. این نقد ادعا می‌کند که شاخص‌های اختلاف نا اطمینانی و نا اطمینانی یکسان نیستند. ریچ و همکاران (۲۰۱۲) معتقد هستند که اختلاف در نا اطمینانی می‌تواند نشان‌دهنده عدم‌توافق باشد و نه نا اطمینانی. گلاس (۲۰۲۰) نیز این روش را معیار ضعیفی برای اندازه‌گیری نا اطمینانی می‌داند. محمدی و همکاران (۲۰۱۸) در پژوهشی، اثر نوسان‌های نرخ ارز واقعی بر تجارت محصولات کشاورزی در ایران را در بازه زمانی ۱۹۸۰-۲۰۱۲ بررسی کرده‌اند. با استفاده از مدل‌های GARCH و EGARCH مشخص شد که نوسان‌های نرخ ارز در بلندمدت اثر منفی و معناداری بر صادرات و واردات محصولات کشاورزی دارد. زمانیان و بهراد امین (۲۰۱۴) به مطالعه اثر نا اطمینانی نرخ ارز واقعی بر تقاضای واردات ایران طی سال‌های ۱۹۸۰-۲۰۱۲ با استفاده از داده‌های سالانه پرداختند. مدل‌های ECARCH، GARCH و ARDL برای تولید لگاریتم سری‌های واریانس و بررسی هم‌انباشتگی استفاده شدند. نتایج این پژوهش نشان داد که تنها متغیرهای نرخ ارز مؤثر واقعی و واردات هم انباشته‌اند و دارای رابطه بلندمدت هستند. نتایج مدل ECM برای کوتاه‌مدت بیانگر آن است که اگرچه بین نا اطمینانی نرخ ارز و واردات در بلندمدت رابطه معناداری وجود ندارد اما اثر منفی نا اطمینانی نرخ ارز بر واردات در کوتاه‌مدت معنادار است. برگی و سینکلر (۲۰۲۱)، کلاوریا (۲۰۲۱) و لاهیری و همکاران (۲۰۲۲) از این روش در پژوهش خود استفاده کرده‌اند. در ایران نیز پژوهش‌هایی با این روش انجام شده است که می‌توان به مطالعه ارشدی (۲۰۱۱) و شکری و همکاران (۲۰۲۰) اشاره کرد.

روش سوم، اندازه‌گیری شاخص نا اطمینانی برمبنای داده‌های گوگل است. در این روش عنوان می‌شود که علت شکل‌گیری نا اطمینانی اقتصادی معرفت‌شناختی یا هستی‌شناسی است. در دیدگاه معرفت‌شناختی نا اطمینانی ادعا می‌شود که علت ندانستن تصمیم‌گیران در مورد آینده، محدودیت‌های شناختی و محاسباتی بشر است. به‌عبارت‌دیگر، چون نمی‌دانیم چه اتفاقی در آینده رخ می‌دهد با پدیده نا اطمینانی روبه‌رو هستیم. مانسکی (۲۰۱۵) معتقد به این رویکرد است و علت شکل‌گیری نا اطمینانی در اقتصاد را به

دلیل کمبود دانش می‌داند. کمبود دانش سبب می‌شود که افراد برای یافتن اطلاعات و از بین بردن شکاف‌های دانشی، اقدامات متناسبی انجام دهند. یکی از این روش‌ها جست‌وجو در منابع اطلاعاتی ازجمله گوگل است. (مانسکی، ۲۰۱۱؛ ۲۰۱۳) درصورتی‌که نا اطمینانی با افزایش اطلاعات کاهش یابد، شدت جست‌وجو برای اطلاعات می‌تواند به‌عنوان شاخصی برای اندازه‌گیری نا اطمینانی به کار گرفته شود. شرکت گوگل برای نمایش میزان جست‌وجو ابزار «گوگل‌ترندز» را معرفی نموده است. بانتمپی (۲۰۲۱) نیز اخیراً از این روش در پژوهش خود استفاده کرده است. این ابزار برای ارزیابی میزان جست‌وجوی واژگان مرتبط با رخدادهای سیاسی و اقتصادی و برای فهم میزان اطلاعاتی که افراد برای افزایش سطح آگاهی خود نیاز دارند، مورداستفاده قرار می‌گیرد. در پژوهش‌های داخلی نیز به مطالعه ابریشمی و همکاران (۲۰۲۰) اشاره کرد.

روش چهارم برای اندازه‌گیری نا اطمینانی اقتصادی، روش‌های مبتنی بر اخبار است. اثرگذارترین مقاله در این حوزه به کار پژوهشی بیکر و همکاران (۲۰۱۶) برمی‌گردد. در این مقاله، نویسندگان فرض می‌کنند که اطلاعات مربوط به نا اطمینانی اقتصادی را می‌توان در اخبار مشاهده کرد. این رویکرد در مقابل دیدگاه دیگری قرار می‌گیرد که پوشش اخبار باعث ایجاد نا اطمینانی می‌شود. یک مزیت این روش چنین است که با استفاده ازاین‌روش می‌توان نا اطمینانی در موضوعات مختلف را اندازه‌گیری کرد. این مزیت در مقابل شاخص‌های مبتنی بر بازارهای سهام مطرح می‌شود. مزیت بعدی این شاخص، دوره زمانی استخراج آن است. به‌گونه‌ای که به‌راحتی می‌توان به‌صورت روزانه این شاخص را اندازه‌گیری کرد. این روش توسط محققان دیگری نظیر مور (۲۰۱۷)، توباک و همکاران (۲۰۱۸)، بوری و همکاران (۲۰۲۱)، توپچو و همکاران (۲۰۲۱)، گوپتا و همکاران (۲۰۲۱) و آرباتلی (۲۰۲۲) مورد استفاده قرار گرفته است.

همان‌طور که در بخش قبل اشاره شد، در این مطالعه از مفهوم نا اطمینانی بنیادین برای اندازه‌گیری استفاده کرده‌ایم. ازاین‌رو، روش مبتنی بر بازارهای مالی مناسب نخواهد بود. روش پراکندگی نظرات نااطمینانی‌کنندگان، از روش‌هایی است که توسط نهادهای دولتی یا حاکمیتی اجرایی می‌شود؛ اما با توجه به داده‌های فراوان شبکه‌های اجتماعی، از با استفاده از روش‌های مبتنی بر اخبار شاخص را اندازه‌گیری کرده و با نتایج شاخص اندازه‌گیری شده با استفاده از داده‌های گوگل مقایسه می‌کنیم.

٤- اندازه‌گیری شاخص نا اطمینانی رسانه‌بنیان

شاخص نا اطمینانی رسانه بنیان بر اساس توسعه روش بیکر و همکاران (۲۰۱۶) اندازه‌گیری شده است.

در این روش، اخبار رسانه‌ها جمع‌آوری شده، بخشی از آن توسط ارزیابان انسانی به دو دسته اثرگذار بر نا اطمینانی و خنثی تقسیم می‌شوند. سپس ماشین با استفاده از الگوریتم استخراج شده از مرحله قبل تمامی اخبار را برچسب‌گذاری می‌کند. در مرحله بعد، تعداد اخباری که تحت عنوان اثرگذار بر نا اطمینانی برچسب‌گذاری شده بودند، شمارش می‌شوند. در انتها، شاخص نا اطمینانی اقتصادی با فرکانس روزانه یا ماهانه، بعد از استانداردکردن، اندازه‌گیری می‌شود (Baker et al., 2016). این فرآیند در ۴ مرحله انجام می‌شود (Han & Kamber, 2006).

مرحله یک، جمع‌آوری اخبار: اخبار موردبررسی در این مطالعه، شامل محتوای ۲۸ کانال تلگرامی خبری پر مخاطب در فاصله ژانویه ۲۰۱۷ تا دسامبر ۲۰۲۰ است (جدول ۱ پیوست). در این بازه زمانی ۳,۱۱۷,۹۶۰ خبر منتشر شده است. علت انتخاب پلتفرم تلگرام این است که در دوره پژوهش بالاترین سهم ترافیک اینترنت مصرفی ایران در تلگرام بود و تلگرام به‌عنوان مرجع اصلی خبر مخاطبان مورد استفاده قرار گرفته است. انتخاب ۲۸ کانال نیز بر اساس دو پارامتر صورت پذیرفته است: اول؛ تعداد مخاطبان که با دو شاخص میزان عضو و میزان بازدید سنجیده می‌شود. دوم؛ کانال‌های خبری رسمی که هویتی فراتر از یک کانال تلگرامی دارند و به خبرگزاری رسمی یا سایت خبری معتبر وابسته هستند. این اخبار با استفاده از کتابخانه مادلین پروتو[۱] جمع‌آوری شده است[۲]. در پایگاه داده MYSQL جمع‌آوری شده، اخبار بر اساس ستون‌های نام کانال، آدرس کانال، تعداد اعضا کانال، تعداد کاراکتر خبر، متن خبر و تعداد بازدید خبر ثبت و ذخیره شدند. برای انتخاب داده‌ها، فیلترهای زیر اعمال شد:

۱. خبرهای فاقد بازدید حذف شد. علت این اقدام بدون اثر بودن محتواهای منتشر شده بود.

۲. پیام‌های فورواردشده از کانال‌های دیگر حذف شد. دلیل اعمال این فیلتر تبلیغاتی بودن پیام‌های فورووارد شده است.

۳. پیام‌های با تعداد کارکتر کمتر از ۷۰ حذف شد. این فیلتر به این دلیل اعمال شد که اخبار کمتر از ۷۰ کاراکتر معمولاً در کنار عکس و ویدئو معنادار می‌شوند؛ بنابراین، بررسی چنین متن‌هایی فاقد ارزش

پژوهشگاه علوم انسانی و مطالعات فرهنگی

[۱] Madeline Proto یک کتابخانه به زبان PHP است که به افراد اجازه می‌دهد مانند یک اپلیکیشن موبایل و بدون نیاز با کارکردن مستقیم با API، با سرور تلگرام ارتباط برقرار کنید و تقریباً همه امکاناتی را که در یک اپلیکیشن موبایل وجود دارد، از طریق کدنویسی به زبان PHP تحت‌وب پیاده‌سازی کنید.

[۲] برای دسترسی به کدهای دستوری جمع‌آوری اخبار می‌توانید با نویسندگان مکاتبه کنید.

معنایی است.

مرحله دو، برچسب‌گذاری انسانی: این شاخص با بهره‌گیری از طبقه‌بندی متن[1] از روش‌های یادگیری ماشین با ناظر[2] نگاشته شده است. فرایند طبقه‌بندی در واقع نوعی یادگیری با ناظر است که در طی دو مرحله انجام می‌شود (Han & Kamber, 2006). در مرحله اول مجموعه‌ای از داده‌ها که در آن هر داده شامل تعدادی خصوصیت دارای مقدار و یک خصوصیت بنام خصوصیت کلاس (دارای بار نا اطمینانی اقتصاد یا خنثی) است، برای ایجاد یک مدل داده بکار می‌روند که این مدل داده در واقع توصیف‌کننده مفهوم و خصوصیات مجموعه داده‌هایی است که این مدل از روی آن‌ها ایجاد شده است. مرحله دوم، فرآیند طبقه‌بندی اعمال یا به‌کارگیری مدل داده ایجاد شده بر روی داده‌هایی است که شامل تمام خصوصیات داده‌هایی است که برای ایجاد مدل داده بکار گرفته شده‌اند، به‌جز خصوصیت کلاس این مقادیر که هدف از عمل طبقه‌بندی نیز تخمین مقدار این خصوصیت است.

در این مرحله ۱۳۴۰۴ خبر به‌صورت نمونه تصادفی در اختیار ارزیابان انسانی[3] قرار گرفت. ارزیابان از بین دانشجویان کارشناسی‌ارشد رشته‌های اقتصاد، مدیریت و علوم‌سیاسی انتخاب شدند. برای آشنایی با هدف پروژه جلسات توجیهی برگزار شد. نتایج برچسب‌گذاری در جدول ۱ آمده است.

جدول (۱): نتایج برچسب‌گذاری انسانی اخبار

عنوان	اخبار برچسب‌گذاری شده	افزایش‌دهنده نا اطمینانی	خنثی
تعداد	۱۳۴۰۴	۱۷۰۱	۱۱۷۰۳
درصد	۱۰۰	۶۹ .۱۲	۳۱ .۸۷

منبع: یافته‌های پژوهش

مرحله سه، پیش‌پردازش متن: بعد از برچسب‌گذاری انسانی، برای انجام طبقه‌بندی اخبار لازم است که

[1] Text classification
[2] Supervised machine learning
[3] مکانیزم انتخاب ارزیابان انسانی و نحوه برچسب‌گذاری در پیوست شماره ۲ آمده است.

اخبار متنی به بردارهای عددی تبدیل شوند. به فرایند تبدیل متن به اعداد، بردارسازی[1] یا بازنمایی برداری[2] متون یا مدل بردار کلمات[3] می‌گویند. دو رویکرد برای بازنمایی برداری متون وجود دارد: رویکرد کوله کلمات[4] و رویکرد بازنمایی توزیع شده[5] (Han & Kamber, 2006). ازآنجایی‌که رویکرد کوله کلمات خروجی ضعیف‌تری نسبت به روش مشابه دارد، از رویکرد بازنمایی توزیع شده و روش word2vec استفاده شده است. مدل word2vec یک مدل نمایش‌دهنده کلمات است که میکالوو و همکاران (۲۰۱۳) آن را بسط داده‌اند. این مدل از دو لایه پنهانی استفاده می‌کند که در شبکه عصبی سطحی برای ساخت یک بردار از هر کلمه استفاده می‌کند در مرحله اول، روش‌های پیش‌پردازش متن انجام شده و فیلترهای نرمال‌سازی، جداسازی جملات و توکنایز کردن، ریشه‌یابی کلمات و حذف کلمات توقف صورت پذیرفته است[6]. در مرحله دوم، اخبار با استفاده از روش vec۲word به بردارهای عددی تبدیل شد. این فرآیند با استفاده از سامانه «متن‌کاوی فارسی‌یار[7]» صورت پذیرفته است.

مرحله چهار، طبقه‌بندی اخبار: در این مرحله ماشین با استفاده از ۴ الگوریتم یادگیری ماشین، اخبار را به دو دسته اثرگذار بر نا اطمینانی و خنثی تقسیم می‌کند. الگوریتم‌های مورد استفاده عبارت‌اند از: رگرسیون لاجستیک[8]، بیز ساده[9]، پرسپترون چندلایه[10]، درخت تصمیم[11] (C4. 5 یا J48). این الگوریتم‌ها با استفاده از نرم‌افزار WEKA صورت پذیرفته است. نتایج حاصل از برچسب‌گذاری توسط این چهار روش در جدول ۲ درج شده است.

---

[1] Vectorization
[2] Vector Representation
[3] Vector Space Model (VSM)
[4] Bag-of-Word (BOW)
[5] Distributed Representation
[6] توضیحات روش‌های پیش‌پردازش متن در پیوست ۳ درج گردیده است.
[7] www.text-mining.ir
[8] Logistic Regression
[9] Naïve Bayes
[10] Multilayer Perceptron
[11] Decision Tree

پژوهشگاه علوم انسانی و مطالعات فرهنگی
پرتال جامع علوم انسانی

جدول (۲): نتایج برچسب‌گذاری روش‌های یادگیری ماشین

عنوان	تعداد کل اخبار	اثرگذار (درصد)	خنثی (درصد)
داده‌های آموزشی	۱۳۴۰۴	۰.۱۲۷	۰.۸۷۳
بیز ساده	۳,۱۱۷,۹۶۰	۰.۰۲۱	۰.۹۷۹
C4. 5	۳,۱۱۷,۹۶۰	۰.۰۲۱	۰.۹۷۹
پرسپترون چندلایه	۳,۱۱۷,۹۶۰	۰.۰۱۲	۰.۹۸۸
لجستیک	۳,۱۱۷,۹۶۰	۰.۳۳۱	۰.۶۶۹

منبع: یافته‌های پژوهش

مهم‌ترین شاخص‌های ارزیابی عملکرد یادگیری ماشین در زیر آمده است:

- **مثبت صحیح (TP):** مدل به‌درستی خبر دارای بار نا اطمینانی را شناسایی کرده است.
- **مثبت کاذب (FP):** مدل به‌اشتباه خبر را دارای بار نا اطمینانی شناسایی کرده است. (خطای نوع اول)
- **صحت:** این شاخص به این دلیل معرفی می‌شود که مدل تمایل به بیش‌برآورد مثبت نداشته باشد. در واقع، صحت مقدار موارد صحیح طبقه‌بندی شده توسط الگوریتم از یک کلاس مشخص، به کل تعداد مواردی است که الگوریتم به‌صورت صحیح و غلط طبقه‌بندی کرده است:

$$Precision = \frac{TP}{TP + FP}$$

- **فراخوانی**[1]: زمانی که ارزش منفی کاذب بالا باشد، معیار فراخوانی اهمیت بالایی پیدا می‌کند. نسبت مقداری موارد صحیح طبقه بندی شده توسط الگوریتم از یک کلاس به

[1] Recall

تعداد موارد حاضر در کلاس مذکور، فراخوانی نامیده می‌شود:

$$Recall = \frac{TP}{TP + FN}$$

- **شاخص F**[1]: این شاخص معیار فراخوانی و دقت را به‌صورت همزمان در نظر می‌گیرد. در بهترین حالت ۱ و در بدترین حالت صفر است:

$$f_measure = \frac{2(recall * precision)}{recall + precision}$$

در بسیاری از موارد که وزن یا حساسیت دو گزینه یکسان است، از میانگین وزنی استفاده می‌شود؛ به‌عبارت‌دیگر، شاخص‌ها برای هر دو حالت اندازه‌گیری شده و سپس با احتساب وزن تعداد مشاهدات، میانگین‌گیری می‌شود. عملکرد هر یک از الگوریتم‌ها در جدول ۳ نمایش داده‌شده است. بهترین عملکرد را الگوریتم C4. 5 از شاخه درخت تصمیم دارد. بعدازاین الگوریتم، پرسپترون چندلایه، لاجستیک و ضعیف‌ترین عملکرد متعلق به روش بیز ساده است.

جدول (۱): مقایسه عملکرد الگوریتم‌های یادگیری ماشین

عنوان	بیز ساده	C4. 5	پرسپترون چندلایه	لجستیک
مثبت صحیح	۰. ۷۳۵	۰. ۹۵۴	۰. ۹۴۶	۰. ۸۸۰
مثبت کاذب	۰. ۲۳۲	۰. ۲۵۴	۰. ۱۹۹	۰. ۶۸۰
صحت	۰. ۸۷۲	۰. ۹۵۲	۰. ۹۴۶	۰. ۸۵۶
فراخوانی	۰۷۳۵	۰. ۹۵۴	۰. ۹۴۶	۰. ۸۸۰
شاخص F	۰. ۷۷۶	۰. ۹۵۱	۰. ۹۴۶	۰. ۸۵۶
درصد شناسایی صحیح	۰. ۷۳۴	۰. ۹۵۳	۰. ۹۴۶	۰. ۸۸۲

منبع: یافته‌های پژوهش

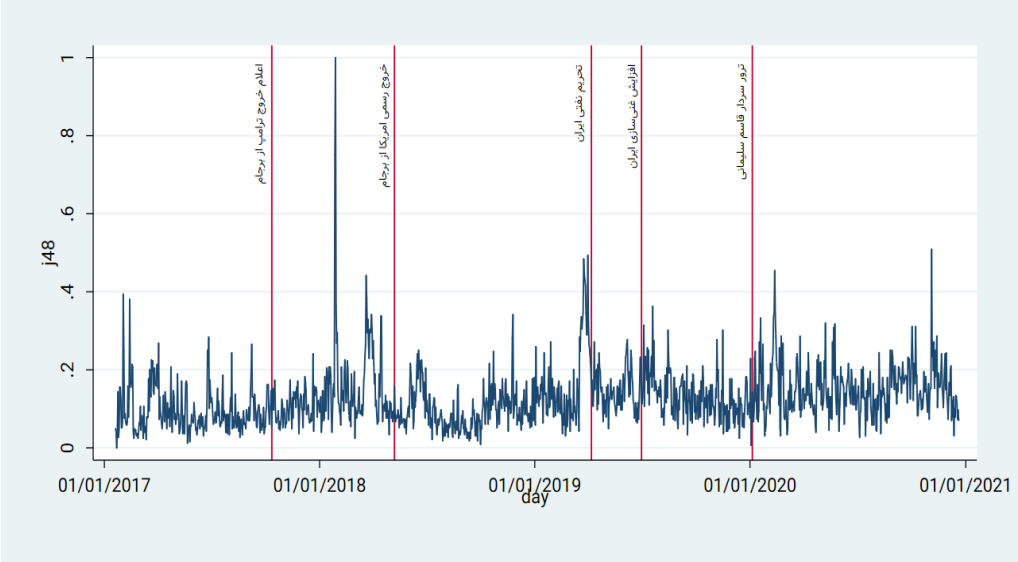
[1] F_measure

مرحله پنج، اندازه‌گیری شاخص نا اطمینانی اقتصادی رسانه‌بنیان: شاخص نا اطمینانی اقتصادی بعد از برچسب‌گذاری تمامی اخبار در دو مرحله اندازه‌گیری می‌شود: در مرحله اول، تعداد اخباری که دارای بار معنایی (اثرگذار) هستند، را در هرروز جمع کردیم (X_{it}). در مرحله دوم، به‌منظور استاندارسازی داده‌ها X_{it} را از حداقل کم کرده و بر فاصله بین حداقل و حداکثر تقسیم می‌کنیم. شاخص نا اطمینانی اقتصادی بین صفر و یک قرار دارد. هر چه شاخص به صفر نزدیک‌تر باشد، نشان‌دهنده سطح پایین‌تری از شاخص نا اطمینانی اقتصادی است و هر چه شاخص به یک نزدیک‌تر باشد، سطح بالاتری از نا اطمینانی اقتصادی را نشان می‌دهد. نام هر یک از شاخص‌های نا اطمینانی اقتصادی برمبنای الگوریتم یادگیری ماشین درج شده است. این چهار شاخص عبارتند از: j48، mlp ، logistic و nb. برای ایجاد درک بهتری از شاخص نا اطمینانی اقتصادی، ۵ اتفاق مهم در حوزه ایجاد نا اطمینانی اقتصادی در نظر گرفته شده است:

۱. اعلام خروج ترامپ از برجام،
۲. خروج رسمی ترامپ از برجام،
۳. تحریم نفتی ایران،
۴. افزایش غنی‌سازی ایران،
۵. ترور سردار قاسم سلیمانی.

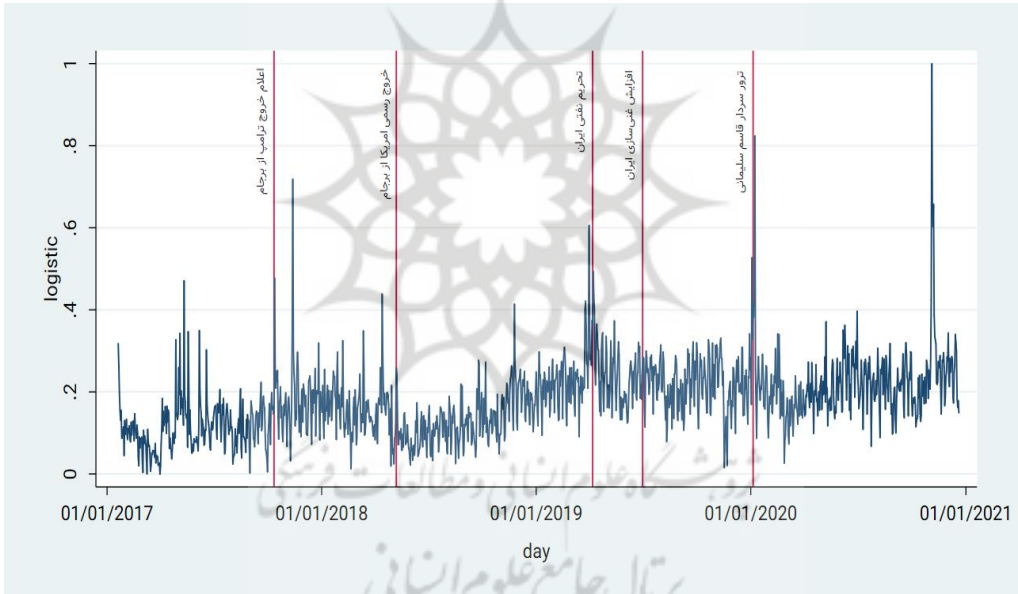
خروجی این دو مرحله، چهار سری‌زمانی شاخص نا اطمینانی اقتصادی به‌صورت روزانه شده است. این سری‌زمانی‌های بر اساس روش‌های مختلف طبقه‌بندی متن است. پنج نمودار زیر روند این سری‌های زمانی به‌صورت روزانه را نشان می‌دهد:

پژوهشگاه علوم انسانی و مطالعات فرهنگی
پرتال جامع علوم انسانی



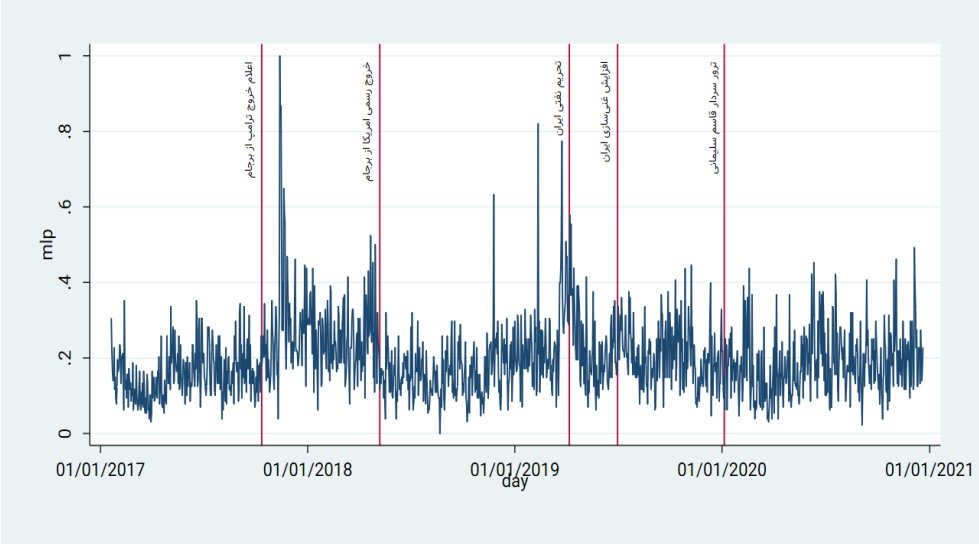
شکل (۱): شاخص نا اطمینانی اقتصادی روزانه بر مبنای روش C4. 5

منبع: یافته‌های پژوهش



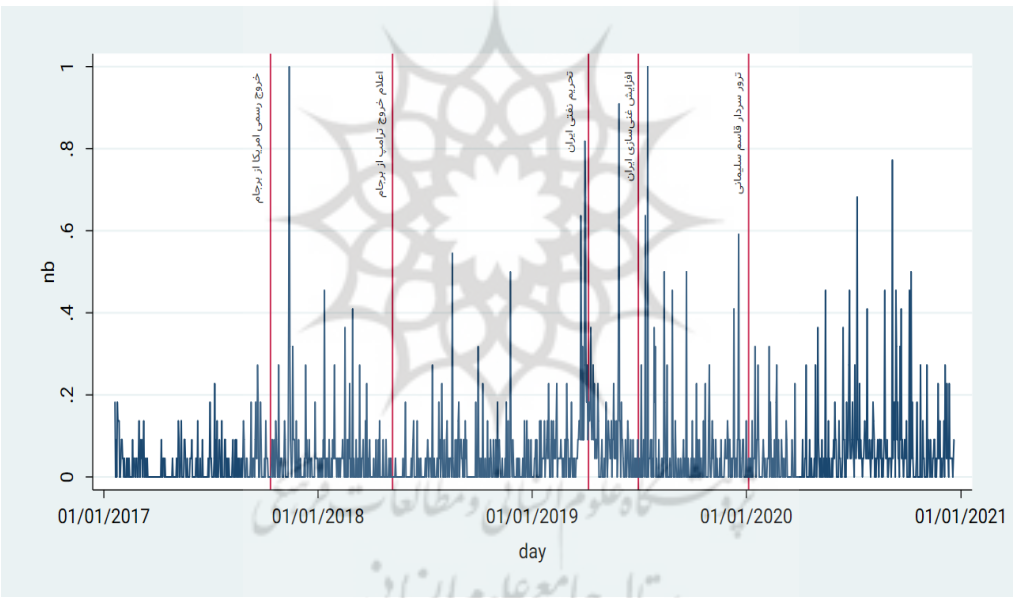
شکل (۲): شاخص نا اطمینانی اقتصادی روزانه بر مبنای روش لاجستیک

منبع: یافته‌های پژوهش



شکل (۳): شاخص نا اطمینانی اقتصادی روزانه بر مبنای روش پرسپترون چندلایه

منبع: یافته‌های پژوهش



شکل (۴): شاخص نا اطمینانی اقتصادی روزانه بر مبنای روش بیز ساده

منبع: یافته‌های پژوهش

۵. ارزیابی شاخص نا اطمینانی اقتصادی رسانه‌بنیان

در این بخش به ارزیابی شاخص نا اطمینانی اقتصادی با استفاده از سه روش شواهد تاریخی، روش برچسب‌گذاری مجدد و مقایسه با شاخص نا اطمینانی مبتنی بر داده‌های گوگل می‌پردازیم. این سه روش از مقاله بیکر و همکاران (۲۰۱۶) استخراج شده است.

ارزیابی با شواهد تاریخی: چنانچه این شاخص بتواند، وقایع تاریخی را توضیح دهد، می‌توان انتظار داشت که شاخص مناسبی است (Baker, 2016). به‌منظور بررسی شواهد تاریخی، ابتدا مهم‌ترین رخدادها را مرور می‌کنیم. ۵ اتفاق مهم در این دوره رخ داده است: اعلام خروج ترامپ از برجام، خروج رسمی ترامپ از برجام، تحریم نفتی ایران، افزایش غنی‌سازی ایران، ترور سردار قاسم سلیمانی.

شکل ۱ بر مبنای روش C4. 5 از روش‌های درخت تصمیم، رسم شده است. بالاترین سطح نا اطمینانی در دوره اعلام خروج ترامپ از برجام تا خروج رسمی آمریکا از برجام اتفاق افتاده است. در این دوره اخبار در مورد برجام و خروج آمریکا از آن شدت گرفته و سبب ایجاد سطح بالایی از نا اطمینانی اقتصادی شده است. ازآنجایی‌که شعار اصلی دولت یازدهم، مذاکرات هسته‌ای بود، افکار عمومی به نتیجه‌رسیدن مذاکرات را معادل بهبود وضعیت اقتصادی و شکست آن را به‌منزله، افول اقتصاد می‌دانست؛ بنابراین، این دوره بالاترین سطح شاخص نا اطمینانی اقتصادی را دارد. شاخص نا اطمینانی اقتصادی در حوالی تحریم نفتی افزایش می‌یابد. این جریان خبری به‌منزله بازگشت کامل تحریم‌ها به کشور بود. در مقابل، افزایش سطح غنی‌سازی ایران به‌عنوان برخورد ایران با خروج آمریکا از برجام، موجب تغییرات جدی در نا اطمینانی اقتصادی نشده است. ترور سردار قاسم سلیمانی دیگر اتفاقی است در حوالی آن شاخص نا اطمینانی اقتصادی رشد یافته است. به‌صورت کلی، می‌توان گفت که نوسانات شاخص نا اطمینانی اقتصادی محدود و دارای چند پرش بوده است که خروج آمریکا از برجام، تحریم نفتی و ترور سردار سلیمانی برمی‌گردد.

شکل ۲ بر مبنای روش لاجستیک ترسیم شده است. بالاترین سطح نا اطمینانی در این نمودار به اواخر سال ۲۰۲۰ برمی‌گردد که در دوره انتخابات ریاست‌جمهوری ترامپ است. سطح نا اطمینانی اقتصادی بعد از خروج رسمی ترامپ از برجام افزایش می‌یابد و با تحریم‌های نفتی به اوج خود می‌رسد. این الگوریتم توانسته است رخدادهای مهم را به‌خوبی توضیح دهد. ترور سردار قاسم سلیمانی و تحریم نفتی ایران به‌وضوح به‌این‌ترتیب است. در مقابل سطح نا اطمینانی در دوره اعلام خروج ترامپ از برجام تا خروج ترامپ از برجام، پایین‌تر از روش C4. 5 است.

شکل ۳ بر مبنای روش پرسپترون چندلایه آماده شده است. خروجی این الگوریتم نشان می‌دهد که

متوسط سطح نا اطمینانی تغییر چشم‌گیری نکرده است و عموماً شاخص بین ۰.۱ تا ۰.۴ در حال نوسان است؛ اما در رخدادها دارای جهش می‌شود و کم‌کم به سطح میانگین بازمی‌گردد. بالاترین سطح نا اطمینانی در دوره‌ای که زمزمه‌های خروج آمریکا از برجام بود، رخ داده است. بعدازاین دوره بالاترین سطح به تحریم نفتی ایران بازمی‌گردد. سطح شاخص در دوره ترور سردار قاسم سلیمانی و افزایش غنی‌سازی ایران تغییر چشم‌گیری نکرده است.

در شکل ۴، شاخص نا اطمینانی اقتصادی بر اساس روش بیز ساده رسم شده است. در این الگوریتم، تعداد روزهایی که شاخص صفر است، به‌مراتب بیشتر از سایر روش‌هاست. این اتفاق به دلیل نامتعادل بودن داده‌ها رخ داده است. از آن‌جایی‌که بخش عمده اخبار فاقد بار معنایی هستند، الگوریتم ترجیح می‌دهد سهم بیشتری از اخبار را بدون بار ارزیابی کند. بااین‌حال، بالاترین سطح در دوره خروج آمریکا از برجام و افزایش غنی‌سازی رقم خورده است. تحریم نفتی رتبه بعدی بالاترین سطح نا اطمینانی را دارد؛ اما اعلام خروج ترامپ از برجام، نتوانسته است شاخص را افزایش دهد که قابل‌تأمل است.

مقایسه با روش ارزیابی انسانی: به‌منظور ارزیابی کیفیت الگوریتم‌ها، ۵۰۰ خبر را که توسط الگوریتم‌ها برچسب‌گذاری شده بود را مجدداً در معرض ارزیابی انسانی قراردادیم. نتایج در جدول زیر منعکس شده است. درصدی از اخبار دارای بار نا اطمینانی که ارزیاب انسانی و الگوریتم یکسان تشخیص دادند در ستون «مثبت صحیح»، اخبار خنثی که هر دو خنثی تشخیص دادند در ستون «منفی صحیح» و اخباری که دارای بار نا اطمینانی بوده و الگوریتم اشتباه تشخیص داده است در «منفی کاذب» و اخباری که خنثی بوده و الگوریتم اشتباه تشخیص داده در «مثبت کاذب» ثبت شده است. نتایج جدول ۴ نشان می‌دهد که الگوریتم‌های یادگیری ماشین، عملکرد نسبتاً قابل قبول‌تری داشته‌اند.

جدول (٤): ارزیابی انسانی الگوریتم‌های یادگیری ماشین

روش	مثبت صحیح	منفی صحیح
C4. 5	۶۹	۹۱
پرسپترون چندلایه	۷۶	۹۸
لجستیک	۸۱	۹۶
بیز ساده	۶۲	۸۹

منبع: یافته‌های پژوهش

مقایسه با سایر شاخص‌های نا اطمینانی اقتصادی: روش دیگر ارزیابی شاخص نا اطمینانی اقتصادی، مقایسه آن با سایر روش‌ها است. با توجه به محدودیت‌های داده‌ها در ایران، از مقایسه با روش مبتنی بر جستجو اینترنتی استفاده کردیم. گوگل میزان جستجو را به‌صورت شاخص میزان جستجو[1] در اختیار عموم قرار می‌دهد. این شاخص میزان جستجوها را در زمان و مکان مشخص ارائه می‌کنند و بر اساس بازه زمانی مورد ارائه به روی عدد ۱۰۰ نرمال می‌کند. به‌این‌ترتیب که بیشترین فراوانی جستجو ۱۰۰ در نظر گرفته می‌شود و سایر مشاهدات بر اساس آن تعیین می‌شود. برای اندازه‌گیری شاخص نا اطمینانی سیاست خارجی، دو اقدام صورت پذیرفت: در گام اول، فهرستی از کلمات مرتبط با سیاست خارجی جمع‌آوری شد که شامل «تحریم»، «ترامپ»، «توافق‌هسته‌ای» و «برجام» می‌شود. در گام دوم، میزان چگالی این واژگان در پایگاه داده‌های گوگل موردبررسی قرارگرفت و «تحریم» و «برجام» به‌عنوان کلیدواژه اصلی در شکل‌گیری شاخص نا اطمینانی انتخاب شد. با توجه به اینکه داده‌های گوگل به‌صورت هفتگی است، برای ساختن شاخص ماهانه ابتدا مجموع مقادیر ماهانه را اندازه‌گیری شده و سپس استاندارد شده است. هرچه این شاخص به صفر نزدیک‌تر باشد نشان می‌دهد که سطح نا اطمینانی پایین‌تر است و هر چه به سمت یک میل کند، نشان‌دهنده سطح بالاتری از نا اطمینانی است.[2]

شکل ۵ و ۶ وضعیت این شاخص برای دو کلیدواژه «تحریم» و «برجام» را در این بازه زمانی نشان می‌دهد. در نمودار شاخص نا اطمینانی سیاست خارجی (کلیدواژه تحریم) اوج شاخص در دوره خروج رسمی آمریکا از برجام تا تحریم نفتی ایران اتفاق افتاده است. از آن‌جایی‌که این روش، مبتنی بر یک کلیدواژه است، وزنی زیادی به کلیدواژه مربوطه می‌دهد. این در حالی است که در الگوریتم‌های یادگیری ماشین، ۹۹ ویژگی برای هر خبر در نظر گرفته می‌شود. مقایسه این شاخص ماهانه در مقابل نمودارهای روش‌های مختلف نشان می‌دهد که نوسانات در دوره‌های مشابه تا اندازه‌ای یکسان است.

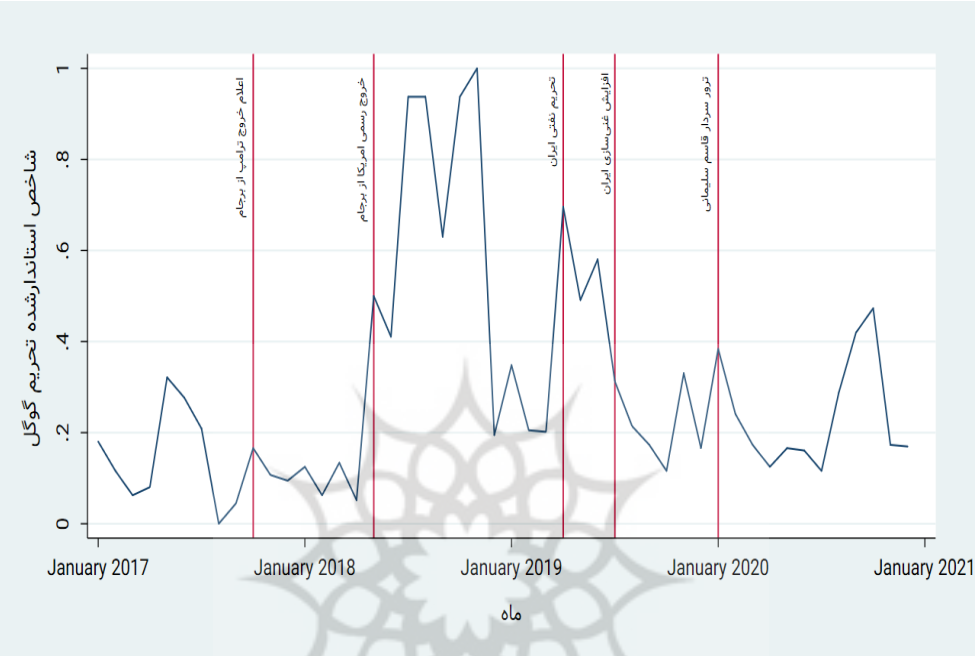
پژوهشگاه علوم انسانی و مطالعات فرهنگی
پرتال جامع علوم انسانی

[1] Search Volume Index

[2] نحوه محاسبه این شاخص در پیوست ۵ مقاله درج شده‌است.

این اتفاق در مورد شاخص نا اطمینانی سیاست خارجی (کلیدواژه برجام) نیز به‌صورت مشابه افتاده است؛ اما به دلیل تک متغیره بودن این روش، نا اطمینانی اقتصادی در دوره خروج رسمی آمریکا از برجام دارای بزرگ‌نمایی است. درنهایت باید گفت که الگوریتم‌های یادگیری ماشین به‌صورت همزمان کلیدواژه‌های متعددی را در نظر گرفته و نوسانات ناشی از هر یک از این موارد را در خود نشان می‌دهد.

شکل (۵): شاخص نا اطمینانی سیاست خارجی (برمبنای میزان جستجوی کلیدواژه تحریم)



منبع: یافته‌های پژوهش

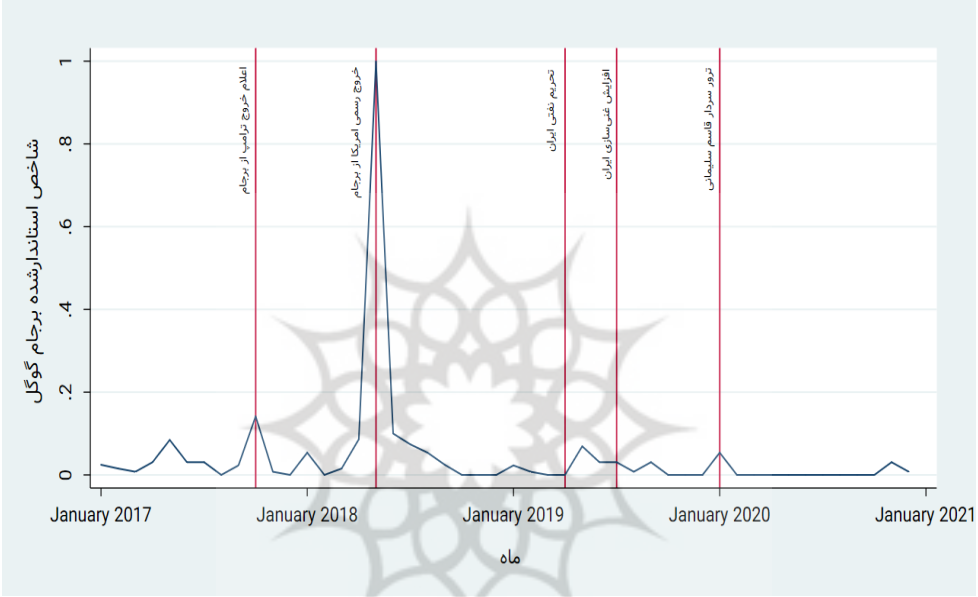
٦. اثر شاخص نااطمینانی‌اقتصادی بر نرخ ارز

به‌منظور بررسی اثر نااطمینانی‌اقتصادی بر نرخ ارز از مدل گارچ استفاده می‌کنیم که ابتدا آزمون‌های لازم برای وجود اثرات گارچ انجام شد. برای برآورد مدل از دو معادله میانگین و واریانس زیر استفاده شده است:

$$Ln(ER)=Bep+ARMA(p,q)+\varepsilon_t$$
$$Var(\varepsilon_t)=A()+\sum\alpha_{1,i}\varepsilon^2_{t-i}+\sum\alpha_{2,i}\sigma^2_{t-i}$$

در معادله فوق نااطمینانی‌اقتصادی در معادله میانگین آورده شده و جز ARMA نیز در معادله میانگین قرار دارد. داده‌های مورد استفاده در این مدل، لگاریتم ارز به‌عنوان متغیر وابسته است. این متغیرها به‌صورت روزانه مورد استفاده قرار گرفته‌اند. در معادله میانگین مدل گارچ شاخص‌های نااطمینانی‌اقتصادی (j48، پرسپترون چندلایه، لاجستیک و بیزساده) بدون عرض از مبدأ وارد شده و با روش‌شناسی باکس جنکینز[1] AR(1) و MA(1) وارد مدل شده است. در مرحله اول، آزمون وجود اثرات آرچ انجام شده است و نتایج نشان می‌دهد در هر چهار معادله اثرات آرچ وجود دارد. سپس معادلات میانگین و واریانس برآورد شده است. این رگرسیون با استفاده از نرم‌افزار استتا برآورد شده است.

شکل (۶): شاخص نا اطمینانی سیاست خارجی (بر مبنای میزان جستجوی کلیدواژه برجام)



منبع: یافته‌های پژوهش

پژوهشگاه علوم انسانی و مطالعات فرهنگی
پرتال جامع علوم انسانی

[1] Box & Jenkins

جدول (۵): نتایج مدل گارچ لگاریتم نرخ ارز

	پرسپترون چندلایه		لاجیستیک		j48		بیز ساده	
متغیر	**میانگین**	**واریانس**	**میانگین**	**واریانس**	**میانگین**	**واریانس**	**میانگین**	**واریانس**
L. ar	1. 000*** (0. 000)		1. 002*** (0. 000)		1. 002*** (0. 000)		1. 003*** (0. 000)	
L. ma	-0. 295*** (0. 014)		-0. 608*** (0. 002)		-0. 608*** (0. 002)		-0. 596*** (0. 002)	
L. arch	1. 972*** (0. 110)		2. 101*** (0. 093)		1. 902*** (0. 084)			1. 795*** (0. 074)
L2. arch	0. 319*** (0. 042)		1. 363*** (0. 065)		1. 254*** (0. 066)			1. 563*** (0. 064)
logistic			0. 005*** (0. 001)					
mlp	0. 059*** (0. 002)							
j48					0. 001 (0. 001)			
nb							0. 010*** (0. 001)	
جمله‌ثابت	10. 545*** (0. 102)	0. 000*** (0. 000)	10. 466*** (0. 013)	0. 000*** (0. 000)	10. 542*** (0. 010)	0. 000*** (0. 000)	10. 542*** (0. 012)	0. 000*** (0. 000)
تعدادمشاهدات	1,431	1,431	1,431	1,431	1,431	1,431	1,431	1,431

*** p<0. 01, ** p<0. 05, * p<0. 1

جملات خطا در پرانتز قرار دارد

منبع: یافته‌های پژوهش

نتایج حاصل از رگرسیون نشان می‌دهد که نااطمینانی‌اقتصادی با روش‌های پرسپترون چندلایه، لاجستیک و بیزساده بر میانگین لگاریتم نرخ دلار اثر مثبت و معنی‌داری دارد. به‌عبارت‌دیگر، با افزایش نااطمینانی‌اقتصادی، لگاریتم نرخ ارز افزایش پیدا می‌کند. ثابت‌بودن علامت ضرایب و معنی‌داری آن، به‌منزله اطمینان از پایداری نتایج مدل است. معادلات ARIMA نیز در تمامی مدل‌ها معنی‌دار است؛ اما معادلات واریانس نیز نشان می‌دهد که دو وقفه آرچ معنی‌دار است.

به‌طور خلاصه، می‌توان گفت که نااطمینانی‌اقتصادی محاسبه شده بر مبنای اخبار بر لگاریتم نرخ ارز

به‌صورت مثبت و معنی‌داری اثرگذار بوده است. این اثر در مدل پرسپترون چندلایه که عملکرد بهتری بر اساس شاخص‌های یادگیری‌ماشین داشت، بزرگ‌تر است. برای اعتبار بخشیدن بیشتر به نتایج پژوهش، محققین می‌توانستند به ارزیابی دقت عملکرد مدل GARCH بپردازند. ازآنجایی‌که استفاده از مدل ساده GARCH (1,1) در یک چارچوب همراه با نا اطمینانی، چندان مناسب به نظر نمی‌رسد، محققین می‌توانستند سایر مدل‌های GARCH را نیز در نظر گرفته و به مقایسه بین آن‌ها بپردازند.

٦. نتیجه‌گیری

نااطمینانی‌اقتصادی به‌عنوان یکی از متغیرهای تأثیرگذار بر تصمیمات اقتصادی بنگاه‌ها و خانوارها به‌حساب می‌آید. در این مقاله تلاش کردیم تا این متغیر مهم را اندازه‌گیری کنیم. در مطالعات تجربی ۴ دسته مدل برای اندازه‌گیری شاخص نا اطمینانی اقتصادی وجود دارد که به دلیل دسترسی به داده‌های شبکه‌های اجتماعی از روش مبتنی بر خبر استفاده شد. در این روش با استفاده از جمع‌آوری اخبار به‌صورت گسترده و برچسب‌گذاری آن‌ها با استفاده از روش‌های یادگیری ماشین، شاخص محاسبه می‌شود. به این منظور، ۳,۱۱۷,۹۶۰ خبر از ۲۸ کانال شبکه تلگرام در بازه زمانی ژانویه ۲۰۱۷ تا دسامبر ۲۰۲۰ جمع‌آوری شد. سپس، ۱۳۴۰۴ خبر توسط ارزیابان انسانی برچسب‌گذاری شد. برچسب‌ها دارای دو حالت «اثرگذار بر نا اطمینانی» و «خنثی» بودند. ارزیابان بر اساس پروتکل‌ها مشخص انتخاب شده و در دو جلسه توجیهی شرکت کردند.

در گام بعد، اخبار با استفاده از الگوریتم word2vec به بردارهای عددی تبدیل شدند تا آماده تحلیل‌های کمی شوند. بعد از برداری شدن اخبار، با استفاده از الگوریتم‌های یادگیری ماشین، اخبار برچسب‌گذاری شدند. روش‌های یادگیری ماشین مورد استفاده در این پژوهش عبارتند از: رگرسیون لاجستیک، بیز ساده، پرسپترون چند لایه، درخت تصمیم (C4. 5 یا J48). بررسی عملکرد این الگوریتم‌ها نشان داد که الگوریتم C4. 5 از شاخه درخت تصمیم، عملکرد بهتری دارد و پرسپترون چندلایه، لاجستیک و بیز ساده در رده‌های بعدی به لحاظ عملکرد قرار دارند. شاخص‌های ارزیابی عملکرد الگوریتم‌های یادگیری ماشین، برآورد مثبت صحیح، صحت، فراخوانی و شاخص F است که نشان می‌دهد تا چه اندازه الگوریتم قادر به تشخیص صحیح برچسب خبر است. به‌منظور ایجاد شاخص نا اطمینانی، تعداد اخبار «اثرگذار بر نا اطمینانی» شمرده و سپس بین صفر و یک استاندارد شد.

بررسی روند تاریخی ۴ شاخص نا اطمینانی رسانه‌بنیان نشان می‌دهد که ۳ شاخص بهتر می‌تواند اتفاقات تاریخی این دوره را توضیح دهد اما عملکرد شاخص مبتنی بر روش بیز ساده ضعیف‌تر است. همچنین،

ارزیابی شاخص نا اطمینانی اقتصادی رسانه‌بنیان، با استفاده از دو روش برچسب‌گذاری مجدد و مقایسه با شاخص نا اطمینانی مبتنی بر داده‌های گوگل انجام شد که نتایج نشان می‌دهد که عملکرد این شاخص مناسب است.

روند شاخص نا اطمینانی اقتصادی رسانه‌بنیان با روش C4. 5 با وقایع مهم دوره مطالعه همخوانی دارد به‌گونه‌ای که بالاترین سطح نا اطمینانی در دوره اعلام خروج ترامپ از برجام تا خروج رسمی آمریکا از برجام گزارش شده است؛ زیرا افکار عمومی به نتیجه‌رسیدن مذاکرات را معادل بهبود وضعیت اقتصادی و شکست آن را به‌منزله، افول اقتصاد می‌دانست. این شاخص در دوره شروع تحریم شدید نفتی ایران افزایش می‌یابد؛ چون که این جریان خبری به‌منزله بازگشت کامل تحریم‌ها به کشور بود. در مقابل، افزایش سطح غنی‌سازی ایران به‌عنوان برخورد ایران با خروج آمریکا از برجام، موجب تغییرات جدی در نا اطمینانی اقتصادی نشده است. ترور سردار قاسم سلیمانی دیگر اتفاقی است در حوالی آن شاخص نا اطمینانی اقتصادی رشد یافته است. در مجموع، روند شاخص نا اطمینانی اقتصادی ایران دارای چند پرش بوده است که به دوره خروج آمریکا از برجام، تحریم نفتی و بالا گرفتن تقابل آمریکا با ایران و بروز ترور سردار سلیمانی برمی‌گردد.

شاخص نا اطمینانی اقتصادی رسانه‌بنیان می‌تواند به‌صورت روزانه، و حتی بازه زمانی کوتاه‌تر، اطلاعات مهمی در اختیار تصمیم‌گیران اقتصادی و سیاسی قرار دهد که با توجه به تجربه وقایع سال‌های اخیر در کشور پیشنهاد می‌شود به‌منظور کاهش نا اطمینانی در مدیریت وقایع آینده استفاده شود. به عبارتی، میزان نا اطمینانی که محتوای اخبار وقایع مختلف نشان می‌دهد در محاسبات سیاست‌گذاران ایران باید در نظر گرفته شود.

مقاله حاضر، اولین تلاش در زمینه استفاده از روش‌های یادگیری ماشین در متن فارسی در حوزه نا اطمینانی اقتصادی است. تحقیقات آتی از چند منظر می‌توانند به غنای این حوزه بیفزایند: ۱. استفاده از سایر منابع خبری مانند پایگاه‌های اینترنتی، شبکه‌های اجتماعی، خبرگزاری‌ها و روزنامه‌ها. ۲. استفاده از الگوریتم‌های شخص‌سازی‌شده. ۳. بهبود روش‌های تبدیل متن به بردار عددی. ۴. محاسبه سایر شاخص‌های نا اطمینانی از قبیل سیاست خارجی و سلامت.

References

Abrishami, H.; Komijani, A.; Nouri, M., & Memarian, M. (2020). Calculation of uncertainty index based on an Internet search: A case study of the foreign exchange market of Iran. *The Journal of Economic Policy*, *12*(23),

99-131. (In Persian)

Aizenman, J.; & Marion, N. P. (1993). Policy uncertainty, persistence and growth. *Review of international economics*, *1*(2), 145-163. Alexopoulos, M., & Cohen, J. (2015).

The power of print: Uncertainty shocks, markets, and the economy. *International Review of Economics & Finance*, *40*, 8-28.

Arbatli, E. C.; Davis, S. J.; Ito, A., & Miake, N. (2022). Policy uncertainty in Japan. Journal of the Japanese and International Economies, measuring economic policy uncertainty, 1192.

Arshadi, A. (2011). Oil Price Volatility Modeling: A Frame for Measuring Uncertainty Index Using the Arima-Garch Model, *Energy Economics Review*, *8*(30); 205-220. (In Persian)

Bachmann, R., & Bayer, C. (2013). 'Wait-and-See'business cycles?. *Journal of Monetary Economics*, *60*(6), 704-719.

Bachmann, R.; Elstner, S., & Sims, E. R., (2010). *Uncertainty and Economic Activity: Evidence from Business Survey Data*. NBER Working Papers 16143, National Bureau of Economic Research, Inc.

Baker, S. R.; Bloom, N., & Davis, S. J. (2016). Measuring economic policy uncertainty. *The quarterly journal of economics*, *131*(4), 1593-1636.

Bateman, B. W. (1996). *Keynes's uncertain revolution*. University of Michigan Press.

Bauer, M.; Lakdawala, A., & Mueller, P. (2019*). Market-based monetary policy uncertainty*. WBS Finance Group Research Paper.

Baum, C. F.; Caglayan, M., & Talavera, O., (2010). On the Sensitivity of Firms' Investment to Cash Flow and Uncertainty. *Oxford Economic Papers*, *62* (2), 286–306.

Beckert, J. (1996). What is sociological about economic sociology? Uncertainty and the embeddedness of economic action. *Theory and society*, *25*(6), 803-840.

Bernanke, B. S., (1983). Irreversibility, Uncertainty, and Cyclical Investment. *The Quarterly Journal of Economics*, *98* (1), 85–106.

BINDER, C. (2017). Measuring uncertainty based on rounding: New method and application to inflation expectations. *Journal of Monetary Economics*, *90* (2017) 1–12.

Bloom, N.; Floetotto, M.; Jaimovich, N.; Saporta-Eksten, I., & Terry, S. J. (2018). Really uncertain business cycles. *Econometrica*, *86*(3), 1031-1065.

Bloom, N., (2007). Uncertainty and the Dynamics of R&D. *AMERICAN*

ECONOMIC REVIEW, 250-255.

Bloom, N., (2009). The impact of uncertainty shocks. *Econometrica*, *77* (3), 623–685.

Bontempi, M. E.; Frigeri, M., Golinelli, R., & Squadrani, M. (2021). EURQ: A New Web Search-based Uncertainty Index. *Economica*.

Bouri, E., & Gupta, R. (2021). Predicting Bitcoin returns: Comparing the roles of newspaper-and internet search-based measures of uncertainty. *Finance Research Letters*, *38*, 101398.

Bürgi, C., & Sinclair, T. M. (2021). What does forecaster disagreement tell us about the state of the economy?. *Applied Economics Letters*, *28*(1), 49-53

Cascaldi-Garcia, D.; Sarisoy, C.; Londono, J. M.; Rogers, J. H.; Datta, D.; RT Ferreira, T., & Zer, I. (2020). What is certain about uncertainty? *International Finance Discussion Paper*, (1294).

Claveria, O. (2021). On the aggregation of survey-based economic uncertainty indicators between different agents and across variables. *Journal of Business Cycle Research*, *17*(1), 1-26.

Davidson, P. (1991). Is probability theory relevant for uncertainty? A post Keynesian perspective. *Journal of Economic Perspectives*, *5*(1), 129-143.

De Bruin, W. B.; Manski, C. F.; Topa, G., & Van Der Klaauw, W. (2011). Measuring consumer uncertainty about future inflation. *Journal of Applied Econometrics*, *26*(3), 454-478.

Glas, A. (2020). Five dimensions of the uncertainty–disagreement linkage. *International Journal of Forecasting*, *36*(2), 607-627.

Gupta, R.; Subramaniam, S.; Bouri, E., & Ji, Q. (2021). Infectious disease-related uncertainty and the safe-haven characteristic of US treasury securities. *International Review of Economics & Finance*, *71*, 289-298

Han, J.; Kamber, M., & Mining, D. (2006). Concepts and techniques. *Morgan Kaufmann*, *340*, 94104-3205.

Jones, P. M.; Olson, E. (2013). The Time-Varying Correlation between Uncertainty, Output, and Inflation: Evidence from a DCC-GARCH Model. *Economics Letters*, *118*(1), 33–37.

Keynes, J. M. (2007). *A treatise on probability*. Cosimo, Inc.

Keynes, J. M. (1936). *the general theory of employment, interest andmoney*. Palgrave Macmillan, London.

Knight, Frank H. (1921). *Risk, Uncertainty and Profit*. University of Illinois at Urbana-Champaign's Academy for Entrepreneurial Leadership

Historical Research Reference in Entrepreneurship.

Lahiri, K.; Peng, H., & Sheng, X. S. (2022). *Measuring uncertainty of a combined forecast and some tests for forecaster heterogeneity*. In Essays in Honor of M. Hashem Pesaran: Prediction and Macro Modeling. Emerald Publishing Limited.

Lawson, T. (1985). *Keynes' economics: methodological issues*. Croom Helm, London

Lawson, T. (1988). Probability and uncertainty in economic analysis. *Journal of post Keynesian economics*, *11*(1), 38-65.

Leduc, S., & Zheng, L. (2016). Uncertainty shocks are aggregate demand shocks. *Journal of Monetary Economics*. *82*, 20–35.

Leduc, S., & Zheng, L. (2016). Uncertainty shocks are aggregate demand shocks. *Journal of Monetary Economics*, *82*, 20–35.

Lei, C.; Lu, Z., & Zhang, C. (2015). News on inflation and the epidemiology of inflation expectations in China. *Economic Systems*, *39*(4), 644-653.

Lucas, R. E. & Gillman, M. (2012). *Expectations and the Neutrality of Money*. In Collected Papers on Monetary Theory (pp. 1-24). Harvard University Press.

Lucas, R. E., & Prescott, E. C. (1971). Investment under Uncertainty. *Econometrica*, *39*(5), 659–681.

Lucas, R. E., & Sargent, T. J. (Eds.). (1981). *Rational expectations and econometric practice* (Vol. 2). University of Minnesota Press.

Ludvigson, S.; Sai Ma, C., & Serena, N. (2021). Uncertainty and Business Cycles: Exogenous Impulse or Endogenous Response?. *American Economic Journal: Macroeconomics*, *13* (4): 369-410.

Machina, M. J. (1987). Choice under uncertainty: Problems solved and unsolved. *Journal of Economic Perspectives*, *1*(1), 121-154.

Manski, C. F. (2010). When consensus choice dominates individualism: Jensen's inequality and collective decisions under uncertainty. *Quantitative Economics*, *1*(1), 187-202.

Manski, C. F. (2013). *Public policy in an uncertain world*. Harvard University Press.

Manski, C. F. (2015). Communicating uncertainty in official economic statistics: An appraisal fifty years after Morgenstern. *Journal of Economic Literature*, *53*(3), 631-53.

Marcus, A. A. (1981). Policy Uncertainty and Technological

Innovation. *The Academy of Management Review*, *6* (3), 443–448.

Menger, K., & Shubik, M. (2015). *The role of uncertainty in economics*. In Essays in Mathematical Economics, in Honor of Oskar Morgenstern (pp. 211-232). Princeton University Press.

Mikolov, T.; Chen, K.; Corrado, G., & Dean, J. (2013). *Efficient estimation of word representations in vector space*, preprint arXiv: 1301. 3781.

Moore, A. (2017). Measuring economic uncertainty and its effects. Economic record, 93(303), 550-575.

Rich, R. W.; Song, J., & Tracy, J. S. (2012). *The measurement and behavior of uncertainty: evidence from the ECB Survey of Professional Forecasters*. FRB of New York Staff Report, (588).

Rodrik, D. (1991). Policy Uncertainty and Private Investment in Developing Countries. *Journal of Development Economics*, *36* (2), 229–242

Rossi, B., &Tatevik Sekhposyan. (2015). Macroeconomic uncertainty indices based on nowcast and forecast error distributions. *American Economic Review*, *105* (5), 650–55.

Sahinoz, s.; Erdogan Cosar, E. (2018). *Economic policy uncertainty and economic activity in Turkey, Applied Economics Letters*

Sakhi, F.; Mohammadi, M., & Mohammadi, H. (2018). The effects of exchange rate volatility on foreign agricultural trade in Iran. *Journal of Agricultural Economics Researches*, *37*(10), 21-40. (In Persian)

Samuelson, PA. (1969*). Classical and neoclassical theory*. Penguin Books, London.

Shokri, M.; Barghi, M..; Motafaker Azad, M., & Salmani Bishak, M. (2020). Investigating the Impact of Economic Sanctions and Exchange Rate Uncertainty on FDI in Iran: A Fuzzy Approach. *Journal of Economics and Modeling*, 11(3), 33-59. (In Persian)

Smith, A. (2010). *The Wealth of Nations: An inquiry into the nature and causes of the Wealth of Nations*. Harriman House Limited.

Tobback, E.; Naudts, H.; Daelemans, W.; de Fortuny, E. J., & Martens, D. (2018). Belgian economic policy uncertainty index: Improvement through text mining. *International journal of forecasting*, *34*(2), 355-365.

Topçu, G., & Oran, J. (2021). Measuring economic policy uncertainty in Turkey. *International Journal of Econmic Policy in Emerging Economies*, *14*(3), 288-305.

Walras, L. (1896). *Éléments d'économie politique pure, ou, Théorie de la richesse sociale*. F. Rouge.

Williamson, O. E. (1993). Calculativeness, trust, and economic organization. *The journal of law and economics*, *36*(1, Part 2), 453-486.

Wu, X.; Kumar, V.; Quinlan, J. R.; Ghosh, J.; Yang, Q.; Motoda, H., & Steinberg, D. (2008). Top 10 algorithms in data mining. *Knowledge and information systems*, *14*(1), 1-37.

Zamanian, G., & Behrad Amin, M. (2015). Effect of Exchange Rate Uncertainty on the Import Demand of Iran Application of Ardl and Egarch Methods. *Quarterly Journal of Applied Economic Studies of Iran*, *3*(12), 129-148. (In Persian)

پژوهشگاه علوم انسانی و مطالعات فرهنگی
پرتال جامع علوم انسانی

پیوست یک- جدول فهرست کانال‌های تلگرامی

جدول ۲ پیوست- فهرست کانال‌های تلگرامی مورد استفاده در مقاله همراه با آدرس کانال و تعداد اخبار استفاده شده

ردیف	نام کانال	آی دی تلگرام	تعداد پیام‌ها
۱	انتخاب	@entekhab_ir	150,604
۲	اقتصاد آنلاین	@eghtesadonline	108,100
۳	عصر ایران	@MyAsriran	261,542
۴	صرفاً جهت اطلاع	@serfan_jahate_ettela	146,951
۵	خبر فوری	@akhbarefori	145,466
۶	اخبار رسمی	@khabari_rasmi	94,704
۷	اخبار روز	@iranfnews	138,222
۸	بی بی سی فارسی	@bbcpersian	71,275
۹	خبر فوری و مهم	@irforinews	62,979
۱۰	ایران خبر	@iranni_news	82,957
۱۱	خبر سراسری	@Anlainkhabari	46,947
۱۲	خبر فوری (قرمز)	@khabarfouri	99,739
۱۳	اخبار موثق	@MOVASAGH_IR	51,594
۱۴	خبر از ما	@KhabarAzMa	31,318
۱۵	اخبار منتخب	@akhbar_montakhab	68,386
۱۶	پارسینه	@parsinehnews	130,367
۱۷	اقتصاد نیوز	@Eghtesadnews_com	60,191
۱۸	دنیای اقتصاد	@den_ir	31,656
۱۹	خبرگزاری ایسنا	@isna94	110,157
۲۰	تابناک	@TeleTabnak	49,118
۲۱	خبرگزاری تسنیم	@Tasnimnews	111,129
۲۲	خبرگزاری فارس	@farsna	121,057
۲۳	خبرگزاری مهر	@mehrnews	38,014
۲۴	خبرآنلاین	@khabaronline_ir	211,760
۲۵	الف	@alef_news	177,945
۲۶	خبرگزاری ایلنا	@ilnair	305,992
۲۷	خبرگزاری ایرنا	@irna_1313	59,678
۲۸	آخرین خبر	@akharinkhabar	150,112
	جمع کل اخبار		3,117,960

منبع: یافته‌های پژوهش

پیوست دو- برچسب‌گذاری انسانی

با توجه به اینکه در این پژوهش از روش‌های یادگیری با نظارت استفاده شده است، ۱۰درصد از مجموع اخبار توسط ارزیابان انسانی برچسب‌گذاری شده است. برای سهولت این فرآیند، نرم‌افزار موبایلی تحت سیستم عامل اندروید طراحی و پیاده‌سازی شد. این نرم‌افزار به ما کمک کرد تا فرآیند برچسب‌گذاری انسانی با دقت بالاتری صورت بپذیرد و ثبت داده‌های آن به‌صورت آنی در پایگاه داده‌ها ثبت شود.

پیوست سوم-روش‌های پیش‌پردازش متن

در مراحل قبل اخبار کانال‌های تلگرامی دریافت، در پایگاه داده ثبت و توسط ارزیابان برچسب‌گذاری شد. به‌منظور تحلیل داده‌های کیفی اخبار، باید این اخبار به بردارهای عددی تبدیل شوند تا بتوان از روش‌های یادگیری‌ماشین بهره برد؛ اما قبل از انجام این تبدیل اقدامات پیش‌پردازش متن انجام پذیرفت تا با متن مناسب‌تری برای تبدیل به بردارهای عددی مواجه شویم. از۴ روش پیش‌پردازش متن در این مقاله استفاده شده است. تمامی این روش‌ها با استفاده از کدهای دستوری موجود در سامانه «متن‌کاوی فارسی‌یار»[1] صورت پذیرفته است:

نرمال‌سازی متن
جداسازی جملات و توکنایز کردن
ریشه‌یابی کلمات
حذف کلمات توقف

پیوست چهار-روش‌های طبقه‌بندی متن

رگرسیون لاجستیک

رگرسیون لاجستیک اولین بار توسط دیوید کاکس[2] (۱۹۵۸) معرفی شد. این روش شبیه رگرسیون خطی است و تلاش می‌کند که یک طبقه‌بندی‌کننده[3] مناسب با استفاده از ضرایب رگرسیون ارائه کند. برخلاف رگرسیون‌های خطی، خروجی این مدل احتمالی است که در آن متغیر ورودی مقادیر خاصی را دریافت می‌کند. این یک روش طبقه‌بندی دو حالتی است.

---

[1] https://text-mining.ir
[2] David Cox
[3] Classifier

قبل از این‌که وارد جزییات مدل رگرسیون لاجستیک شویم، علامت‌گذاری‌هایی که در آینده استفاده می‌کنیم را معرفی می‌نماییم. فرض کنید$X(X_1, X_2,\ldots,X_K)$ فضای ویژگی‌های k بعدی است که شامل متغیرهای توضیحی است. همچنین فرض کنید که$Y=\{0,1\}$ نشان‌دهنده کلاس‌های امکان‌پذیر است که می‌تواند برچسب صفر یا یک را دریافت کند. مجموعه داده‌های آموزشی مجموعه‌ای از (X_i,Y_i) است که برمبنای فرض توزیع‌شدگی یکسان (i. i. d) از فضای برداری برچسب‌گذاری شده (X,Y) انتخاب شده است. بر اساس داده‌های آموزشT، هدف یک طبقه‌بندی‌کننده دو حالتی پیداکردن تابع احتمالی مانند h است که بتواند با صحت خوبی کلاس‌های سایر داده‌ها را برچسب‌گذاری کند. این تابع مپینگ[1] از فضای ویژگی‌ها (X)به مجموعه کلاس‌ها (Y) است.

$$Y=\begin{cases}0 \text{ if inclass } 1\\ 1 \text{ if inclass } 2\end{cases} \qquad (۱)$$

در رگرسیون لاجستیک، تابع احتمال $Y=1$ به‌صورت زیر تعریف می‌شود:

$$H(X;\theta)=g(\theta^T X)=\frac{1}{1+e^{-\theta^T x}} \qquad (۲)$$

در این رابطه θ بردار پارامترهایی است که فضای تابعی مپینگ ازX به Y را پارامتری می‌کند و $g(z)=\frac{1}{1+e^{\bar{z}}}$

را تابع لاجستیک یا تابع سیگموئید[2]می‌نامند.

وظیفه رگرسیون لاجستیک پیداکردن بردارθ است. به‌گونه‌ای که h(x;θ) بتواند توزیع احتمالY را بر اساس داده‌های آموزشی مدل‌سازی کند. برای پیداکردنθ، می‌توان یک سری فرض در مورد تابع احتمال در نظر گرفت و پارامترθ با استفاده از روش حداکثر درست‌نمایی[3] برآورد کرد. به‌صورت دقیق‌تر، فرض می‌کنیم که برچسب کلاس‌ها و ویژگی‌های ورودی به‌وسیله تابع زیر با هم مرتبط هستند:

پژوهشگاه علوم انسانی و مطالعات فرهنگی
پرتال جامع علوم انسانی

[1] Mapping
[2] Sigmoid Function
[3] Maximum Likelihood

(۳)

$$\begin{cases} p(y=1)=p \\ p(y=0)=q \end{cases} \quad \text{where } p=h(x),\ q=1-p$$

تابع حداکثر درست‌نمایی نیز این‌گونه تعریف می‌شود:

$$L(\theta)=\prod_{i=1}^{m} p(y_i \mid x_i;\theta) \qquad (٤)$$

در تابع فوق، $p(y_i|x_i;\theta)$، احتمال وقوع Y_i به شرط x_i که توسط θ پارامتری شده است؛ بنابراین، باید θ انتخاب شود که تابع $L(\theta)$ را حداکثر کند.

بیز ساده

الگوریتم بیز ساده یکی از قدیمی‌ترین روش‌های رسمی طبقه‌بندی است که در عین سادگی از کارایی به‌شدت بالایی برخوردار است (وو و همکاران[1]، ۲۰۰۸). این الگوریتم طبقه‌بندی بر پایه احتمالات و برمبنای قضیه بیز شکل گرفته است. برای هر کلاس، احتمال پسین[2] با استفاده از داده‌های آموزشی محاسبه می‌شود و کلاس نا اطمینانی شده بر اساس بالاترین احتمال پسین انتخاب می‌شود. الگوریتم بیز ساده احتمال پسین $P(Y|X)$ بر اساس برآورد درست‌نمایی $P(X|Y)$ و احتمال $P(Y)$ محاسبه می‌کند. در این روابط $X=X_1,X_2,\ldots,X_K$ نشان‌دهنده بردار ویژگی‌هاست و Y کلاس داده را مشخص می‌کند. برای محاسبه $P(Y|X)$ فرض بر این است که هر یک از ویژگی‌ها به‌صورت مشروط مستقل از یکدیگر هستند تا برآورد $P(X|Y)$ آسان‌تر شود. بر اساس این فرض $P(X|Y)$ را می‌توان این‌گونه برآورد کرد:

$$P(X|Y)=P(X_1,X_2,\ldots,X_K|y)=\prod_{i=1}^{k} p(x_i|k) \qquad (٥)$$

پژوهشگاه علوم انسانی و مطالعات فرهنگی
پرتال جامع علوم انسانی

[1] Wu et al
[2] Posterior Probability

بر اساس رابطه فوق، الگوریتم بیز ساده احتمال‌های پسین برای هر کلاس در Y را این‌گونه محاسبه می‌کند:

$$P(Y=y_l|X_1,...,X_k)=\frac{P(Y=y_l)\prod_{i=1}^{k}P(X_i|Y=y_l)}{\sum_j P(Y=y_j)\prod_{i=1}^{k}P(X_i|Y=y_l)} \quad (۶)$$

علاوه بر محاسبه احتمال‌های پسینی برای هر کلاس، الگوریتم طبقه‌بندی بیز ساده کلاس با بالاترین احتمال پسینی را نیز نشان می‌دهد:

$$\arg\max P(Y=y_l)\prod_{i=1}^{k}P(X_i|Y=y_l) \quad (۷)$$

روش‌های برآورد $P(X_i|Y),i\in\{1,2,...,k\}$ بر اساس نوع ویژگی‌ها (پیوسته یا گسسته) تعیین می‌گردد. بااین‌حال، در هر دو حالت روش حداکثر درست‌نمایی معمولاً برای برآورد $P(X_i|Y)$ به کار می‌رود (Barber, 2012). زمانی که ویژگی‌ها به‌صورت گسسته هستند، برآورد حداکثر درست‌نمایی به شیوه زیر محاسبه می‌گردد:

$$P(X_i=x_{ij}|Y=y_l)=\frac{number\ of\ times\ X_i=x_{ij}\ for\ class\ y_l}{number\ of\ instances\ in\ class\ y_l} \quad (۸)$$

زمانی که x_i متغیر پیوسته است، معمولاً از روش توزیع گاوسی برای برآورد استفاده می‌شود (Michael, 1997). در این شرایط، طبقه‌بندی‌کننده بیز ساده را «طبقه‌بندی کننده بیز ساده گاوسی» می‌نامند. در این شرایط، ما باید میانگین ($\widehat{\mu}_{il}$) و واریانس ($\sigma^2{}_{il}$) هر یک از $y_l\in Y$ را برآورد کنیم. معمولاً از برآوردکننده‌های حداکثر درستنمایی برای برآورد $\widehat{\mu}_{il}$ و $\sigma^2{}_{il}$ استفاده می‌کنند. برآوردکننده $\widehat{\mu}_{il}$ به شکل زیر محاسبه می‌شود:

$$\hat{\mu}_{il} = \frac{sum\ x_i\ values\ of\ class\ y_i}{number\ of\ ins\ tance\ s\ in\ class\ y_i} \quad (۹)$$

و برآوردگر σ^2_{il} به‌صورت زیر به دست می‌آید:

$$\sigma^2_{il} = \frac{sum\ of\ square\ of\ differences\ between\ x_i\ values\ and\ \hat{\mu}_{il}}{number\ of\ ins\ tance\ s\ in class\ y_i} \quad (۱۰)$$

محاسبه $\hat{\mu}_{il}$ و $\sigma^2_{il}, P(X_i = x_{ij} | Y = y_i)$ از طریق برآورد تابع چگالی احتمال زیر می‌تواند انجام شود:

$$P(X_i = x_{ij} | Y = y_i) = \frac{1}{\sqrt{2\sigma^2_{il}\pi}} e^{-\frac{(x_{ij}-\hat{\mu}_{il})^2}{2\sigma^2_{il}}} \quad (۱۱)$$

در انتها $P(Y = y_i)$ را به‌وسیله برآورد ماکزیمم راست‌نمایی به شرح ذیل به دست می‌آوریم:

$$P(Y = y_i) = \frac{number\ of\ ins\ tances\ in\ class\ y_i}{number\ of\ ins\ tances\ in\ class\ y_i\ in\ the\ training\ set} \quad (۱۲)$$

توجه کنید که با NB می‌توانیم $P(X|Y)$ و $P(Y)$ را مدل‌سازی کنیم. مادامی‌که توزیع مشترک p(xly) با p(xly)p(y) برابر باشد می‌توانیم توزیع مشترک با NB را مدل‌سازی کنیم. به‌عنوان نتیجه مثال‌های دیتاهای جدید می‌توانند نمونه‌ای از NB باشند. بدین دلیل NB به‌عنوان مدل تولیدکننده[1] شناخته می‌شود (Flach, 2012).

پژوهشگاه علوم انسانی و مطالعات فرهنگی
پرتال جامع علوم انسانی

[1] Generative

درخت‌تصمیم (C4. 5)

درخت طبقه‌بندی و رگرسیون (CART[1]) یکی از الگوریتم‌های مفید در یادگیری‌ماشین برای نا اطمینانی است. CART اولین بار توسط بریمن و همکاران (۱۹۸۴) معرفی شد و گامی بزرگ در پیشبرد علم آمار غیرپارامتریک، داده‌کاوی، یادگیری‌ماشین و هوش مصنوعی بود. CART یک الگوریتم قدرتمند و ساده درخت‌تصمیم بود. ایده اصلی در اینجا یافتن شکاف‌های باینری روی متغیرها به روش از بالا به پایین است؛ به‌گونه‌ای که هر شکاف یک فرزند راست و یک فرزند چپ تولید می‌کند. بهترین شکاف در هر مرحله به شکلی انتخاب می‌شود که ناخالصی هر گره فرزند را در مقایسه با گره والدین کاهش دهد. تمام متغیرها برای یافتن بهترین تقسیم فردی در هر مرحله آزمایش می‌شوند. انشعابات موفقیت‌آمیز منجر به ساخت درخت می‌شود. اگر این فرآیند برای همه موارد تکرار شود تا اینکه تنها یک مقدار در برگ آخر باقی بماند، با نا اطمینانی بیش‌ازاندازه و نادرست مواجه می‌شویم. برای جلوگیری از نا اطمینانی بیش‌ازاندازه، از معیار توقف برای حداقل تعداد برگ استفاده می‌شود. انحراف یا آنتروپی ممکن است برای تعیین ناخالصی استفاده شود. ناخالصی در برگ این‌گونه محاسبه می‌شود:

$$I_t = -2\sum_{j\ at\ t} n_{tj} \log p_{j|t} \qquad (۱۳)$$

در رابطه بالا $p_{j|t} = \frac{p_{tj}}{p_t}$ و $p_{j|t}$ احتمال رسیدن یک مورد به برگ t در کلاس j است و p_t احتمال رسیدن به برگ t است. همچنین $p_{j|t}$ به‌وسیله $\frac{n_{tj}}{n_t}$ برآورد می‌شود که در آن n_{tj} تعداد کلاس J است که به برگ t می‌رسد و n_t تعداد کل مواردی است که به t می‌رسد.

بنابراین ناخالصی برآوردشده برابر خواهد بود با:

$$\hat{I}_t = -2\sum_{j\ at\ t} n_{tj} \log \frac{n_{tj}}{n_t} \qquad (۱۴)$$

پژوهشگاه علوم انسانی و مطالعات فرهنگی
پرتال جامع علوم انسانی

[1] Classification and Regression Tree

و انحراف در یک گره T به‌این‌ترتیب محاسبه می‌گردد:

$$I(T) = \sum_{leaves\ t \in T} I_t \qquad (۱۵)$$

درخت‌تصمیم $C_{4.5}$ یکی از روش‌های بسیار متداول یادگیری‌ماشین برای طبقه‌بندی است (Quinlan، 1993). این روش از رویکرد تقسیم و حل[1] برای ساختن درخت‌تصمیم استفاده می‌کند. $C_{4.5}$ از چند منظر با CART متفاوت است. وو و همکاران (۲۰۰۸) مهم‌ترین تفاوت را این‌گونه ذکر کرده‌اند: « $C_{4.5}$ اجازه آزمون‌های دوگانه یا چندگانه را می‌دهد». این در حالی است که CART فقط آزمون‌های دوجانبه را پشتیبانی می‌کند.

پرسپترون چندلایه

مک کالوک و همکاران (۱۹۴۰) روش شبکه عصبی را به‌عنوان ابزار نا اطمینانی با ساخت مدل ریاضی از یک سیستم ناشناخته ارائه کردند. یکی از معروف‌ترین روش‌های شبکه‌های عصبی، شبکه‌های عصبی پرسپترون چندلایه (MLP) است. شبکه MLP شامل یک لایه ورودی و خروجی و حداقل یک لایه مخفی است. هرکدام از این لایه‌ها دارای تعدادی نورون و شامل واحدهای پردازش است. سیگنال‌های ورودی به وسیلهٔ ضریب‌های به‌هنجارکننده به مقدار یک نرمالیزه شده و بعد از محاسبات، خروجی به مقدار واقعی برگردانده می‌شود. همچنین مقادیر اولیهٔ وزن‌ها به‌صورت اتفاقی در نظر گرفته شده‌اند. نحوه آموزش پرسپترون چند لایه از طریق الگوریتم پس‌انتشار است. به این ترتیب که خروجی‌های واقعی با خروجی‌های دلخواه مقایسه می‌شوند و وزن‌ها به‌وسیله الگوریتم پس‌انتشار، به‌صورت تحت نظارت تنظیم می‌گردند تا الگوی مناسب به وجود آید. برای الگوری ورودی pام، مربع خطای خروجی برای تمامی سلول‌های لایه خروجی شبکه به‌صورت زیر در می‌آید:

$$E_p = \frac{1}{2}(d^p - y^p)^2 = \frac{1}{2}\sum_{j=1}^{s}((d_j{}^p - y_j{}^p)^2 \qquad (۱۶)$$

[1] Divide-Conquer Approach

که در آن $d_j{}^p$ خروجی دلخواه برای jامین سلول در لایه خروجی و خروجی واقعی برای jامین سلول در لایه خروجی، S ابعاد بردار خروجی، y^p بردار خروجی واقعی، d^p بردار خروجی دلخواه است. مرجع خطای کل E برای P الگو به‌صورت زیر درمی‌آید:

$$E = \sum_{p=1}^{P} E_p = \frac{1}{2}\sum_{p=1}^{P}\sum_{j=1}^{s}(d_j{}^p - y_j{}^p)^2 \qquad (۱۷)$$

وزن‌ها با هدف کاهش تابع هزینه E به مقدار حداقل کردن به روش گرادیان نزولی تنظیم می‌گردند. معادله به روزدرآوردن وزن‌ها به‌صورت زیر است:

$$w_{ij}(t+1) = w_{ij}(t) + \eta\Delta w_{ij}(t) + \alpha\Delta w_{ij}(t-1) \qquad (۱۸)$$

که در آن $\Delta w_{ij} = -(\frac{\partial E_p}{\partial w_{ij}(t)})$ و η ضریب یادگیری، α ضریب لحظه‌ای $w_{ij}(t+1)$ وزن جدید و $w_{ij}(t)$ وزن قبلی است.

روند یادگیری زمانی متوقف می‌شود که مجموع کل خطا، برای p الگو از مقدار آستانه تعیین شده کمتر شود یا تعداد کل دوره‌های تعلیم به پایان برسد.

پژوهشگاه علوم انسانی و مطالعات فرهنگی
پرتال جامع علوم انسانی

پیوست پنجم- محاسبه شاخص نا اطمینانی سیاست خارجی

گوگل میزان جستجو را به‌صورت شاخص میزان جستجو[1] در اختیار عموم قرار می‌دهد. این شاخص میزان جستجوها را در زمان و مکان مشخص ارائه می‌کنند و بر اساس بازه زمانی مورد ارائه به روی عدد ۱۰۰ نرمال می‌کند. به این ترتیب که بیشترین فراوانی جستجو ۱۰۰ در نظر گرفته می‌شود و سایر مشاهدات بر اساس آن تعیین می‌شود.

به‌عبارت‌دیگر، فرمول محاسبه این شاخص بر یک واژه مورد جستجو S در زمان t به شرح زیر است:

$$SVI_{st} = \frac{sv_{st}}{sv_{Gt}.MSV_{[0,T]}}(100) = \frac{sv_{st}}{sv_{Gt}.\max_{t=[0,T]}\{\frac{sv_{st}}{sv_{Gt}}\}}(100) \qquad (۱۹)$$

معادله ۱۹- محاسبه شاخص حجم جستجو اینترنتی/ منبع: گوگل ترندز

$SVst$ میزان جستجو واژه s در بازه زمانی t و SV_{Jt} بیانگر تمامی جستجوها در بازه زمانی t است. $MSV_{O,T}$نشان‌دهنده حداکثر مقدار نسبت $\frac{SV_{st}}{SV_{Gt}}$ در طی دوره زمانی ۰ تا t است و با قرارگرفتن آن در مخرج کسر و ضرب کل کسر در عدد ۱۰۰، مقدار شاخص SVI_{st} (شاخص حجم جستجوی اینترنتی) همواره در بازه ۰ تا ۱۰۰ می‌گیرد.

برای محاسبه شاخص نا اطمینانی سیاست خارجی، دو اقدام صورت پذیرفت: در گام اول، فهرستی از کلمات مرتبط با سیاست خارجی جمع‌آوری شد که شامل «تحریم»، «ترامپ»، «توافق‌هسته‌ای» و «برجام» می‌شود. در گام دوم، میزان چگالی این واژگان در پایگاه داده‌های گوگل موردبررسی قرارگرفت و «تحریم» و «برجام» به‌عنوان کلیدواژه اصلی در شکل‌گیری شاخص نا اطمینانی انتخاب شد. با توجه به اینکه داده‌های گوگل به‌صورت هفتگی است، برای ساختن شاخص ماهانه ابتدا مجموع مقادیر ماهانه را محاسبه شده و سپس استانداردشده است. برای استانداردکردن داده‌ها بین صفر و یک، از رابطه زیر استفاده شده است:

$$FPU = \frac{SVI - Min}{Max - Min} \qquad (۲۰)$$

پژوهشگاه علوم انسانی و مطالعات فرهنگی
پرتال جامع علوم انسانی

[1] Search Volume Index

معادله ۲۰- استانداردکردن شاخص نا اطمینانی سیاست خارجی/ منبع: محاسبات پژوهش

شاخص نا اطمینانی سیاست خارجی ایران (FPU) در بازه زمانی پژوهش و به‌صورت هفتگی محاسبه شده است. همچنین SVI، شاخص حجم جستجوی اینترنتی است. هرچه این شاخص به صفر نزدیک‌تر باشد نشان می‌دهد که سطح نا اطمینانی پایین‌تر است و هر چه به سمت یک میل کند، نشان‌دهنده سطح بالاتری از نا اطمینانی است.

پژوهشگاه علوم انسانی و مطالعات فرهنگی
پرتال جامع علوم انسانی