Analyzing Twitter Users Posts and Presenting a Pattern for Application of Data in Policy Process

Somayeh Labafi*
PhD in Media Management; Assistant Professor; Iranian Research Institute for Information Science and Technology (IranDoc); Tehran, Iran Email: labafi@irandoc.ac.ir

Mohammad Mahdi Kavousi
MA in Computer engineering; Institute of Higher Education of Ershad Damavand, Damavand, Iran Email: kavousi.mm@gmail.com

Iranian Journal of Information Processing and Management

Received: 26, Jan. 2022 **Accepted: 17, Oct. 2022**

Iranian Research Institute for Information Science and Technology (IranDoc)
ISSN 2251-8223
eISSN 2251-8231
Indexed by SCOPUS, ISC, & LISTA
Vol. 38 | No. 4 | pp. ??-??
Summer 2023
https://doi.org/jipm.38.4

Abstract: In this research, the activity of Twitter's users has been analyzed. The sample of the research was the users who posted on the topic of technology. The purpose of this research was to present the process of using the data of Twitter's users in the policy making process. This process has suggested how to use data analysis in each step of the policy making process. Algorithms presented in this process were tested with 6-month data of all Persian language Twitter's users who have published tweets on the topic of science and technology policy. The proposed classification model was able to distinguish and separate evidence tweets from non-evidence tweets with acceptable accuracy. Also, the text of the tweets of the most influential Twitter users who posted tweets and the number of connections their tweets created were extracted and analyzed. The selection criterion of these users was the number of connections they had made (more than 300 connections), which shows that these users have paid the most attention to information filtering, applications and social networks. The innovation of this research was the proposal to use the data of social network users in the policy making process.

Abstract: Science and Technology, Social Media, User Analysis, Twitter

* Corresponding Author

شناسایی و تحلیل رفتار کاربران شبکه‌های اجتماعی متمرکز بر فناوری و ارائه فرایند کاربست داده‌ها در سیاست‌گذاری شواهد مبنا

سمیه لبافی
دکتری مدیریت رسانه؛ استادیار؛ پژوهشگاه علوم و فناوری اطلاعات ایران (ایرانداک)؛ تهران، ایران؛
پدیدآور رابط labafi@irandoc.ac.ir

محمدمهدی کاووسی
کارشناسی ارشد نرم‌افزار؛ دانشگاه ارشاد دماوند؛ دماوند، ایران kavousi.mm@gmail.com

پژوهش‌نامهٔ پردازش و مدیریت اطلاعات

دریافت: ۱۴۰۰/۱۱/۰۶ | پذیرش: ۱۴۰۱/۰۷/۲۵ | مقاله برای اصلاح به مدت ۴ ماه نزد پدیدآوران بوده است.

نشریهٔ علمی | رتبهٔ بین‌المللی
پژوهشگاه علوم و فناوری اطلاعات ایران (ایرانداک)
شاپا (چاپی) ۸۲۲۳-۲۲۵۱
شاپا (الکترونیکی) ۸۲۳۱-۲۲۵۱
نمایه در LISTA، ISC، و SCOPUS
jipm.irandoc.ac.ir
دورهٔ ۳۸ | شمارهٔ ٤ | ص‌ص ۱۲۲۷-۱۲۵۶
تابستان ۱۴۰۲
https://doi.org/jipm.38.4

چکیده: در این پژوهش رفتار کاربران شبکه اجتماعی توئیتر که به مسائل سیاستی حوزه فناوری حساس بوده‌اند و در رابطه با آن محتوا تولید کرده‌اند، تحلیل شده است. در این پژوهش شیوه کاربست داده‌های کاربران رسانه‌های اجتماعی در فرایند سیاست‌گذاری و تحلیل رفتار این کاربران مورد نظر بوده است. ازاین‌رو، فرایند کاربست داده‌های رسانه‌های اجتماعی در سیاست‌گذاری علم و فناوری ارائه شده است. این فرایند چگونگی کاربست تحلیل داده‌ها در هر مرحله از فرایند سیاست‌گذاری را پیشنهاد داده است. الگوریتم‌های ارائه‌شده در این فرایند با داده‌های ۶ ماهه، کلیه کاربران فارسی‌زبان توئیتر که با موضوع سیاست‌های حوزه علم و فناوری توئیت منتشر کرده‌اند، آزمون شد و مدل دسته‌بندی پیشنهادی توانست شواهد از غیرشواهد را با دقت خوبی تشخیص و جدا سازد. متن توئیت‌های نفوذدارترین کاربران توئیتر که در کلیدواژه‌های استخراج‌شده توئیت ارسال بودند، به همراه تعداد ارتباطاتی که توئیت آن‌ها ایجاد کرده بود، استخراج و تحلیل شد. معیار انتخاب این کاربران، تعداد ارتباطاتی بود که ایجاد کرده بودند (بیش از ۳۰۰ ارتباط) که نشان می‌دهد این کاربران بیش از همه، موضوع فیلترینگ اطلاعات، اپلیکیشن‌ها و شبکه‌های اجتماعی را مد نظر داشتند.

کلیدواژه‌ها: سیاست‌گذاریِ مبتنی بر شواهد، علم و فناوری، رسانه‌های اجتماعی، تحلیل کاربران

۱. مقدمه

یکی از بهترین راه‌های شناخت و درک ویژگی‌های کاربران برای مد نظر قرار دادن دیدگاه‌های آن‌ها در فرایند حکمرانی و سیاست‌گذاری، استفاده از رسانه‌های اجتماعی است. رسانه‌های اجتماعی می‌تواند کانالی میان کاربران و سیاست‌گذاران باشد و به‌عنوان منبعی جدید در درگیرسازی شهروندان در تدوین و اجرای سیاست‌ها به‌کار رود (Mellouli, Driss & Trabelsi 2019). این پدیده فرصتی بزرگ برای دولت‌ها ایجاد می‌کند تا در مورد شهروندان یاد بگیرند و به‌طور مؤثر با آن‌ها ارتباط برقرار کنند. اگر دولت‌ها بخواهند یاد بگیرند چگونه از مشارکت کاربران در رسانه‌های اجتماعی بهره‌برداری کنند، پایش رسانه‌های اجتماعی یا به‌طور کلی، نظارت اجتماعی در این فضا باید مورد توجه قرار گیرد (Panagiotopoulos, Bowen & Brooker 2017).

از آنجا که کاربران رسانه‌های اجتماعی، محتوای کاربرساخته راجع به سیاست‌های علم و فناوری تولید می‌کنند (Simonofski, Fink & Burnay 2021; Lee, Lee & Choi 2020; Driss, Mellouli & Trabelsi 2019; Gintova 2019; Panayiotopoulos, Bowen & Brooker 2017; Fernandez et al. 2014) و همچنین، سیاست‌گذاران نسبت به محتوای تولیدشده در رسانه‌های اجتماعی حساسیت بیشتری دارند، بستر رسانه‌های اجتماعی مناسب‌ترین بستر است که مورد مطالعه این پژوهش بوده است. سیاست‌گذاران نشان داده‌اند که به‌دلیل تأثیرگذاری رسانه‌های اجتماعی (Valenzuela et al. 2017; Bucher & Helmond 2017)، نسبت به محتوای تولیدشده در آن حساسیت دارند. شاهد این مدعا پیشینه پژوهش‌های حوزه سیاست‌گذاری و رسانه اجتماعی است و همچنین سامانه‌های تحلیل دادهٔ رسانه اجتماعیِ داخل کشور که همگی تحلیل دادهٔ رسانه‌های اجتماعی را با هدف اصلیِ مشتریِ دولتی توسعه داده‌اند. اما سیاست‌گذاران در ایران تاکنون به‌طور نظام‌مند از داده‌های رسانه‌های اجتماعی در فرایند سیاست‌گذاری و از جمله سیاست‌گذاری حوزه علم و فناوری استفاده نکرده‌اند و در حد رصدهای مقطعی به این موضوع کفایت شده است. ناکارآمدی و شکست برخی از اسناد سیاستی در حوزه فناوری‌های نوین که عدم رضایت کاربران را در پی داشته، شاهد این مدعاست که سیاست‌گذار از دانش زمینه‌ای تولیدشده توسط کاربران در پلتفرم‌های رسانه اجتماعی استفاده نکرده است. این عدم حضور دیدگاه‌های کاربران رسانه‌های اجتماعی که نسبت به موضوعات سیاستی حوزه علم و فناوری در حوزه‌های مختلف

حساس هستند، می‌توانند فرایند تدوین، اجرا و ارزیابی سیاست‌ها را در آینده با چالش روبه‌رو سازد؛ چرا که دیدگاه‌های این کاربران با ارائه دانش زمینه‌ای نسبت به مشکلات و مسائل می‌توانند مسیر سیاست‌گذاری مبتنی بر شواهد در این حوزه را تسهیل نماید. این سؤال که داده‌های رسانه‌های اجتماعی چگونه می‌توانند در فرایند سیاست‌گذاری علم و فناوری دخیل شود، تاکنون پاسخ روشنی نیافته است (Mellouli, Driss & Trabelsi 2019). جنبه‌های تاریک این موضوع، پژوهشگران را به این‌سو هدایت کرده است که چگونه می‌توان داده‌های رسانه‌های اجتماعی را در سیاست‌گذاری علم و فناوری به‌طور نظام‌مند به‌کار بست؟ بدین منظور در مرحله اول، فرایند کاربست داده‌های رسانه‌های اجتماعی در فرایند سیاست‌گذاری ارائه شده است. سپس، رفتار کاربران منتشرکننده توئیت تحلیل و روش‌های استخراج دانش از این توئیت‌ها متناسب با حوزه سیاست‌گذاری علم و فناوری، شناسایی شده و شیوه کاربست آن در فرایند سیاست‌گذاری پیشنهاد داده شده است.

۲. مبانی نظری

۲-۱. سیاست‌گذاری مبتنی بر شواهد

استفاده از اطلاعات، ابزار قدرتمندی برای کمک به سیاست‌گذاران و دستیابی به موفقیت در سیاست‌گذاری‌هاست که آغاز آن به دوره حکمرانی کارآمد[1] در اوایل دوران مدرنیته در اروپا بازمی‌گردد. پس از این دوران بود که رویکرد سیاست‌گذاری مبتنی بر شواهد ظهور کرد. سیاست‌گذاری مبتنی بر شواهد، استفاده از نظام‌مند از شواهد در شکل‌دهی به سیاست‌های عمومی است؛ رویکردی که منشأ آن در حوزه پزشکی و درمان مبتنی بر شواهد[2] است (Howlett 2009; Parsons 2002; Sanderson 2002). سیاست‌گذاری مبتنی بر شواهد زمانی فراگیر شد که مدرن‌سازی دولت‌ها[3] در دستور کار کشورها قرار گرفت و تمایلات برای سیاست‌گذاری بر مبنای تحلیل‌های علوم اجتماعی و ... بیشتر شد (Parsons 2002; Sanderson 2002). سیاست‌گذاران و مدیران دولتی در پاسخ به نیازها و فشارهای شهروندانی که خدمات برای آن‌ها ناکافی یا نامناسب بود، تلاش می‌کردند. سرخوردگی مدیران سازمان‌های عمومی در مورد نرخ پایین بازگشت سرمایه‌گذاری‌ها در برنامه‌های اجتماعی، آن‌ها را به جست‌وجوی رویکردهای جدید سیاست‌گذاری سوق داد

1. efficient governance 2. evidence-based medicine 3. modernizing government

(Cairney 2016). این بیداری سیاست‌گذاران، فرصت‌هایی را برای علوم رفتاری و اجتماعی ایجاد کرد تا راه‌حل‌هایی به شکل رویکردهای جدید برای به‌دست آوردن کنترل بیشتر بر واقعیت‌های مبهم و آشفته ارائه دهد. راه‌حل‌های مبتنی بر دستورالعمل‌های ایدیولوژیک قدیمی دیگر کارایی نداشت و سیاست‌گذاران از رویکردهای جدیدتر استقبال می‌کردند (Parkhurst 2017). این جهت‌گیری‌های جدید با تمایل سازمان‌های مردم‌نهاد بزرگ، اصناف و دیگر ذی‌نفعان برای مشارکت یا رویکردهای مشارکتی برای پرداختن به مسائل اصلی همراه شد (ibid). سیاست‌گذاران در پی حل مشکلات خود و افزایش کارایی سیاست‌ها، به رویکرد سیاست‌گذاری مبتنی بر شواهد[1] که خروجی تحقیقات حوزه علوم اجتماعی در دهه‌های ۶۰ و ۷۰ بود، روی آوردند. این رویکرد جدید در سیاست‌گذاری، نتیجه نمایان شدن ناکارآمدی‌های سیاست‌گذاری‌های دولتی و اجرای آن در زمینه‌های مختلف بود. در این رویکرد تلاش شده است تمرکزگرایی در رابطه با تدوین سیاست‌ها کم‌رنگ‌تر شود و دایره تنگ سیاست‌گذاران و کنش‌های سیاست‌گذار گسترده‌تر گردد.

رویکرد سیاست‌گذاری مبتنی بر شواهد، هم یک مجموعه فعالیت‌های تاکتیکی در سیاست‌گذاری را ارائه می‌دهد و هم به‌دنبال تبیین شیوه‌های مشروع‌تر تصمیم‌گیری توسط سیاست‌گذاران است که می‌تواند جایگزین سیاست‌گذاری‌های مبتنی بر شهود[2]، عقیده و ایدئولوژی شود (Cartwright & Hardie 2012). محققان این حوزه معتقدند، رویکرد سیاست‌گذاری مبتنی بر شواهد دو اصل اساسی دارد و در واقع، مبتنی بر دو زمینه‌ای است که تا وجود نداشته باشد نمی‌توان به پیاده‌سازی درست این نوع سیاست‌گذاری امید داشت. اولین زمینه اساسی، یک فرهنگ سیاسی مطلوب است که اجازه ورود عناصر اساسی شفافیت و عقلانیت را در فرایند سیاست‌گذاری می‌دهد. این موضوع به‌نوبه خود ممکن است اولویت سیاست‌گذاران را برای استفاده از دانش در سیاست‌گذاری‌ها تسهیل کند. دوم، فرهنگ تحقیقاتی متعهد به تحقیقات تحلیلی با استفاده از روش‌های دقیق علمی که طیف وسیعی از شواهد را برای سیاست‌گذاری تولید می‌کند. رویکرد سیاست‌گذاری مبتنی بر شواهد در رابطه با استفاده از روش‌های بهتر برای جمع‌آوری نظام‌مند شواهد و استفاده دقیق از آن به‌عنوان پایه و اساس تصمیم‌گیری در بخش عمومی در سه دههٔ پیش تاکنون مورد توجه قرار گرفته است. در نتیجه، در حال حاضر، بدنه بزرگی از مطالعات در این حوزه وجود دارد که با تحلیل پیچیدگی‌های عوامل زمینه‌ای کمک می‌کند تا شرایط تأثیرگذار بر اثربخشی سیاست‌ها را توضیح دهند (Head 2010). ایجاد این فرهنگ تحقیقاتی

1. evidence-based policy (EBP) 2. intuition-based policy (IBP)

که پاسخ‌گوی نیازهای سیاست‌گذار در زمینه‌های مختلف باشد، هنوز در همه زمینه‌های سیاست‌گذاری ایجاد نشده و نمی‌توان با کمبود این فرهنگ تحقیقاتی که بتواند اطلاعات مورد نیاز سیاست‌گذار را تهیه کند، امید به ایجاد و توسعه سیاست‌گذاری مبتنی بر شواهد در جوامع داشت.

۲-۲. رسانه‌های اجتماعی و سیاست‌گذاری

استفاده از داده‌های رسانه‌های اجتماعی و در کل، کلان‌داده‌ها در سیاست‌گذاری برای اولین‌بار در کشور چین مطرح شد. در سال ۲۰۱۲، دفتر سیاست‌گذاری علم و فناوری کاخ سفید[1] نیز «برنامه تحقیق و توسعه کلان‌داده‌ها» را مطرح کرد، که از آن پس کلان‌داده‌ها در همه سیاست‌گذاری‌های دولت فدرال آمریکا مورد استفاده قرار گرفت. در سال ۲۰۱۳، دولت استرالیا «راهبرد خدمات عمومی فناوری اطلاعاتی و ارتباطاتی ۲۰۱۲ تا ۲۰۱۵ استرالیا»[2] را منتشر کرد. در سال ۲۰۱۳، انگلیس سند «فرصت‌ها برای سود: راهبرد قابلیت کلان‌داده انگلیس»[3] را منتشر ساخت. فرانسه نیز در سال ۲۰۱۵ برای اولین‌بار «برنامه اقدام جهت تسهیل جمهوری دیجیتال در ۲۰۱۵»[4] را راه‌اندازی کرد. در همان سال، دولت چین نیز «بیانیه توسعه برنامه اقدامی برای ترویج توسعهٔ کلان‌داده‌ها»[5] را منتشر کرد که کلان‌داده‌های چین را به‌صورت ملّی مورد استفاده قرار می‌داد. پس از این اسناد، تاکنون کشورهای مختلف در قالب اسناد سیاستی مختلف، بهره‌برداری از داده‌های رسانه‌های اجتماعی و به‌طور کلی، کلان‌داده‌ها را در سیاست‌گذاری‌های ملی مورد نظر قرار داده‌اند (Yan 2018).

«نپاناگاپلوس، باون و بروکر» در قالب پژوهشی که در خصوص تعیین ارزش داده‌های رسانه‌های اجتماعی به‌عنوان شواهد سیاستی از منظر سیاست‌گذاران انجام دادند، ویژگی‌های داده‌های رسانه‌های اجتماعی در مقابل منابع سنتی شواهد را از منظر چارچوب قابلیت‌های جمعی معرفی کردند (Panayiotopoulos, Bowen & Brooker 2017). کار آن‌ها نشان می‌دهد که در مقایسه با روش‌های ثابت و سنتی پیشین، پیچیدگی‌های فنی تجزیه‌وتحلیل داده‌های رسانه‌های اجتماعی به انواع جدیدی از قابلیت‌های علوم داده (به‌عنوان مثال، ابزارها و مهارت‌های ادغام/ دستکاری مجموعه‌های کلان‌داده) نیاز دارد. به‌عنوان مثال، روش‌هایی مانند تجزیه‌وتحلیل هشتگ‌ها، به‌طور کلی، در تحقیقات علوم

1. White House Science and Technology Policy Office (OS-TP)
2. Australian Public Service Information and Communication Technology Strategy 2012- 2015
3. Opportunities for Profits: The UK Data Capability Strategy
4. Action Plan to Facilitate Digital Republic In 2015
5. Notice on The Development of Action Plan for The Promotion of Big Data Development

اجتماعی ایجاد شده‌اند. برای نمونه کارهای Barnett et al. (2012) و Highfield, Harrington & Bruns (2013)، در این خصوص بوده است. اما تاکنون برای فعالیت‌های تجاری یا سیاستی که تمرکز آن‌ها روی تحلیل آمارها و اندازه‌گیری‌های محتوای منتشر شده در رسانه‌های اجتماعی بوده، اعمال نشده است. این موضوع پیچیدگی استفاده از این داده‌ها در سیاست‌گذاری‌ها را نشان می‌دهد که نیازمند تحلیل‌های متفاوتی نسبت به دیگر زمینه‌های استفاده از داده‌های رسانه‌های اجتماعی است.

به‌دلیل جدید بودن موضوع استفاده از داده‌های رسانه‌های اجتماعی در سیاست‌گذاری، تاکنون پژوهش‌های کمی به این حوزه پرداخته‌اند. تحقیقات فعلی هنوز به ترکیب داده‌های مخاطبان دیجیتال با فرایند سیاست‌گذاری و اینکه چگونه این منابع می‌توانند به سیاست‌گذاران کمک کند تا بینش جدیدی داشته باشند، توجه کافی نشان نداده است. اگرچه فرصت‌های مشخصی برای تعامل با جوامع جدید در رسانه‌های اجتماعی وجود دارد، با این حال، ارتباط بین مخاطبان سنتی و رسانه‌های اجتماعی می‌تواند به‌ویژه برای نهادهای سیاست‌گذار چالش‌برانگیز باشد (Panagiotopoulos et al. 2015). سیاست‌گذاران ابتدا باید فرضیاتی مناسب در مورد شهروندان دیجیتال بپذیرند تا بتوانند سودمندی ورود داده‌های جمعی را که توسط مخاطبان رسانه‌های اجتماعی تولید می‌شود، به‌درستی ارزیابی کنند. ایجاد چنین فهمی از نظرات عمومی به‌طور معمول، تأثیر زیادی در رویه‌های سیاست‌گذاری دارد و به‌نوبه خود تغییراتی در رابطه با تصمیمات مهم سیاستی به‌دنبال می‌آورد (Barnett et al. 2012; Walker et al. 2010).

Driss, Mellouli & Trabelsi (2019) با این پیش‌فرض که رسانه‌های اجتماعی می‌توانند کانالی میان سیاست‌مداران و شهروندان باشند، به ارائه چارچوبی جهت استخراج دانش از رسانه اجتماعی «فیس‌بوک» پرداخته‌اند. این چارچوب بر اساس ابزار تحلیل معنایی متن ارائه شده است که داده‌ها را از رسانه‌های اجتماعی جمع‌آوری و داده‌های ارزشمند را جهت ارائه به سیاست‌گذاران انتخاب می‌کند. «ناپولی»، استدلال می‌کند که رسانه‌های اجتماعی بنا بر چارچوب نظری «منابع عمومی»، یک منبع عمومی محسوب می‌شود که می‌توان و باید آن را در جهت توسعه منافع عمومی مورد استفاده قرار داد. این پژوهش استدلال می‌کند که داده‌های جمع‌آوری شدهٔ کاربر در رسانه‌های اجتماعی می‌توانند به‌عنوان یک منبع عمومی باشد و در توسعه و نظارت بر اجرای سیاست‌ها در راستای منافع عمومی قابل بهره‌برداری است (Napoli 2019).

«استملاتوس» و همکاران، در پژوهش خود نشان می‌دهند که ویژگی‌های ساختاری رسانه اجتماعی

آنلاین توئیتر در یونان می‌تواند اطلاعات ارزشمندی در مورد وابستگی سیاسی گره‌های کاربران ارائه کند. به بیان دیگر، آن‌ها نشان می‌دهند که از روی مشخصات ساختاری کاربران توئیتر می‌توان برای پیش‌بینی تمایل سیاسی گره‌های مهم کاربری استفاده کرد. آن‌ها از توسعه الگوریتم‌های تشخیص‌دهنده استفاده کردند تا بتوانند از روی این سیگنال‌های ارائه‌شده از سوی کاربران، پیش‌بینی‌هایی را انجام دهند (Stamatelatos et al. 2020). «پوی و شولر» در پژوهشی تأثیر رسانه‌های اجتماعی در مشارکت شهروندان در زمینه‌های سیاستی را بررسی کرده و نتیجه گرفتند که هرچه دسترسی به رسانه‌های اجتماعی بیشتر باشد و افراد بیشتر در این رسانه‌ها فعالیت کنند، در حوزه‌های سیاستی نیز مشارکت بیشتری خواهند داشت (Poy & Schuller 2020). از آنجا که استفاده از پلتفرم‌های رسانه‌های اجتماعی یک منبع جدید از اطلاعات را که مورد نیاز دولتمردان است، ایجاد کرده است، باید به‌دقت توسط دولتمردان تحلیل شود و داده‌های ارزشمند از آن استخراج گردد.

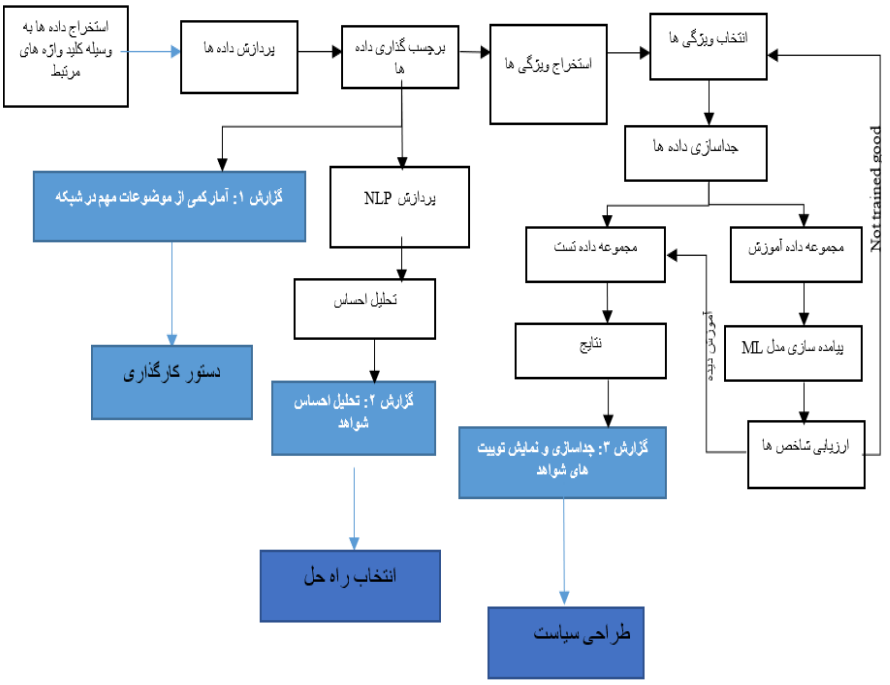
۳. فرایند پیشنهادی کاربست داده‌های رسانه‌های اجتماعی در سیاست‌گذاری

پژوهش حاضر به‌دنبال طراحی فرایندی است که بتوان تحلیل‌های مناسبی از کاربران رسانه‌های اجتماعی و داده‌های آن‌ها را توسعه داده و آن را متناسب با بخش‌های مختلف فرایند سیاست‌گذاری حوزه علم و فناوری ارائه داد. فرایند ارائه‌شده (نمودار ۱) با به‌کارگیری روش‌ها و تکنیک‌های تحلیل داده‌های رسانه‌های اجتماعی به سیاست‌گذار حوزه فناوری در خصوص ارائه شواهد معتبر از طریق رسانه‌های اجتماعی کمک می‌کند. بر اساس مدل چرخه‌ای سیاست‌گذاری از Young et al. (2002) سیاست‌گذاری در فرایندی شش‌مرحله‌ای (تنظیم دستور کار، تدوین گزینه‌های سیاستی، انتخاب گزینه‌های سیاستی، تدوین سند سیاستی، اجرای سیاست‌ها، و ارزیابی سیاست‌ها) صورت می‌گیرد. بر اساس این مدل که به فعالیت‌های سیاست‌گذار در هر مرحله از فرایند سیاست‌گذاری پرداخته است، سیاست‌گذاری متشکل از اقدامات و فعالیت‌هایی است که بایستی در مراحل مختلف انجام شود. اولین مرحله که در این مدل و از جمله در بسیاری از مدل‌های سیاست‌گذاری وجود دارد، مرحله تعیین دستور کار است و اینکه چه موضوعات و مسائلی برای مداخله سیاست‌گذارانه مد نظر قرار گیرد. با توجه به روش‌های تحلیل داده‌های توئیتر، نتایجی که پس از استخراج داده‌ها، تمیزسازی داده‌ها و برچسب‌گذاری موضوعی داده‌ها، انجام می‌شود، می‌تواند در قالب نمودار شماره ۱، مسائل مهم را در زمینه سیاست‌گذاری علم و فناوری برای سیاست‌گذار مشخص سازد. این خروجی اولیه در رابطه با انتخاب مسائل

سیاستی که باید مداخله سیاست‌گذارانه روی آن انجام گیرد، مناسب است؛ چرا که این خروجی میزان محتوای تولیدشده در توئیتر در رابطه با موضوعات مختلف سیاست‌گذاری علم و فناوری را تدارک می‌بینند.

دومین خروجی که این فرایند تحلیل داده‌های توئیتر تدارک می‌بینند، تحلیل احساس داده‌های استخراج‌شده و برچسب‌گذاری‌شده در رابطه با کلیدواژه‌های مختلف علم و فناوری است. تحلیل احساس، میزان قطبیت محتوای تولیدشده (توئیت‌های تولیدشده) در رابطه با کلیدواژه‌های سیاست‌گذاری علم و فناوری را مشخص می‌سازد. حجم و قطبیت محتوای تولیدشده در رابطه با کلیدواژه‌های مختلف می‌تواند به سیاست‌گذار در انتخاب راه‌حل‌های پیشنهادی که در دست دارد، کمک کند. سیاست‌گذار در این مرحله نیازمند کسب اطلاعات در مورد نظرات ذی‌نفعان مختلف سیاست‌گذاری در رابطه با راه‌حل‌های مختلف سیاستی است که یکی از کانال‌های اطلاع از نظرات ذی‌نفعان در رابطه با گزینه‌های سیاستی مطرح، داده‌های توئیتر و استفاده از تحلیل احساس این داده‌هاست. این خروجی می‌تواند به‌عنوان دومین خروجی فرعی، مورد استفاده سیاست‌گذار علم و فناوری قرار گیرد.

سومین خروجی و خروجی اصلی که این فرایند با به‌کارگیری الگوریتم‌های یادشده پس از تحلیل داده‌های توئیتر ارائه می‌کند، تشخیص و جداسازی توئیت‌هایی است که می‌تواند به‌عنوان شواهد در تدوین سیاست‌ها مورد استفاده قرار گیرد. از آنجا که سیاست‌گذار برای استفاده از داده‌های توئیتر، مشکل تشخیص شواهد از غیرشواهد را در میان انبوهی از توئیت‌های مرتبط با حوزه علم و فناوری دارد، توسعه الگوریتمی که بتواند به او در این رابطه کمک کند، از جمله اهداف این پژوهش است. در این راستا یک الگوریتم یادگیرنده طراحی شده است. در این الگوریتم که بخشی از فرایند اصلی است، داده‌ها را پس از پاک‌سازی و انتخاب بر اساس ویژگی‌های مشخص، به دو قسمت داده‌های آزمون و داده‌های آموزش تقسیم می‌شود و با استفاده از فرایند یادگیری و پیاده‌سازی آن روی حجمی از توئیت‌ها، الگوریتم، توئیت‌هایی را که می‌تواند به‌عنوان شواهد مورد استفاده قرار گیرد، از غیر شواهد تشخیص خواهد داد. پس از این مرحله، خروجی به‌دست‌آمده که بر اساس تحلیل شواهد موجود در توئیتر در مورد سیاست‌گذاری علم و فناوری است، می‌تواند در تدوین سیاست‌ها مورد استفاده سیاست‌گذار قرار گیرد.



نمودار ۱. فرایند کاربست داده‌های رسانه‌های اجتماعی در سیاست‌گذاری علم و فناوری

٤. روش پژوهش

٤-١. ارزیابی فرایند پیشنهادی

در فرایند پیشنهادی، نحوه تأثیر تحلیل‌های داده‌های شبکه اجتماعی توئیتر در فرایند تصمیم‌گیری سیاست‌گذاران ارائه شده است. همان‌گونه که در فرایند ارائه‌شده مشاهده گردید، این فرایند بر مبنای توسعه سه بخش تأثیرگذار در فرایند سیاست‌گذاری طراحی شده است. در بخش اول، گزارش آماری موضوعات پرتکرار در پیام‌های کاربران ارائه می‌شود؛ بخش دوم، تحلیل رفتار کاربران، میزان قطبیت و احساس درون متن‌ها را با استفاده از پردازش زبان طبیعی استخراج کرده و به‌عنوان گرایش نظرات کاربران در موضوع‌های سیاستی در حوزه علم و فناوری ارائه می‌دهد؛ در بخش سوم یک الگوریتم پیشنهادی با استفاده از تکنیک‌های یادگیری ماشین که پیام‌های شواهد را از پیام‌های غیرشواهد تشخیص می‌دهد، ارائه شده است.

٤-۱-۱. جمع‌آوری داده[1]

در راستای ارزیابی فرایند پیشنهادی، اولین مرحله‌ای که به‌منظور تشخیص توئیت‌های شواهد از غیرشواهد انجام شد، اقدام به جمع‌آوری داده از شبکه اجتماعی توئیتر با استفاده از ابزار توئیتر API[2] بود. ازاین‌رو، ۳۹ کلمه کلیدی در حوزهٔ «سیاست‌گذاری علم و فناوری» با بررسی ادبیات موضوع در زبان فارسی (قانعی راد، محمدی و بیگدلو ۱۳۹۰؛ قاضی نوری و قاضی نوری ۱۳۹۱؛ قاضی نوری و رضایی نیک ۱۳۹۳؛ احمدیان و همکاران، ۱۳۹۷؛ کلانتری، منتظر، قاضی نوری ۱۳۹۸؛ قدیمی و حجازی، ۱۳۹۸)، انتخاب و استخراج توئیت‌ها بر اساس این کلمات کلیدی آغاز شد. از آنجا که برای تعدادی از کلمات کلیدی به حد کافی توئیت وجود نداشت، آن کلمات کلیدی در همان ابتدا حذف و برای بقیه موارد داده‌ها استخراج گردید. در جدول ۱، برخی از کلمات کلیدی بر اساس اهمیتشان و یا تعداد توئیت‌های مرتبط، هم با هشتگ و هم بدون هشتگ در توئیت‌ها جست‌وجو شده‌اند. بر اساس کلمات کلیدی انتخاب‌شده، تعداد ۲۸۲۷۷ توئیت جمع‌آوری شد. پس از حذف بخشی از توئیت‌ها در حین پاک‌سازی اولیه، تعداد ۱۴۰۲۹ توئیت باقی ماند. در مجموع، در مرحله اول حدود ۱۴۰۰۰ توئیت (پیش از پاک‌سازی داده‌ها) در این بازه زمانی استخراج و جمع‌آوری گردید.

جدول ۱. کلمات کلیدی استخراج‌شده از ادبیات موضوع

سواد اطلاعاتی (#)	دسترسی به اطلاعات	آزادی اطلاعات	جامعه اطلاعاتی (#)
حریم خصوصی (#)	بازارهای اطلاعاتی	فیلترینگ (#)	افشای اطلاعات
داده #	کالاهای اطلاعاتی	اشتراک دانش (#)	شکاف دیجیتال
متخصص اطلاعات	قیمت‌گذاری اطلاعات	مدیریت دانش	توسعه ICT
حقوق مالکیت فکری	سیاست‌های اطلاعات	دسترسی آزاد (#)	پردازش اطلاعات (#)
حفاظت از اطلاعات (#)	سیاست‌های ارتباطات	کیفیت داده	سیاست‌گذاری علم (#)
اطلاعات علمی فناورانه	سیاست‌گذاری علم و فناوری	محرمانگی داده	سیاست‌گذاری فناوری
زیرساخت اطلاعاتی	اطلاعات شخصی	ارتباطات راه دور	همانندجویی اطلاعات
زیرساخت ارتباطی	سرقت ادبی (#)	صنعت اطلاعات	دسترسی آزاد
شبکه‌سازی علم	بیطرفی شبکه	زیرساخت اطلاعات	

1. data collection
2. https://developer.twitter.com

نمودار ۲، تعداد توئیت‌های دانلودشده را نشان می‌دهد که از میان موضوعات مربوط به حوزه علم و فناوری موضوع فیلترینگ با ۸۸۱۷ توئیت و #فیلترینگ با ۱۲۴۸ توئیت مهم‌ترین موضوع از دید کاربران توئیتر در میان کلیدواژه‌های حوزه سیاست‌گذاری علم و فناوری در میان کاربران فارسی‌زبان بوده است. در این توئیت‌ها قریب به اتفاق کاربران، رویکرد اعتراضی نسبت به فیلتر شدن سرویس‌های فناورانه داشتند. خشم کاربران از سیاست‌گذاران، در این توئیت‌ها قابل مشاهده بود. پس از آن، توئیت‌هایی که کلمات کلیدی اطلاعات شخصی، افشای اطلاعات، دسترسی به اطلاعات و دسترسی آزاد داشتند، به‌ترتیب، پراهمیت‌ترین موضوعات از دید کاربران بوده است. کاربران به موضوع افشای اطلاعات آن‌ها از سوی سامانه‌های دولتی و خصوصی به‌شدت معترض بودند و گاهی سرویس‌های یکپارچه‌ای که اطلاعات شخصی آن‌ها را در دسترس دیگر سازمان‌ها قرار می‌داد، مورد اعتراض بودند. موضوعاتی مانند سرقت ادبی، و نیاز به سیاست‌گذاری در این حوزه، جزء خواست‌های کاربران توئیتر بود که در رابطه با این موضوع توئیت زده بودند و از بی‌سیاستی و عدم اجرای سیاست‌های این حوزه شکایت می‌کردند. #حریم_خصوصی و جامعه اطلاعاتی، آزادی اطلاعات از اهمیت نسبی برخوردار بوده و موضوعاتی از قبیل اشتراک دانش، ارتباطات راه دور، زیرساخت اطلاعاتی و غیره کم‌اهمیت بوده و به برخی از موضوعات مانند زیرساخت ارتباطاتی، شبکه‌سازی علم، بی‌طرفی شبکه، همانندجویی اطلاعات و مدارک و غیره در توئیت‌ها از سوی کاربران پرداخته نشده است.

نمایش موضوعات پرتکرار از دید کاربران در قالب ابرکلمات در شکل زیر آورده شده است. این ابرکلمات اهمیت موضوعات مختلف در حوزه علم و فناوری را در میان کاربران ایرانی شبکه اجتماعی توئیتر نشان می‌دهد که چه موضوعاتی در این دورهٔ زمانی بیشتر مورد توجه کاربران بوده است. همان‌گونه که در ابرکلمات هم مشخص است، کاربران فارسی‌زبان از میان کلیدواژه‌های استخراج‌شده، فیلترینگ و رفع آن را مهم‌ترین خواسته خود مطرح می‌کردند. دسترسی آزاد به اطلاعات و مدارک نیز از موضوعات مهمی است که مورد نظر کاربران بوده و بر اساس ابرکلمات نمایش داده‌شده، حجم زیادی از توئیت‌ها مرتبط با این کلیدواژه است.

نمودار ۲. ابرکلمات مرتبط با کلمات کلیدی

٤-۱-۲. پیش‌پردازش داده[1]

به‌منظور پردازش داده‌ها، ابتدا تمام پیام‌ها می‌بایست به‌عنوان شواهد و غیرشواهد برچسب دریافت می‌کرد. این مرحله با بررسی ادبیات موضوع بر اساس دارا بودن معیارها (معیارهای شواهد) انجام شد و به تمامی نمونه‌ها برچسب شواهد و غیرشواهد تعلق گرفت. منطق برچسب‌زنی بر اساس پیشینه پژوهش در حوزه سیاست‌گذاری مبتنی بر شواهد، شناسایی کلیدواژه‌های سیاست‌گذاری علم و فناوری استوار بود. بر اساس پیش‌فرض‌های روش سیاست‌گذاری مبتنی بر شواهد، چالش‌هایی وجود دارد که باعث می‌شود میان محققان و سیاست‌گذاران در خصوص استفاده از انواع شواهد، اختلاف نظر به‌وجود بیاید. اول اینکه، تعریف آنچه که به‌عنوان «شواهد» از آن یاد می‌شود، تعریفی ذهنی است و معیارهای عینی برای آن تعریف نمی‌شود. پژوهشگرانی که از روش سیاست‌گذاری مبتنی بر شواهد دفاع می‌کنند، به‌طور معمول، تعریفِ به نسبت دقیقی از شواهد دارند که صرفاً تحقیقات آکادمیک دقیق را دربرمی‌گیرد (Head 2008). در مقابل، هنگامی که سیاست‌گذاران از روش سیاست‌گذاری مبتنی بر شواهد سخن می‌گویند، به‌طور معمول، دایره تعریفشان گسترده‌تر است و تجربیات حرفه‌ای، دانش سیاسی، نظر ذی‌نفعان، و ارزیابی‌های دولتی را نیز دربرمی‌گیرد (ibid). این تمایز نشان می‌دهد که علت اصلی

1. data preprocessing

اختلاف نظر میان پژوهشگران و سیاست‌مداران در رابطه با روش سیاست‌گذاری مبتنی بر شواهد از کجا نشأت می‌گیرد. ازاین‌رو، در این پژوهش تلاش شد که معیارهای مناسبی در حین عمل برچسب‌زنی توئیت‌ها و تقسیم آن‌ها به شواهد و غیرشواهد، مد نظر قرار گیرد.

بر اساس مواردی که در بالا بیان شد، در این مرحله پژوهشگر تمامی توئیت‌های استخراج‌شده را تک‌تک مطالعه و بر حسب محتوای آن، برچسب شواهد و غیرشواهد به آن زده است.

٤-٢. شناسایی رفتار کاربران منتشرکننده توئیت شواهد

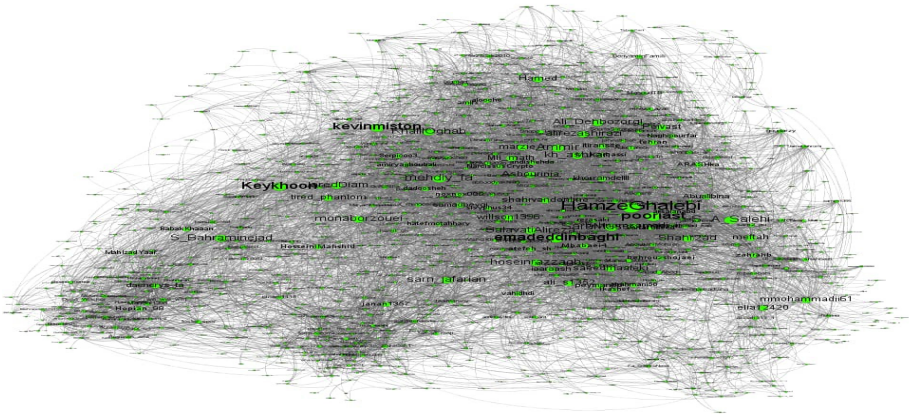
برای رسیدن به یکی از اهداف پژوهش، یعنی کشف الگوها و رفتارهای کاربرانی که پیام‌های شواهد ارسال می‌کنند، تلاش شد تا رفتارهایی که این کاربران در تولید شواهد از خود نشان می‌دهند، مشخص گردد. همان‌گونه که تحلیل داده‌ها نشان می‌دهد، شواهدی که کاربران توئیتر تولید می‌کنند، دارای ویژگی‌هایی است. اکثر توئیت‌های شواهد حاوی هشتگ نبوده و تعداد بسیار کمی صرفاً دارای یک هشتگ هستند. توئیت‌هایی که می‌توان آن‌ها را به‌عنوان شواهد قلمداد کرد، دارای هشتگ نبودند یا هشتگ‌های کمی داشتند. این موضوع نشان می‌دهد که متن‌هایی که می‌توان آن‌ها را به‌عنوان شواهد در این حوزه قلمداد کرد، بیشتر توسط کاربر هشتگ‌گذاری نمی‌شود. اکثر توئیت‌های شواهد، هیچ کاراکتر ایموجی وجود نداشته است. یعنی کاربرانی که می‌توان توئیت آن‌ها را به‌عنوان شواهد قلمداد کرد، در توئیت خود در بیشتر موارد از ایموجی استفاده نمی‌کنند. عدم استفاده از ایموجی در اغلب توئیت‌های شواهد، می‌تواند نشان‌دهنده فضای رسمی انتشار این توئیت‌ها قلمداد شود. می‌توان گفت که در ۹۶ درصد از توئیت‌های شواهد، عدد به‌کار نرفته است. در واقع، می‌توان گفت که کاربران منتشرکننده شواهد از عددها و ارقام بسیار کم استفاده کرده‌اند. توئیت‌های شواهد به‌جای ارائه اعداد و ارقام، در پی توضیح، تبیین و یا انتقاد از موضوعات بوده‌اند. همچنین، اکثر کاربران تولیدکننده شواهد، تمایلی به استفاده از بیش از یک خطاب[1] در توئیت‌هایشان ندارند که به نظر می‌رسد به این دلیل باشد که کاربر به‌دنبال انتشار نظر شخصی خود یا پاسخ به توئیت کاربر

1. Mention (@)

مشخص دیگری باشد که نظر خود را به اشتراک گذاشته است و به‌دنبال خطاب و جلب توجه دیگر کاربران نیست. عدم وجود لینک URL در بیش از ۸۷ درصد از توئیت‌های شواهد دیده شده است و ۱۳ درصد دیگر فقط حاوی یک لینک URL هستند. این موضوع نشان می‌دهد که استناد به متن‌های خارج از پلتفرم توئیتر در تولید شواهد، کمتر اتفاق افتاده است و شواهد درون پلتفرمی تولید و توزیع شده‌اند. همچنین، می‌توان مشاهده نمود که کاربرانِ شواهد بیشتر به نوشتن نظر خود در یک جمله یا کمتر از ۳ جمله تمایل دارند و این نشان از مختصرگویی توئیت‌های شواهد است. افزون بر این، از الگوهای به‌دست‌آمده در رفتار کاربران می‌توان به استفاده کمتر از تعداد علائم نگارشی، استفاده کمتر از کلمات با طول بیشتر از پنج حرف، و همچنین به‌کارگیری محدود تعداد کلمات با طول کمتر از ۳ حرف اشاره کرد.

یکی از فعالیت‌های تحلیلی که در رابطه با شناسایی رفتار کاربران و اجتماعات ایجاد شده و توسط آن‌ها در شبکه‌های اجتماعی انجام می‌شود، استفاده از اطلاعات مربوط به گراف شبکهٔ ساخته‌شده از کاربران و نوع فعالیت‌های آن‌هاست. با استفاده از ابزار API توئیتر اقدام به جمع‌آوری نام کاربری دنبال‌کنندگان[1] هر یک از کاربرانی که پیام شواهد ارسال کرده بودند، گردید. پس از استخراج روابط بین آن‌ها گراف شبکه ایجاد شد. در این شبکهٔ کاربران، گره‌های گراف و ارتباط بین آن‌ها به‌عنوان یال در نظر گرفته شده است. می‌توان با استفاده از یال‌های یک گره یا یال‌های موجود بین هر گره میزان ارتباط بین آن‌ها را به‌دست آورد. در این شبکه کاربرانی هستند که تعداد یال‌های آن‌ها از دیگر گره‌ها بیشتر است. این افراد به‌عنوان کاربران مهم و تأثیرگذار در شبکه شناخته می‌شوند. تأثیرگذارترین این افراد با شناخت کاربرانی که بیشترین تعداد یال را در شبکه داشتند، تشخیص داد شد و پیام توئیت آن‌ها از نظر موضوع مورد بررسی قرار گرفت. در زیر جهت شناسایی شبکه و ارتباطات موجود در شبکه ارسال‌کنندگان توئیت شواهد، گراف آن‌ها ترسیم شد. در این گراف، تأثیرگذاران شبکه و میزان ارتباطات آن‌ها مشخص شده است.

1. following



شکل ۱. گراف شبکه کاربران منتشرکننده توئیت شواهد

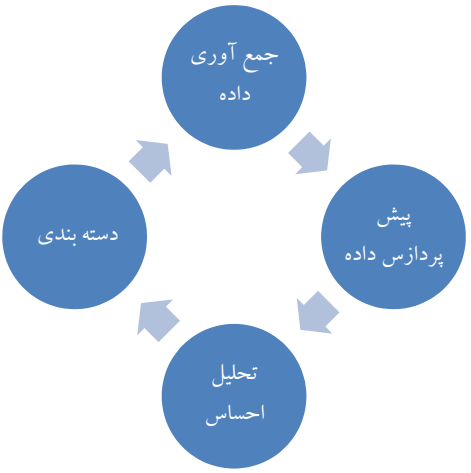
همان‌گونه که در گراف بالا مشخص است، تأثیرگذاران شبکه توئیت‌هایی ارسال کرده‌اند که باعث برانگیخته شدن دیگر کاربران و در نتیجه، اشتراک زیاد آن متن در یک شبکه بزرگ از توئیتر شده است. متن توئیت‌های ۱۱ نفر از بانفوذترین کاربران توئیتر که در کلیدواژه‌های استخراج‌شده توئیت ارسال کرده‌اند، به همراه تعداد ارتباطاتی که توئیت آن‌ها برانگیخته است، استخراج شده است. معیار انتخاب این کاربران، تعداد ارتباطاتی بوده که ایجاد کرده بودند (بیش از ۳۰۰ ارتباط). برخی از این کاربران نیز بیش از یک توئیت منتشر کرده‌اند که همگی توانسته‌اند شبکه خوبی از ارتباطات را ایجاد کنند. همان‌گونه که از متن پرنفوذترین توئیت‌ها مشخص است، این کاربران بیش از همه موضوع فیلترینگ اطلاعات، اپلیکیشن‌ها و شبکه‌های اجتماعی را مد نظر داشته‌اند. موضوع بعدی دسترسی به اطلاعات و حریم خصوصی اطلاعات بوده است که مد نظر کاربران قرار گرفته است.

٤-٣. تحلیل احساس[1]

در این بخش برای استخراج میزان گرایش و احساس موجود در نظرات کاربران نسبت به یک موضوع خاص در حوزه علم و فناوری و ارائه آن به سیاست‌گذاران، از تکنیک‌های پردازش زبان طبیعی استفاده کرده‌ایم. تحلیل احساس یکی از روش‌های پردازش زبان

1. sentiment analysis

طبیعی است و روش‌های محاسبه گوناگونی دارد (Mittal & Patidar 2019). تحلیل احساس در توئیتر به‌دلایل زیر با برخی چالش‌ها همراه است: ۱) محدودیت طول متن توئیت که دارای محدودیت ۲۸۰ کاراکتری است منجر به تولید متن‌های فشرده[1] می‌شود، ۲) استفاده از کلمات ناسزا باعث ایجاد دشواری در تشخیص معنای آن‌ها برای ماشین می‌شود، ۳) در توئیتر استفاده از URLها، هشتگ‌ها یا خطاب‌ها مجاز است و این در فرایند تحلیل احساس مشکل ایجاد می‌کند. نمودار ۳، فرایند تحلیل احساس را نشان می‌دهد.



نمودار ۳. فرایند تحلیل احساس توئیت‌ها

پس از استخراج داده‌ها، فرایند تحلیل احساس شامل ۳ بخش پیش‌پردازش، تشخیص احساس و دسته‌بندی است. در پیش‌پردازش گام‌هایی صورت گرفت: ۱) حذف توئیت‌های تکراری و ریتوئیت‌ها. در این رابطه تمامی توئیت‌ها بررسی و تمامی توئیت‌های تکراری و ریتوئیت‌های تکراری از مجموعه داده‌ها حذف شدند؛ ۲) حذف کاراکترهای خاص مانند URLها، اعداد و علائم نگارشی یکی از مراحلی بود که تلاش شد روی مجموعه داده‌های شواهد انجام شود؛ ۳) حذف استاپ‌وردها[2] که شامل حروف ربط، ضمایر و غیره می‌شود نیز از جمله مراحلی بود که انجام شد؛ چرا که این حروف ارزش تحلیلی در خصوص تحلیل احساس را ندارند؛ ۴) حذف کلمات غیر فارسی در مرحله بعد انجام شد؛ ۵). بازگرداندن هر کلمه به شکل اصلی آن کلمه[3]. به‌عنوان مثال، کلمه «کتاب‌ها» در این

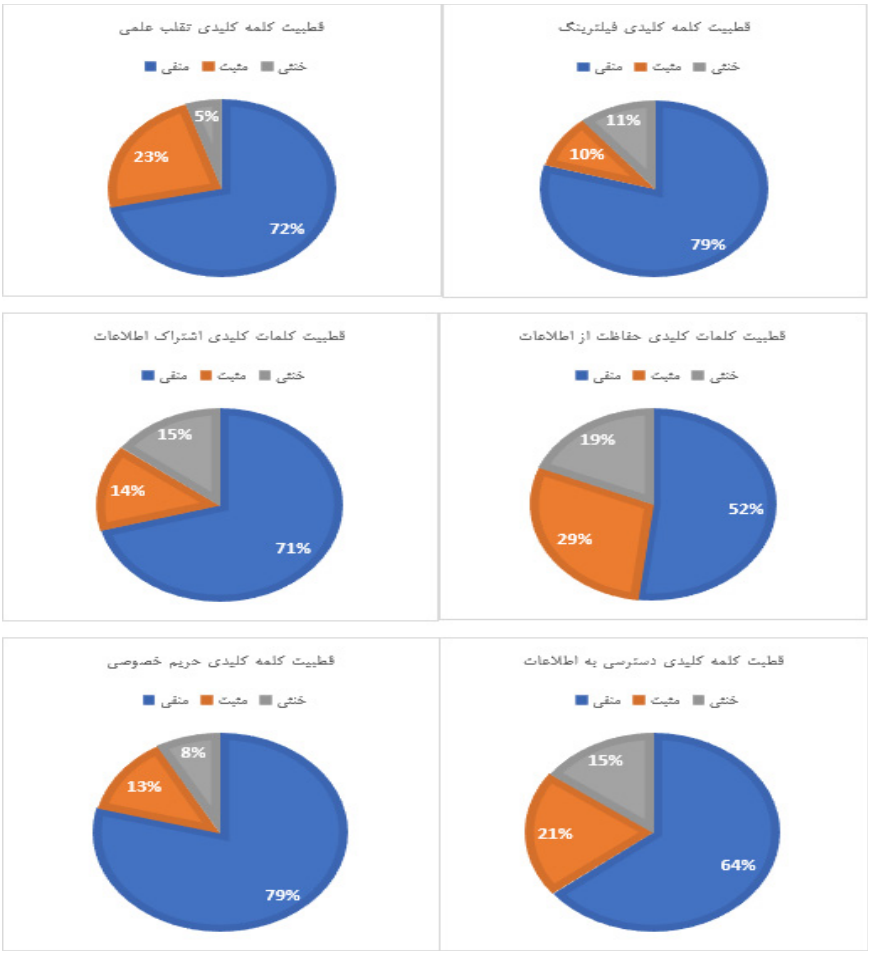
1. intensive statement 2. stop words 3. stemming

مرحله به کلمه «کتاب» تبدیل شود؛ و ۶) متن به‌صورت توکن‌های کوچکی از کلمات درآمد[1].

پس از مراحل پیش‌پردازش، با استفاده از محاسبه میزان قطبیت کلمات موجود در متن احساس و عواطف موجود در متن در سه دسته مثبت، منفی و خنثی دسته‌بندی می‌شوند. میزان قطبیت کلمه با استفاده از فرهنگ لغتی که در آن به بار معنایی کلمات امتیازدهی شده است، تعیین می‌گردد. به‌عنوان مثال، قطبیت مثبت کلمه «عالی» از کلمه «خوب» بیشتر است، در حالی که کلمه «بدتر» از کلمه «بد» قطبیت منفی بیشتری دارد. قبل از محاسبه قطبیت توئیت باید اینکه کلمات یک توئیت تا چه حد عقیده و نظر شخصی کاربر است[2] و اینکه چقدر بر اساس یک حقیقت و واقعیت بیرونی و بر اساس مشاهدات یا اندازه‌گیری‌ها[3]ست، مشخص گردد. با محاسبه این دو میزان احساس یک توئیت محاسبه می‌گردید و در گام سوم، توئیت‌ها در قالب سه دسته مثبت، منفی و خنثی دسته‌بندی شدند. در زیر قطبیت برخی از کلمات کلیدی پرتکرار آورده شده است. همان‌گونه که مشاهده می‌شود، فیلترینگ و حریم خصوصی بیشترین قطبیت منفی را داشتند و توئیت‌های مرتبط با حفاظت از اطلاعات بیشترین قطبیت مثبت را دارا هستند.

مشاهده و بررسی مجموعه توئیت‌های استخراج‌شده و همچنین، توئیت‌هایی که به‌عنوان شواهد جدا شدند، نشان‌دهنده حساسیت بالای کاربرانی است که در رابطه با کلیدواژه‌های استخراج‌شده، توئیت منتشر کرده‌اند. میزان قطبیت استخراج‌شده در بیشتر کلیدواژه‌ها منفی بوده و نیازمند مداخله فوری سیاست‌گذار در این حوزه‌هاست. در زیر قطبیت برخی از کلیدواژه‌ها نشان داده شده است.

1. tokenization 2. subjective 3. objective



نمودار ٤. نمودار قطبیت کلمات کلیدی

٤-٤. استخراج ویژگی‌ها[1]

در پژوهش‌های پیشین و مرتبط با حوزه تحلیل رفتار کاربران شبکه‌های اجتماعی، به‌منظور تحلیل داده‌های کاربران از ویژگی‌های مبتنی بر حساب کاربری و ویژگی‌های مبتنی بر متن استفاده گردیده است که نشان از موفقیت این ویژگی‌ها دارد. به‌عنوان مثال، برای تشخیص اسپم[2] در بسیاری از مطالعات پیشین (مانند Chen et al. 2016; Sedhai & Sun 2018)، از ویژگی‌های مبتنی بر حساب کاربری برای بررسی خصوصیات پروفایل

1. Feature Extraction 2. spaam

کاربـران اسـتفاده گردیـده و در برخـی دیگـر از ویژگی‌هـای مبتنـی بـر متـن اسـتفاده شـده اسـت (Hijawi et al. 2017). بـا توجـه بـه موفقیـت ایـن ویژگی‌هـا در مطالعـات مختلـف، ایـن پژوهـش نیـز به‌منظـور تشـخیص شـواهد از غیرشـواهد اقـدام بـه اسـتفاده از برخـی از ایـن ویژگی‌هـا و همچنیـن، اسـتخراج ویژگی‌هـای جدیـد در حـوزه تشـخیص پیام‌هـای شـواهد از غیرشـواهد نمـود. در ایـن پژوهـش هـر دو دسـته از ویژگی‌هـای مبتنـی بـر نـام کاربـری، و مبتنـی بـر متـن مـورد اسـتفاده قـرار گرفتـه اسـت تـا به‌وسـیله آن‌هـا الگـو و رفتـار کاربرانـی کـه پیام‌هـای شـواهد ارسـال می‌کننـد، نسـبت بـه کسـانی کـه پیام‌هایشـان شـواهد محسـوب نمی‌شـود، مـورد بررسـی قـرار گیـرد. ویژگی‌هـای مبتنـی بـر متـن استخراج‌شـده از ادبیـات موضـوع، به‌منظـور بررسـی الگوهـای موجـود در متـن پیام‌هـای شـواهد اسـتخراج شـد. از ویژگی‌هـای اسـتفاده شـده مربـوط بـه متـن، توئیت‌هایـی کـه حـاوی کلمـات ناسـزا و رکیـک هسـتند، تعـداد جمـلات هـر توئیـت، تعـداد خط‌هـا، تعـداد فاصله‌هـا، طـول توئیـت، تعـداد ایموجی‌هـا، تعـداد لینک‌هـا، تعـداد علائـم نگارشـی، تعـداد کاراکترهـا، طـول کلمـات، زمـان ارسـال توئیـت، تعـداد هشـتگ، تعـداد URLهـا، تعـداد خطاب‌هـای موجـود در متـن و غیـره مـورد اسـتفاده قـرار گرفتـه اسـت.

٤-٥. انتخاب ویژگی[1]

ایـن مرحلـه، مرحلـه‌ای اسـت کـه بایسـتی ویژگی‌هـای استخراج‌شـده، ارزیابـی و مناسـب‌ترین آن‌هـا انتخـاب شـود. ویژگی‌هایـی مناسـب‌تر هسـتند کـه بتواننـد یادگیـری بیشـتری بـرای الگوریتـم ایجـاد کننـد. از مجموعـه ویژگی‌هـای ارائه‌شـده در جـداول ۲ و ۳، برخـی از ویژگی‌هـا بـرای مـدل یادگیـری ماشـین از اهمیـت بالاتـر و تأثیـر بیشـتری برخـوردار بودنـد و برخـی تأثیـر کمتـری بـر روی یادگیـری مـدل داشـتند. ویژگی‌هـای دارای اهمیـت کمتـر می‌بایسـت از مجموعـه ویژگی‌هـا حـذف می‌گردیـد و به‌منظـور افزایـش دقـت و کارایـی مـدل پیشـنهادی لازم بـود زیرمجموعـه‌ای از بهتریـن ویژگی‌هـا از مجموعـه کامـل ویژگی‌هـای ارائه‌شـده انتخـاب می‌شـد. محاسـبه بهـره اطلاعاتـی[2] یکـی از روش‌هـای موفقیت‌آمیـز در حـوزه انتخـاب ویژگـی اسـت کـه در بسـیاری از مطالعـات پیشـین مـورد اسـتفاده قـرار گرفتـه اسـت. در ایـن پژوهـش نیـز از ایـن شـاخص بـرای انتخـاب ویژگی‌هـای مناسـب اسـتفاده شـد.

به‌منظـور پیاده‌سـازی مـدل پیشـنهادی، ابتـدا تمامـی مجموعـه توئیت‌هـای دانلودشـده در

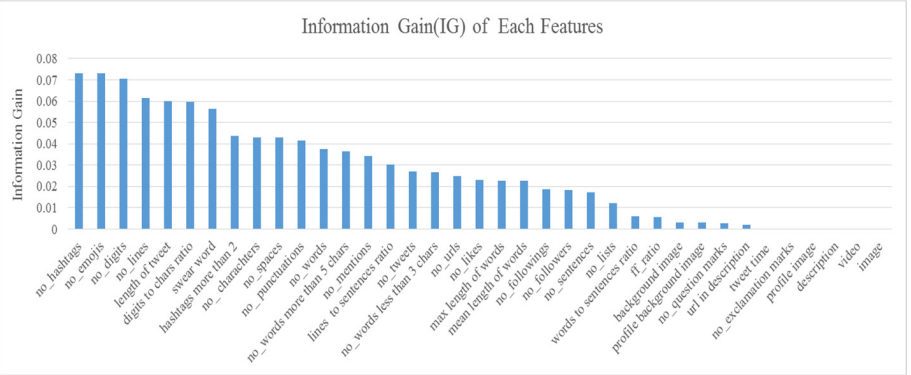
1. feature Selection 2. information gain (IG)

هر موضوع بدون در نظر گرفتن دسته‌بندی موضوعات آن‌ها، به‌صورت یکپارچه به یک مجموعه داده تبدیل شد و از میان آن‌ها تعداد ۵۹۰۰ توئیت به‌صورت اتفاقی انتخاب گردید. در حین فرایند تعیین برچسب دسته‌ها، پس از حذف توئیت‌هایی که کلمات کلیدی در آن‌ها در معنای دیگری به‌کار رفته یا متن توئیت مرتبط با سیاست‌های داخلی کشور نبود، تعداد ۴۵۶۰ توئیت باقی ماند. در میان ویژگی‌های استخراج‌شده، در مدل تشخیص شواهد، برخی از ویژگی‌ها از اهمیت بیشتر و برخی از اهمیت کمتری برخوردار هستند.

در محاسبه میزان بهره اطلاعات هر ویژگی، ارزش بهره اطلاعات برخی از ویژگی‌ها صفر بود. این امر به این دلیل اتفاق افتاد که برخی از ویژگی‌ها را همه حساب‌های کاربری دارا بودند و در نتیجه، بهره اطلاعات آن صفر گردید. به‌عنوان مثال، در مجموعه داده مطالعه‌شده، تمامی حساب‌های کاربری موجود در مجموعه داده، دارای توضیحات و عکس پروفایل در حساب کاربری خود بودند. ازاین‌رو، بهره اطلاعات این ویژگی‌ها برابر با صفر است؛ چون نمی‌تواند الگوریتم را برای تشخیص شواهد از غیرشواهد آموزش دهد. همچنین، مشخص گردید که زمان ارسال توئیت‌ها در تمامی ساعات شبانه‌روز بین دسته‌های شواهد و غیرشواهد به‌طور تقریباً یکسان توزیع شده بود. ازاین‌رو، ویژگی‌ای مانند زمان انتشار توئیت‌ها ویژگی خوبی برای یادگیری الگوریتم محسوب نمی‌شد و حذف گردید. همچنین، هیچ‌یک از پست‌های کاربرانی که متن‌های شواهد و غیرشواهد توئیت کرده‌اند، حاوی عکس و ویدئو نبوده است. بنابراین، این امر باعث شد که ارزش این ویژگی نیز در این کار برابر با صفر باشد.

با در نظر گرفتن مقادیر بهره اطلاعاتِ ویژگی‌ها، زیرمجموعه‌های مختلفی از داده‌ها برای پیاده‌سازی مدل دسته‌بندی[1] انتخاب گردیدند و مشخص گردید که برخی از ویژگی‌ها در کنار ویژگی‌های دیگر برای مدل یادگیری ماشین مفید هستند. بر این اساس، ویژگی‌های شماره ۱ تا ۲۵ به‌عنوان زیرمجموعه ویژگی‌های مناسب، در این پژوهش انتخاب شد که می‌تواند الگوریتم را جهت تشخیص توئیت شواهد از غیرشواهد آموزش دهد. نمودار ۵، مقادیر مربوط به بهره اطلاعات هر ویژگی را نشان می‌دهد و ویژگی‌های با بهره اطلاعات کمتر از ۰/۰۱ از مجموعه ویژگی‌ها حذف گردیدند. از میان زیرمجموعه انتخاب‌شده تعداد ۲۰ ویژگی مبتنی بر متن[2] و ۵ ویژگی مبتنی بر حساب کاربری[3] است.

1. classification model 2. content based 3. account based



نمودار ۵. مقادیر مربوط به بهره اطلاعات هر یک از ویژگی‌ها

٤-٦. دسته‌بندی[1]

پس از مرحله انتخاب ویژگی‌های مناسب که بتواند بر اساس بهره اطلاعات پیش‌بینی خوبی برای الگوریتم ایجاد کند، فرایند تشخیص پیام‌های شواهد به‌وسیله الگوریتم‌های یادگیری ماشین آغاز شد. از آنجا که در تشخیص شواهد به‌دنبال خروجی مشخص و برچسب هر نمونه هستیم، رویکرد پیشنهادی از نوع یادگیری با نظارت و دسته‌بندی بود. همان‌گونه که ملاحظه شد، مدل اصلی ارائه شده (نمودار ۱)، شامل ساختاری با یک دسته‌بند نظارت‌شده که پیام‌های شواهد را تشخیص می‌دهد، طراحی شده است. بدین منظور نیازمند یادگیری دسته‌بندها هستیم. فرایند یادگیری دسته‌بندها ابتدا با بخشی از داده‌های برچسب‌خورده اتفاق می‌افتد و سپس، با دریافت دانش، الگوریتم عمل پیش‌بینی برچسب را برای نمونه‌های ورودی جدید انجام خواهد داد. برای طی کردن این دو گام که شامل بخش یادگیری و آزمون است، در آغاز، ویژگی‌های استخراج‌شده مربوط به نمونه S به‌صورت مجموعه $S = \{f_1,f_2,f_3,\ldots,f_n\}$ و برچسب هر نمونه که با نماد L نمایش داده می‌شود، به هر نمونه انتصاب داده می‌شود. بنابراین، هر نمونه در مجموعه داده به شکل $\{(S_1,L_1), (S_2,L_2), (S_3,L_3), \ldots, (S_n,L_n)\}$ =Data درمی‌آید. برای مرحله آموزش، بخش عمده‌ای از مجموعه داده‌ها انتخاب شده و به‌منظور یادگیری دسته‌بندها، داده‌های آموزشی به همراه برچسب آن‌ها به شکل مجموعه داده به الگوریتم داده می‌شود تا یادگیری خود را بر روی مجموعه آموزش انجام دهد. سپس، برای پیش‌بینی نمونه‌های جدید، بخش دیگری از مجموعه داده (مجموعه آزمون) را که الگوریتم دسته‌بند آن نمونه‌ها را ندیده

1. classification

است، بدون برچسب هر نمونه به شکل $Test=\{S_1,S_2,...,S_n\}$ برای پیش‌بینی برچسب آن‌ها به دسته‌بند داده می‌شود. سپس، توسط دسته‌بند آموزش دیده شده به هر نمونه برچسب دسته‌ای مشخص تعلق می‌گیرد. سرانجام، دسته‌های پیش‌بینی‌شده هر نمونه با برچسب اصلی آن نمونه مقایسه شده و عملکرد دسته‌بند توسط معیارهای ارزیابی مورد سنجش قرار می‌گیرد.

٤-۷. معیارهای ارزیابی[1]

در راستای ارزیابی عملکرد مدل دسته‌بندی جهت تشخیص شواهد، از معیارهایی که به‌صورتی گسترده در حوزه ارزیابی دسته‌بندها توسط محققان به‌کار رفته، استفاده گردید (Chen et al. 2016). به‌طور معمول، عملکرد یک دسته‌بند توسط معیارهای دقت[2]، صحت[3]، فراخوانی[4] و میانگین هارمونیک[5] قابل ارزیابی هستند. همه این چهار معیار ارزیابی مدل دسته‌بندی انتخاب‌شده جهت ارزیابی این مدل از اهمیت زیادی برخوردار بودند که محاسبه شدند.

در این بخش در راستای تشخیص شواهد[6]، عملکرد ۶ دسته‌بند[7] شامل ماشین بردار پشتیبان[8]، درخت تصمیم[9]، X-GBoost، K نزدیک‌ترین همسایه[10]، آنالیز تشخیصی خطی[11] و رگرسیون منطقی[12] با هم مقایسه گردید. ابتدا مدل‌های دسته‌بند بر روی مجموعه داده آموزش پیاده‌سازی شدند و پس از فرایند یادگیری، دسته‌بندهای یادگیری‌شده[13] روی مجموعه آزمون پیاده‌سازی شد و نتایج عملکرد آن‌ها مطابق با جدول زیر بر اساس معیارهای ارزیابی برای هر کدام از الگوریتم‌ها نشان داده شده است.

1. evaluation metrics
2. accuracy
3. precision
4. recall
5. F-measure
6. evidence detection
7. classifiers
8. support vector machine (SVM)
9. decision tree (DT)
10. K nearest neighbor (KNN)
11. linear discriminant analysis (LDA)
12. logistic regression (LR)
13. trained classifiers

جدول ۲. معیارهای ارزیابی الگوریتم‌های مورد استفاده در پژوهش

Algorithm	Precision	Recall	F_Measure
Decision Tree (DT)	79.13	90.10	84.26
XGBoost	80.00	83.17	81.55
K-Nearest Neighbor(KNN)	75.54	68.81	72.02
Logistic Regression(LR)	72.96	70.79	71.86
Linear Discriminant Analysis(LDA)	73.85	47.52	57.83
Support Vector Machine(SVM)	62.80	50.99	56.28

معیار میانگین هارمونیک F در میان دسته‌بندها برای درخت تصمیم و XGBoost از دیگر دسته‌بندها بالاتر بود. همان‌طور که مشاهده می‌شود، درخت تصمیم و XGBoost توانسته‌اند به‌ترتیب، مقادیر ۸۴/۲۶ و ۸۱/۵۵ درصد را کسب کنند. پس از آن‌ها الگوریتم‌های دسته‌بند K نزدیک‌ترین همسایه، رگرسیون منطقی، LDA و ماشین بردار پشتیبان قرار گرفتند. از جدول زیر می‌توان نتیجه پیاده‌سازی دسته‌بندها را بر روی مجموعه داده‌های آزمون مشاهده کرد که درخت تصمیم با ۸۵/۰۹ درصد عملکرد بهتری در مقایسه با دیگر الگوریتم‌های دسته‌بند داشته است.

در راستای تشخیص توئیت‌های شواهد از غیرشواهد، مدل پیشنهادی این پژوهش توانست با دقت ۸۵/۰۹ درصد توئیت‌های شواهد را از غیرشواهد تشخیص دهد. این میزان از دقت، نسبت به مطالعات پیشین در این حوزه قابل قبول محسوب می‌شود. همچنین، با توجه به اینکه مدل ارائه‌شده، جدید بوده و ویژگی‌ها و دسته‌بندی‌های ارائه‌شده نیز جدید بودند، این میزان از دقت کاملاً مطلوب است. تحلیل بیشتر راجع به روش ارائه‌شده نشان می‌دهد که ویژگی‌های مبتنی بر متن بهتر از ویژگی‌های مبتنی بر حساب کاربری در این کار تأثیر دارد. این موضوع را می‌توان این‌گونه توضیح داد که کاربرانی که توئیت‌هایی از جنس شواهد ارسال می‌کنند، می‌توانند در مقطعی پیام شواهد و در مقطعی نیز پیام غیر شواهد ارسال کنند. در مورد کاربرانی که توئیت‌های غیر شواهد ارسال کرده‌اند نیز به همین صورت است. به همین دلیل، ویژگی‌های مبتنی بر متن در این کار از اهمیت بالاتری برخوردار بودند که در نتیجه، می‌توان ادعا کرد که ترکیب این دو دسته از ویژگی‌ها می‌تواند دقت مدل را افزایش دهد.

الگوریتم انتخاب‌شده برای مدل پیشنهادی که اکنون یادگیری در مورد آن اتفاق

افتاده، می‌تواند با هر مجموعه داده‌ای از شبکه اجتماعی توئیتر با دقت حدود ۸۵ درصد کار کند و توئیت‌های شواهد را از غیر شواهد جدا سازد. این مدل در واقع، کاربست داده‌های یکی از مهم‌ترین و تأثیرگذارترین شبکه‌های اجتماعی فارسی‌زبان را در فرایند سیاست‌گذاری علم و فناوری، با رویکرد سیاست‌گذاری مبتنی بر شواهد مدل کرده است. سیاست‌گذار با کاربست فرایند سیاست‌گذاری پیشنهادی و همچنین، مدل پیشنهادی جهت تشخیص شواهد می‌تواند نظرات، علایق، نیازها، ظرفیت‌ها و ... را در رابطه با کاربران دریابد و با استفاده از این کانال از شواهد ایجادشده، تصمیمات سیاستی مناسبی اتخاذ کند.

۵. بحث و نتیجه‌گیری

اولین مانع توسعه رویکرد «سیاست‌گذاری مبتنی بر شواهد» این است که این رویکرد در حد یک گفتمان باقی مانده و برای کاربست آن جهت اتخاذ تصمیم‌های سیاستی سنجه‌ها و یا مدل‌های عملیاتی در دست نیست. اینکه باید در سیاست‌گذاری از داده‌ها و شواهد استفاده شود، امر جدیدی نیست. در واقع، اصلاً نمی‌توان یک فرایند تصمیم‌گیری را تصور کرد که در آن به نحوی از اشکالی از شواهد موجود استفاده نشود. البته، مخالفان جنبش سیاست‌گذاری مبتنی بر شواهد نیز به‌شدت معتقدند که تصمیمات نباید مبتنی بر تعصب، دلبخواه و بر اساس کلیشه‌های موجود اتخاذ شوند. یکی از جنبه‌های مهم و قابل بحث این رویکرد از سیاست‌گذاری، طرح چارچوبی است که بتواند اطلاعات خوب (مناسب) را از اطلاعات بد (نامناسب) تفکیک کند. بنابراین، گرچه موضوع سیاست‌گذاری مبتنی بر شواهد فی‌نفسه یک ظاهر عینیت‌گرایانه دارد، اما اینکه کدام شواهد را می‌توان معتبر ارزیابی کرد، مقوله‌ای است که مستلزم ارائه نوآوری‌ها در این خصوص خواهد بود. ازاین‌رو، هدف این پژوهش نیز ارائه چارچوبی فرایندی برای کاربست داده‌های رسانه‌های اجتماعی در فرایند سیاست‌گذاری بوده است. قدرت شواهد به اعتبار روش استخراج آن از میدان مورد نظر بستگی دارد. آنچه که به‌عنوان شواهد توسط پژوهشگران توصیه می‌شود، هر آن چیزی است که قابلیت استخراج داشته و به‌صورت ملموس قابل بررسی باشد.

از آنجا که رسانه‌های اجتماعی از جمله توئیتر، فرصت شناخت و درک افکار عمومی را برای سیاست‌گذاران ایجاد می‌کنند، استفاده از این رسانه‌ها یک فرصت بزرگ برای سیاست‌گذاری‌ها محسوب می‌شوند. بر اساس رویکرد سیاست‌گذاری مبتنی بر

شـواهد کـه در سـال‌های اخیـر از اقبـال خوبـی در میـان سیاسـت‌گذاران برخـوردار شـده اسـت، سیاسـت‌گذار نیازمنـد داشـتن شـواهد مناسـب از کانال‌هـای مختلـف اسـت تـا بـا کاربسـت آن‌هـا در فراینـد سیاسـت‌گذاری بتوانـد سیاسـت‌های مؤثـر جهـت حـل مشـکلات عمومـی را طراحـی کنـد. یکـی از زمینه‌هـای مهـم سیاسـت‌گذاری عمومـی، سیاسـت‌گذاری علـم و فنـاوری اسـت کـه در سـال‌های اخیـر به‌دلیـل فروریختـن مرزهـای فناوری‌هـای مختلـف بـا چالش‌هـای زیـادی مواجـه شـده اسـت. سیاسـت‌گذار جهـت فائـق آمـدن بـر ایـن چالش‌هـا نیازمنـد اسـتفاده از شـواهد متنـوع در فراینـد سیاسـت‌گذاری اسـت. در کنـار شـواهد رسـمی، شـواهد غیررسـمی نیـز امـروزه معتبـر شـناخته شـده اسـت. ایـن پژوهـش جهـت پاسـخ‌گویی بـه ایـن پرسـش کـه آیـا می‌تـوان بـر اسـاس پیش‌فرض‌هـا و اسـتانداردهایی کـه در رابطـه بـا شـواهد وجـود دارد، از محتـوای کاربرسـاخته در شـبکه اجتماعـی توئیتـر، به‌عنـوان شـواهد مناسـب بـرای سیاسـت‌گذاری اسـتفاده کـرد؟ اگـر چنیـن کاری بـه لحـاظ پیش‌فرض‌هـای نظـری امکان‌پذیـر اسـت، چـه تحلیل‌هایـی اکنـون در رابطـه بـا داده‌هـای رسـانه‌های اجتماعـی صـورت گیـرد و مناسـب‌ترین آن‌هـا بـرای به‌کارگیـری در توئیتـر کـدام اسـت کـه بتـوان محتـوای ایـن رسـانه اجتماعـی را بـه شـواهد مناسـب بـرای سیاسـت‌گذار علـم و فنـاوری تبدیـل کـرد؟

بدیـن منظـور، پژوهشـگران جهـت فراهم‌سـازی پیش‌فرض‌هـای نظـری، کار را از مـرور مفهـوم سیاسـت‌گذاری مبتنـی بـر شـواهد آغـاز و سـپس، کارهـای انجام‌شـده در ایـن خصـوص و در رابطـه بـا ویژگی‌هـای شـواهد را مـرور کرده‌انـد. پـس از آن، ایـن پژوهـش بـه بررسـی ادبیـات سیاسـت‌گذاری علـم و فنـاوری پرداختـه اسـت. در ادامـه، فراینـد مفهومـی کاربسـت داده‌هـای رسـانه اجتماعـی توئیتـر در سیاسـت‌گذاری علـم و فنـاوری بـر اسـاس حوزه‌هـای موضوعـی ایـن حـوزه از سیاسـت‌گذاری، ارائـه شـده اسـت. پژوهشـگران بـا بررسـی الگوریتم‌هـای تحلیـل داده‌هـای رسـانه‌های اجتماعـی و اسـتفاده از مشـاوره‌های خبرگانـی، فراینـد مناسـب جهـت تحلیـل داده‌هـای رسـانه‌های اجتماعـی را انتخـاب و جهـت کاربـرد در سیاسـت‌گذاری سـه خروجـی بـرای آن در نظـر گرفتـه اسـت؛ دو خروجـی فرعـی و یـک خروجـی اصلـی کـه همگـی در سـه بخـش از مـدل شـش عنصـری سیاسـت‌گذاری از Young et al. (2002) می‌توانـد مـورد اسـتفاده قـرار گیـرد.

در خروجـی اولِ فراینـد ارائه‌شـده، نظـر بـر آن بـوده اسـت کـه فراهـم‌آوری اطلاعاتـی راجـع بـه کمیت‌هـای (توصیف‌هـای کمـی) ایجادشـده در خصـوص کلیدواژه‌هـای اصلـی در شـبکه اجتماعـی توئیتـر می‌توانـد بـه سیاسـت‌گذار در خصـوص انتخـاب دسـتور کار مناسـب بـرای

آغاز فرایند سیاست‌گذاری کمک کند. سیاست‌گذار مجبور به انتخاب یک دستور کار از میان انبوهی از مسائل است که موضوعاتی که کاربر در توئیتر حساسیت بیشتری نسبت به آن‌ها داشته، می‌تواند در انتخاب این مسئله درست، کمک‌کننده باشد. در خروجی دوم که تحلیل احساس و در واقع، تحلیل رفتارِ کاربرانی که در رابطه با کلیدواژه‌های مورد نظر توئیت شواهد منتشر می‌کنند، ارائه می‌شود، قصد کمک به اتخاذ راه‌حل‌های مناسب جهت حل مسئله در کنار دیگر کانال‌های اطلاعات بوده است.

در خروجی سومِ فرایند، پژوهشگران در صدد آزمون خروجی‌های تعریف‌شده برای فرایند پیشنهادی برآمدند. مهم‌ترین خروجی این الگوی پیشنهادی یعنی تشخیص شواهد از غیرشواهد، به‌منظور در اختیار قرار دادن کانالی از شواهد در اختیار سیاست‌گذار علم و فناوری است. بدین منظور الگوریتم انتخاب‌شده با داده‌های شش‌ماههٔ توئیتر فارسی آزمون شد و با دقت قابل قبولی (۸۵/۹)، مورد تأیید قرار گرفت. هدف ارائه این خروجی از فرایند پیشنهادی، از میان بردن سردرگمی سیاست‌گذار علم و فناوری در میان انبوهی از توئیت‌هایی است که در هر لحظه در این پلتفرم منتشر می‌شود. این الگوریتم در واقع، توئیت‌هایی از جنس شواهد را جدا ساخته و تنها آن‌ها را در اختیار سیاست‌گذار قرار می‌دهد. پس از آن، این سیاست‌گذار است که با در دست داشتن شواهد معتبری از کانالی مانند شبکه اجتماعی توئیتر می‌تواند دست به تدوین سیاست‌های آن‌ها بزند. با استفاده از این فرایند پیشنهادی، سیاست‌گذاران می‌توانند به‌طور نظام‌مند از داده‌های توئیتر در فرایند سیاست‌گذاری علم و فناوری استفاده کنند. استفاده از دیدگاه‌های کاربران توئیتر که نسبت به موضوعات مختلف سیاستی در حوزه علم و فناوری حساس هستند، می‌تواند فرایند تدوین، اجرا و ارزیابی سیاست‌ها را تسهیل کند؛ چرا که دیدگاه‌ها، نیازها، و ظرفیت‌های این کاربران برخاسته از دانش زمینه‌ای آن‌ها نسبت به مشکلات حوزه علم و فناوری است. پژوهشگر در تمام فرایند پژوهش خود این نکته را مد نظر داشته است که شبکه‌های اجتماعی و از جمله توئیتر تنها می‌توانند بخشی از شواهد مورد نیاز سیاست‌گذار را فراهم آورد و هیچ‌گاه ادعا نشده است که تنها استفاده از این‌گونه شواهد و چشم‌پوشی از دیگر شواهد، می‌تواند سیاست‌گذار را به اتخاذ سیاست‌های درست رهنمون باشد. اما بایستی تأکید کرد که از آنجا که امکانات سیاست‌گذار برای پیمایش‌های گسترده و با روایی بالا، جهت دستیابی به نظرات شهروندان، کم است، استفاده از داده‌های رسانه‌های اجتماعی که در حجمی انبوه تولید می‌شود و تحلیل آن‌ها

بسیار ارزان‌تر و سریع از دیگر داده‌هاست، نباید در فرایند سیاست‌گذاری به‌عنوان یکی از کانال‌های شواهد، مورد غفلت واقع شود.

در راستای مقایسه با کارهای پیشین می‌توان گفت که گرچه کارهای تحقیقاتی زیادی در چند سال گذشته از کاربست داده رسانه‌های اجتماعی در فرایند سیاست‌گذاری پشتیبانی می‌کند، اما می‌توان گرایشی را نیز در تحقیقات شناسایی کرد که این کاربست را دارای مشکلاتی می‌دانند. بیشتر مشکلاتی که در خصوص کاربست این نوع داده‌ها به‌عنوان شواهد در سیاست‌گذاری یاد می‌شود، عبارت‌اند از: نبودِ ابزارهای تحلیلی مناسب برای استخراج شواهد از میان پست‌های رسانه‌های اجتماعی (Castelló, Morsing & Schultz 2013; Janssen & Helbig 2016; Driss, Mellouli & Trabelsi 2019)؛ ابزارهای تحلیلی که تاکنون توسعه داده شده، بیشتر مناسب توسعه برند شرکت‌ها و یا دریافت نظر مشتریان در مورد محصولات و خدمات شرکت‌هاست و ممکن است در رابطه با سیاست‌گذاری به‌خوبی کار نکند (Schniederjans, Cao & Schniederjans 2013; Panagiotopoulosa, Bowen & Brooker 2017)؛ اظهار نظر کاربران غیرمتخصص در قالب پست‌های رسانه اجتماعی در مورد موضوعات تخصصیِ سیاستی که ممکن است آن را از دایره شواهد خارج کند و ابزاری هم برای جداسازی این‌گونه داده‌ها از هم وجود ندارد (Boyd & Crawford 2012; Davies, Nutley & Smith 2001; Simonofski, Fink & Burnay 2021)؛ حجم بالای کلان‌دادهٔ رسانه اجتماعی که ممکن است دچار سوگیری‌هایی هم شده باشد که پیچیدگی تحلیل آن را افزایش می‌دهد (Schintler & Kulkarni 2014; Gintova 2019). مشارکت‌های کاربران در تولید پست‌های رسانه‌های اجتماعی به‌طور معمول، با اهداف تحلیل‌کنندگان این داده‌ها متفاوت است (Cartwright & Hardie 2012; Head 2008; Parsons 2002; Sanderson 2002; Napoli 2019)؛ نمی‌توان رضایت هر کاربر را هنگام جمع‌آوری داده‌های او به‌دست آورد که این باعث ایجاد مسائل حریم خواهد شد (Boyd & Crawford 2012; Bekkers et al. 2013).

فهرست منابع

احمدیان، محمدمهدی، حسنعلی آقاجانی، میثم شیرخدایی، و امیرمنصور طهرانچیان. ۱۳۹۷. طراحی مدل سیاست‌گذاری علم و فناوری مبتنی بر رویکرد پیچیدگی اقتصادی. *سیاست‌گذاری عمومی* ۴(۴): ۹-۲۷.

قاضی نوری، سپهر، و سروش قاضی نوری. ۱۳۹۱. سیاست‌گذاری علم و فناوری در قالب سیاست‌های عام و خاص. *رهیافت* ۵۰: ۴۹- ۵۵.

قاضی نوری، سروش، و نفیسه رضایی‌نیک. ۱۳۹۳. بررسی الزامات، چالش‌ها و قابلیت‌های شبکه اجتماعی کنشگران مدیریت فناوری و نوآوری ایران. *تحقیقات فرهنگی ایران* (۲): ۴- ۷۳.

قانعی‌راد، محمد امین، محمدی، علیرضا، بیگدلو، نسرین. (۱۳۹۰). بررسی الگوهای تعاملی نهادهای پشتیبان پژوهشی و اجرایی با شوراهای عالی سیاست‌گذاری علم و فناوری، *رهیافت*. ۴۹: ۵- ۱۷.

قدیمی، اکرم، و الهه حجازی. ۱۳۹۸. سیاست‌گذاری علم، فناوری و ترویج علم در ایران: یک ضرورت ملی. *ترویج علم* ۱۷: ۵- ۳۱.

کلانتری اسماعیل، منتظر غلامعلی، سید سپهر قاضی نوری. ۱۳۹۸. تدوین سناریوهای گذار به وضعیت بهبودیافتهٔ ساختار سیاست‌گذاری علم و فناوری در ایران. *پژوهش‌های مدیریت راهبردی* ۷۴: ۷۵- ۱۰۲.

References

Barnett, J., K. Burningham, G. Walker, & N. Cass. 2012. Imagined publics and engagement around renewable energy technologies in the UK. *Public Understanding of Science* 21 (1): 36–50.

Bekkers, V., A. Edwards, & D. de Kool. 2013. Social media monitoring: Responsive governance in the shadow of surveillance? *Government Information Quarterly* 30 (4): 335–342.

Boyd, D., & K. Crawford. 2012. Critical questions for big data. *Information, Communication & Society* 15 (5): 662–679.

Bucher, T. & A. Helmond. 2017. The affordances of social media platforms. In: J. Burgess., A. Marwick & T. Poell, eds. *The SAGE handbook of social media.* London: SAGE Publications Ltd, pp. 233-253.

Cairney, P. 2016. *The Politics of Evidence-Based Policy Making*. London: Palgrave Macmillan.

Cartwright, N., J. Hardie. 2012. *Evidence-Based Policy A Practical Guide to Doing It Better*. New York: Oxford.

Castelló, I., M. Morsing, & F. Schultz. 2013. Communicative dynamics and the polyphony of corporate social responsibility in the network society. *Journal of Business Ethics* 118 (4): 683–694.

Chen, C., Y. Wang, J. Zhang, Y. Xiang, W. Zhou, & G. Min. 2016. Statistical Features-Based Real-Time Detection of Drifted Twitter Spam. *IEEE Transactions on Computational Social Systems* 12 (4): 914 – 925.

Davies, H., S. Nutley, & P. Smith. 2001. *What works: Evidence-based policy and practice in public services.* Bristol: Policy Press.

Driss, O. B., S. Mellouli, & Z. Trabelsi. 2019. From citizens to government policy-makers: Social media data analysis. *Government Information Quarterly* 36: 560–570.

Evans, S., H. Ginnis & J. Bartlett. 2015. *A Guide to Embedding Ethics in Social Media Research.* Bristol: Palgrave.

Fern´andez, M., T. Wandhoefer, B. Allen, A. Cano Basave, & H. Alani. 2014. Using social media to inform policy making: to whom are we listening? In: European Conference on Social Media (ECSM 2014), 10-11 Jul 2014, Brighton, UK.

Gintova, M. 2019. Understanding government social media users: an analysis of interactions on Immigration, Refugees and Citizenship Canada Twitter and Facebook. *Government Information Quarterly* 36: 1-10.

Head, B. 2010. Reconsidering evidence-based policy: Key issues and challenges. *Policy and Society* 29: 77–94.

_____. 2008. Three lenses of evidence-based policy. *Australian Journal of Public Administration* 67 (1): 1–11.

Highfield, T., S. Harrington, & A. Bruns. 2013. Twitter as a technology for audiencing and fandom. *Information, Communication & Society* 16 (3): 315–339.

Hijawi, W, H. Faris, J. Alqatawna, A. Al-Zoubi, & I. Aljarah. 2017. Improving Email Spam Detection Using Content Based Feature Engineering Approach. Jordan Conference on Applied Electrical Engineering and Computing Technologies (AEECT). Jordan.

Howlett, M. 2009. Policy analytical capacity and evidence-based policy-making: Lessons from Canada. *Canadian Public Administration* 52 (2): 153–175.

Janssen, M., & N. Helbig. 2016. Innovating and changing the policy-cycle: Policy-makers be prepared! *Government Information Quarterly* 12 (3): 120- 132.

Lee, E. J., H. Y. Lee, & S. Choi. 2020. Is the message the medium? How politicians' Twitter blunders affect perceived authenticity of Twitter communication. *Computers in Human Behavior* 104: 106-188.

Mittal, A., & S. Patidar. 2019. *Sentiment Analysis on Twitter Data: A Survey*. Presented in "Association for Computing Machinery". Bangkok, Thailand. July 27–29.

Napoli, P. N. 2019. User Data as Public Resource: Implications for Social Media Regulation. *Policy and Internet*. doi: 10.1002/poi3.216.

Panayiotopoulos, P., F. Bowen, & P. Brooker. 2017. The value of social media data: Integrating crowd capabilities in evidence-based policy. *Government Information Quarterly* 34: 601–612.

Panagiotopoulos, P., L. C. Shan, J. Barnett, A. Regan, & A. McConnon. 2015. A framework of social media engagement: Case studies with food and consumer organisations in the UK and Ireland. *International Journal of Information Management* 35 (4): 394–402.

Parkhurst, J. 2017. *The politics of evidence; from evidence-based policy to the good governance of evidence*. London: Routledge.

Parsons, W. 2002. From muddling through to muddling up - evidence based policy making and the modernisation of British government. *Public Policy and Administration* 17 (3): 43–60.

Poy, S., & S. Schüller. 2020. Internet and voting in the social media era: Evidence from a local broadband policy. *Research Policy* 49 (1): 103-121.

Sanderson, I. 2002. Evaluation, policy learning and evidence-based policy making. *Public Administration* 80 (1): 1–22.

Schintler, L. A., & R. Kulkarni. 2014. Big data for policy analysis: The good, the bad, and the ugly. *Review of Policy Research* 31 (4): 343–348.

Schniederjans, D., E. S. Cao, & M. Schniederjans. 2013. Enhancing financial performance with social media: An impression management perspective. *Decision Support Systems* 55 (4): 911–918.

Sedhai, S., & A. Sun. 2017. Semi-Supervised Spam Detection in Twitter Stream. *Computational Social Systems* 5 (1): 169 – 175.

Sedhai, S., and A. Sun. 2018. Semi-Supervised Spam Detection in Twitter Stream. *Computational Social Systems* 5 (1): 169 – 175.

Simonofski, A, J. Fink, & C. Burnay. 2021. Supporting policy-making with social media and e-participation platforms data: A policy analytics framework, *Government Information Quarterly* 38 (3): 101-122.

Stamatelatos, G., S. Gyftopoulos, G. Drosatos, & P. S. Efraimidis. 2020. Revealing the political affinity of online entities through their Twitter followers. *Information Processing and Management* 57: 102-172.

Valenzuela, S., M. Piña, & J. Ramírez. 2017. Behavioral Effects of Framing on Social Media Users: How Conflict, Economic, Human Interest, and Morality Frames Drive News Sharing. *Journal of Communication* 67 (5): 803–826.

Walker, G., N. Cass, K. Burningham, & J. Barnett. 2010. Renewable energy and sociotechnical change: Imagined subjectivities of "the public" and their implications. *Environment and Planning* 42 (4): 931–947.

Yan, Z. 2018. Big Data and Government Governance. *International Conference on Information Management and Processing*. IEEE. 110-114. China.

Young, K., D. Ashby, A. Boaz, & L. Grayson. 2002. Social Science and the Evidence-based Policy Movement. *Social Policy & Society* 1 (3): 215 -224.

سمیه لبافی

متولد ۱۳۶۳، دارای مدرک تحصیلی دکتری در رشته مدیریت رسانه از دانشگاه تهران است. ایشان هم‌اکنون استادیار پژوهشگاه علوم و فناوری اطلاعات ایران (ایرانداک) است. بیشتر فعالیت‌های او در این پژوهشگاه پیرامون رسانه‌های اجتماعی و سیاست‌گذاری رسانه‌ای است.

علایق پژوهشی وی در حوزه سیاست‌گذاری رسانه‌های نوین، رسانه‌های اجتماعی و روش‌های تحلیل شبکه است. او بنیان‌گذار آزمایشگاه «رسانه‌های اجتماعی و حکمرانی داده‌ها»ست.

محمدمهدی کاوسی

متولد ۱۳۶۴، دارای مدرک تحصیلی کارشناسی ارشد مهندسی نرم‌افزار از دانشگاه ارشاد دماوند است. تجربه کار در حوزه صنعت و دانشگاه به‌عنوان تحلیل گر داده، مدیر پروژه نرم‌افزار، تحلیل گر نیروی انسانی از سوابق او است.

علاقه‌های پژوهشی او حوزه یادگیری ماشین، یادگیری عمیق، علم داده و تحلیل رسانه‌های اجتماعی هستند.