
بسم الله الرحمن الرحیم

وزارت علوم، تحقیقات و فناوری

پژوهشگاه علوم و فناوری اطلاعات ایران (ایرانداک)

طرح پژوهشی

ارائه الگوی کاربست داده های رسانه های اجتماعی در سیاستگذاری علم و فناوری

مجری

سمیه لبافی

خرداد ۱۴۰۰

وزارت علوم، تحقیقات و فناوری
پژوهشگاه علوم و فناوری اطلاعات ایران
ایرانداک

شماره: ۳۳/۱۷۱۰
تاریخ: ۱۴۰۰/۰۳/۰۵
پیوست: ندارد

پژوهش . مدیریت دانش . آموزش

به نام خدا

ایرانداک در نیم‌قرنِ دوم کارِ خود، همچنان جوان و کوشا، همراه هر پژوهش، پژوهشگر، و سیاست‌گذار

۱۳۱۴۷ – ۱۴۰۰

گواهی

انجام طرح پژوهشی:

- **عنوان:** ارائه الگوی کاربست داده‌های رسانه‌های اجتماعی در سیاستگذاری علم و فناوری
- **مجری:** دکتر سمیه لبافی
- **همکار(ان):** ـ
- **شمارهٔ قرارداد:** ۳۳/۱۰۷۳
- **تاریخ قرارداد:** ۱۳۹۹/۲/۱۳
- **تاریخ شروع:** ۱۳۹۹/۲/۱۳
- **تاریخ پایان:** ۱۴۰۰/۲/۲۸
- **ناظر:** دکتر لیلا نامداریان

در پژوهشگاه علوم و فناوری اطلاعات ایران گواهی می‌شود. گزارش پایانی این طرح، بر پایهٔ داوری در نشست ۴۱۱ (تاریخ ۱۴۰۰/۳/۵) شورای پژوهشی پژوهشگاه به تصویب رسیده است. لازم می‌دانم مراتب قدردانی خود را به مجری، و ناظر این طرح تقدیم دارم.

با آرزوی بهروزی
پرویز شهریاری
معاون پژوهش و آموزش

تهران . خیابان انقلاب اسلامی . چهارراه فلسطین . شمارهٔ ۱۰۹۰ صندوق پستی ۱۳۷۱-۱۳۱۸۵
تلفن: ۶۶۴۹۴۹۸۰ - ۶۶۹۵۱۴۳۰ دورنگار: ۶۶۴۳۶۲۵۴ تلفن گویا: ۶۶۴۹۴۹۵۴
irandoc.ac.ir
info@irandoc.ac.ir

ارائه الگوی کاربست داده های رسانه های اجتماعی در سیاستگذاری علم و فناوری

مدیر طرح پژوهشی: سمیه لبافی، پژوهشگاه علوم و فناوری اطلاعات ایران.

نشانی: تهران، خیابان انقلاب، چهارراه فلسطین، شماره ۱۰۹۰، طبقه ۳ ، اتاق ۳۱۶

صندوق پستی: ۱۳۱۵۷۷۳۳۱۴ تلفن: ۶۶۴۹۴۹۸۰-۰۲۱(داخلی۳۶۶)

رایانامه: labafi@irandoc.ac.ir

چکیده

یکی از بهترین فرصت هایِ شناخت و درک افکار عمومی به منظور بهره گیری در فرایند سیاست گذاری، رسانه های اجتماعی هستند. این رسانه ها یک فرصت بزرگ برای دولت‌ها ایجاد میکنند تا در مورد شهروندان یاد بگیرند و به طور موثر با آنها ارتباط برقرار کنند. اگر دولت ها بخواهند از فرصتِ مشارکت کاربران در رسانه های اجتماعی بهره‌برداری کنند، پایش رسانه های اجتماعی، یا بطور کلی نظارت اجتماعی از طریق رسانه های اجتماعی یکی از بهترین راه حل ها است. سیاستگذاران علم و فناوری در ایران به عنوان بخشی از سیاستگذاری عمومی، تاکنون به طور سیستماتیک از داده های رسانه های اجتماعی در فرایند سیاستگذاری علم و فناوری استفاده نکرده اند. این عدم حضور دیدگاههای کاربران رسانه های اجتماعی که نسبت به موضوعات سیاستی در حوزه علم و فناوری حساس هستند می تواند فرایند تدوین، اجرا و ارزیابی سیاست ها را در آینده با چالش مواجه سازد. چرا که دیدگاههای این کاربران با ارائه دانش زمینه ای نسبت به مشکلات حوزه علم و فناوری می تواند مسیر سیاستگذاری و اجرای آن را تسهیل نماید. از این رو در این پژوهش شیوه کاربست داده های کاربران رسانه های اجتماعی در فرایند سیاستگذاری علم و فناوری مورد نظر بوده است. از آنجا که این سوال که داده های رسانه های اجتماعی چگونه می تواند در فرایند سیاستگذاری علم و فناوری دخیل شود؟ تاکنون پاسخ روشنی نیافته است، در این پژوهش، به دنبال پاسخ به این سوال بوده ایم که چگونه می توان از داده های رسانه های اجتماعی در سیاستگذاری علم و فناوری استفاده کرد؟ بدین منظور پژوهشگر در مرحله اول، به دنبال تبیین موضوع استفاده از داده های رسانه های اجتماعی به عنوان بخشی از شواهد معتبر، در فرایند سیاستگذاری علم و فناوری بوده است. سپس شناسایی تحلیل های مناسب جهت استخراج دانش از داده های رسانه اجتماعی توئیتر، متناسب با نیازهای سیاستگذار علم و فناوری مد نظر بوده است. شناسایی رفتار غالبِ کاربرانی که توئیت شواهد منتشر می کنند و همچنین شناسایی قطبیت توئیت های منتشر شده، از جمله این تحلیل ها بوده است. سپس الگوی مناسب کاربست داده های رسانه های اجتماعی در سیاستگذاری علم و فناوری ارائه شده است. این الگو با داده های ۶ ماهه کلیه کاربران فارسی زبان توئیتر که با موضوع سیاستگذاری علم و فناوری توئیت منتشر کرده اند، آزمون شده و نتایج آن ارائه شده است.

کلمات کلیدی: سیاستگذاریِ مبتنی بر شواهد؛ سیاستگذاریِ علم و فناوری؛ داده هایِ رسانه های اجتماعی؛ توئیتر

فهرست مطالب

عنوان **شماره صفحه**

فهرست جدول ها

عنوان **شماره صفحه**

فهرست نمودارها

عنوان **شماره صفحه**

سپاس‌گزاری

در پایان جای دارد از همه کسانی که بنده را با مشورت های خوب خود در راه انجام این پژوهش یاری دادند سپاس‌گزاری کنم. از پژوهشگاه ایرانداک و همه همکاران عزیز آن که امکان انجام این کار را برای بنده فراهم کردند بسیار سپاس‌گزارم. از جناب آقای مهندس کاووسی به خاطر کمک های فنی ایشان به این پروژه بسیار سپاس‌گزارم. همچنین از سرکار خانم دکتر نامداریان که به عنوان ناظر محترم پروژه، در این کار یاری رسان بودند بسیار متشکرم.

فصل اول

پیشینه پژوهش

مقدمه و بیان مسئله

رسانه های اجتماعی جایگاه ویژه ای در زندگی اجتماعی افراد یافته اند. این رسانه ها که بر پایه ایدئولوژی و تکنولوژی وب ۲ بنا شده اند(Kaplan & Heinlein, 2010)، تاثیر به سزایی بر افراد گذاشته و نوع ارتباط و تعامل آنها را تحت تاثیر قرار داده اند(Tao & Moon, 2015). این رسانه ها بر اساس مشارکت کاربران در تولید محتوا کار می کنند و باعث ظهور کاربرانی فعال و پویا به جای مخاطبانی منفعل شده اند، تا جایی که کاربران در فرآیندهای مختلف اجتماعی بیش از پیش مشارکت می کنند(Olmsted & Wolter, 2018). استفاده از این رسانه های اجتماعی برای ارائه نظرات در رابطه با مسائل مختلف اجتماعی هر روزه در حال گسترش است. در واقع با ظهور رسانه‌های اجتماعیِ مشارکتی، اکوسیستم جدیدی برای مشارکت کاربران در رویدادهای اجتماعی بوجود آمده است(Fuchs, 2017). مبادله محتوا توسط کاربران می تواند ارزش هایی نظیر ساخت یک گفتمان اجتماعی، تولید آگاهی در رابطه با موضوعی خاص، تغییرات سیاسی، اقتصادی و ... را ایجاد کند(Edom, Hwang & Kim, 2018) . استفاده از ارزش های ایجاد شده در این فضا توسط حاکمیت ها، می تواند باعث شکل گیری حکمرانی های مطلوب تری در جوامع گردد.

یکی از بهترین راه های شناخت و درک ویژگی های کاربران برای مد نظر قرار دادن دیدگاه های آنها در فرایند حکمرانی و سیاست گذاری، استفاده از رسانه های اجتماعی است. رسانه‌های اجتماعی با بیش از میلیون‌ها کاربر فعال، به یک منبع پویا از اطلاعات، در مورد علایق، نیازها و عقاید مردم تبدیل شده‌اند و بعنوان یک منبع بسیار غنی از داده ها، برای دست یافتن به نظرات میلیون ها کاربر در نظر گرفته می‌شوند. رسانه های اجتماعی دارای داده هایی از قبیل متن، ویدئو و ... است که این حجم از داده های کاربر ساخته[1] می تواند سیاستگذاران را به سمت انتخاب های آگاهانه تر رهنمون سازد. گفته می شود که رسانه های اجتماعی می تواند کانالی میان کاربران و سیاستگذاران باشد و به عنوان منبعی جدید در درگیر سازی شهروندان در تدوین و اجرای سیاست ها به کار رود(Mellouli, Driss & , 2019 Trabelsi). این پدیده فرصتی بزرگ برای دولت‌ها ایجاد می‌کند تا در مورد شهروندان یاد بگیرند و به طور موثر با آنها ارتباط برقرار کنند(Miriam, 2014). اگر دولت ها بخواهند یاد بگیرند چگونه از مشارکت کاربران در رسانه های اجتماعی بهره‌برداری کنند، پایش رسانه های اجتماعی، یا بطور کلی نظارت اجتماعی در این فضا، باید مورد توجه قرار گیرد(Panagiotopoulos, Bowen, & Brooker, 2017).

[1] User Generated Content

با انقلاب پلتفرمی رخ داده، نقش سیاستگذاران تغییر کرده است و یکی از این نقش های جدید، تحلیل و استخراج دانش از نظرات شهروندانی است که در این پلتفرم ها فعالیت می کنند(Janssen & Helbig, 2016). در واقع مفهوم سیاستگذاری مبتنی بر شواهد[۲]، در اینجا مجال طرح می یابد. سیاستگذار که مجبور به استفاده از انواع شواهد است، نمی تواند شواهدی را که به طور گسترده در موضوعات مختلف توسط شهروندان در حال تولید است، نادیده بگیرد. سیاستگذاران نیازمند آگاهی از نظرات کاربران در قالب پست های رسانه های اجتماعی هستند چرا که کاربران با کمترین دروازه بانی نظرات خود را منتشر می کنند(Dekker, van den Brink& Meijer, 2019). کاربران نظرات مثبت و یا منفی خود را راجع به مسائل مختلف در پلتفرم های رسانه های اجتماعی منتشر می سازند. این یک فرصت استثنایی برای سیاستگذاران است که ارتباط خود را با شهروندان بیشتر و از نیازها و نظرات آنها مطلع شوند و شواهد تولید شده از سوی آنها را در سیاستگذاری های خود به کار گیرند(DePaula, Dincelli & Harrison, 2018). ناپولی(۲۰۱۹) در توضیح فراگیری این پدیده، منطق استفاده از منابع عمومی که پیش از این بر منابع سنتی در جامعه حاکم بود را به مفهوم داده های رسانه های اجتماعی تعمیم داده است. چرا که مفهوم سازی جمع آوری داده های کاربر و پذیرش آن بعنوان یک منبع عمومی برای سیاستگذار، راه حلی برای مداخله کاربر در سیاستگذاری در جهت تامین منافع عمومی را فراهم می کند(Napoli, 2019).

سیاستگذاران علم و فناوری در ایران تاکنون به طور سیستماتیک از داده های رسانه های اجتماعی در فرایند سیاستگذاری استفاده نکرده اند. این عدم حضور دیدگاه‌های کاربران رسانه های اجتماعی که نسبت به موضوعات سیاستی در حوزه علم و فناوری حساس هستند می تواند فرایند تدوین، اجرا و ارزیابی سیاست ها را در آینده با چالش روبرو سازد. چرا که دیدگاه‌های این کاربران با ارائه دانش زمینه ای نسبت به مشکلات حوزه علم و فناوری می تواند مسیر سیاستگذاری مبتنی بر شواهد در این حوزه را تسهیل نماید. از این رو در این پژوهش بررسی شیوه کاربست داده های کاربران رسانه های اجتماعی در فرایند سیاستگذاری علم و فناوری مورد نظر بوده است.

از آنجا که این سوال که داده های رسانه های اجتماعی چگونه می تواند در فرایند سیاستگذاری علم و فناوری دخیل شود تاکنون پاسخ روشنی نیافته است (Mellouli, Driss & Trabelsi , 2019)، پژوهشگر به دنبال پاسخ به این سوال خواهد بود که آیا می توان داده های رسانه های اجتماعی را در سیاستگذاری علم و فناوری به کار بست؟ و چگونه این کاربست باید صورت گیرد؟ بدین منظور در مرحله اول، پژوهشگر به دنبال ارائه یک چارچوب مفهومی برای پشتیبانی نظری از این ایده بوده است که آیا می توان از داده های رسانه های اجتماعی در فرایند سیاستگذاری علم و فناوری استفاده کرد؟ سپس، شناسایی روش های استخراج دانش از داده های رسانه اجتماعی توئیتر متناسب با حوزه های موضوعی سیاستگذاری علم و فناوری، پیگیری شده است. همچنین الگوی مناسب کاربست داده های رسانه های

[2] Evidence Base Policy

اجتماعی در سیاستگذاری علم و فناوری طراحی و ارائه گردیده است. در مرحله دوم از پژوهش، این الگو با داده های ۶ ماهه دوم سال ۹۸ آزمون شده است. در این پژوهش، پژوهشگر الگو را به مفهوم شناسایی عناصر لازم در چگونگی بکار گیری داده های رسانه های اجتماعی در فرایند سیاستگذاری علم و فناوری به کار برده است. از آنجا که موضوع کاربست داده های رسانه های اجتماعی در سیاستگذاری علم و فناوری، موضوعی جدید بوده و حدود و مرزهای آن نامشخص است، کار اصلی پژوهشگر شناسایی عناصر لازم جهت تحقق این امر و تعیین حدود مفهومی آن است.

برای شکل دهی به این الگو، نیازمند این بوده ایم که مبانی نظری مرتبط با این حوزه را که می تواند پشتیبانی نظری مناسبی از این الگو فراهم کند، مرور کنیم. از اینرو، پژوهشگر با این پیش فرض که داده های رسانه های اجتماعی می تواند به عنوان نوعی از شواهد در فرایند سیاستگذاری علم و فناوری به کار گرفته شود، نقطه شروع کار خود را بر مبنای مرور ادبیات نظری سیاستگذاری مبتنی بر شواهد و سپس اقامه منطق مناسب جهت استفاده از داده های رسانه های اجتماعی در سیاستگذاری، قرار داده است. مرور از سیاستگذاری مبتنی بر شواهد شروع می شود و سپس به بررسی انواع شواهد می پردازد. پس از آن، کارهای مرتبط با داده های رسانه های اجتماعی در ادبیات موضوع، مرور می شود و تلاش شده تا تحلیل های مناسب این داده ها، و شیوه کاربست آن در فرایند سیاستگذاری بررسی شود.

مبانی نظری

۱.۱.سیاستگذاری مبتنی بر شواهد

استفاده از اطلاعات، ابزار قدرتمندی برای کمک به سیاستگذاران و دستیابی به موفقیت در سیاستگذاری ها است که آغاز آن به دوره حکمرانی کارآمد[3] در اوایل دوران مدرنیته در اروپا بازمی گردد. در دوره‌های بعد، رابطه اطلاعات/قدرت با ظهور علوم اجتماعی اثباتی که در قرن نوزدهم ظهور کرد و به سرعت در قرن بیستم گسترش یافت، ارتباط نزدیکی پیدا کرد. از لحاظ تاریخی، هدف اساسی در علوم اجتماعی — بر اساس گزاره های ارزشی آن – همیشه در روشنگری و بهبود زندگی انسان ها بوده‌است(McDonald, 1993). در کشورهای توسعه یافته، تحقیقات علوم اجتماعی به سرعت در دوران پس از جنگ در دهه‌های ۱۹۴۰ و ۱۹۵۰ توسعه یافت، که خروجی آن اقتصاد کینزی و برنامه‌ریزی های رفاه محور بود. مفهوم سیاستگذاری نیز که توسط لاسول و همکارانش ارائه و توسعه یافت بود به وضوح با مبانی ارزشی دموکراسی و رفاه اجتماعی در ارتباط بود (Head, 2010).

پس از این دوران بود که رویکرد سیاستگذاری مبتنی بر شواهد ظهور کرد. سیاستگذاری مبتنی بر شواهد، استفاده سیستماتیک از شواهد در شکل دهی به سیاست های عمومی است. رویکردی که منشا آن در حوزه پزشکی و درمان

[3] Efficient Governance

مبتنی بر شواهد[۴] است(Howlett, 2009;Parsons, 2002; Sanderson, 2002) . سیاستگذاری مبتنی بر شواهد زمانی فراگیر شد که مدرن سازی دولت ها[۵] در دستور کار کشورها قرار گرفت و تمایلات برای سیاستگذاری بر مبنای تحلیل های 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 سازمان های عمومی در مورد نرخ پایین بازگشت سرمایه گذاری ها در برنامه های اجتماعی، آن‌ها را به جستجوی رویکردهای جدید سیاستگذاری سوق داد(Cairney, 2016). این بیداری سیاستگذاران، فرصت‌هایی را برای علوم رفتاری و اجتماعی ایجاد کرد تا راه‌حل‌هایی به شکل رویکردهای جدید برای به دست آوردن کنترل بیشتر بر واقعیت‌های مبهم و آشفته ارائه دهد. راه‌حل‌های مبتنی بر دستور العمل‌های ایدیولوژیکی قدیمی دیگر کارایی نداشت و سیاستگذاران از رویکردهای جدید تر استقبال می کردند(Parkhurst, 2017). این جهت گیری های جدید با تمایل سازمان های مردم نهاد بزرگ، اصناف و دیگر ذینفعان برای مشارکت یا رویکردهای مشارکتی برای پرداختن به مسائل اصلی، همراه شد(Parkhurst, 2017). سیاستگذاران در پی حل مشکلات خود و افزایش کارایی سیاست ها، به رویکرد سیاستگذاری مبتنی بر شواهد[۶] که خروجی تحقیقات حوزه علوم اجتماعی در دهه های ۶۰ و ۷۰ بود روی آوردند. این رویکرد جدید در سیاستگذاری، نتیجه نمایان شدن ناکارآمدی های سیاستگذاری های دولتی و اجرای آن در زمینه های مختلف بود. در این رویکرد تلاش شده است تمرکزگرایی در رابطه با تدوین سیاست ها کم رنگ تر شود و دایره تنگ سیاستگذاران و کنش های سیاستگذار، گسترده تر شود.

رویکرد سیاستگذاری مبتنی بر شواهد به دنبال توسعه گزینه‌های سیاستی به منظور بهبود کیفیت تصمیم‌گیری های سیاستگذاران است. جنبش مبتنی بر شواهد در سیاستگذاری مدرن، آخرین نسخه استفاده از دانش کاربردی برای کمک به حل و فصل مشکلات عمومی است. استدلال این است که این رویکرد، برای حل منطقی مسائل، با تمرکز بر دانش حاصل از اطلاعات دقیق، کار می کند. همچنین گفته می شود که این رویکرد توانایی هماهنگی و پاسخ به ریسک های محیطی را در فرایند اجرای سیاست ها دارد. در نسخه تکنوکراتیک ترِ رویکرد مبتنی بر شواهد، هدف، تولید دانش مورد نیاز برای برنامه‌های تنظیم گری و ایجاد دستورالعمل‌ها و ابزاری برای مقابله با مشکلات شناخته‌شده اس(Roberts, 2005). در این رویکرد، سیاستگذار به گزینه های بیشتری توجه می کند و در جهت مداخله دیدگاههای رقیب نیز در فرایند سیاستگذاری، قدم برخواهد داشت.

رویکرد سیاستگذاری مبتنی بر شواهد هم یک مجموعه فعالیت‌های تاکتیکی در سیاستگذاری را ارائه می‌دهد؛ و هم به دنبال تبیین شیوه های مشروع تر تصمیم‌گیری توسط سیاستگذاران است که می تواند جایگزین سیاستگذاری های

[4] Evidence-Based Medicine
[5] Modernizing Government
[6] Evidence Base Policy (EBP)

مبتنی بر شهود[۷]، عقیده و ایدئولوژی شود(Hardie & Cartwright, 2012). محققان این حوزه معتقدند، رویکرد سیاستگذاری مبتنی بر شواهد دو اصل اساسی دارد و در واقع مبتنی بر دو زمینه ای است که تا وجود نداشته باشد نمی توان به پیاده سازی درست این نوع سیاستگذاری امید داشت. اولین زمینه اساسی، یک فرهنگ سیاسی مطلوب است که اجازه ورود عناصر اساسی شفافیت و عقلانیت را در فرآیند سیاستگذاری می دهد. این موضوع به نوبه خود ممکن است اولویت سیاستگذاران را برای استفاده از دانش در سیاستگذاری ها تسهیل کند. دوم، فرهنگ تحقیقاتی متعهد به تحقیقات تحلیلی با استفاده از روش‌های دقیق علمی که طیف وسیعی از شواهد را برای سیاستگذاری تولید می کند. رویکرد سیاستگذاری مبتنی بر شواهد، در رابطه با استفاده از روش‌های بهتر برای جمع‌آوری سیستماتیک شواهد و استفاده دقیق از آن به عنوان پایه و اساس تصمیم‌گیری در بخش عمومی در سه دهه پیش تاکنون مورد توجه قرار گرفت. در نتیجه در حال حاضر بدنه بزرگی از مطالعات در این حوزه وجود دارد که با تحلیل پیچیدگی های عوامل زمینه‌ای کمک می‌کند تا شرایط تاثیر گذار بر اثربخشی سیاست ها را توضیح دهند(Head, 2010). ایجاد این فرهنگ تحقیقاتی که پاسخ گوی نیازهای سیاستگذار در زمینه های مختلفی باشد، هنوز در همه زمینه های سیاستگذاری ایجاد نشده است و نمی توان 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.

مفهوم «مدیریت مبتنی بر شواهد» نیز در این سال ها پژوهش های قابل‌توجهی را به خود اختصاص داده‌است و این مفهوم نیز به عنوان مفهومی نزدیک به سیاستگذاری مبتنی بر شواهد، موضوع تحقیقاتِ مختلفی بوده است. این مفهوم براساس این ایده است که برنامه هایی که با اطلاعات منسجم پشتیبانی می‌شوند، نسبت به برنامه هایی که بدون اطلاعات هستند برتری دارند. مدیران سازمانی به شواهد دقیق در مورد عملکرد، استانداردها و شرایط محیطی وابسته هستند؛ کسب و کارهای موفق بیش از آنکه متکی بر سابقه، قدرت و شهود شخصی باشند، متکی بر اطلاعات قابل‌اعتماد هستند (Sutton, 2006; Rousseau, 2006 & Argyris) حتی در این رابطه، مفهوم «عمل مبتنی بر شواهد» به طور گسترده ای در حوزه های پزشکی رشد کرده است(Fitzgerald , 2005 ; Yeager & Dopson, 2006). شهرت تحقیقات حوزه پزشکی با رویکرد «مبتنی بر شواهد» آنقدر زیاد است که بیشتر ادبیات حوزه سیاستگذاری مبتنی بر شواهد، با این پرسش آغاز می‌شود که آیا این رویکرد جدید سیاستگذاری مبتنی بر شواهد در حوزه پزشکی و سلامت، برای حوزه‌های دیگر سیاستگذاری، مانند رفاه اجتماعی، آموزش، ایجاد عدالت اجتماعی، سلامت جامعه و مدیریت منابع طبیعی و ... قابل‌اجرا است؟(Cairney, 2016). از اینرو باید کارهای اصلی و قدیمی تر در رابطه با سیاستگذاری مبتنی بر شواهد را در زمینه های علوم پزشکی تعقیب کرد.

[7] Intuition Base Policy (IBP)

گرایش به رویکرد «سیاستگذاری مبتنی بر شواهد» از سوی سیاستگذاران و محققان، هم فرصت و هم چالش هایی ایجاد کرده است. این رویکرد برای مدیران و سیاستگذاران، فرصت بهبود مستمر در سیاستگذاری براساس ارزیابی منطقی و مقایسه آگاهانه ی گزینه‌های سیاستی، را در پی داشته است. در این رویکرد همچنین مزایای متقابل زیادی برای محققان و شهروندان وجود دارد (Hardie & Cartwright, 2012). نیاز به اطلاعات خوب یکی از پایه‌های سیاستگذاری خوب است (Shaxson, 2005). جنبش سیاستگذاری مبتنی بر شواهد به دنبال ارتقا تجزیه و تحلیل های دقیق گزینه‌های سیاستی است، در نتیجه ورودی‌های مفیدی را برای سیاست گذاران در توسعه سیاست ها فراهم می‌کند. با این حال، اکنون روشن است که امیدها برای پیشرفت‌های بزرگ و سریع در سیاست‌ها، به عنوان نتیجه بهره گیری از تحقیقات دقیق، به آن راحتی که انتظار می‌رفت تحقق نیافته است. اگر واقع گرا باشیم درخواهیم یافت که بهبودهای مبتنی بر شواهد مطلوب و ممکن هستند اما نمی‌توانیم انتظار داشته باشیم که یک سیستم سیاست گذاری بسازیم که با یافته‌های تحقیقات عینی کاملا منطبق باشد. چالش اینجاست که ممکن است سیاستگذاری مبتنی بر شواهد فضای امنی ایجاد کند که سیاستگذاران بدور از مشورت با ذی نفعان، صرفا بر اساس تحقیقاتی با دستکاری در نتایج، سیاست های مورد نظر خود را تدوین کند. از اینرو تحقیقات و نتایج معتبر تحقیقات، گمشده سیاستگذاری مبتنی بر شواهد است.

ایجاد ارتباط سیستماتیک بین تحقیقات دقیق و فرآیندهای سیاست گذاری نیازمند تلاش طولانی مدت و پیچیده در همه سطوح است. اکنون شواهد صرفا می‌تواند سیاستگذار را آگاه سازد و نمی توان انتظار داشت ارتباط سیستماتیک بین آن و فرآیند سیاستگذاری وجود داشته باشد. دلایل متعددی برای این امر وجود دارد. اول، امکان تحقیقات دقیق، با یافته‌های تحقیقاتی دقیق در مورد مسایل کلیدی، در بسیاری از حوزه‌ها برای اطلاع‌رسانی به سیاستگذاران وجود ندارد. علاوه بر این، این پارادوکس وجود دارد که هرچه بیشتر در مورد مسایل اجتماعی کاوش کنیم، بیشتر احتمال دارد از شکاف‌ها و محدودیت‌های دانش خود آگاه شویم. دوم اینکه، سیاستگذاران و رهبران سیاسی اغلب توسط عوامل بسیاری بجز شواهد تحقیق، تحریک و تحت‌تاثیر قرار می‌گیرند. بنابراین می توان گفت، تنها در دسترس بودن تحقیقات قابل‌اعتماد تاثیر آن را تضمین نمی‌کند. سیاستگذاران اغلب با مسائلی مانند حفظ حمایت طرفداران خود، لزوم پاسخگویی به رسانه‌ها، بهبود اعتبار خود، و ... درگیر هستند. سوم، حتی در جایی که شواهد معتبر مستند شده‌است، اغلب «تناسب» ضعیفی بین چگونگی گردآوری این اطلاعات توسط محققان(به عنوان مثال گزارش‌های علمی) و نیازهای عملی سیاستگذاران وجود دارد. محققان خود اغلب در ارائه یافته‌های خود ماهر نیستند و ترجیح می‌دهند از تعامل مستقیم با بحث‌های عملی در مورد مسائل کلیدی فاصله بگیرند و تنها به حوزه نظر بپردازند. چهارم اینکه، ارزش دانش حرفه‌ای، متمایز از دانش تجربی مبتنی بر تحقیق، هر روز حیاتی تر می شود. سیاستگذاران چگونه باید مزایای نسبی تکیه بر یافته‌های دقیق تحقیقات علمی/سیستماتیک و تکیه‌بر تجربه عملی ارائه کنندگان خدمات

حرفه‌ای را بسنجند؟ این موضوع یکی از مشکلاتی است که در رابطه با این رویکرد وجود دارد. پنجم، به نظر می‌رسد سیاستگذاری مبتنی بر شواهد در مسائلی که دارای محیط های پیچیده ای هستند و یا در معرض تغییرات سریع قرار دارند، کمتر موفقیت آمیز بوده است(Head, 2008). گفته می شود، این رویکرد از سیاستگذاری در بهترین حالت می‌تواند مجموعه ای مورد توافق و متقاعد کننده از شواهد تحقیقاتی را ارائه داده و آن را در اختیار سیاستگذار قرار دهد و نمی تواند به حل مسائل کمک مستقیم داشته باشد. سیاستگذاری نیازمند شواهد مختلفی در حوزه‌های سیاستی پیچیده، متغیر و دارای ریسک بالا است. از اینرو تولید شواهد به روز، چند بعدی و چند سطحی، می تواند کمک بیشتری به سیاستگذار در این حوزه دهد.

۲.۱. شواهد چیست؟

همان گونه که گذشت، توسعه سیاستگذاری مبتنی بر شواهد، براساس استفاده از تحقیقات دقیق و تلفیق این یافته‌ها با فرآیندهای سیاستگذاری، در حوزه سیاستگذاری عمومی در حال گسترش است. جستجوی شواهد دقیق و قابل اطمینان، و ترویج استفاده از آن در فرآیند سیاستگذاری، جزء اصول اساسی رویکرد سیاستگذاری مبتنی بر شواهد هستند. سیاستگذاری مبتنی بر شواهد بدون شواهد دقیق، کافی و مناسب نمی تواند به درستی کار کند و تاثیرات مورد انتظار را داشته باشد. متخصصان اکنون علایق پژوهشی قابل توجهی در خصوص تعیین کیفیت داده‌ها جهت استفاده در تحقیقات کاربردی از خود نشان داده اند و پژوهش های زیادی وجود دارند که به سوالات روش‌شناختی نظیر اعتبار، قابلیت اطمینان و عینیت شواهد تولید شده، پاسخ می‌دهند(Head, 2008). اما نظراتی هم وجود دارد که معتقدند در سیاستگذاری مبتنی بر شواهد، دایره شواهد باید گسترش یابد تا انواع مهم دیگر شواهد را نیز در بر بگیرد.

به طور سنتی، مبنای «شواهد» به عنوان پایه و اساس سیاستگذاری مبتنی بر شواهد، دانشی است که با تحقیقات کاربردی، چه در داخل و چه خارج از سازمان های دولتی، تولید می‌شود. این شواهد، شامل شواهد کلی در مورد روندهای گسترده و توضیحات پدیده‌های اجتماعی و سازمانی، و همچنین شواهد خاص تولید شده از طریق شاخص‌های عملکرد و ارزیابی‌های برنامه است (Nutley, Davies & Walter, 2002). اشکال متعددی از شواهد مرتبط با سیاست ها وجود دارد که برای درک مسایل و چشم اندازهای موفقیت مداخلات سیاست گذارانه، حیاتی هستند. شواهد سیاستی می تواند هم از منابع علمی مانند آمارها و آزمایش ها و همچنین در قالب علوم اجتماعی مانند پیمایش ها و گروه های کانونی بدست آید. بسته به ماهیت موضوعی که مورد نظر سیاستگذار است و توانایی جمع آوری داده ها و قابلیت های سیاستگذاران(Howlett, 2009; Parsons, 2002) ، دولت ها برای انجام فعالیت های مرتبط با سیاستگذاری، نیازمند سطح مشخصی از قابلیت های تحلیل شواهد هستند. اما گفته می شود، چنین موردی در عمل

وجود ندارد و دولت ها نمی توانند انبوهی از شواهد رسمی و غیر رسمی را به طور سیستماتیک تحلیل و وارد فرایند سیاستگذاری کنند (Newman, Cherney, & Head, 2017).

سیاستگذار نیازمند شواهدی در ابعاد مختلف برای طراحی سیاست های کارا و موثر می باشد. یک اصل مهم در سیاستگذاری مبتنی شواهد، تاکید بر این است که نیاز به ساختن مبنایی برای جمع آوری شواهد بر مبنای درگیرسازی عموم و همه ذی نفعان وجود دارد که در واقع مبنای تولید شواهد از جنس علوم اجتماعی محسوب می شود (Head, 2008). درگیر شدن با عموم و همه ذی نفعان یک گزینه سیاستی، معمولا به عنوان بخشی از ارتباطات بلند مدت سیاستگذار با آنها و همچنین توسعه قابلیت های درگیری با عموم تلقی می شود (Pang, Lee, & DeLone, 2014). داده های رسانه های اجتماعی به عنوان منابع لحظه ای و مفید اطلاعات از ذی نفعان مهم و عموم مردم است که نمی توان در این رابطه آن را نادیده گرفت. شواهد سیاستی تاکنون بیش از همه از نتایج تحقیقات رسمی مرتبط با سیاست ها تامین شده است. وقتی که شواهد فقط از منابع قدیمی مثل پیمایش یا مشاوره با متخصصان و ... اخذ می شود، گستره ورودی های سیاستگذاری محدود می شود، گرچه ممکن است به صرفه اقتصادی باشد و قابلیت جمع آوری داشته باشد. اکنون شاهد افزایش استفاده از فن‌آوری‌های جدید برای جمع‌آوری داده‌ها و تجزیه و تحلیل داده‌ها به منظور اندازه‌گیری ماهیت و وسعت مسائل سیاستی، ارزیابی اثرات فعلی سیستم‌های خدماتی و ارائه معیارهایی برای پیش بینی اثربخشی آنها در آینده هستیم. این ابزار ها به عنوان مکمل شیوه های سنتی جمع آوری و تحلیل داده ها، عمل می کنند.

از دیدگاه سیاستگذاری مبتنی بر شواهد، این سوال مطرح گردید که چه نوع ای داده/ اطلاعاتی برای تولید شواهد، هم برای درک مشکلات پیچیده و هم برای ایجاد راه‌حل‌های پایدار، مورد نیاز است. برخی از محققان علوم اجتماعی و تحلیل گران سیاستگذاری شروع به طرح این پرسش کردند که آیا تداوم مشکلات پیچیده اجتماعی بیش از همه به فقدان اطلاعات، یعنی «شکاف» در پایگاه های داده در دسترسِ سیاستگذار، مربوط می‌شود یا خیر(Head, 2008). اما اکنون، محققان معتقدند که به دست آوردن داده‌های بیشتر برای پر کردن شکاف‌های شناخته‌شده لزوما ما را به سمت راه‌حل‌های سیاستی خوب سوق نخواهد داد، زیرا عمده فعالیتی که در سیاستگذاری باید صورت گیرد، آشتی دادن دیدگاه‌های ارزشی و سیاسی متفاوت است. این محققان بر روی کیفیت داده ها و همچنین سیستم سیاستگذاری کارا تاکید دارند. سیستمی از سیاستگذاری که بتواند میان شواهد و نظرات ذی نفعان همبستگی ایجاد کند.

از اینرو بر سر اینکه چه نوع شواهدی و با چه کیفیتی می تواند به چه میزان به بهبود سیاستگذاری ها کمک کند، اختلاف نظر وجود دارد. می توان گفت در دیدگاهی وسیع‌تر، یک مبنا واحد جهت اینکه شواهد چیست؟ وجود ندارد. در واقع، ساختارهای نامتجانس دانشی به مجموعه‌ای از شواهد تبدیل می‌شوند که به سیاستگذاران نه در تعیین

سیاست ها بلکه در آگاه سازی آنها کمک می کند(Parkhurst, 2017). در این بررسی گسترده از دانش مربوط به سیاستگذاری، اعتبار و سودمندی «شواهد علمی»، که توسط استانداردهای روش‌شناسی علمی تایید شده و یک ورودی بسیار مهم برای توسعه سیاستگذاری است، محل سوال است. بنابراین، تحقیقات دقیق و سیستماتیک ارزش زیادی دارد، اما باید در یک زمینه وسیع‌تر از شواهد مختلف، دیده شود تا قابل استفاده باشد(Cairney, 2016). یعنی نیازمند این هستیم که شواهد علمی در رابطه با سیاست ها با دیگر شواهد همراه شود تا نیازهای سیاستگذاری را برآورده سازد. از این رو، استدلال می‌شود که سیاستگذاری موثر در طراحی، اجرا و ارزیابی، به چندین نوع شواهد وابسته است. همه این شواهد مختلف می توانند به طور مستقیم یا غیر مستقیم در توسعه سیاست های خوب، دخیل باشند و به ما کمک کنند تا «اثربخشی سیاست ها» را در یک محیط سیاست گذاری شبکه‌ای درک کنیم. سیاست ها، ذی نفعان مختلف با منافع متفاوتی دارد از اینرو شواهد گوناگونی نیاز است تا در فرایند سیاستگذاری ها وارد شود.

محیط عمومی که در آن سیاست‌های دولت ها اجرا می شود، حوزه عمومی و در واقع حوزه افکار عمومی است که همه ذی نفعان سیاست ها در آن حضور دارند. این محیط نسبت به سیاست های مختلف واکنش نشان می دهد و موفقیت و یا شکست سیاست ها در این محیط رقم می خورد. سیاستگذار باید به این محیط پیچیده که ذی نفعان در آن حضور دارند پاسخگو باشد و آنها را برای اجرای سیاست ها ترغیب کند. از اینرو سیاستگذاری مبتنی بر شواهد به این موضوع هم توجه دارد که چگونه فرهنگ سیاسی در یک محیط، جهت گیری های سیاستگذاری را شکل می‌دهد. سیاستگذاران شواهد را جهت طراحی و تدوین سیاست ها بر اساس ارزیابی های سیستماتیک و ایدئولوژیک خود، جمع آوری، فیلتر، ترجمه و تفسیر می کنند(Ingold & Monaghan, 2016) . استفاده از نظرات عموم شهروندان در مورد سیاست ها و کاربست آنها در سیاستگذاری در قالب شواهد رسمی، امر جدیدی نیستند اما تازگی، حجم بالا، پیچیدگی و تنوع[8] داده های رسانه های اجتماعی یک دیدگاه چالش برانگیز جدید در این خصوص ایجاد می کند. نیازمندی های سیاستگذاران در استفاده از داده های تعاملات دیجیتال شهروندان، برای اطلاع از دیدگاهای عمومی و اینکه چگونه گفتمان های عمومی راجع به سیاست ها شکل می گیرند و ذی نفعان اصلی چه کسانی هستند و گروههای تخصصی و ارتباطات آن ها به چه صورت است هر روزه گسترش می یابد (Castelló et al., 2013; Kasabov, 2008; Sanderson, 2002). این پدیده سوال در مورد اینکه چطور این مخاطبان جدید می تواند شواهد مفیدی برای سیاستگذاران ارائه دهند را افزایش می دهد.

گفته می شود(Parkhurst, 2017)، سه نوع دانش به عنوان شواهد، وجود دارد که برای سیاستگذاری مهم هستند. این انواع دانش از این قرار است: دانش سیاسی؛ دانش تحلیلی و فنی؛ و تجربه میدانی و حرفه‌ای. این سه نوع دانش، سه

[8] Novelty, Volume, Complexity and Diversity

نوع لنز شواهدی را برای سیاستگذار فراهم می سازد. همه این سه نوع شواهد، با تفاسیر، قیدها و محدودیت‌هایی همراه هستند که در زیر توضیح داده شده است(Cairney, 2016):

۱- دانش سیاسی در واقع، دانش فنی، تحلیلی و قضاوت کنشگران سیاسی است. این فعالیت‌های تحلیلی و قضاوتی شامل چندین عنصر حیاتی مربوط به سیاستگذاری مانند: قضاوت های تحلیلی کنشگران سیاستگذاری، در نظر گرفتن بستر تنظیم سیاست ها؛ در نظر گرفتن دستور کار سیاست ها؛ تعیین اولویت‌ها؛ تدارک ارتباطات لازم و کلیدی برای اجرای سیاست ها، در نظر گرفتن ایدئولوژی های اساسی حاکم بر فضای سیاسی؛ تشکیل گروه‌های ائتلافی؛ و مذاکره بر سر پذیرش سیاست ها می شود.

این نوع از دانش «سیاسی» در درجه اول در گروه هایی مانند سیاستمداران، احزاب، گروه‌های سازمان‌یافته و رسانه‌های عمومی ذاتا وجود دارد. اگر چه بخشی از این دانش خصوصی و محرمانه است اما بخش بزرگی از آن به طور گسترده در میان مردم و به خصوص توسط رسانه‌ها منتشر می‌شود. این دانش، پراکنده، بسیار سیال و به شدت بر مبنای زمینه های ایدئولوژیک و گرایشات حزبی است. سیاستگذاری که با استفاده از چنین دانش سیاسی انجام می شود، بیشتر بر جذب حمایت کنشگران در شبکه سیاستگذاری تمرکز می کند تا اینکه بر عینیت گرایی تمرکز داشته باشد.

جهت دار بودن و تعصب در پذیرش اطلاعات، اظهارات ایدئولوژیک و ... بخشی از ویژگی های زمینه ای سیاستگذاری است که در آن این نوع از دانش و شواهد سیاسی وجود دارد. به عبارت ساده تر در این نوع از شواهد، ممکن است یک بخش از شواهد مورد استدلال قرار گیرد؛ و بخش‌های بزرگی از شواهد دیگر یا نادیده گرفته شوند یا جهت دار به نفع گرایشات سیاسی، تفسیر شوند. این استفاده جهت دار از شواهد اغلب رفتار سیاسی و یا بازی سیاسی نامیده می شود. در بازی سیاسی، ادعا و فریب، عادی است، گاهی اوقات استفاده جهت دار از شواهد بصورت تاکتیکی و گذرا اتفاق می افتد؛ اما گاهی اوقات این ادعاها و فریب ها به طور سیستماتیک بر اساس اهداف ایدئولوژیک سیاستگذار است که اتفاق می افتد. به این نوع از سیاستگذاری برخی «سیاستگذاری مبتنی بر ایمان» نیز گفته اند.

برخی از حوزه‌های سیاستگذاری وجود دارند که موضوع تعهدات دولت در میان است. این تعهدات، دیگر برای بحث درباره ماهیت مشکل، بهترین راه‌حل و ارزیابی طیف شواهد مربوط به اثربخشی سیاست ها، باز نیستند. این به این معنی است که برخی از مواضع سیاسی «ضد داده» یا «ضد شواهد» هستند، به این معنا که «مبنای» شواهد آن‌ها محدود است و صرفا بر اساس تعهدات سیاسی شان است. این تعهدات، جزئی از ارزش‌ها و مواضع ایدئولوژیک رهبران یا احزاب سیاسی است. برخی از اولویت‌های سیاسی سیاستگذاران، تنها اجازه می‌دهد انواع خاصی از «شواهد» در فرایند سیاستگذاری مورد توجه قرار گیرند و تفسیرهای انتقادی در این

شرایط برای سیاستگذار ناخوشایند است. برخی دیگر از مشکلات در این حوزه نیز از این قرارند که تمایل به امتیاز دادن به برخی شواهد به عنوان شواهد مرتبط و رد کردن شواهد دیگر به استدلال شواهد غیر مرتبط یا صرفا ایدئولوژیکی، در میان سیاستگذاران وجود دارد. تعداد کمی از پروژه‌های تحقیقاتی در حوزه سیاستگذاری، بدون در نظر گرفتن این انتظارات سیاسی انجام می‌شوند و اکثر پروژه‌ها با خواست سیاستگذار به تقویت یک دیدگاه خاص کمک کرده اند.

۲- دانش مبتنی بر تحقیقات میدانی، محصول تجزیه و تحلیل سیستماتیک عوامل و روندهای فعلی و گذشته، و تحلیل روابط علّی در یک پدیده است که عوامل و روندها را توضیح می‌دهد. در رابطه با بررسی و ارزیابی سیاست ها، طیفی از دانش نظام مند و میان رشته‌ای (اقتصادی، حقوقی، جامعه‌شناسی، مدیریتی و غیره)وجود دارد که کمک‌های بسیاری به درک و بهبود سیاست ها می‌کند. به ندرت اتفاق آرا در مورد ماهیت مشکلات، علل گرایش‌ها یا روابط و بهترین رویکرد برای راه‌حل، میان محققان علوم اجتماعی وجود دارد. رشته‌های علمی مختلف ممکن است رویکردهای روش‌شناختی متفاوتی داشته باشند و دیدگاه‌های مکمل یا گاهی رقیب در مورد مسائل پیچیده ارایه دهند. برای همین عجیب نیست که رویکردهای بین رشته‌ای چند سطحی در دهه‌های اخیر برای رسیدگی به مشکلات اجتماعی مطرح شده‌اند.

قالب های علمی (مبتنی بر تحقیق) شواهد، محصول کار محققان و متخصصان روش‌های سیستماتیک جمع‌آوری و تجزیه و تحلیل داده ها است. گفته می شود توجه به کیفیت و ثبات داده‌ها در این رویکرد علمی، اساسی است. در یک سوی طیف، در علوم رفتاری و اجتماعی جهت انجام بررسی‌های سیستماتیک، استانداردهای جدی به کار گرفته می شود و تنها مطالعاتی که بر ارزیابی اثرات علّی عوامل یک پدیده تمرکز دارند، به رسمیت می شناسد. این رویکرد به عنوان سختگیرانه‌ترین روش برای درک و ارزیابی آنچه در حوزه سیاستگذاری اتفاق می افتد، شناخته می‌شود. گاهی تمایل به داده‌های کمی در رابطه با این دسته از شواهد دیده می شود، اگرچه داده‌های کیفی که مرتبط با نگرش های ذی نفعان حوزه سیاستگذاری است و می تواند به توضیح شرایط و ماهیت آن کمک کند(دیویس ۲۰۰۴؛ پرسی اسمیت ۲۰۰۵). در سوی دیگر این طیف، برخی روش‌های مرتبط با رویکرد هرمنوتیک تمایل دارند سیاست و ارزیابی آن ها را بیشتر شبیه پروژه‌های یادگیری اجتماعی در نظر بگیرند و در تعامل با زمینه مربوطه نتایج تحقیق را تکمیل کنند. این رویکرد در مقابل رویکری قرار می گیرد که داده ها را بر اساس روش های علوم تجربی تحلیل می کنند.

۳- دانش عملی در زمینه سیاستگذاری، عبارت از دانشی است که متخصصان بر اثر کار کردن در صحنه عمل کسب کرده اند و در واقع دانش سازمانی مرتبط با اجرای سیاست ها است. این شبکه های حرفه‌ای و مدیریتی اغلب مجری سیاست ها هستند و طیف متنوعی از کارکنان سازمان ها شامل مدیران و متخصصان است که به

طور مستقیم در اجرا و ارزیابی سیاست ها درگیر هستند (پاوسون و همکاران ۲۰۰۳). دانش این دسته از افراد به لحاظ تجره عملی که در پروژه های سیاستگذاری دارند می تواند به عنوان منبعی برای شواهد مورد نیاز در زمینه سیاستگذاری مبتنی بر شواهد مورد استفاده قرار گیرد. شاید بیش از همه این دانش عملی کارشناسان باشد که بتواند به شفاف شدن پیش بینی ها در مورد اجرای سیاست ها در میدان عمل کمک کند.

همانگونه که در بالا پیشنهاد شد، سه نوع از دانشِ شواهد ساز وجود دارد که در طراحی، اجرا و ارزیابی سیاست‌ها نقش محوری دارند. از اینرو می توان گفت، با یک پایگاه شواهد مواجه نیستیم بلکه چندین پایگاه شواهد وجود دارد. این ساختارهای نامتجانس دانش و شواهد به مجموعه‌ای چندگانه از شواهد تبدیل می‌شوند که به جای تعیین سیاستگذاری، صرفا آگاهی دهنده و تحت‌تاثیر قراردهنده هستند. سه نوع شواهد مربوط به سیاستگذاری شامل سه دیدگاه در مورد اطلاعات مفید و قابل‌استفاده برای سیاستگذار هستند. هر یک از این انواع، دارای پروتکل‌های مشخص دانشی، تخصصی، سیاسی، در مورد آنچه که به عنوان «شواهد» به حساب می‌آید، می‌باشد. در این رابطه سوالاتی نیز از سوی محققان سیاستگذاری مطرح شده است؛ آیا در میان این کانال های مختلف شواهد، شکاف وجود دارد؟ و اگر چنین است، آیا مکانیزم‌ها / مشوق‌هایی برای ترویج هم‌کاری نزدیک‌تر آنها با یکدیگر وجود دارد؟ آیا یکی از این کانال های دانشی بر دیگری ارجحیت دارد؟ اینها پرسش هایی است که در رابطه با کانال های شواهدی مختلف وجود دارد. مسلما همه اینها می تواند در کنار هم کمک کننده باشد اما اینکه سیاستگذار چگونه از هر کدام بهره ببرد، مسئله ای است که می تواند این شواهد را کارا و یا ناکارآمد سازد.

یک مساله کلیدی دیگر اهمیت شواهد «کیفی» است، به عنوان مثال شواهد مربوط به ارزش‌ها، نگرش‌ها و ادراکات ذینفعان و سیاستگذاران بسیار مهم است. اما اینکه چگونه باید از این شواهد استفاده نمود؟ در این خصوص نظرات متفاوت است. برخی از متخصصان تمایل دارند از روش‌های ترکیبی مناسب استفاده کنند و پلی در میان شکاف شواهد کیفی/کمی ایجاد کنند. روش‌های ترکیبی توسط محققانی که تلاش می‌کنند مشکلات پیچیده را توضیح دهند و مداخلات پیچیده را ارزیابی کنند، بسیار زیاد مورد استفاده قرار می گیرد (Woolcock , 2009). شواهد مختلف براساس دقت روش‌شناختی جمع‌آوری و تحلیل آنها، رتبه‌بندی می شوند و در واقع ما با «سلسله‌مراتب شواهد» روبرو هستیم. پژوهشگران در مورد روش‌ها و ابزارهای مختلف جمع آوری و تحلیل داده نظرات مختلفی دارند. در اینجا سوال اساسی میزان اعتماد به اعتبار یافته‌های تحقیق است. استانداردهای سیاستگذاری مبتنی بر شواهد از روش‌های علمی برای ایجاد اعتبار داده‌ها استفاده می کند(Young., et , 2002). این استانداردها کمک می کند داده های معتبرتری وارد فرایند سیاستگذاری گردد.

سرمایه‌گذاری در ایجاد بانک‌های اطلاعاتی برای محققان علوم اجتماعی و سیاستگذاران همچنان برای دولت ها موضوع مهمی است. اما چه نوع اطلاعاتی برای ذینفعان و تصمیم گیرانی که با چالش های مشکلات پیچیده مرتبط با بسیاری از ذینفعان سروکار دارند، ارزشمند خواهد بود؟(Parkhurst, 2017). با در نظر گرفتن این نگرانی‌ها، به نظر می‌رسد که چهار عامل تعیین‌کننده وجود دارد که می تواند به کارایی بیشتر سیاستگذاری مبتنی بر شواهد کمک کند و می تواند به عنوان پایه‌ای برای طراحی و توسعه یک سیستم سیاستگذاری مبتنی بر شواهد قرار گیرد:

(۱) پایگاه‌های اطلاعاتی با کیفیت بالا در حوزه‌های موضوعی مربوطه. دولت‌ها سرمایه گذاران اصلی در جمع‌آوری و انتشار داده‌های سیستماتیک در مورد روندها و مسائل مهم اقتصادی، اجتماعی و زیست‌محیطی بوده‌اند. اصل شفافیت نیاز به سرمایه‌گذاری در اطلاعات با کیفیت بالا و قابل‌دسترس را تقویت می‌کند. و این دولت ها هستند که بایستی در این زمینه اقدام و سرمایه گذاری های لازم را انجام دهند. اما به هر حال ساختن پایگاه‌های اطلاعاتی و سرمایه‌گذاری در این حوزه پرهزینه است.

(۲) جذب متخصصان با مهارت بالا در تحلیل داده‌ها و سیاست ها. داده‌ها باید به اطلاعات مفیدی در مورد الگوهای علی، روندها در طول زمان و درک اثرات احتمالی ابزارهای سیاستی مختلف تبدیل شوند. حجم انبوهی از داده ها بدون تحلیل های مناسب نمی تواند کمک موثری برای سیاستگذار باشد. این متخصصان تحلیل داده هستند که با تحلیل این داده ها به اطلاعات ارزشمند باعث ایجاد سیاست های مناسب در حوزه های مختلف خواهند شد.

(۳) مشوق‌های سیاستی و سازمانی برای استفاده از تجزیه و تحلیل های مبتنی بر شواهد. آنچه مسلم است، فرآیندهای تصمیم‌گیری دولتی به طور تاریخی با استفاده از تحلیل های مبتنی بر شواهد تنظیم نشده اند؛ این امر مستلزم توجه به جو سازمانی و فرهنگ پیرامون تولید تحلیل و مشاوره است. برخی مدیران سیاسی، پوپولیسم و ضد شفافت هستند و مشکلات بزرگی را در راه ترویج سیاستگذاری مبتنی بر شواهد ایجاد می‌کنند. از سوی دیگر، برخی از مدیران و دولت‌ها، از چشم‌انداز تحلیل‌های شفاف‌تر و دقیق تر برای بهبود سیاستگذاری ها استقبال می کنند. ایجاد مشوق هایی برای سیاستگذاران و مدیران، برای استفاده از داده های تولید شده در فرایند سیاستگذاری می تواند در این مسیر کمک کننده باشد.

(۴) ایجاد درک متقابل بین مدیران اجرایی سیاستگذاری، محققان و سیاستگذاران. تحقیقات در مورد سیاستگذاری مبتنی بر شواهد در هر دو بخش «عرضه» (کار محققان یا تحلیلگران) و «بخش تقاضا» (رفتار سیاستگذاران در استفاده از اطلاعات) را مد نظر دارد. اگر قرار باشد پژوهش ها برای تولید شواهد، بهبود یابد تمام این گروه‌ها باید فعالیت‌ها و اولویت‌های خود را با هم تطبیق دهند. محققان باید بر روی مسائل اولویت دار کار کنند، سوالات درست بپرسند و

یافته‌های مرتبط را به روشی نظام مند در دسترس سیاستگذار قرار دهند (Nutley et al. , 2007). و از سوی دیگر سیاستگذاران، شواهد مورد نیاز سیاستگذاری را به کار گیرد.

۳.۱. چالش های استفاده از شواهد در سیاستگذاری

مطالبات مربوط به دولت کارآمد و موثر، نیاز به اطلاعات را افزایش داد. این امر، اهرمی برای تحقیقات اجتماعی کاربردی، مرتبط با ارزیابی سیاست ها، اثربخشی اجرا، و مدل‌های جدید برای مقابله با مسایل پیچیده با استفاده از ابزارها و فرآیندهای سیاستگذاری جدید فراهم کرد. این امر مفهوم کلی سیاستگذاری مبتنی بر شواهد را ترویج داد. هدف دیدگاه سیاستگذاری مبتنی بر شواهد، دستیابی به دانش جمعی، به عنوان مبنایی برای یادگیری سیاست ها بویژه سیاست های تکراری در یک فرآیند سیاست گذاری است که به طور دقیق یافته‌های تحقیقات و ارزیابی ها را مورد استفاده قرار می دهد (Head, 2010).

همیشه کارآمدی استفاده از تحقیقات علوم اجتماعی در سیاستگذاری مورد سوال بوده است چرا که بسیاری از آن‌ها از کیفیت لازم برخوردار نبوده و یا با اجرای ضعیف نتایج آن در سازمان‌های دولتی همراه بوده اند. نتایج تحقیقات علوم اجتماعی چگونه می‌تواند دقیق‌تر و قابل‌اعتمادتر شوند؟ تا بتواند به سیاستگذار در تدوین سیاست ها کمک کند. این پرسش تا سال ها یکی از پرسش های اساسی این حوزه بود. تا دهه ۱۹۶۰ این اعتقاد به وجود آمد که روش‌های کمی می تواند این دغدغه را برطرف سازد و اصرار زیادی وجود داشت که تحقیقات علوم اجتماعی استانداردهای روش شناسانه خود را ارتقاء دهد. استفاده از داده‌های کمی و روش‌های تجربی به عنوان ابزاری برای ارایه شواهد دقیق‌تر و قابل‌اطمینان برای تصمیم گیرندگان در این دوره مورد حمایت قرار گرفت. از اینرو در این دهه‌ها، تمرکز بر مزایای دقت بالا در تحقیقات و استفاده بیشتر از روش‌های کمی و تحلیلی بوجود آمد(Head, 2010).

اما در مقابل گرایش هایی به تحقیقات کیفی نیز در این رابطه بوجود آمده است. در اینجا به استدلال‌هایی که در مخالفت با تقدم داده های کمی بر داده های کیفی می شود، ارائه شده است: اول مشکل اجرای تحقیقات کمی در حوزه‌های حساس سیاستگذاری ؛ دوم مشکل پیاده سازی نتایج تحقیقات تجربی به دنیای تعاملی و پیچیده واقعی؛ (Deaton, 2009) و سوم تمایل نتایج تحقیقات تجربی به کم‌اثر کردن تجربه حرفه‌ای‌ها که دارای دانش زمینه‌ای در حوزه سیاستگذاری مربوطه هستند نیز از جمله این موارد است(Henrik, 2014). این موارد استدلال های بسیار مهمی است که باعث شد، انواع دیگر تحقیقات و شواهد نیز در فرایند سیاستگذاری معتبر شمرده شود. همچنین این احتمال وجود دارد که سیاستمداران، مدیران سیاسی، محققان و کاربران دیدگاه‌های بسیار متفاوتی در مورد اینکه چه نوع

شواهدی قابل‌اعتماد هستند داشته(Henrik, 2014). بنابراین چالش اصلی این است که چگونه می توان از دانش عملی مدیران و تجربه کاربران استفاده کرد.

باید گفت، تصمیمات سیاستگذاری در دنیای واقعی از مدل‌های تحلیلی – تجربی تبعیت نمی‌کنند، بلکه از سیاست ها، قضاوت ها و دانش عملی حاصل می‌شوند. فرآیند سیاست گذاری حتی در کشورهای دموکراتیک نیز از عباراتی مانند حل مساله منطقی و ... استفاده می‌کند، اما در واقع فرآیند سیاست گذاری در این کشورها نیز گنگ، سیاسی و گاهی متناقض با پارادایم های اصلی است. به جای رسیدن به «راه‌حل‌ها» که می توان آن را با استانداردهای عقلانیت انجام داد، سیاست گذاری به توافق، مصالحه، تعدیل و توافق نیاز دارد که تنها می توان آن را با استانداردهایی مانند مقبولیت، تعامل، باز بودن، پاسخگویی به منافع دیگر بازیگران ارزیابی کرد(Katerina , 2013).

موضوع دیگر اجرای سیاست ها و پایبندی به این سیاست های مبتنی بر شواهد است. در اینجا، نگرانی ها به اهمیت تعهدات بلند مدت یا پایدار دولتی هم برای اجرای سیاست ها و هم برای فرآیندهای ارزیابی آن، باز می گردد. در حالت ایده‌آل، سیاستگذاری مبتنی بر شواهد در مورد راه‌حل‌های سریع برای مسائل تاکتیکی ساده نیست، بلکه مناسب مسائل پیچیده تری است که در آن‌ها یافته‌های تحقیقات و ارزیابی ها می‌توانند باعث درک الگوها و روندها در مسائل مختلف شوند. سیاستگذاران اغلب اصرار دارند که نتایج قابل‌اندازه‌گیری در یک چارچوب زمانی کوتاه در دسترس باشند، که منجر به تمرکز بیشتر بر فعالیت‌های قابل‌مشاهده به جای ایجاد مبانی برای منافع پایدار می‌شود. این تناقضی است که در خصوص تدوین و اجرای چنین سیاست هایی میان تولید کنندگان شواهد و سیاستگذاران وجود دارد.

از چالش های دیگر این است که چگونه سیاستگذاری مبتنی بر شواهد می‌تواند مشکلات بزرگ و پیچیده را تحلیل کند. در مسایل نسبتا ساده که در آن تمام متغیرها را می توان مشخص و کنترل کرد، دقت روش احتمالا مهم نیست و راحت می توان عوامل علی را مشخص کرد. اما در سیاست های با اهداف چندگانه، یا در جایی که ذی نفعان در معرض تاثیرات زیادی فراتر از دامنه سیاست ها قرار دارند، چالش درک دقیق مسئله پیچیده‌تر می‌شود(, Katerina 2013). از جمله این موارد مسائل و مشکلاتی است که دولت‌ها در استفاده از شواهد دارند. مسائلی مانند: وجود بحران یا فوریت؛ اولویت‌های سیاسی؛ تغییر ارزش‌های اجتماعی در افکار عمومی. اینها مواردی از مسائل پیچیده پیش روی سیاستگذار است که در رویکرد سیاستگذاری مبتنی بر شواهد، با مشکلات تولید شواهد مناسب و چند بعدی مواجه خواهد شد.

از سوی دیگر، برخی از عرصه‌های سیاستگذاری واگرا هستند و به شدت در آن میان ذی نفعان اختلاف نظر وجود دارد. در این حوزه‌های سیاستگذاری بحث برانگیز، ممکن است تمایل به ایجاد شفافیت نداشته باشند. چرا که شرایط

ارائه شواهد ممکن است با وجود بهترین تلاش‌های کسانی که می‌خواهند موضع «عینی» را در سیاستگذاری حفظ کنند، اما تحت‌تاثیر رویکردهای حزبی و سیاسی قرار گیرند. نمونه‌های سیاستگذاری هایی با این شرایط را می توان در حوزه های مسائل ارزشی و اخلاقی باشد. محققان که به دنبال کمک به ایجاد عینیت در چنین حوزه‌هایی هستند، ممکن است خطر به کار گیری مواضع پیشنهاد شده توسط طرفداران در یک سوی بحث و بر این اساس بی‌اعتبار شدن توسط دیگران را در پیش بگیرند(Cartwright & Hardie, 2012). این از چالش های عمده ای است که تولید کنندگان شواهد را تهدید می کند.

پژوهشگران معتقدند، سه نوع چالش اصلی برای به کار گیری رویکرد سیاستگذاری مبتنی بر شواهد وجود دارد. یکی از آن‌ها ناشی از ماهیت ذاتا سیاسی و ارزش–محور سیاستگذاری و تصمیم‌گیری است. تصمیمات سیاستگذار در درجه اول از حقایق و مدل‌های تجربی استنتاج نمی‌شوند، بلکه از سیاست ها، قضاوت و ارزش ها استنتاج می‌شوند. دامنه سیاستگذاری ذاتا با تعامل بین حقایق، هنجارها و اقدامات مطلوب شکل می گیرد. از اینرو برخی از تنظیمات سیاستگذارانه به دلیل تعهدات دولتی در برابر ورود داده‌ها به سیاستگذاری مقاوم هستند (Cartwright & Hardie, 2012). دوم، اطلاعات به روش‌های مختلف توسط کنشگرانی که از طریق لنز های مختلف نگاه می‌کنند، درک می‌شوند و مورد استفاده قرار می‌گیرند. بر اساس این دیدگاه، بیش از یک نوع از «شواهد» وجود دارد. همه این شواهد کمک به توسعه سیاستگذاری دارند، اگر چه زمینه تصمیم‌گیری، پویا و مورد مذاکره و بازیگران کلیدی، ایجاد دیدگاه‌های مشترک را دشوار می‌سازند. سومین چالش برای ارائه یک مفهوم مشترک از شواهد در سیاستگذاری مبتنی بر شواهد این است که تنظیم شبکه‌های پیچیده سیاستگذاری، استفاده از شواهد مختلف در بخش های مختلف را دشوار ساخته است. شبکه‌های سیاستگذاری مختلف، تنوع تجربیات و بنابراین تنوع شواهد(اطلاعات، تفاسیر، اولویت‌ها و دیدگاه‌ها) را نه تنها در مورد آنچه که اثرگذار است بلکه در مورد آنچه ارزشمند و معنادار است، ارائه می‌کنند(Cartwright & Hardie, 2012). بنابراین استفاده از این شواهد مختلف، چالش های زیادی را در بر خواهد داشت. یکی از مسیرهایی که می توان از این چالش ها کاست، استفاده از شواهد مختلف در مراحل مختلف فرایند سیاستگذاری ها است.

٤.١. داده های رسانه های اجتماعی

رسانه های اجتماعی با ویژگی اصلی خود نسبت به دیگر رسانه ها یعنی ایجاد بستر تولید محتوا و توزیع توسط کاربران شناخته می شوند(Kaplan, 2018). به دلیل رشد و تعدد رسانه های اجتماعی، آنها را در قالب دسته بندی های مختلفی مشخص ساخته اند. یکی از این دسته بندی ها که انواع مختلف رسانه اجتماعی را به شش دسته جداگانه تقسیم کرده

است، کار کاپلان و مازورک(۲۰۱۸) است. آنها با کاربرد مفاهیمی از حوزه علوم ارتباطات و رسانه (غنای رسانه‌ها [۹]، خودافشاگری [۱۰]، خودمعرفی[۱۱]، حضور اجتماعی[۱۲])، رسانه های اجتماعی را به شش گروه متفاوت دسته بندی کرده اند. که عبارت است از: (۱) پروژه‌های مشارکتی[۱۳] (مانند ویکیپدیا)، (۲) (میکرو) بلاگ‌ها (مانند توئیتر)، (۳) جوامع محتوایی (مانند فلیکر)، (۴) رسانه های اجتماعی (مانند فیسوک)، (۵) دنیای بازی‌های مجازی[۱۴] (مانند اورکوئست)، و (۶) دنیای اجتماعی مجازی [۱۵] (مانند سکندلایف). این دسته‌های مختلف رسانه‌های اجتماعی، از نظر شیوه خودمعرفی (مقصود طریقی است که به موجب آن، افراد قادر به کنترل تصویری می‌شوند که سایرین از ایشان می‌سازند)، خودافشاگری(مقصود از آن آشکارسازی آگاهانه یا ناآگاهانه اطلاعات شخصی ست که سازگار با وجهه‌ای است که فرد ترجیح می‌دهد از خود انتقال دهد)، غنای رسانه‌ای(نشانگر میزان اطلاعاتی ست که در یک دوره زمانی مشخص منتقل می‌شود) و از جنبه شدت حضور اجتماعی(میزانی که این رسانه ها تماس صوتی، بصری، و فیزیکی را تسهیل می‌نمایند) متفاوت هستند. جدول۱ دسته بندی ویژگی های مختلف این شش گروه از رسانه‌های اجتماعی را نشان می دهد(Kaplan, Mazurek, 2018).

جدول ۱.۱. دسته بندی رسانه های اجتماعی

		غنای رسانه ای/ حضور اجتماعی		
		کم	متوسط	زیاد
خودافشاگری	زیاد	بلاگ ها	رسانه های اجتماعی	دنیای اجتماعی مجازی
خودمعرفی	کم	پروژه های مشارکتی	جوامع محتوایی	دنیای بازی های مجازی

از طریق دسته های مختلف این رسانه ها به طور روزانه، میلیاردها داده(متن، ویدئو و ...) توسط کاربران تولید و منتشر می شود. این داده ها که دارای ویژگی هایی از جمله حجم بالا، سرعت انتشار زیاد و ... هستند بر اساس علایق و نظرات کاربران تولید می شود. این داده های تولید شده توسط کاربران، درباره موضوعات مختلف و با زمینه های متفاوتی تولید می شود. کاربران تلاش می کنند با حداقل دروازه بانی در مورد علایق، نیازها و نظرات خود در رسانه های اجتماعی، داده تولید کنند. این داده ها قابلیت ایجاد گفتمان هایی حول موضوعات مختلف داشته و همانگونه که تاکنون مشاهده شده، توانسته است تاثیرات اجتماعی، سیاستی و ... بر جوامع و حکمرانی آنها داشته باشد(Fuchs,

[9] Media Richness
[10] Self- Disclosure
[11] Self-Presentation
[12] Social Presence
[13] Collaborative Projects
[14] Virtual Game Worlds
[15] Virtual Social Worlds

2017). این داده ها دارای ویژگی های مهمی همچون زمینه ای بودن هستند. این داده ها با سرعت بسیار بالا و فوری در خصوص پدیده های مختلف توسط کاربران تولید می شود. این داده ها، نسبت به دیگر داده هایی که معمولا سیاستگذاران در فرایند سیاستگذاری به کار می گیرند، رسمیت کمتری دارد، و به صورت پراکنده توسط تعداد زیادی کاربر تولید می شود که ناشی از ویژگی هایی است که این داده ها نسبت به دیگر داده ها دارند.

همانگونه که در بخش های پیشین توضیح داده شد، شواهد در فرایند سیاستگذاری مبتنی بر شواهد، می تواند شواهدی رسمی حاصل از تحقیقات علمی باشد و همچنین می تواند شواهدی غیر رسمی که از کانال های مختلفی جمع آوری می شود، باشد. یکی از مسیرهایی که اکنون به واسطه تحول تکنولوژی در دسترس سیاستگذار برای جمع آوری شواهد میدانی است، رسانه های اجتماعی می باشند. رسانه های اجتماعی در سال های اخیر توسط دولت ها برای انتشار اطلاعات مورد پذیرش قرار گرفته است اما ارزش منابع اطلاعاتی که در قالب داده های رسانه های اجتماعی، می تواند در اختیار دولت ها قرار گیرد، کمتر شناخته شده است. داده های رسانه های اجتماعی به عنوان شواهد غیر رسمی می تواند در بخش هایی از فرایند سیاستگذاری مورد استفاده قرار گیرد. این داده ها به صورت کاملا خودجوش و آگاهانه توسط کاربران در زمینه های مختلف تولید می شود و از آنجا که برخی از زمینه های سیاستگذاری دارای پیچیدگی های میدانی زیادی است و سیاستگذار به لحاظ نو بودن آن، اطلاعات کمتری از آن زمینه دارد، در تدوین سیاست ها می تواند بسیار موثر باشد. از این منظر که اکنون شهروندان(کاربران) از ذی نفعان اصلی سیاست ها محسوب می شوند، اطلاع از تمایلات آنها در زمینه های مختلف می تواند به تدوین و اجرای سیاست ها کمک کند.

کاربران رسانه های اجتماعی هر لحظه اطلاعاتی درباره فعالیت های شخصی و حرفه ای خود، همچنین بازخورد هایی درباره سیاست های مختلف را منتشر می کنند و در بحث هایی شرکت می کنند که می تواند سیاستگذاران را راجع به موضوعات مختلف سیاستی آگاه کند. محتوای جمع سپاری شده رسانه های اجتماعی، می تواند به طور گسترده ای در توسعه سیاست گذاری ها مورد استفاده قرار گیرد اما اکنون مشخص نیست که چطور این داده های رسانه های اجتماعی، می تواند بخشی از شواهد در سیاستگذاری مبتنی بر شواهد، قرار گیرد. با توجه به پیچیدگی های فنی و سیاستی مربوط به نظارت اجتماعی، چنین شیوه های استفاده از داده های رسانه های اجتماعی در سیاستگذاری را می توان به عنوان یک نوع از جمع سپاری سیاستگذاری در نظر گرفت(Prpić, Taeihagh, & Melton, 2015) . در مقایسه با پلتفرم های جمع سپاری، جمع آوری داده ها از رسانه های اجتماعی معمولا بدون هدف و کاربرد مشخصی برای استفاده در سیاستگذاری ها انجام می شود.

سیاستگذاران به فوری بودن، مقرون به صرفه بودن و تنوع[16] داده های رسانه های اجتماعی که می تواند ورودی های خوبی باشد که از منابع آنلاین گرفته می شود، توافق دارند. اما محدودیت هایی در رابطه با داده های رسانه های اجتماعی وجود دارد شامل: میزان نمایندگی این داده ها از نظرات عموم مردم و میزان شمول[17] نظرات همه ذی نفعان سیاستگذاری، در تولید این داده های کلانی است که از این پلتفرم ها فراهم می شود. داده‌های رسانه های اجتماعی، دارای ویژگی‌های متمایزی نسبت به داده‌های موجود در پایگاه های داده سنتی که معمولا سیاستگذاران برای تصمیم گیری ها از آن استفاده می کنند، هستند. از جمله این ویژگی ها در زیر آمده است:

- داده‌های رسانه های اجتماعی حجیم هستند. در حال حاضر، به دلیل رشد نمایی داده‌های پلتفرم های رسانه اجتماعی، به سختی می‌توان حجم دقیق داده های موجود در این اکوسیستم را تخمین زد. بررسی‌ها بیانگر آنند که فیسبوک، تقریباً دارای ۲ میلیارد کاربر ماهانه فعّال است. در یوتیوب نیز بیش از ۴۰۰ ساعت محتوای بارگذاری[18] شده در دقیقه، و ۱ میلیارد ساعت محتوای مشاهده شده در روز وجود دارد. میکروبلاگ توئیتر (که سال ۲۰۰۶ راه‌اندازی گردید) دارای بیش از ۳۰۰ میلیون کاربر ماهانه فعّال[19] است که تقریباً نیم میلیارد توییت در روز منتشر می‌کنند(Kaplan, 2018). حجم عظیمی داده‌های رسانه های اجتماعی، مدیریت درست این داده‌ها را از طریق تکنیک‌های سنتی پایگاه داده دشوار می‌کند.
- داده‌های رسانه های اجتماعی، پراکنده و ناهمگن است. با توجه به خاصیت اصلی رسانه های اجتماعی که اتصال گره‌های مختلف از طریق این پلتفرم هاست، داده‌های رسانه های اجتماعی معمولاً در طیف گسترده‌ای از ابزارهای شخصی و سرورها توزیع می‌شوند، که در نقاط مختلف دنیا قرار دارند. ضمنا، داده‌های رسانه های اجتماعی، غالباً ماهیت ذاتی چندرسانه‌ای دارند، یعنی علاوه بر داده های متنی، که بیشتر برای بیان نظرات استفاده می‌شود، بسیاری از انواع دیگر داده ها، مانند تصاویر، فایل‌های صوتی و ویدیویی اغلب در رسانه های اجتماعی وجود دارد. از اینرو پردازش این داده‌ها و توانایی مقابله با ناهمگونی داده‌های چندرسانه‌ای به تکنیک‌های توسعه یافته تری از تحلیل داده، نیاز دارد(Low & Brown, 2016).
- داده‌های رسانه های اجتماعی بدون ساختار است. تاکنون هیچ ساختار داده ای ثابت و یکپارچه ای و یا مدل‌هایی که صفحات رسانه های اجتماعی دقیقا از آن پیروی کند، وجود ندارد، که اینها خود الزامات رایجی در مدیریت داده‌ها است. درعوض، طراحان رسانه های اجتماعی قادر هستند تا به طور دلخواه، داده های درون این رسانه ها را، با روش ها و الگوریتم های خود ترتیب دهند. ترتیب و روش های دسته بندی

[16] Immediacy, Cost-Effectiveness and Diversity
[17] Representation and Inclusion
[18] Content Uploaded
[19] Monthly Active Users

داده ها، در هر کدام از این اپلیکیشن های رسانه های اجتماعی از الزامات فنی و خلاقانه همان اپلیکیشن پیروی می کند. مثلا برخی از آنها باز بوده و می توان با استفاده از لینک ها به دیگر رسانه ها دسترسی داشت(توئیتر، فیسبوک و ...) و برخی از آنها بسته بوده و نمی توان از طریق لینک ها با دیگر رسانه های اجتماعی ارتباط برقرار کرد(اینستاگرام) و داده ها در همان رسانه اجتماعی مدفون می شود. این مؤلفه‌های ساختاری رسانه های اجتماعی، در درجه اول می‌توانند به جای آشکار کردن معانی موجود در داده، از کیفیت ارائه کلان داده ها سود برند(Napoli, 2016). در نتیجه، برای مقابله بهتر با ماهیت بدون ساختار این داده ها و استخراج روابط متقابل پنهان شده در این داده‌ها، برای تسهیل کاربران برای یافتن اطلاعات یا خدمات مورد نیاز در آن، اکنون نیاز فزاینده ای وجود دارد.

- داده های رسانه های اجتماعی پویاست. ساختار ضمنی و صریح داده‌های رسانه های اجتماعی، مدام به روز می‌شود. به خصوص، با توجه به نیازها و دسترسی های کاربران، انواع ابزارهای انتشار داده ها توسط این رسانه ها، مدام تغییر می کند. این ویژگی منجر به تغییراتی مکرر در ابزارهای تحلیل داده در رسانه های اجتماعی می شود، که اغلب متفاوت از بازیابی سنتی داده ها در پایگاههای اطلاعاتی است.

ویژگی‌های فوق نشان می‌دهد که داده‌های رسانه های اجتماعی، نوع خاصی از داده‌ها هستند که متفاوت از داده‌های موجود در سیستم‌های سنتی پایگاه داده ای است. در نتیجه، نیاز فزاینده‌ای برای توسعه تکنیک‌های پیشرفته‌تر تحلیل این داده ها جهت تبدیل آن به شواهد مناسب برای استفاده در فرایند های سیاستگذاری مبتنی بر شواهد وجود دارد. ابزارهای مختلفی جهت جمع آوری و تحلیل و مصور سازی محتوای رسانه های اجتماعی هر روزه توسعه پیدا می کند (Bekkers, Edwards, & de Kool, 2013; Panagiotopoulos, Shan, Barnett, Regan, & McConnon, 2015; Williams et al., 2013).

پاناگاپلوس، باون و بروکر(۲۰۱۷)، در قالب پژوهشی که در خصوص تعیین ارزش داده های رسانه های اجتماعی به عنوان شواهد سیاستی از منظر سیاستگذاران انجام دادند، ویژگی های داده های رسانه های اجتماعی در مقابل منابع سنتی شواهد را از منظر چارچوب قابلیت های جمعی، معرفی کردند. در زیر، جدول۲ این ویژگی ها را بر اساس چارچوب یاد شده نشان می دهد.

جدول ۲.۱. منابع سنتی داده در مقابل داده های رسانه های اجتماعی در سیاستگذاری مبتنی بر شواهد

	منابع سنتی شواهد	داده های رسانه های اجتماعی
اکتساب مخاطب	روشهای جمع آوری داده مانند نظرسنجی ها، گروه های کانونی، مصاحبه با خبرگان و غیره	رویکردهای مبتنی بر استفاده از کلمات کلیدی برای جمع آوری داده و تحلیل

		شبکه برای آماده سازی محتوا استفاده می شود.
	نمونه گیری و فرایند انتخاب مشارکت کنندگان مشخص است	نمونه گیری مبتنی بر مشارکت و درگیرشدن خود کاربران صورت می گیرد.
	اطلاعات مشارکت کنندگان به طور مستقیم از آنها دریافت می شود.	مخاطب با توجه به فراداده موجود هر پلتفرم رسانه اجتماعی در مورد کاربران، ساخته می شود.
	شرکت کنندگان موافقت و مشارکت فعال در تولید داده ها دارند.	شرکت کنندگان منفعلانه و با رضایت ضمنی[۲۰] مشارکت می کنند.
	ورودی داده ها بسته به روش جمع آوری آن، می تواند خصوصی یا عمومی باشد.	داده ها از فضاهای عمومی دیجیتال، تأمین می شود.
مقیاس پذیری ورودی	محدودیت های هزینه و زمان در جمع آوری بازخوردهای مشارکت کنندگان وجود دارد.	داده ها در لحظه در دسترس است.
	تأمین مجموعه های اطلاعاتی بزرگتر، معمولاً با صرف منابع زیادی امکان پذیر است.	تامین داده ها صرف نظر از تعداد مشارکت ها تا حد زیادی از ویژگی صرفه به مقیاس[۲۱] سود می برد.
قابلیت های ارتباطی	روابط پایدار و طولانی مدت با ذی نفعان سیاست ها وجود دارد.	روابط میان تولید کنندگان شواهد با پویایی بیشتری در شبکه های دیجیتال ایجاد می شود.

[20] Implicit Consent
[21] Scalable

	تأثیرگذاران سنتی بر یک حوزه سیاستی، شناخته شده و رفتارهای آنها قابل پیش بینی است.	ایجاد تأثیرگذاران و نقاط کانونی، بیشتر به روند ها و موضوعات مختلف در شبکه، وابسته اند.
روش های ادغام	از طریق تجزیه و تحلیل آماری، تجزیه و تحلیل های اقتصاد سنجی و روشهای علوم اجتماعی، شواهد «سخت» جمع آوری می شود.	با استفاده از تجزیه و تحلیل درگیری کاربر، متن کاوی و سایر تکنیک های تحلیل مبتنی بر شبکه، «شواهد نرم» جمع آوری می شود.

جدول بالا به وضوح نشان می دهد که در مقایسه با روشهای ثابت و سنتین پیشی، پیچیدگی های فنی تجزیه و تحلیل داده های رسانه های اجتماعی به انواع جدیدی از قابلیت های علوم داده(به عنوان مثال ابزارها و مهارت های ادغام/دستکاری مجموعه های کلان داده) نیاز دارد. به عنوان مثال، روش هایی مانند تجزیه و تحلیل هشتگ ها، به طور کلی در تحقیقات علوم اجتماعی ایجاد شده اند(به عنوان مثال برنت و همکاران، ۲۰۱۲ ؛ هیفیلد و همکاران ، ۲۰۱۳) اما برای فعالیت های تجاری یا سیاستی که تمرکز آنها روی خلاصه آمار ها و اندازه گیری های وایرال شده ها است، اعمال نشده است. این موضوع پیچیدگی استفاده از این داده ها در سیاستگذاری ها را نشان می دهد که نیازمند تحلیل های متفاوتی نسبت به دیگر زمینه های استفاده از داده های رسانه های اجتماعی است.

مطالعات مختلفی منافع تحلیل و مصورسازی این نوع جدید از تعاملات الکترونیک و اینکه دولت ها چگونه می توانند از آن استفاده کنند را مورد توجه قرار داده اند (Mergel & Bretschneider, 2013) . روند رصد و استخراج این تعاملات دیجیتال، می تواند برای فرایند های مختلف سیاستگذاری مفید باشد اما ممکن است با نیازهای سیاست گذاران برای توسعه رویکردهای مبتنی بر شواهد همسو نباشد. (Cartwright & Hardie, 2012; Head, 2008; Parsons, 2002; Sanderson, 2002). این بدان دلیل است که داده های رسانه های اجتماعی با منابع فعلی و رسمی شواهد که از عموم شهروندان و ذی نفعان در قالب نظرسنجی ها و مشاوره های تخصصی جمع آوری می شود، متفاوت است. علاوه بر این، استفاده از ارزش داده های رسانه های اجتماعی برای سیاست گذاران، نیازمند ایجاد قابلیت های جدیدی برای حل چالش هایی مانند درک چگونگی تکامل گفتمان های اجتماعی توسط ذینفعان اصلی در این فضا دارد (Castelló, Morsing, & Schultz, 2013; Janssen & Helbig, 2016). از اینرو نیازمند حل چالش هایی هستیم که بتوان داده های رسانه های اجتماعی را به درستی تحلیل و در جای درست خود در سیاستگذاری ها مورد استفاده قرار داد.

۵.۱. داده های رسانه های اجتماعی به عنوان منبع عمومی

اکنون نیازمند تبیین این موضوع هستیم که آیا می توان از داده هایی که کاربران در رسانه های اجتماعی و از جمله توئیتر تولید می کنند، جهت پیشبرد هدفی خاص(سیاستگذاری) استفاده کرد. با توجه به محدودیت هایی که بر اساس مفاهیمی مانند حریم خصوصی و محافظت از داده ها برای استفاده از داده های کاربران وجود دارد، چگونه می توان از این داده ها بدون کسب اجازه از کاربران استفاده کرد. برای این موضوع نیازمند فراهم آوری پشتیبان های نظری هستیم.

از زمانی که اخبار مربوط به انتشار داده های خصوصی افراد از طریق فیس بوک در سال ۲۰۱۴، منتشر شد، طی شش سال گذشته مفهومی، که آنانی و گیلسپی(۲۰۱۷) آن را «شوک عمومی[۲۲]» مربوط به پلتفرم های رسانه اجتماعی نامیدند، معرفی شده است. منظور از مفهوم شوک عمومی، تاثیر مهم عملکرد این پلتفرم های بظاهر خصوصی در زندگی عمومی مردم است که می تواند زندگی آنها را مختل کند، چرا که این پلتفرم ها و رسانه های اجتماعی به شکل گسترده ای بستر فعالیت های عموم مردم شده اند. شاید برجسته‌ترین شوک عمومی در سال ۲۰۱۶ بود که از طریق رسانه‌های اجتماعی در انتخابات ریاست‌جمهوری آمریکا مداخله ای آشکار صورت گرفت. این رسوایی مربوط به، واگذاری داده‌های کاربران از سوی فیس بوک به شرکت کمبریج آنالایتیکا بود که بعد از انتخابات افشا شد از این طریق، مداخله ای در نظرات کاربران و در نتیجه در نتایج انتخابات صورت گرفته است(Napoli, 2019). در سال ۲۰۱۹ نیز، استفاده از پلتفرم‌های رسانه اجتماعی به وسیله یک سفید پوست نژاد پرست برای قتل عام دسته‌جمعی در یکی از مساجد نیوزلند باعث ایجاد یک شوک عمومی دیگر گردید و افزایش نگرانی‌ها در مورد عملکرد پلتفرم‌های رسانه اجتماعی را بیشتر کرد(Cheng, 2019).

این رویدادها بحث های گسترده ای در جهان بر سر نیاز به نظارت بیشتر بر داده های پلتفرم های رسانه اجتماعی را برانگیخته است (Applebaum , 2019 ; Hughes , 2019) . بحث هایی در رابطه با حریم خصوصی و امنیت داده‌های کاربر تا بحث هایی که در مورد امکان اثرگذاری تعصبات سیاسی بر عملکرد پلتفرم های رسانه های اجتماعی است، از این دست بوده اند (ناپولی , ۲۰۱۹). به هر حال در دو سه سال اخیر، شاهد جنبش های نو ظهوری با عنوان «عدالت داده» و درخواست برای قرار دادن مسایل داده‌های کاربر در چارچوب مسئولیت اجتماعی بوده ایم. بر اساس این خواست عمومی، تلاش های محققان سیاستگذاری و سیاستگذاران در تعیین شیوه های استفاده از داده های رسانه های اجتماعی، به طور چشم گیری افزایش یافته است.

[22] Public Shock

بدون شک، دوره های گذار و تحول در حوزه ارتباطات و رسانه باعث ایجاد بی‌ثباتی در فرایند کار آنها می‌شود. در چنین زمانی لازم است سیاست گذاران و محققان سیاستگذاری به دنبال پارادایم‌های مناسبی باشند که حوزه ارتباطات و رسانه را به تعادلی که در راستای اهداف اجتماعی باشد، بازگردانند. استفاده از داده های رسانه های اجتماعی نیز از این دست است که اکنون سیاستگذاران در پی ایجاد مبناهای قانونی و سیاستی هستند که چگونه آن را به بهترین شکل تنظیم کنند.

در رویکردهای تنظیمی در حوزه رسانه ها، یکی از منظق های مورد استفاده، منطق ضرورت های ایجاد شده بر اساس ویژگی های تکنولوژی های جدید است که سیاستگذار با تغییرات تکنولوژیکی به دنبال تنظیم مقررات در این حوزه بوده است. اما شاید متداول ترین منطق مورد استفاده برای سیاستگذاری و تنظیم حوزه رسانه ها، استفاده از منطق «منابع عمومی کمیاب» بوده باشد که در رابطه با رسانه های مختلف به دلیل تاثیرگذاری و فراگیری گسترده شان مورد استفاده قرار گرفته است. اخیرا نیز در رابطه با تنظیم رسانه های اجتماعی، این منطق «منابع عمومی» مورد استفاده قرار گرفته است. منطق منابع عمومی که به طور تاریخی برای رسانه های سنتی مورد استفاده قرار گرفته است، اخیرا برای رسانه های اجتماعی هم مورد نظر سیاستگذار بوده است. مفهوم سازی جمع آوری، تحلیل و استفاده از داده‌های کاربر به عنوان یک منبع عمومی یک راه‌حل منطقی برای بحث در مورد این است که چگونه مفهوم حریم خصوصی/ امنیت داده ها، در استفاده از این داده های ارزشمند حفظ شود. و به اسم حریم خصوصی و یا امنیت داده ها، استفاده دولت ها و سیاستگذاران از این منبع ارزشمند محدود نشود. مفهوم سازی داده‌های کاربر با این روش نه تنها یک منطق زیربنایی برای مداخله دولت در نحوه گردآوری داده‌ها، تحلیل داده ها و استفاده از چنین داده‌هایی فراهم می‌کند، بلکه مبنایی برای تحمیل تعهدات منافع عمومی مرتبط با محتوا فراهم می‌کند. بدین معنا که اگر قرار است دولت ها از این منبع ارزشمند عمومی استفاده کنند بر اساس منطق منابع عمومی باید در مسیر حفظ و توسعه منافع عمومی این کار انجام شود و از اینرو تعهدات حفظ منافع عمومی در این حوزه برای دولت ها ایجاد می شود.

در اصل، منابع عمومی شامل دارایی های مشترک یک ملت است که به نمایندگی از طرف آنها در اختیار پادشاه یا حکمران قرار دارد. بر اساس منطق منابع عمومی، هر آنچه که متعلق به تمامی جامعه است پس منافع بهره برداری از آن نیز متعلق به همه مردم می باشد که به نمایندگی از آنها حکومت ها و دولت ها این وظیفه را بر عهده می گیرند. حکومت ها در این خصوص وظیفه دارند که مزایای بهره برداری از این منابع عمومی را در اختیار همه قرار دهند و برای مواردی استفاده کنند که نفع آن به همه به طور یکسان خواهد رسید. به عبارت دیگر، در منطق منابع عمومی، الزامات مربوط به منافع عمومی کاملا مشخص است. یک نکته کلیدی در این منطق این است که کسانی که حق استفاده از این منابع عمومی را بدست می گیرند از بخشی از آزادی های فردی و سازمانی در استفاده از این منابع، محروم می شوند و بایستی منافع آن را تماما در راستای منفعت عموم صرف نمایند(Napoli, 2019). منطق منابع

عمومی ادعا می کند که وقتی از یک منبع عمومی استفاده می کنند مانند حق انتشار محتوا که اکنون رسانه های اجتماعی از آن برخوردار هستند، باید برخی از تعهدات مربوط به منافع عمومی را که ممکن است حقوق اولیه آنان را نقض کند، نیز بپذیرند. یعنی دسترسی دولت ها برای استفاده از این داده ها به صورت کلی و نه داده های فردی، را برای استفاده در جهت منافع عمومی تسهیل کنند.

طی سال‌های گذشته سیاستگذاران در تنظیم رسانه های اجتماعی از این منطق بهره برده اند و رسانه های اجتماعی را از بهره گیری از داده های کاربران صرفا برای منافع شرکتی منع کرده اند و در این راستا محدودیت هایی تدوین کرده اند. محققانی با سابقه طولانی در مخالفت با مداخله دولت ها در حوزه رسانه ها نیز در تحقیقاتشان نتیجه گرفته اند که منطق منابع عمومی تنها منطق و قوی‌ترین منطق قابل دفاع برای نظارت دولت ها بر رسانه های اجتماعی و حفظ منافع عمومی می باشد. منطق منابع عمومی که در مقررات گذاری رسانه های سنتی هم مورد استفاده قرار گرفته بحث می کند که چطور این منطق می تواند در مورد رسانه های اجتماعی نیز به کار گرفته شده است(Zuckerberg, 2019). استدلال می شود که کل داده های کاربران(و نه داده های تک کاربر) می تواند به عنوان یک منبع عمومی مفهوم سازی شود که باعث اعمال یک چارچوب نظارتی بر مبنای منافع عمومی در پلتفرم های رسانه های اجتماعی شود که درآمد خود را از جمع آوری، تجمیع و به اشتراک گذاری داده های های کاربران کسب می کنند.

می توان گفت در این حوزه، بحث ها اغلب به اینکه آیا منطق پذیرفته شده ای برای استفاده از داده های رسانه های اجتماعی ایجاد شده است یا نه، باز می گردد(Turow & Couldry, 2018). این موضوع می تواند به بحث ما در مورد استفاده سیاستگذار دولتی از داده های رسانه های اجتماعی به عنوان شواهد، کمک کند. ایده این است که، اگر داده های تجمیع شده کابران در رسانه های اجتماعی را بر اساس منطق منابع عمومی، یک منبع عمومی به حساب بیاوریم، استفاده از آن در اختیار دولت ها قرار می گیرد. دولت ها ملزم به استفاده از این داده ها در راستای منافع عمومی هستند و عدم استفاده و یا استفاده نامناسب می تواند دولت ها را متهم به هدر دادن منابع عمومی کند. در این راستا پژوهشگر معتقد است، تولید شواهد از طریق جمع آوری، تحلیل و دسته بندی داده های کاربران، و استفاده از این شواهد در راستای سیاستگذاری می تواند یکی از بهترین کاربردهایی باشد که سیاستگذار دولتی از این منابع عمومی در راستای منافع عمومی کرده است. داده های رسانه های اجتماعی به عنوان یک منبع عمومی که بازتاب دهنده علایق و نیازهای شهروندان است، می تواند به عنوان دسته ای از شواهد که می تواند به سیاستگذار در تصمیم گیری ها کمک کند، وارد فرایند سیاستگذاری شود. اما اینکه چگونه می توان از این داده ها در این راستا بهره برد، نیازمند بحث در مورد چگونگی تحلیل این داده ها است که به بهترین شکل بتواند در فرایند سیاستگذاری، مورد استفاده قرار گیرد.

٦.١. چارچوب مفهومی قابلیت های جمعی و نیازهای سیاستگذاران

اکثر خبرگان سیاستگذاری – حتی کسانی که تجارب حرفه ای آنها با رسانه های اجتماعی محدود بوده است – بر این باورند که رسانه های اجتماعی از ارزش و اعتبار بالقوه ای در سیاستگذاری ها برخوردارند(Panagiotopoulosa, Bowena ,Brooker, 2017). با توجه به جذابیت داده های رسانه های اجتماعی به عنوان منابع اطلاعاتی باز و حجیم، یکی از راه های مقابله با عدم استفاده از این نوع داده ها در سیاستگذاری، بررسی امکان استفاده از رویکردهای سیاستگذاری سازگار با این موضوع است.

استفاده از داده های رسانه های اجتماعی در سیاستگذاری می تواند با منطق استفاده از ارزش هایی که اجتماعات می توانند بیافرینند، توجیه شود. برای پاسخ به این موضوع، از چارچوب مفهومی قابلیت های جمعی از Prpić, Taeihagh, et al. (2015)می توان استفاده کرد که خود از یک کار گسترده تر بر روی قابلیت های جذب[۲۳] از (Culnan et al., 2010; Zahra & George, 2002) گرفته شده است. قابلیت های جمعی اشاره به توانایی نهادها برای جمع آوری گسترده مشارکت مخاطبان دیجیتال ناشناس(تملک) و توسعه یک جریان اطلاعاتی مناسب برای استفاده در اهداف سیاستگذاری (ادغام) است. کدام قابلیت های داده های رسانه های اجتماعی می تواند نیاز سیاست گذاران را برای توسعه ورودی های سیاستگذاری پشتیبانی کند؟ با فهم اینکه چطور سیاستگذاران تملک و ادغام بالقوه داده های رسانه های اجتماعی را درک می کنند، می توان میان قابلیت های جذب شده از طریق داده های رسانه های اجتماعی لینک برقرار کرده است و از این لنز تئوریک در خصوص تبیین مسئله استفاده کرد.

در پاسخ به نیاز سیاستگذاران و ضرورت استفاده از داده های رسانه های جمعی، مفهوم قابلیت های جمعی[۲۴] یک رویکرد جدید برای هماهنگ کردن تعاملات دیجیتال(از جمله فعالیت های کاربران در رسانه های اجتماعی) با فرایندهای سیاستگذاری است.(Prpic & Shukla, 2014; Prpić, Shukla, Kietzmann, & McCarthy, 2015) که محققان به آن اشاره داشته اند. Prpić، Taeihagh و دیگران (۲۰۱۵) قابلیت های جمعی را به عنوان بخش اصلی سیاستگذاری در به دست آوردن سرمایه جمعی یا منابع جدید از طریق جمع سپاری، قاب بندی می کنند. در واقع، قابلیت های جمعی به تصمیمات یک نهاد در مورد چگونگی «(۱) به دست آوردن منابع پراکنده از یک اجتماع و (۲) همسو کردن مشارکت های جمعی آنها با نیازهای مورد نظر» اشاره دارد (Prpić, Shukla, et al., 2015, p. 81). این دو مرحله متمایز، از دو عنصر ضروری و مکمل در مفهوم قابلیت های جمعی که توسط زاهرا و جوزج(۲۰۰۲) شناسایی و ارائه شد یعنی: تملک[۲۵] و ادغام[۲۶] دانش خارجی با نیازهای داخلی سازمان ها و نهادها، گرفته شده است.

---

[23] Absorptive Capacities

[24]

[25]

[26]

در اینجا و در زمینه سیاستگذاری مبتنی بر شواهد، تملک به فعالیت های جمع آوری محتوای رسانه های اجتماعی اشاره دارد که دو چالش پیش رو دارد: (۱) درک انواع تعاملات دیجیتال در رسانه های اجتماعی که می توانند بالقوه مفید باشند (به عنوان مثال کاربران مختلف، پلتفرم های متفاوت، تکرار و دامنه رصد ها)، اما نیازمند کاری فراتر از اندازه گیری تأثیرات اجتماعی یا نظارت منظم بر روی مجموعه ای از کلمات کلیدی است و (۲) انتخاب و توسعه زیرساخت مناسب برای تامین محتوا و توانایی ایجاد جریان ورودی اطلاعات برای سیاستگذار (مانند ابزارهای رصد و نظارت رسانه های اجتماعی).

ادغام نیز به فرآیندهای تغییراتی گفته می شود که در آن تعاملات دیجیتال کاربران جمع آوری شده، فیلتر، مصور و یا به شکل دیگر تبدیل می شوند تا جریان اطلاعات مفیدی برای اهداف مورد نظر امکان پذیر شود. در زمینه سیاستگذاری، این مربوط به ترکیبی از تجزیه و تحلیل های داده های رسانه های اجتماعی به عنوان ورودی به مراحل مختلف سیاست گذاری است. این فرایندهای جمع بندی و استفاده از مجموعه کلان داده ها برای پاسخ به یک سوال خاص(به عنوان مثال قطبیت افکار عمومی در مورد یک سیاست جدید)، همیشه آسان نیست. همانطور که Prpić، Taeihagh و دیگران (۲۰۱۵) توجه داده اند، برخی از اشکال مشارکت مردم نیاز به تجمیع مشارکت های جمعی دارند در حالی که برخی دیگر بیشتر تحلیل محور هستند و به چندین سطح از فیلتر و پالایش، نیاز دارند.

تجزیه و تحلیل ها به این واقعیت اشاره می کند که کسب مقادیر زیادی داده از رسانه های اجتماعی همیشه به معنای تولید «اجتماع[۲۷]» ای نیست که بتواند بازخورد مفیدی در تدوین سیاست ها ارائه دهد. اگرچه این یک نتیجه گیری ثابت شده از تحقیقات در حوزه کلان داده ها است(به عنوان مثال کار بوید و کراوفورد، ۲۰۱۲؛ شنتلر و کولکارنی، ۲۰۱۴) ، مطالعات دیگر نیز یافته هایی در مورد چگونگی تفسیر این محدودیت ها توسط سیاست گذاران، ارائه می دهد. از منظر ادغام، مشارکت های جمع سپاری شده هنگامی که به طور انعطاف پذیر و مناسبی، خلاصه و تحلیل شوند، می توانند بینش های مفیدی راجع به یک موضوع ارائه دهند.

در مقیاس گسترده تر، تامین نیازهای سیاست گذاران با استفاده از دو روش اکتساب و ادغام محتوای رسانه های اجتماعی، به الزام توسعه ابزارهایی پیشرفته جهت بازخورد ویژگی های[۲۸] این شبکه ها، نیاز دارد. این ویژگی ها از نیاز به درک چگونگی تکامل شبکه ها در رسانه های اجتماعی هنگامی که گروه های مختلف کاربر درگیر بحث هستند، و انعکاس وضعیت اجتماعات خاص، بوجود می آیند. چنین ویژگی هایی می تواند مکمل جستجوهای مبتنی بر

27

28

کلمات کلیدی موجود باشد که نشان می دهد چگونه کاربران رسانه های اجتماعی در مورد مسائل مربوط به کار دولت ها تعامل دارند(اما نه اینکه تمرکز آنها صرفا بر نحوه «لایک» یا «اشتراک» محتوا باشد).

به عنوان یک مفهوم نظری، قابلیت های جمعی توجه را به تامین منابع مورد نیاز برای تهیه و ترکیب داده های مخاطبان دیجیتال، جلب می کند. فرایندهای اکتساب و ادغام به عنوان فرایندهای جداگانه اما بهم پیوسته با همدیگر نشان می دهد که نظارت اجتماعی چیزی فراتر از یک مشکل فنی خزش، استخراج و جمع آوری داده ها از منابع ممکن است. در مقابل، این موضوع چالشی استراتژیک است که به عنوان نیازمندی های توسعه توانایی های سیاستگذاران در نظر گرفته می شود(به عنوان مثال افزایش مهارت های علم داده، ابزارها و روش های جدید تجزیه و تحلیل داده، سازگاری آن با فرایندهای سیاست گذاری). تجزیه و تحلیل قابلیت های جمعی با گسترش کارهای قبلی در مورد ظرفیت جذب و فرآیند های ایجاد ارزش برای سیاستگذار، پیوندی با ادبیات پیشین برقرار می کند (Culnan et al.، 2010؛ Ooms et al.، 2015؛ Roberts et al.، 2012).

قابلیت های جمعی نقطه شروع مدل های تحلیلی مربوط به اپلیکیشن های جمع سپاری انبوه است که می تواند در زمینه بخش دولتی نیز بیشتر توسعه یافته و شرح داده شود (به عنوان مثال Chiu ، Liang، Turban &، 2014؛ Estelles-Arolas & Gonzalez-Ladron-de-Guevara، 2012). در این میان، قابلیت های جمعی به تملک و ادغام بستگی دارد، که دو مولفه اول شناسایی شده توسط زاهرا و جورج (۲۰۰۲) هستند - اما سایر موارد شناسایی شده شامل: تحول[۲۹](تغییر شکل فرایندهای موجود برای تطبیق دانش جدید) و بهره برداری[۳۰](ایجاد صلاحیت های جدید از دانش جذب شده) است. انتقال از ادغام به تحول و در نهایت بهره برداری[۳۱] می تواند با عمق بیشتری مورد تحلیل و بحث قرار گیرد(به عنوان مثال، ارتباط داده های رسانه های اجتماعی در توسعه توانایی های ارتباطی و تعاملی در فرایند سیاستگذاری). همچنین است رابطه بین محتوای رسانه های اجتماعی و مخاطبان که قبلاً به عنوان یک فرایند نظارت و جذب دیده شده است [۳۲] (Panagiotopoulos et al, 2015).

در همه موضوعات سیاستی، ممکن است به طور بالقوه، دیدگاههای انبوه وجود نداشته باشد یا یکنواختی کاربران در یک موضوع سیاستی وجود نداشته باشد. به همین ترتیب، در سایر مباحث مربوط به سیاست ها، افراد جامعه ممکن است انگیزه هایی داشته باشند تا حجم زیادی از محتوا را در رسانه های اجتماعی، به ویژه برای مسائل حساس و دو قطبی ارائه دهند. در واقع در یک موضوع سیاستی باید توجه کرد که کاربران: (۱) چقدر کانالهایی مانند توئیتر را برای بحث در مورد منافع حرفه ای خودشان ارزشمند می دانند و (۲) میزان تمایل آنها به مشارکت در تولید داده جهت

[29] Transformation
[30] Exploitation
[31] Assimilation to Transformation and Eventually Exploitation
[32] Iterative Process of Monitoring and Engaging

استفاده در فرایند ها سیاستگذاری، چقدر است؟ این سوالات می تواند مسیر را برای استفاده و یا عدم استفاده از این داده ها در سیاستگذاری ها برای سیاستگذاران، روشن کند.

۷.۱. تحلیل داده های رسانه های اجتماعی

رسانه‌های اجتماعی نتیجه ظهور تکنولوژی و ایدئولوژی وب ۲ هستند که دسترسی به داده ها و کانال های انتشار داده را بهترین روش هم آفرینی در پیشبرد پروژه های مشارکتی می داند. این رسانه‌ها کاربران را قادر می‌سازند که در هر جا و هر زمانی بتوانند باهم در ارتباط باشند(ایروانچی‌زاده،۱۳۹۴). این رسانه‌ها باعث فراهم کردن بستری هستند که در آن رفتار افراد را در مقیاسی بی‌سابقه، بدون هیچ حد و مرز جغرافیایی، با دید و رویکردهای جدید، قابل مشاهده است(Liu, et al, 2014). با رشد سریع و انفجاری این داده ها، رسانه های اجتماعی به یک بستر قدرتمند برای ذخیره، انتشار و بازیابی داده ها و همچنین به منبع دانشی مفید تبدیل شده است. با توجه به ماهیت بسیار زیاد، متنوع، پویا و بدون ساختار داده‌های رسانه های اجتماعی، بهره برداری از این داده‌های با چالش‌های زیادی از جمله ساختار ناهمگن، مسائل مربوط به محل اقامت و قابلیت مقیاس پذیری و غیره روبرو است. در نتیجه، کاربران رسانه های اجتماعی همیشه غرق در اقیانوس اطلاعات و با مشکل اضافه بار اطلاعات هنگام تعامل در این رسانه ها مواجه هستند. به طور معمول، مشکلات زیر اغلب در تحقیقات و کاربردهای داده های رسانه های اجتماعی وجود دارد:

۱) یافتن اطلاعات مرتبط: کاربر برای یافتن اطلاعات خاص در رسانه های اجتماعی، دیگر مانند وب ۱ و سایت های اینترنتی دسترسی به موتورهای جستجو برای بازیابی به اطلاعات ندارد و در واقع پلتفرم های رسانه های اجتماعی، سیاهچاله های اطلاعات هستند(خواجه ئیان، ۱۳۹۸). حجم این داده ها و از سوی دیگر عدم دسترسی کامل به پیشینه این داده های برای کاربران، مسئله از دست دادن داده های ارزشمند گسترده ای را ایجاد خواهد کرد.

۲) یافتن اطلاعات مورد نیاز: از آنجا که بسیاری از جستجو های تعریف شده در این پلتفرم ها به روشی پرس و جو را انجام می‌دهند که عمدتاً بر اساس یک کلمه کلیدی یا چندین کلمه کلیدی وارد شده انجام می‌شود. بعضی اوقات نتایج بدست آمده دقیقاً مطابق با آنچه کاربر به دلیل وجود هموگرافی (کلمات با ظاهر یکسان ولی معنای متفاوت) نیاز دارد، مطابقت ندارد.

۳) یادگیری دانش مفید: در سرویس های جستجو، نتایج مربوط به جستجو در قالب یک لیست از صفحات رتبه‌بندی شده به کاربران باز می‌گردد. در بعضی موارد، کاربران نیاز دارند که نه تنها مجموعه برگشتی صفحات وب را مرور

کنند بلکه دانش بالقوه مفیدی را نیز از آنها استخراج کنند. نیازی که رویکردهای داده کاوی را در میان محققان ایجاد کرد.

۴) شخصی‌سازی اطلاعات: در حالی که کاربر در این فضا در حال تعامل است، نیازمند دریافت اطلاعات مرتبط تری با علایق و نیازهای خود است. الگوریتم های تحلیل داده ها می تواند این نیاز کاربران را مرتفع سازد.

(۵) تحلیل روابط در رسانه های‌های اجتماعی: وجود ارتباطات ذاتی در میان کاربران یک پدیده مهم و متمایز در رسانه های اجتماعی بشمار می‌آید. چنین روابطی می‌تواند در مدل های گرافیکی نمایش داده شود و این مدل ها مبنای بسیاری از تحلیل ها قرار گیرد(Missaoui, sarr, 2014).

رسانه‌های اجتماعی سهم بسزایی در تولید و انتشار داده ها دارند، از اینرو استخراج و تحلیل این داده ها از اهمیت ویژه‌ای برخوردار است. رسانه‌های اجتماعی خروجی فراتر از متن دارند، آن‌ها به محتوای گوناگونی، نظیر وب سایت‌ها، عکس‌ها، ویدئوها و سایر رسانه‌ها پیوند می‌دهند و این موضوع اهمیت آنان را افزایش می دهد (اصنافی، ۱۳۹۵). با تحلیل این داده‌های انبوه می‌توان به دانشی قابل اتکا در حوزه های تصمیم‌گیری و سیاستگذاری دست‌یافت(Roy, et al, 2015). داده‌های موجود در رسانه‌های اجتماعی قابلیت استخراج و تحلیل های متناسب با نیاز سیاستگذار را دارند. در حالی که در حوزه سیاستگذاری، استفاده از داده های رسانه‌های اجتماعی هنوز موضوعی جدید است. از اینرو تلاش برای توسعه تحلیل های مناسب جهت آماده سازی داده های رسانه های اجتماعی برای ورود آنها به فرایند سیاستگذاری، یکی از دغدغه های پژوهشی جدید محسوب می شود. یکی از پرسش های اصلی این پژوهش نیز شناسایی تحلیل هایی است که بتواند داده های رسانه های اجتماعی را برای استفاده در فرایند سیاستگذاری مناسب گرداند. تاکنون پژوهشی در خصوص قابلیت استفاده از داده های تحلیل شده رسانه های اجتماعی، به عنوان شواهد در فرایند سیاستگذاری در ایران انجام نشده است.

در سال‌های اخیر الگوریتم های گوناگونی جهت تحلیل داده های رسانه های اجتماعی ایجاد شدند. این درحالی است که امروزه رسانه‌های اجتماعی سهم بسزایی در انتقال، انتشار و ایجاد اطلاعات دارند و بهره‌مندی از آن‌ها در آینده از اهمیت ویژه ای برخوردار خواهد بود. پایگاه های داده رسانه های اجتماعی به همراه توسعه تحلیل های مناسب، بیشترین میزان کمک به شواهد محور شدن سیاستگذاری را می تواند داشته باشد. عدم دسترسی سیاستگذار به داده های رسانه های اجتماعی، می تواند منجر به سیاست هایی بر خلاف خواست ها و نیازهای شهروندان شود (Missaoui, sarr, 2014).

براساس بررسی پیشینه تحقیقات انجام شده، بیشترین حجم مطالعات در خصوص استخراج و تحلیل داده های رسانه های اجتماعی متعلق به توئیتر است(Fernando, ets, 2019; Şerban, 2019; Amador, ets, 2016). تاکنون بیشترین مطالعات در خصوص کاربرد داده های رسانه های اجتماعی در زمینه های گوناگون علوم اجتماعی، علوم سیاسی و ... بر روی رسانه اجتماعی توئیتر انجام شده است(Bik & Goldstein, 2013; Spina, 2019). قابلیت هایی نظیر سهولت، دسترس پذیری و هزینه ی پایین داده کاوی درتوئیتر نسبت به سایر شبکه های اجتماعی سبب شده است تا بیشتر پژوهش های این حوزه بر بستر رسانه های اجتماعی توئیتر انجام گیرد(Tenopir, valentine & King, 2013, Park & Keys, 2017). هم چنین براساس شواهد میدانی، کاربران فارسی‌زبان توئیتر در حال افزایش بوده و مخاطبان ایرانی در حال استفاده‌ی روزافزون از آن به عنوان بستری برای طی کردن فرآیند اجتماعی شدن و برقراری ارتباط هستند(Zhou, Bandari, Kong & Qian, 2010; Stamatelatos, ets, 2020). از این رو استفاده از رسانه اجتماعی توئیتر جهت استخراج داده ها و کاربست آن در فرایند سیاستگذاری علم و فناوری، مد نظر این پژوهش بوده است.

تجزیه و تحلیل رسانه های اجتماعی اشاره به آنالیز داده های ساخت یافته و بی ساخت رسانه های اجتماعی دارد(Gandomi & Haider, 2015). رسانه اجتماعی اصطلاح گسترده ای است که دربرگیرنده پایگاه های آنلاین متنوعی است که به کاربران اجازه می دهد تا محتوا و مضامین را ایجاد و مبادله نمایند. اگرچه، تاریخ تحقیق بر روی شبکه های اجتماعی واقعی به اوایل سال ۱۹۲۰ برمی گردد، با این وجود، تحلیل رسانه های اجتماعی حوزه در حال تولدی است که پس از ظهور Web 2.0 در اوایل سال ۲۰۰۰، پدیدار شده است. خصوصیت اصلی تحلیل گری های مدرن رسانه های اجتماعی، ماهیت داده های آن است. تحقیق در مورد تجزیه و تحلیل داده های رسانه های اجتماعی یک میان حوزه ای در بین چندین حوزه علمی مختلف شامل روان شناسی، جامعه شناسی، انسان شناسی، علوم کامپیوتر، ریاضیات، فیزیک، و اقتصاد محسوی می شود. بازاریابی، کاربرد اولیه تحلیل های رسانه های اجتماعی در سال های اخیر بوده است. این می تواند به پذیرش گسترده و رو به رشد رسانه های اجتماعی توسط مشتریان نسبت داده شود (He, Zha,& Li, 2013)، اما اکنون مصارف دیگری نیز یافته اند(Gandomi & Haider, 2015). کاربردهایی از جمله استفاده از این تحلیل ها در ارائه خدمات عمومی و تصمیم گیری های دولتی از جمله این کاربردها هستند.

تاکنون در خصوص دسته بندی تحلیل های توسعه داده شده جهت داده های رسانه های اجتماعی، پژوهش هایی انجام شده است. این پژوهش ها دو دسته عمده از تحلیل ها را در این خصوص شناسایی کرده اند. با توجه به اینکه محتوای تولید شده توسط کاربر (برای مثال، احساسات، تصاویر، ویدئوها، و ...) و روابط و تعاملات میان موجودیت های درون شبکه ها (برای مثال، مردم، سازمان ها، و محصولات) دو منبع داده ای در رسانه های اجتماعی

هستند(Saif, He, and Alani, 2017). براساس این دسته بندی، تحلیل های رسانه های اجتماعی را می توان به دو گروه دسته بندی کرد:

- تحلیل های مبتنی بر محتوا[۳۳]: تجزیه و تحلیل های مبتنی بر محتوا بر داده های منتشر شده توسط کاربران در رسانه های اجتماعی (همچون فیدبک مشترکان و شهروندان، نظرات در مورد محصولات و خدمات، تصاویر، و ویدئوها) متمرکز می شود. چنین محتوایی در رسانه های اجتماعی اغلب حجیم[۳۴]، بی ساخت[۳۵]، پیچیده[۳۶]، و دینامیک[۳۷] است. تحلیل متن، صدا، ویدئو، می تواند موجب ایجاد فهم و بینش از چنین داده هایی شود. همچنین، تکنولوژی های کلان داده موجود در رسانه های اجتماعی، می تواند برای حل چالش های تحلیل داده ها، مورد استفاده قرار گیرد. در زیر تکنیک های تحلیل مبتنی بر متن مرور شده است.

در سال های اخیر، تحلیل های گوناگونی با رویکرد مبتنی بر محتوا، جهت بهره برداری مناسب از داده های رسانه های اجتماعی توسعه یافته است. تحلیل احساسات و عقاید کاربران رسانه های اجتماعی یکی از پرکاربرد ترین این تحلیل هاست. تحلیل احساسات و عقاید کاربران رسانه های اجتماعی در ابعاد و با اهداف مختلفی مورد توجه پژوهشگران بوده است. از میان حوزه های تحلیل احساسات در رسانه های اجتماعی که بیش از همه محققان را به سمت خود جلب کرده است می توان از بازخورد نظرات کاربران توئیتر، ردیابی احساسات در طول زمان، شناسایی قطبیت (مثبت، منفی و خنثی) تشخیص طعنه و کنایه[۳۸] و تشخیص و اندازه‌گیری احساسات اشاره نمود. محققان معتقدند که توئیتر حاوی داده های ارزشمندی برای آگاهی نسبت به رفتارها و نیازهای کابران درباره موضوعات مختلف است (Giachanou and Crestani 2016) و این موضوع داده های توئیتر را جهت انجام تحلیل ها مناسب می کند.

اکنون محققان الگوریتم طبقه‌بندی احساسات موثری را طراحی و توسعه داده اند که در سطح هشتگ‌ها و صرف نظر از موضوع توئیت‌ها به تعیین قطبیت احساسی[۳۹] آنها کمک می‌کرد. یافته‌ها نشان می دهد که هشتگ‌ها دارای سه نوع اطلاعات مفید شامل قطبیت معنایی توییت‌ها؛ رابطه هم رخدادی هشتگ هاو معنای لفظی هشتگ ها هستند (Wang, Wei, Liu, Zhou, and Zhang 2018). یافته‌های تحقیقات نشان می دهد، رویکرد معنایی[۴۰] برای تجزیه و تحلیل کلان داده‌ها[۴۱] مناسب‌تر است و طیف گسترده‌ای از مباحث را پوشش می‌دهد، در حالی که رویکرد شناختی بیشتر برای مجموعه داده‌های نسبی یا مجموعه‌های موضوعی خاص مناسب است (Saif, He, and Alani, 2017).

---

[33] Content-Based Analytics
[34] Voluminous
[35] Unstructured
[36] Noisy
[37] Dynamic
[38] Irony Detection
[39] Sentiment Polarity
[40] Semantic Approach
[41] Big Data

با آن که تحقیقات زیادی درباره تحلیل توییتهای مرتبط با موضوعات در حوزه سیاستگذاری علم و فناوری انجام نگرفته است، اما توییتها درباره سیاست های علم و فناوری کم نیستند. توئیتر، علاوه بر کاربران عمومی، شمار زیادی از کاربران دانشگاهی را به خود جذب کرده است (Mohammadi et al. 2015؛ Bornmann 2014؛ Holmberg and Thelwall 2014). آلن[۴۲] (۲۰۱۸) معتقد است در دنیای تحلیل های سیاستگذاری، رسانه‌های‌اجتماعی هنوز جدید هستند. برای سیاستگذاران و محققان، چگونگی تعریف و تشخیص تحلیل های مناسب برای داده های رسانه های اجتماعی، یک چالش به شمارمی‌رود. سیاستگذاران در درک با ارزش بودن داده‌های رسانه‌های اجتماعی تردید دارند. هارور[۴۳] و رمضان‌زاده هروی (۲۰۱۵) به مطالعهٔ امکان سنجی برای تحلیل داده های رسانه‌های اجتماعی در ایرلند پرداختند. آن‌ها عقیده داشتند که بین رویدادها و هشتگ‌ها رابطه وجود دارد و برای تحلیل رویدادها می‌توان از آن‌ها بهره برد. تمرکز کار آن‌ها بر روی پست های توئیتر بوده است. نتیجه این پژوهش می تواند شواهدی در خصوص استفاده از تحلیل های مبتنی بر محتوا، از داده های رسانه های اجتماعی جهت کاربرد در سیاستگذاری را بدست دهد. لیثام[۴۴](۲۰۱۴)، ابزارهایی فنی در مواجهه با تحلیل متن داده های رسانه‌های اجتماعی را دسته بندی کرده است. این پژوهش یک طیف از راهبردهای مدیریت اطلاعات را ارائه داده است که می‌توان آن را به رسانه‌های‌اجتماعی براساس نیازهای خاص سیاستگذاری اعمال کرد.

- تحلیل های مبتنی بر ساختار[۴۵]: این نوع تجزیه و تحلیل ها با ترکیب خصوصیات ساختاری رسانه های اجتماعی و استخراج دانش و آگاهی از روابط میان کاربران آن است که اهمیت می یابد. ساختار رسانه های اجتماعی از طریق دسته ای از گره ها و یال ها شکل می گیرد که به ترتیب نشان دهنده کاربران و روابط میان آنها می باشد. این ساختار بصورت گراف ترکیب شده ای از گره ها و یال ها تصویر می شود. دو نوع گراف شبکه با نام گراف های اجتماعی[۴۶] و گراف های فعالیت[۴۷] وجود دارد (Heidemann, Klier, & Probst, 2012). در گراف های اجتماعی، یک یال میان یک جفت گره تنها به معنی وجود یک لینک (برای مثال، دوستی) میان کاربران است. چنین گراف هایی می تواند برای شناسایی ارتباطات یا تعیین مراکز فعالیت در نظر گرفته شود (یعنی کاربرانی با نسبتا تعداد زیادی از لینک های اجتماعی غیرمستقیم). با استفاده از این گراف های اجتماعی، می توان روابط میان افراد و اینکه کدام یک از کاربران به عنوان کاربر مرکزی در یک شبکه است را کشف کرد. در گراف های فعالیت، هرچند یال ها تعاملات واقعی میان هر

[42] Allen
[43] Harrower
[44] Latham
[45] Structure-Based Analytics
[46] Social Graphs
[47] Activity Graphs

جفت گره را نشـــان می دهند. اما تعاملات شـــامل تبادل اطلاعات (برای مثال، لینک ها و توضـــیحات) نیز هست. این گراف ها اطلاعات بیشتری در خصوص نوع فعالیت کاربران در اختیار می گذارند. گراف های فعالیت بر گراف های اجتماعی دارای ارجحیت هستند، زیرا یک رابطه فعال نسبت به یک ارتباط محض، بین دو کاربر وجود دارد و این برای ما جهت کسب اطلاعات بیشتر از این کاربران مناسب تر است.

تکنیک های مختلفی اخیرا برای اسـتخراج اطلاعات از سـاختار شـبکه های اجتماعی، پدید آمده اند که می توانند اطلاعات زیادی در مورد روابط کاربران در یک شبکه شکل گرفته در رسانه های اجتماعی به دست دهند. اطلاع از این شبکه ها و احتمالا کاربران مرکزی که در واقع هدایت کننده این شبکه ها هستند و بر سر موضوعات مختلفی نیز شـکل می گیرند، می تواند در تصـمیم گیری های دولتی و شـرکتی موثر باشـد. در این میان شـبکه هایی که حول موضوعات سیاستگذاری علم و فناوری شکل می گیرند، شامل اطلاعات خوبی در مورد کنشگران آن موضوع و تمایل آن شبکه در خصوص آن موضوع است. این تحلیل های مبتنی بر ساختار شبکه در ادامه مرور شده است.

- شناسایی اجتماعات[۴۸]، که بصورت کشف اجتماعات نیز به آن اشاره می شود، جوامع ضمنی در داخل شبکه را استخراج می کند. برای شبکه های اجتماعی آنلاین، اجتماع به زیر شبکه ای از کاربران اشاره می کند که با وسعت زیادی نسبت به بقیه شبکه با همدیگر در تعامل هستند. اغلب با محاسبه میلیون ها گره و یال، شبکه های اجتماعی آنلاین اندازه های بسیار بزرگی دارند. شناسایی اجتماعات به خلاصه و کوچک تر نمودن شبکه های بزرگ کمک می کند، که به آشکار سازی الگوهای رفتاری موجود در این شبکه ها و پیش بینی ویژگی های مختلف دسته های متفاوت کاربر، کمک خواهد کرد. شـناسـایی اجتماعات مانند دسـته بندی در تکنیک داده کاوی (Aggarwal, 2011)، برای تقسـیم بندی داده ها به زیرمجموعه های کوچک تر براساس شباهت ها مورد استفاده قرار می گیرد. شناسایی اجتماعات، در آغاز جهت کاربرد در بازاریابی و توسعه شبکه ارتباطی جهانی مورد استفاده قرار گرفت (Parthasarathy, Ruan, & Satuluri, 2011). برای مثال، شناسایی این اجتماعات شرکت ها را قادر می سازد تا به توسعه موثر سیستم های توصیه و پیشنهاد محصول خود به مشتریان، دست یابند. یا به سیاستگذاران دولتی کمک خواهد کرد تا دستور کارهای مهم مورد نظر این شبکه ها را شناسایی و به طور موثر بر روی آن بر اساس منافع عمومی کار کنند.
- آنالیز نفوذ اجتماعی[۴۹] به تکنیک هایی اشـاره می کند که مربوط به مدل سـازی و ارزیابی تاثیر عاملان (کنشـگران) و ارتباط ها در شـبکه های اجتماعی می شـود. بدیهی اسـت که رفتار یک عامل در شـبکه اجتماعی توسط دیگران تحت تاثیر قرار می گیرد. بنابراین، بسیار پسندیده است که تاثیر کاربران و دنبال کنندگان را ارزیابی نمود، نقاط قوت این ارتباطات را تعیین نمود، و الگوهای تاثیر انتشار در شبکه را کشف

[48] Community Detection
[49] Social Influence Analysis

کرد. گفته می شـــود تکنیک های آنالیز نفوذ اجتماعی می تواند در بازاریابی برای افزایش موثر اگاهی از برند و پذیرش، نفوذ نموده و مورد استفاده قرار گیرد. ولی بیش از همه این سیاستگذاران دولتی هستند که می توانند ترجیحات و علایق خود را از این طریق در میان طرفدارانشـــان یا جمعیت های خنثی بســـط دهند(Yan, 2018).

جنبه نمایان آنالیز نفوذ اجتماعی، تعیین کمیت اهمیت گره های موجود در درون یک شـــبکه اســـت. معیار های مختلفی برای این هدف توســـعه یافته اند، که شـــامل درجه مرکزیت[۵۰]، رابطه مرکزیت[۵۱]، مرکزیت نزدیکی[۵۲]، و مرکزیت بردار ویژه[۵۳] است(Tang & Liu, 2010). این مقیاس ها، نقاط قوت اتصالات ایجاد شـده بوسـیله یال ها یا مدل وسـعت نفوذ در شـبکه های اجتماعی را ارزیابی می کنند. مدل آسـتانه خطی[۵۴] (LTM) و مدل آبشاری مستقل[۵۵] (ICM) دو مثال معروف از چنین معیارهایی در شبکه ها هستند که قادر به پیش بینی میزان تاثیر گذاری و نفوذ افراد تاثیر گذار در این شـبکه ها حول موضـوعات مختلف اجتماعی و سیاسی هستند (Sun & Tang, 2011).

- پیش بینی لینک[۵۶] تکنیکی است که لینک های بعدی میان گره های موجود در شبکه اصلی را پیش بینی و تعیین میکند. بطور نمونه، ساختار رسانه های اجتماعی ایستا نبوده و بطور پیوسته از طریق ایجاد گره ها و یال های جدید رشد می کند، این رشد و چگونگی آن می تواند توسط تکنیک پیش بینی لینک در آینده تعیین شـود. بنابراین، بدیهی اسـت که هدف، فهم و پیش بینی دینامیک شـبکه ای اسـت که در آینده در رسـانه اجتماعی شـکل خواهد گرفت. در واقع تکنیک پیش بینی لینک، وقوع تعاملات، همکاری، یا نفوذ در میان شـــبکه ها را در فاصـــله زمانی معین، پیش بینی می کنند. تکنیک های پیش بینی لینک که با انتخاب عوامل مناسب می تواند عالی عمل کند، پیشنهاد می کند که ساختار فعلی شبکه مطمئنا دارای اطلاعات ناپیدایی در مورد لینک های آینده اســـت (Gandomi & Haider, 2015) و این اطلاعات را می توان از الان کشـــف کرده و جهت بهره برداری مناســـب از آن برنامه ریزی کرد. این شـــبکه ها که ممکن اســـت در آینده بر اثر موضـــوعات مختلف اجتماعی شـــکل بگیرد ممکن اســـت با هدایت افکار و کنش های کاربران، تاثیرات اجتماعی زیادی از خود بر جای بگذارد.

---

[50] Degree Centrality
[51] Betweenness Centrality
[52] Closeness Centrality
[53] Eigenvector Centrality
[54] Linear Threshold Model
[55] Independent Cascade Model
[56] Link Prediction

برای اولین بار از تکنیک های پیش بینی لینک در زیست شناسی و برای کشف لینک ها یا همبستگی ها در شبکه های زیستی (برای مثال شبکه های تعامل پروتئین-پروتئین)، استفاده شد. که با رفع نیاز به آزمایشات گرانقیمت، این پیش بینی را ممکن ساخت (Hasan & Zaki, 2011). پس از آن در بحث های امنیتی نیز، پیش بینی لینک به آشکار کردن همکاری های بالقوه در شبکه های تروریستی یا جنایی کمک کرد. در رابطه با رسانه های اجتماعی آنلاین، کاربردهای اولیه، پیش بینی لینک در توسعه سیستم های پیشنهاد یا توصیه گر بوده است (برای مثال، فیس بوک «افرادی که ممکن است بشناسید[۵۷]»، یوتیوب «پیشنهادی برای شما[۵۸]»، و همچنین موتورهای توصیه گر Netflix و Amazon).

مجموع تحلیل های توسعه داده شده برای داده های رسانه های اجتماعی در پیشینه تحقیقات در دست، نشان دهنده طراحی الگوریتم های تحلیل داده های رسانه های اجتماعی با دو رویکرد «تحلیل محتوای رسانه های اجتماعی» و «تحلیل ساختار شبکه های موجود در رسانه های اجتماعی» است. همانگونه که گفته شد، این دو شیوه تحلیل و تکنیک های موجود آن، می تواند داده های این بستر رسانه ای را مناسب سیاستگذاری گرداند. در بالا تکنیک هایی مطرح و مرور شد که می تواند با کمترین هزینه در اختیار سیاستگذار دولتی قرار گیرد و در حد کفایت، دانش زمینه ای مناسب در خصوص موضوعات سیاستگذاری را به او ارائه دهد.

۸.۱. رسانه های اجتماعی و سیاستگذاری

استفاده از داده های رسانه های اجتماعی و در کل کلان داده ها در سیاستگذاری برای اولین بار در کشور چین مطرح شد. اگر تاریخچه تدوین اسناد ملی درباره استفاده از کلان داده در امر سیاستگذاری را مرور کنیم به اسنادی برخواهیم خورد که از سال ۲۰۱۲ در کشورهای توسعه یافته در این خصوص تدوین شده است. در سال ۲۰۱۲، دفتر سیاستگذاری علم و فناوری کاخ سفید[۵۹]، «برنامه تحقیق و توسعه کلان داده ها» را مطرح کرد، که از آن پس کلان داده ها در همه سیاستگذاری های دولت فدرال آمریکا مورد استفاده قرار گرفت. در سال ۲۰۱۳، دولت استرالیا «راهبرد خدمات عمومی فناوری اطلاعاتی و ارتباطاتی ۲۰۱۲ تا ۲۰۱۵ استرالیا[۶۰]» را منتشر کرد. در سال ۲۰۱۳، انگلیس سند «فرصت ها برای سود: راهبرد قابلیت کلان داده انگلیس[۶۱]» را منتشر ساخت. فرانسه نیز در سال ۲۰۱۵ برای اولین بار «برنامه اقدام جهت تسهیل جمهوری دیجیتال در ۲۰۱۵[۶۲]» را راه اندازی کرد. در همان سال ۲۰۱۵، دولت چین نیز «بیانیه توسعه

[57] People You May Know
[58] Recommended for You
[59] White House Science and Technology Policy Office (OS-TP)
[60] Australian Public Service Information and Communication Technology Strategy 2012- 2015
[61] Opportunities for Profits: The UK Data Capability Strategy
[62] Action Plan to Facilitate Digital Republic In 2015

برنامه اقدامی برای ترویج توسعه کلان داده ها[۶۳]» را منتشر کرد که کلان داده های چین را به صورت ملّی مورد استفاده قرار می داد. پس از این اسناد تاکنون کشورهای مختلف در قالب اسناد سیاستی مختلف، بهره برداری از داده های رسانه های اجتماعی و به طور کلی کلان داده ها را در سیاستگذاری های ملی مورد نظر قرار داده اند(Yan, 2018).

- فرصت های استفاده از داده های رسانه های اجتماعی برای سیاستگذاری

رسانه های اجتماعی شامل اکوسیستم بزرگی از شبکه ها و سیستم های اشتراک اطلاعات با قابلیت های متنوعی هستند (Kietzmann, Hermkens, McCarthy, & Silvestre, 2011). داده های رسانه های اجتماعی نماینده مناسبی از کلان داده پیچیده و دارای رشد سریع در هر کشور هستند، که فرصت های زیادی را برای حوزه های زیادی به ارمغان آورده است. ا در عصر کلان داده ها، مدیریت داده های رسانه های اجتماعی، بسیار مهم است چون در توسعه تحقیقات و همچنین مدیریت اجتماع مفید و کمک کننده است. داده های رسانه های اجتماعی، پیش بینی های دقیق و صحیحی را امکانپذیر می کنند، اطلاعات را با منابع مختلف یکپارچه می کنند، دانش خلق می کنند، خدمات جدید تولید میکنند و می توانند منجر به ارائه تسهیلات بهتری به شهروندان شوند. داده های رسانه های اجتماعی حامل دانش و اطلاعات مهمی هستند. چنین ویژگی هایی در این داده ها می تواند پشتیبانی قدرتمندی از سیاستگذاران در ارتقا نظارت بر تعاملات دیجیتال شهروندان، ارائه دهند. ترکیب موثر داده های رسانه های اجتماعی با دانش زمینه ای سیاستگذاران، اطلاعات بیشتر و مناسب تری به دست سیاستگذار خواهد داد. سیاستگذاران با سروکار داشتن با داده های رسانه های اجتماعی می توانند هزینه های سیاستگذاری را کم کرده و در زمینه های مختلف به توسعه پایدار دست یابند. داده های رسانه های اجتماعی می تواند شفافیت سیاستگذاری را ارتقا داده، اعتبار دولت ها و سیاستگذاران را افزایش داده، عملکرد نظارت دولتی را بالاتر برده و شکاف نظارت دولتی و واقعیات جامعه را کمتر کند(Yan, 2018).

باید تاکید کرد که اقدامات مورد نیاز سیاست گذاران در رابطه با رسانه های اجتماعی، فراتر از رصد های لحظه ای افکار عمومی است. نظارت اجتماعی در دولت ها در قالب دو روند مهم پیش رفته است. نخست، سیاست گذاران در حالی که از استانداردهای رسمی بالایی در خصوص جمع آوری شواهد برای سیاستگذاری پیروی می کنند، اما به دنبال ابزارهای جایگزین، مقرون به صرفه و مستقیم تری برای توسعه سیاست ها نیز هستند. جمع سپاری سیاست ها و فعالیت های درگیرسازی شهروندان در سیاست ها به صورت دیجیتال، می تواند از این کار حمایت کند. دوم، از نظر توانایی های فنی و آگاهی از اهمیت داده های رسانه های اجتماعی، پیشرفت های چشمگیری اکنون در زمینه تجزیه و تحلیل کلان داده ها صورت گرفته است(Janssen & Helbig, 2016; Schintler & Kulkarni, 2014) .

[63] Notice on The Development of Action Plan for The Promotion of Big Data Development

- چالش های استفاده از داده های رسانه های اجتماعی در سیاستگذاری

داده های رسانه های اجتماعی علاوه بر ارزش بالقوه خود، با مجموعه ای از مسائل اخلاقی، حریم خصوصی و ... برای استفاده گسترده تر در بخش عمومی مواجه هستند(Picazo-Vela et al, 2012). در واقع برخی از اصلی ترین پیچیدگی های داده های رسانه های اجتماعی عبارت اند از : (۱) مشارکت کاربران معمولاً با اهداف مختلف و از سوی کاربران متفاوتی انجام می شود و (۲) در صورت جمع آوری مقادیر زیادی از این نوع از مشارکت ها در قالب داده های رسانه های اجتماعی، باز هم نمی توان اطمینان داشت، اکثریت ذی نفعان یک سیاست در این خصوص درگیر شده و داده تولید کرده اند(Boyd & Crawford, 2012). ایوانز و همکاران(۲۰۱۵) پیشنهاد می کنند که در خصوص تغییرات در فن آوری های تجزینه و تحلیل داده ها، تغییرات قانونگذاری و تغییر در انتظارات کاربران، ملاحظات اخلاقی باید به طور مستمر مورد بررسی قرار گیرد.

با وجود این چالش ها، ارزش داده های رسانه های اجتماعی به عنوان بخشی از چرخه سیاستگذاری و کاربرد جمع آوری شواهد از این فضا آنقدر زیاد است که نمی توان از این ارزش نهفته در رسانه های اجتماعی چشم پوشی کرد. بررسی بیشتر و حل این چالش ها مهم است زیرا رسانه های اجتماعی می توانند مقادیر زیادی داده تولید کنند که به منابع سنتی شباهت ندارند و می توانند حاوی بینش های جدیدی در مورد نحوه تفکر گروه های عمومی در مورد موضوعات سیاستی باشند. اکنون جای دارد که داده های رسانه های اجتماعی، در حالی که برای سیاستگذاری مبتنی بر شواهد مورد استفاده قرار می گیرد، چالش های آن هم مورد بحث قرار گیرد.

الف) مسائل حریم خصوصی[۶۴]. بزرگترین چالش برای دولتهایی که قصد استفاده از داده های رسانه های اجتماعی را در سیاستگذاری ها دارند، چالش حریم خصوصی کاربران صاحب داده ها می باشد. مسئله حریم خصوصی داده ها در فرایند جمع آوری، ترکیب کردن و تحلیل داده ها وجود دارد. کاربران داده های شخصی خود را منتشر کرده اند و معمولا از پتانسیل پیگیری شدن از طریق این داده ها، بی اطلاع هستند. دستیابی به استفاده مجدد از این داده ها در رسانه های اجتماعی، خطر آشکار شدن اطلاعات حساس کاربران را افزایش می دهد. که تاکنون دولت ها تلاشهایی برای مبارزه با استفاده تجاری غیر قانونی از اطلاعات فردی کاربران را داشته اند.

نگرانی های مربوط به حریم خصوصی کاربران عمدتا در رشد سریع اکوسیستم کلان داده ریشه دارد. به عنوان حقوق افرادی که داده های آنها به دیگران منتقل می شود ، مفهوم حریم خصوصی داده ها و محافظت از داده های شخصی مدتهاست که به عنوان حقوق اساسی بشر مورد توجه قرار گرفته است. در حال حاضر، به رسمیت شناختن حق اشخاص

[64] Privacy Issues

یا «اطلاعات قابل شناسایی شخصی[۶۵]» به عنوان شرط آستانه قانونی برای از دست دادن هویت یا حریم خصوصی داده ها در نظر گرفته می شود (Schwartz & Solove, 2011). با این حال، ماهیت ارتباطات دیجیتال در عصر مدرن، نیاز به بازنگری در این تعریف را آشکار ساخته است. هویت دیجیتالی یک فرد طیف گسترده ای از ویژگی های آفلاین قابل ردیابی (مثلاً سن ، محل سکونت ، درآمد و غیره) را علاوه بر انواع ویژگی های های آنلاین (گذرواژه ها ، شماره های پین ، کدهای دسترسی و رفتارها در پلتفرم ها) است و همه این موارد، ثابت می کند پیوندهای مشخصی بین درک اجتماعی و تکنولوژیکی از هویت افراد نیاز است (Wessels, 2012). مصرف کننده دیجیتال امروزی دیگر کاملاً ناشناس نیست زیرا تقریباً هر شکلی از ارتباطات و رفتارهای دیجیتال، داده هایی را تولید می کند که می توان جمع آوری، تجمیع و تجزیه و تحلیل کرد. بدین ترتیب، اطلاعات جمع آوری شده برای یک هدف، به راحتی می توانند برای اهداف دیگر بازیابی شوند، و ارتباط بین مقدار داده های جمع شده در مورد یک فرد و میزان پیش بینی رفتار او، قابل تصور است.

تحقیقات شش عامل تعیین کننده را در ایجاد حریم خصوصی داده های افراد در فضای آنلاین شناسایی می کنند. جمع آوری داده ها[۶۶]، کنترل داده ها[۶۷]، استفاده ثانویه غیر مجاز[۶۸]، دسترسی نامناسب[۶۹]، ردیابی مکان[۷۰] و آگاهی مربوط به این روش ها[۷۱]. هنگامی که داده های شخصی کاربران جمع آوری می شود، مصرف کنندگان می خواهند بر استفاده و توزیع آن کنترل کنند، به ویژه هنگامی که پتانسیل زیادی برای رفتار فرصت طلبانه و نقض قرارداد اجتماعی در یک مبادله آنلاین وجود دارد. بنابراین، نشان داده شده است که آگاهی از شیوه های حفظ حریم خصوصی مربوط به جمع آوری، کنترل، دسترسی و استفاده از داده ها، یک مولفه فعال در ایجاد مرز حریم خصوصی در بین مصرف کنندگان آنلاین ایجاد می شود و واسطه ای مهم برای ایجاد اعتماد در بین کاربران است (Eastin, Brinson, Doorey & Wilcox, 2016).

ب) استفاده ناکافی از داده های رسانه های اجتماعی[۷۲]. در این رابطه سه چالش مهم داده های رسانه های اجتماعی عبارتند از «چه داده هایی پردازش شوند، چگونه تحلیل شوند، چگونه تغییر یابند؟». موسسه مک کینسی[۷۳] اعلام کرده است که دولت ها از بیش از ۹۰ درصد داده های تولیدی فقط یک درصد آنها را تحلیل می کنند. به علاوه، «تحلیل داده ها فاقد زمینه استفاده و فاقد مدیریت سیتماتیک هستند». یعنی فرآیند تولید تحلیل داده با هدف تحقیقات همخوانی

[65] Personally-Identifiable Information
[66] Data Collection
[67] Data Control
[68] Unauthorized Secondary Use
[69] Improper Access
[70] Location Tracking
[71] Awareness Related to These Practices
[72] Inadequate Usage of Big Data
[73] Mckinsey Institute

ندارند. و همچنین «کیفیت تحلیل داده ها دقیق و پایش شده نیست». در واقع اتصال بین داده ها و مسائل باعث رسیدن به راه حل های موثر نشده است.

ج) صحت داده های جمع آوری شده توسط دولت ها[۷۴]. داده های جمع آوری شده توسط سیاستگذار دولتی مشکوک هستند. برطبق مطالعات انجام شده، گروههای اجتماعی مختلف اعتماد پایینی به این داده ها دارند؛ و به دلیل فقدان آگاهی، داده های جمع آوری شده شایسته اعتماد نیست. زنان و افرادی که به مدارک تحصیلی بالاتری دست یافته اند محتاط تر هستند و اعتماد پایینی به داده های جمع آوری شده توسط دولت ها دارند(Yan, 2018).

به دلیل جدید بودن موضوع استفاده از داده های رسانه های اجتماعی در سیاستگذاری، تاکنون پژوهش های کمی به این حوزه پرداخته اند. تحقیقات فعلی هنوز به ترکیب داده های مخاطبان دیجیتال با فرایند سیاستگذاری و اینکه چگونه این منابع می تواند به سیاست گذاران کمک کند تا بینش جدیدی داشته باشند، توجه کافی نشان نداده است. اگرچه فرصت های مشخصی برای تعامل با جوامع جدید در رسانه های اجتماعی وجود دارد (Murthy , 2015، Gross ,Pensavalle) ، ارتباط بین مخاطبان سنتی و رسانه های اجتماعی می تواند به ویژه برای نهادهای سیاستگذار چالش برانگیز باشد (Panagiotopoulos et al., 2015). سیاست گذاران در ابتدا باید فرضیاتی مناسب را در مورد این شهروندان دیجیتال بپذیرند تا بتوانند سودمندی ورود این داده های جمعی را که توسط مخاطبان رسانه های اجتماعی تولید می شود را به درستی ارزیابی کنند. ایجاد چنین فهمی از نظرات عمومی معمولاً تأثیر زیادی در رویه های سیاستگذاری دارند و به نوبه خود تغییراتی در رابطه با تصمیمات مهم سیاستی به دنبال دارند (Barnett, Burningham, Walker, & Cass, 2012; Walker, Cass, Burningham, & Barnett, 2010).

دریس، ملولی و ترابلسی[۷۵](۲۰۱۹)، با این پیش فرض که رسانه های اجتماعی می توانند کانالی میان سیاستمداران و شهروندان باشند، به ارائه چارچوبی جهت استخراج دانش از رسانه اجتماعی فیس بوک پرداخته اند. این چارچوب بر اساس ابزار تحلیل معنایی متن ارائه شده است که داده ها را از رسانه های اجتماعی جمع آوری و داده های ارزشمند را جهت ارائه به سیاستگذاران انتخاب می کند. هدف اصلی این پژوهش ارائه یک چارچوب مفهومی برای کمک به سیاستگذاران جهت استخراج دانش از داده های تولید شده توسط شهروندان در پلتفرم های رسانه های اجتماعی است تا سیاستگذاران از اینکه چه مباحثی میان شهروندان بر سر موضوعات مختلف اجتماعی و سیاسی و ... شکل گرفته آگاه شوند و بتوانند تصمیمات شفاف تری اتخاذ کنند. این محققان معتقدند با این شیوه شهروندان در سیاستگذاری سهیم خواهند شد. این پژوهش نشان می دهد یکی از ابزارهایی که می تواند به سیاستگذاران کمک کند تا سیاست هایی مبتنی بر نظرات شهروندان را تدوین کنند، رسانه های اجتماعی است. از این رو سوال اصلی این است که چگونه

[74] The Authenticity of Government Data
[75] Drissa, Mellouli, Trabelsi

می توان از داده های رسانه اجتماعی در فرایند سیاستگذاری استفاده کرد؟ که بر اساس نتایج این پژوهش داده های رسانه های اجتماعی در دو مرحله از چرخه سیاستگذاری یعنی تعریف مسئله و ارزیابی سیاست ها می تواند مورد استفاده قرار گیرد. ناپولی (۲۰۱۹)، استدلال می کند که رسانه های اجتماعی بنا بر چارچوب نظری «منابع عمومی»، یک منبع عمومی محسوب می شود که می توان و باید آن را در جهت توسعه منافع عمومی مورد استفاده قرار داد. این پژوهش استدلال می کند که داده های جمع آوری شده کاربر در رسانه های اجتماعی می تواند به عنوان یک منبع عمومی باشد که در توسعه و نظارت بر اجرای سیاست ها در راستای منافع عمومی قابل بهره برداری است. او به توسعه یک الگو در حوزه تنظیم گری از طریق کاربست داده های کاربران رسانه های اجتماعی پرداخته است. آنها معتقدند، این الگو کمک می کند تا بین استخراج و تحلیل داده های رسانه های اجتماعی و بهبود کیفیت زندگی عمومی رابطه مثبتی برقرار شود. درواقع منطق استفاده از منابع عمومی که پیش از این بر منابع سنتی در جامعه حاکم بود به مفهوم داده های رسانه های اجتماعی تعمیم داده می‌شود.

فرناندو و همکاران[۷۶](۲۰۱۹)، در پژوهش خود در پی تحلیل و مصور سازی داده های توئیتر بوده اند تا رویکردی کاربردی جهت استخراج دانش مرتبط با رویداده های سیاسی را ارائه دهند. آنها معتقدند از آنجا که استفاده از رسانه های اجتماعی در سال های اخیر بسیار رشد داشته است و اکثر کاربران در آن به خود اظهاری راجع به رویکرد سیاسی شان می پردازند، داده های تولید شده در این رسانه های اجتماعی به ویژه توئیتر می تواند جهت تحلیل در علوم سیاسی مورد استفاده قرار گیرد. در این پژوهش ابزار مصور سازی کلان داده ها مورد استفاده قرار گرفته است تا چالش استخراج دانش از این داده ها را برای محققان علوم سیاسی حل کند. پاناجوتوپولس، باون و بروکر[۷۷](۲۰۱۷)، در پژوهش خود اظهار می کنند که رسانه های اجتماعی تاکنون برای دولت ها ابزاری برای کنترل کاربران بوده است اما از این منبع داده کمتر به خاطر ارزش اطلاعاتی آن، بهره برداری شده است. محتوای کاربر ساخته که در رسانه های اجتماعی تولید می شود می تواند سیاستگذاری ها در همه زمینه ها را ارتقاء دهد. این پژوهش از چارچوب سیاستگذاری مبتنی بر شواهد استفاده می کند و داده های رسانه های اجتماعی را به عنوان شواهد جهت سیاستگذاری در نظر می گیرد. از اینرو اول از چارچوب مفهومی قابلیت های جمعی[۷۸] ، برای آزمون ارزش داده های رسانه های اجتماعی در سیاستگذاری مبتنی بر شواهد استفاده می کند. سوال اصلی این بوده است که سیاستگذاران چگونه ارزش داده های رسانه های اجتماعی را درک می کنند؟ و کدام یک از قابلیت های جمعی داده های رسانه های اجتماعی می تواند نیازهای سیاست گذاران را برای ورودی داده در فرایند سیاستگذاری پشتیبانی کند؟ این پژوهش در حوزه سیاستگذاری کشاورزی در انگلستان انجام شده است و از تحلیل داده های شبکه کشاورزان انگلستانی در توئیتر و

[76] Fernando, Díaz López, Şerban,Gómez-Romero

[77] Panagiotopoulosa, Bowen, Brooker

[78] Crowd capabilities

تحلیل داده های کیفی حاصل از مصاحبه با سیاستگذاران این حوزه سوالات پژوهش را پاسخ داده است. سیاستگذاران در این پژوهش به موقع بودن[۷۹]، به صرفه بودن [۸۰] و تنوع داده های رسانه های اجتماعی را به عنوان ورودی فرایند سیاستگذاری کشاورزی مفید دانسته اند.

پارک و کی[۸۱](۲۰۱۷)، رهبری افکار در توئیتر و ویژگی‌های رهبران افکار توئیتری را مورد بررسی قرار داده اند. درواقع با درنظر گرفتن مفهوم «رهبری افکار» که تا پیش از این در فضای حقیقی و در ارتباطات حقیقی انسانی وجود داشته، این مفهوم مهم ارتباطی را در یک بافت جدید یعنی فضای مجازی و رسانه اجتماعی توئیتر قرار می‌دهند و بررسی می‌کندکه آیا اشکال جدید رهبری افکار در رسانه اجتماعی توئیتر مانند ارتباطات انسانی به وجود می‌آیند یا نه؟ محققان معتقدند، با توجه به اینکه مفهوم رهبری افکار یکی از قدیمی‌ترین رویکردهای نظری در شکل گیری افکار عمومی است، به طور سنتی رهبران افکار ویژگی هایی نظیر طبقه اجتماعی بالاتر، فعالیت بیشتر در گروه های دوستی، تماس‌های اجتماعی زیاد را دارا بوده و همچنین بیشتر از افراد عادی رسانه‌های خبری را پیگیری و در آنها مشارکت می‌کنند. اما با وجود این که بستر شکل گیری بسیاری از این تعاملات و کنش ها در فضای توئیتر بعنوان یک رسانه اجتماعی وجود دارد با این حال، مشخص نیست که آیا مفهوم اصلی رهبری افکار در توئیتر صادق است یا خیر؟ نتایج این پژوهش نشان می دهد، بر خلاف رهبران افکار سنتی، رهبران افکار در توئیتر، بدون در نظر گرفتن محدودیت های اجتماعی، اقتصادی و یا سیاسی، نقش خود را به عنوان نوعی جدیدی از تولیدکنندگان دستور کار[۸۲] و موضوعات مورد بحث برای کاربران و سیاستگذاران ایفا می کنند. هدف پژوهش حاضر شناسایی ویژگی‌های نظری رهبران افکار در توئیتر برای تعمیق در درک مفهوم رهبری افکار در رسانه‌های اجتماعی و تاثیر آنها بر کاربران و سیاست ها است. به طور خاص، درجه‌ای از رهبری افکار را مورد بررسی قرار می‌دهد که در آن رهبری افکار در توئیتر با ویژگی‌های جمعیت شناختی کاربران توئیتر، اندازه شبکه ارتباطی آنلاین و آفلاین آنها، رفتارها و مشارکت های مدنی و فراوانی توییت ها و ریتوییت های آنها توضیح داده می‌شود.

استملاتوس و همکاران[۸۳](۲۰۲۰)، در پژوهش خود نشان می دهند که ویژگی های ساختاری رسانه اجتماعی آنلاین توئیتر در یونان می تواند اطلاعات ارزشمندی در مورد وابستگی سیاسی گره های کاربران ارائه کند. به بیان دیگر آنها نشان می دهند که از روی مشخصات ساختاری کاربران توئیتر می توان برای پیش بینی تمایل سیاسی گره های مهم کاربری استفاده کرد. آنها از توسعه الگوریتم های تشخیص دهنده استفاده کردند تا بتوانند از روی این سیگنال های ارائه شده از سوی کاربران، پیش بینی هایی را انجام دهند. از آنجا که دولت ها بیش از پیش در حال یکی کردن داده

[79] Immediacy
[80] Cost-effectiveness
[81] Park, Kaye
[82] Agenda Setting
[83] Stamatelatos, Gyftopoulos, Drosatos, Efraimidis

های ارتباطات دیجیتال و دیگر داده های زمینه ای در رویکرد جدید سیاستگذاری و مقررات گذاری هستند، داده های رسانه اجتماعی اهمیت می یابد. کاربران رسانه های اجتماعی اطلاعاتی انتشار می دهند که در موضوعات مختلف سیاستگذاری و توسعه سیاست ها از جمله رویکر سیاسی افراد، می تواند به کار گرفته شود.

لوندبرگ و لاتینن[۸۴](۲۰۱۹)، به بررسی گفتمان ضد دموکراتیک در توئیتر می پردازند و مشخصات زبانی آن را بررسی می کنند. آنها ۳ و نیم میلیون پیام توئیتری به زبان انگلیسی در سال ۲۰۱۸ را جمع آوری کردند که در آنها گفتمان ضد دموکراتیک تشخیص داده شده بود. این پژوهش با هدف شناسایی عملیات هایی که باهدف ایجاد اختلاف در جوامع غربی راه اندازی شده بود، انجام گرفته است. محققان اعتقاد دارند شناسایی این عملیات ها و ارائه ویژگی های آن به دولت ها به عنوان ورودی سیاستگذاری امنیتی می تواند به آنها در جهت کاهش شکل گیری این گونه عملیات ها کمک کند. لی، لی و چوی[۸۵](۲۰۲۰)، با این پیش فرض که سیاستگذاران در سراسر جهان به طور فزاینده ای رسانه های اجتماعی را به عنوان کانال ارتباط مستقیم با مردم مورد استفاده قرار می دهند، تاثیر ارتباطات توئیتری سیاستگذاران را بر پیشبرد سیاست های آنها آزمودند. در این راستا دو آزمایش طراحی شده است که تاثیر اشتباهات ارتباطی در ارتباطات توئیتری سیاستگذاران را بررسی کردند که چگونه اشتباهات توئیتری سیاستگذاران ممکن است بر نظر افراد درباره اجرای سیاست های مختلف تأثیر بگذارد. نتایج این پژوهش حاکی از آن است که اشتباهات ارتباطاتی در رسانه های اجتماعی از سوی سیاستمداران، در مرحله اجرای سیاست ها از فرایند سیاستگذاری بیشترین تاثیر را خواهد داشت و کاربران را نسبت به اجرای سیاست ها بی اعتماد خواهد کرد.

فرناندز و همکاران[۸۶](۲۰۱۴) بیان می کنند که رسانه های اجتماعی محلی برای بیان نظرات و به اشتراک گذاری آنها در اختیار کاربران قرار داده است و این یک فرصت بی نظیر برای دولت ها ایجاد می کند تا درباره نظرات شهروندان بیش از پیش بیاموزند و به طور مؤثر از آنها در سیاستگذاری استفاده کنند. آنها از چارچوب مفهومی مشارکت شهروندان[۸۷] استفاده می کنند و آن را به مشارکت دیجیتالی که از طریق رسانه های اجتماعی ایجاد می شود، پیوند زده اند. علیرغم پتانسیل بسیار زیاد این رسانه ها ، نگرانی هایی در مورد سوء استفاده از رسانه های اجتماعی بوجود آمده است، به ویژه هنگامی که برای اطلاع رسانی در مورد سیاست گذاری مورد استفاده قرار می گیرد. از جمله این موارد می توان به عدم آگاهی از ویژگیهای کاربرنی است که در مورد موضوعات سیاستی در رسانه های اجتماعی مشارکت فعال دارند. اگرچه در چند سال گذشته برخی از مطالعات انجام شده است که هدف آنها جمع آوری مشخصات جمعیت شناختی کاربران رسانه های اجتماعی (به عنوان مثال ، جنسیت ، سن ، مکانهای جغرافیایی) بوده است، اما کمتر تمایلی در جهت شناسایی آن دسته از کاربران خاص که در بحث های سیاستی شرکت می کنند وجود

[84] Lundberg, Laitinen
[85] Lee, Lee, Choi
[86] Fernandez, Wandhoefer, Allen, Elisabeth Cano, Alani1
[87] CP

داشته است. آگاهی از اطلاعات کاربرانی که در زمینه های سیاستی در رسانه های اجتماعی بحث می کنند و چگونگی بحث آنها در مورد مباحث مربوط به سیاستگذاری ها، می تواند به ارزیابی و چگونگی وزن دهی نظرات در رسانه های اجتماعی در زمینه سیاست گذاری کمک کند، که این مطالعه با این هدف انجام گرفته است. گینتوا[88](۲۰۱۹) معتقدند که رسانه های اجتماعی با سرعتی سریع توسط دولت های سراسر جهان جهت بهره برداری های سیاسی و اجتماعی پذیرفته می شوند. از جمله دولت کانادا، که در قالب دولت های فدرالی، حساب های رسانه های اجتماعی زیادی ایجاد کرده است و اکنون از آنها برای تعامل با مردم استفاده می کند. تاکنون مطالعات کمتر به رفتارها و دیدگاه های کاربران رسانه های اجتماعی دولتی توجه کرده است و از این رو، این محقق، تجربیات کاربران رسانه های اجتماعی و نحوه تعامل آنها در حساب های توئیتر و فیس بوک یک آژانس دولتی فدرال کانادا را تحلیل می کند. یافته ها نشان می دهد که این آژانس های سیاستگذاری دولتی در کانادا از رسانه های اجتماعی به عنوان ابزاری برای ارزیابی سیاست ها استفاده می کند و داده های رسانه های اجتماعی به عنوان ورودی فرایند سیاستگذاری به طور سیستماتیک مورد استفاده قرار نمی گیرد.

پوی و شولر[89](۲۰۲۰)، در پژوهشی تاثیر رسانه های اجتماعی در مشارکت شهروندان در زمینه های سیاستی را بررسی کرده اند. آنها نتیجه گرفتند که هر چه دسترسی به رسانه های اجتماعی بیشتر باشد و افراد بیشتر در این رسانه ها فعالیت کنند در حوزه های سیاستی نیز مشارکت بیشتری خواهند داشت. از آنجایی که استفاده از پلتفرم های رسانه های اجتماعی یک منبع جدیدی از اطلاعات را که دولتمردان نیاز دارند ایجاد کرده است، باید به دقت توسط دولتمردان تحلیل شود و داده های ارزشمند از آن استخراج گردد. بنابراین چالش دولتمردان پیدا کردن راههای انتخاب و استخراج و تحلیل داده های ارزشمند است که می تواند در فرایند های تصمیم گیری کمک کننده باشد. بنابراین سوال اصلی تحقیق این است که تاثیر میزان مشارکت شهروندان در رسانه های اجتماعی بر مشارکت آنها در زمینه های سیاستی چقدر است؟ ایم، هوانگ و کیم[90](۲۰۱۸)، نقش هایی که اجتماعات شکل گرفته در رسانه های اجتماعی می تواند در بالا بردن مسئولیت اجتماعی دولت ها ایفا کند را بررسی کرده اند. آنها یک تیپولوژی ارتباطات دولت در رسانه های اجتماعی ارائه دادند که در آن داده های رسانه های اجتماعی تبدیل به منبع جدیدی برای سیاستگذاران شده است. این تیپولوژی می تواند ارتباط میان سیاستگذاران و شهروندان را افزایش دهد و همچنین شهروندان را در سیاستگذاری و توسعه سیاست ها با هدف یافتن راه حل های جدید مشارکت می دهد.

از نظر استفاده از داده های رسانه های اجتماعی در سیاستگذاری، نمی توان تصور کرد که این مقدار از مشارکت های آنلاین، همیشه حاوی بینش مفیدی باشند (Boyd & Crawford, 2012; Davies, Nutley, & Smith, 2001).

---

88 Gintova
89 Poy, Schuller
90 Eom, Hwang & Kim

واقعیت های استفاده از اطلاعات در سیاستگذاری حاکی از آن است که مجموعه کلان داده ها دارای سوگیری هایی هستند که به طور چشمگیری در فعالیت ها و مراحل سیاست گذاری تفاوت ایجاد می کنند (Schintler & Kulkarni, 2014). به طور کلی از فناوری اطلاعات و ارتباطات انتظار می رود شواهد محکمی را از طریق داده های «سخت[۹۱]» و «بزرگ[۹۲]» ذخیره شده در سیستم های اطلاعاتی ارائه دهند (Misuraca, Codagnone, & Rossel, 2013). با این حال، همانطور که کوم، جوی استوارت، رز و دانکن(۲۰۱۵) با استفاده از یک مورد تجربی در رابطه با رفاه کودکان نشان می دهند، در ترجمه و تفسیر مجموعه های بزرگ داده به شواهد سیاستی، پیچیدگی های ذاتی وجود دارد.

بنابراین، چالش های عملی و نظری تأمین محتوا از مخاطبان شبکه های اجتماعی و همچنین قابلیت های جدید برای حمایت از این تلاش وجود دارد. اکنون مفهوم قابلیت های جمعی به عنوان یک رویکرد مفید برای قاب بندی داده های رسانه های اجتماعی به عنوان شواهد سیاستگذاری و درک توانایی های مورد نیاز سیاستگذاران برای استفاده از این داده ها به عنوان شواهد، مفید به نظر می رسد.

۹.۱ رسانه های اجتماعی و سیاستگذاری علم و فناوری

سیاستگذاری علم و فناوری[۹۳]، بررسی روابط بین علم و فناوری، دولت و جامعه است. در واقع، سیاستگذاری علم و فناوری با مسئله سیاستگذاری در زمینه های فناورانه و علم و روابط متقابل بین ارزشهای فرهنگی و اهداف اجتماعی، در ارتباط است. سیاستگذاری علم و فناوری اصطلاحی است که به سیاستگذاری عمومی علم و فناوری اشاره دارد چه از نظر سیاست هایی که فعالیت های مربوط به علم و فناوری را مورد حمایت قرار می دهد و چه سیاست هایی که آن دو را به سمت اهداف دولت ها هدایت میکند(کلانتری، منتظر و قاضی نوری، ۱۳۹۸). سیاستگذاری علم و فناوری به مجموع تصمیمات و اقدامات کلیدی گفته میشود که دولت ها برای تشویق و هدایت توسعه تحقیقات علمی و فناورانه از یکسو و بهره گیری از نتایج این تحقیقات برای دستیابی به اهداف کلی اجتماعی، اقتصادی و سیاسی از سوی دیگر، انجام میدهند.

تاریخچه سیاستگذاری علم و فناوری در سایر کشورها حکایت از معیارهای گوناگونی برای سیاستگذاری علم و فناوری دارد. بعضی از کشورها صرفا به جنبه های اقتصادی و تجاری در سیاستگذاری علم و فناوری تکیه داشته اند، بعضی از کشورها به انگیزه های سیاسی برای سیاستگذاری در علم و فناوری توجه کرده اند و بعضی از کشورها سیاستگذاری علم و فناوری را بر محور توسعه دانش و بالا بردن توان علمی و فنی کشور قرار داده اند(قاضی نوری،

91

92

93 Technology & Science Policy

قاضی نوری، ۱۳۹۱).دانش سیاستگذاری علم و فناوری اگر چه عمر کوتاهی دارد، اما دولتها از آن به عنوان اقدامی عملی برای حل مسائل عمومی حوزه علم و فناوری در همه جوامع استفاده کرده اند.

پس از مرور کوتاهی بر حوزه سیاستگذاری علم و فناوری می توان گفت گرچه، به طور عام بیش از چند دهه از عمر سیاستگذاری های کلان و برنامه های توسعه مبتنی بر این سیاستگذاری ها می گذرد اما به طور خاص سیاستگذاری در حوزه علم و فناوری سابقه ای حدودا دو دهه ای دارد. برای اولین بار در برنامه توسعه سوم فصلی به توسعه علم و فناوری اختصاص داده شد. در طول این سالها چندین سند بالادستی برای سیاستگذاری کلان درحوزه علم و فناوری(سند چشم انداز بیست ساله، نقشه جامع علمی کشور و سیاست های کلی علم و فناوری ابلاغی مقام معظم رهبری) تدوین و ابلاغ شده است(کلانتری، منتظر و قاضی نوری، ۱۳۹۸). بعد از گذشت قریب به دو دهه از سیاست گذاری های علم و فناوری، متخصصان معتقدند، نتایج و خروجی ها با آنچه انتظار میرفت، تفاوت زیادی دارند. بر اساس اسناد کلان نظام علم و فناوری، یکی از موضوع های اساسی، تبدیل علم و فناوری به گفتمان اصلی جامعه است که بالطبع فضای مساعدی برای تولید علم و فناوری ایجاد شود. بدیهی ترین هزینه ای که باید در خصوص توجه جامعه به علم و فناوری پرداخته شود هزینه اجتماعی آن است، به این معنا که جلب توجه به این موضوع، مستلزم تدوین سیاستگذاری است که در آن پیش نیازهای سیاستگذاری علم و فناوری در دستور کار قرار گیرد.

مشخص است که دگرگونی و تحول در هر سیاستگذاری از عوامل مؤثر بر پویایی و بالندگی آن حوزه است(Williamson & Piattoeva, 2018). از آنجا که حوزه سیاستگذاری علم و فناوری با چالش های زیادی در سال های اخیر روبرو است، نیازمند طراحی سیاست های متناسب با حل چالش های پیش رو است. از جمله این چالش ها، چالش های مرتبط با از بین رفتن مرزها میان بخش های مختلف علوم و فناوری هاست. بر خلاف رویکردهای سنتی، سیاستگذاری در رابطه با این پدیده همگرایی، فاقد مرزهای روشن است. منظور از مرزهای روشن هم مرزهای روشن در حوزه چاچوب ها و فرایند ها و هم ابزارهای سیاستگذاری است. با تغییر نسبت علم و فناوری در جامعه و همگرایی بخش های مختلف آن، این حوزه سیاستگذار را با چالش های زیادی مواجه خواهد ساخت. در این رابطه چالش هایی مانند چالش داده ها و مالکیت آن و دسترسی آزاد، چالش حریم حریم خصوصی و امنیت داده ها و اطلاعات، چالش اخلاقیات و دسترسی آزاد، چالش انحصار پلتفرم ها و تاثیرات آن بر رقابت، شفافیت و حتی مکانیزم های مشارکت(محمدی، ۱۳۹۹)، وجود دارد که سیاستگذار جهت تدوین سیاست ها بایستی با این چالش ها مواجه مناسبی داشته باشد. در این فضا، سیاستگذار مجبور به تغییر رفتارهای تنظیم گری خود جهت تسهیل فرایند ها در این حوزه است. رفتارهایی از جنس مشارکت با همه ذی نفعان که نیازمندی های همه در تدوین سیاست ها لحاظ گردد. از اینرو در رابطه با سیاستگذاری علم و فناوری نیازمند کنش جمعی، نقش آفرینی ذی نفعان مختلف در سیاستگذاری، در پیش گرفتن مسیری تعاملی در حوزه سیاستگذاری هستیم.

گفته می شود در این فضا نیازمند حرکت از مدل های متمرکز سیاستگذاری به مدل های چند مرکزی می باشیم(محمدی، ۱۳۹۹). پیش نیاز شکل گیری چنین مدل هایی، گسترش و تقویت ارتباطات، ایجاد اعتماد میان ذی نفعان مختلف، توانمند سازی و مشارکت همه ذی نفعان، ایجاد اهداف، هنجارها و ارزش های مشترک است(محمدی،۱۳۹۹). در واقع کار مهم در این رابطه، مشروعیت بخشی به مداخله همه ذی نفعان در سیاستگذاری است. این موضوع همیشه مورد نظر بوده است اما مهم سطحی از مداخله ذی نفعان در سیاستگذاری است که سیاستگذار آن را مشروع می شمارد. سیاستگذار در سطوح مختلفی می تواند همه ذی نفعان را در سیاستگذاری مشارکت دهد. اولین و محدود ترین سطح مشارکت، اطلاع رسانی است. در این سطح از مشارکت، سیاستگذار دیگر ذی نفعان را در رابطه با مداخله مستقیم در سیاستگذاری مشروع نمی شناسد و صرفا در مورد مسائل محدودی به او اطلاع رسانی می کند. دومین سطح از مشارکت مشورت گیری است که سیاستگذار در تدوین سیاست ها با دیگر ذی نفعان مشورت و نیازهای آنان را در نظر می گیرد. و سومین و بالاترین سطح از مشارکت، مشارکت و مسئولیت پذیری همزمان همه ذی نفعان است. این نوع از سیاستگذاری، عالی ترین سطح از مشارکت ذی نفعان در رابطه با تدوین و اجرای سیاست ها است.

استفاده از داده های رسانه های اجتماعی در سیاستگذاری و از جمله سیاستگذاری علم و فناوری همان گونه که در بخش های پیشین شرح آن رفت، به معنای مشارکت دادن ذی نفعان سیاستگذاری علم و فناوری یعنی کاربران مختلف در این نوع از سیاستگذاری است. سیاستگذاری با استفاده از داده های رسانه های اجتماعی، شامل طراحی چارچوب، فرایند ها و ابزارهایی است تا تعامل مشترک در خصوص همه ذی نفعان در حوزه سیاستگذاری ایجاد شود. از اینرو این پژوهش به دنبال ارائه مدلی مفهومی جهت کاربست داده های رسانه اجتماعی توئیتر در سیاستگذاری علم و فناوری بوده است که در بخش بعدی این الگو ارائه شده است.

نتیجه گیری

در این فصل تلاش شد تا منطق استفاده از داده های رسانه های اجتماعی جهت کاربست در سیاستگذاری ها به عنوان شواهد، بررسی شود. بدین منظور، انواع شواهد و ویژگی های آنها و همچنین ویژگی های داده های رسانه های اجتماعی، مرور شد. منطق منابع عمومی و اینکه داده های رسانه های اجتماعی می توانند به مثابه منابع عمومی محسوب شوند که استفاده و حکمرانی بر آن می تواند در اختیار دولت ها قرار گیرد، بررسی گردید. همچنین از چارچوب قابلیت های جمعی استفاده شد و کاربست داده های رسانه های اجتماعی در سیاستگذاری با منطق استفاده از ارزش هایی که اجتماعات می توانند بیافرینند، توجیه شد. قابلیت های جمعی اشاره به توانایی دولت ها برای جمع آوری گسترده مشارکت مخاطبان دیجیتال ناشناس و توسعه یک جریان اطلاعاتی مناسب برای استفاده در اهداف سیاستگذاری است. اینکه کدام قابلیت های داده های رسانه های اجتماعی می تواند نیاز سیاست گذاران را برای توسعه ورودی های سیاستگذاری پشتیبانی کند؟ یک سوال مهم در این حوزه است که تلاش شد در این فصل پاسخ داده شود. می توان گفت، کاربست داده های رسانه های اجتماعی در سیاستگذاری از موضوعاتی است که می تواند بوسیله نظریه های مختلفی توسط پژوهشگران پشتیبانی گردد و منطق مناسب به خود بگیرد تا در نتیجه در میان سیاستگذاران نیز رواج یابد.

فصل دوم

الگوی پیشنهادی پژوهش

مقدمه:

در این فصل تلاش شده است تا بر اساس منطق های پشتیبانی که در فصل پیشین ارائه شد، الگویی برای کاربست داده های رسانه های اجتماعی در سیاستگذاری ارائه شود. در این الگو از الگوریتم های یادگیری ماشین استفاده شده است و تلاش شده تا خروجی های مختلف آن با الگوی چرخه ای سیاستگذاری از یانگ و کوئین(۲۰۰۲)، ترکیب شود. سه عنصر از الگوی یانگ و کوئین(۲۰۰۲)، یعنی تنظیم دستور کار، انتخاب گزینه های سیاستی و تدوین سند سیاستی، با الگوریتم یادگیری ماشین، ترکیب شده است و در نتیجه یک الگو با سه خروجی(تعداد کلمات کلیدی و هشتگ های موضوعی، تحلیل احساس و رفتار کاربر، جداسازی شواهد از غیر شواهد) ارائه گردیده است. هر کدام از این خروجی ها قابلیت کاربست در یکی از عناصر الگوی یانگ و کوئین(۲۰۰۲)، را دارد.

۱.۲. ارائه الگوی مفهومی

همانگونه که پیش از این نیز آمد، پژوهش حاضر به دنبال طراحی الگویی است که بتوان تحلیل های مناسبی از داده های رسانه های اجتماعی را توسعه داده و آن را متناسب با بخش های مختلف فرایند سیاستگذاری علم و فناوری ارائه داد. از اینرو در اولین قدم نیازمند این هستیم که در رابطه با حوزه های تخصصی علم و فناوری آگاهی داشته باشیم. جهت کاربست داده های رسانه اجتماعی در سیاستگذاری علم و فناوری بایستی، کلید واژه های این نوع از سیاستگذاری مشخص شود. بدین منظور، پژوهشگر جهت انتخاب کلید واژه های سیاستگذاری علم و فناوری از پژوهش های مختلفی استفاده کرده است(قانعی راد و همکاران، ۱۳۹۰؛ قاضی نوری و قاضی نوری، ۱۳۹۱؛ قاضی نوری و همکاران، ۱۳۹۳؛ احمدیان و همکاران، ۱۳۹۷؛ نامداریان، ۱۳۹۸؛ کلانتری و همکاران، ۱۳۹۸؛ قدیمی و حجازی، ۱۳۹۸). پژوهشگر بر اساس مطالعه پژوهش های پیشین و استفاده از نظرات خبرگانی در خصوص سیاستگذاری علم و فناوری و میزان داده های موجود در شبکه اجتماعی توئیتر، موضوعات مرتبط با حوزه را انتخاب کرده است. در رابطه با موضوعات سیاستگذاری علم و فناوری، حوزه های موضوعی مانند: اکوسیستم علم و فناوری، فیلترینگ، کپی رایت، حفاظت از داده ها، بی طرفی شبکه، دسترسی به اطلاعات، تسهیم دانش فناورانه، تقلب علمی، آزادی اطلاعات و ارتباطات دوربرد، مورد نظر قرار گرفت.

بسیاری از سیاستگذاری ها در حوزه علم و فناوری در یکی از این کلید واژه های یاد شده، خواهد گنجید. این دسته بندی و تعیین کلید واژه های به پژوهشگر کمک می کرد تا در رابطه با استخراج داده ها و تمرکز بر محتوای تولید شده مرتبط، الگوی مورد نظر خود را ارائه دهد. پس از بررسی های انجام شده(ادبیات موضوع و مصاحبه با خبرگان حوزه تحلیل داده) در خصوص تحلیل های مناسب که بتوان بواسطه آن شواهد مناسبی برای سیاستگذاری علم و

فناوری با استفاده از داده های توئیتر تولید کرد، پژوهشگر موفق به طراحی الگوی مفهومی در این خصوص شد. این تحلیل ها در فاز اول پژوهش مرور شدند و در اینجا در الگوی ارائه شده به کار گرفته شده است.

در این الگو با توجه به اینکه هدف اولیه یعنی تشخیص توئیت ها در قالب شواهد و غیر شواهد، تعریف شده و مشخص است، از الگوریتم های یادگیری ماشین نظارت شده[۹۴] استفاده گردید. الگوریتم های یادگیری ماشین نظارت شده سالهاست که یکی از موفق ترین مدل های یادگیری در حوزه یادگیری ماشین به ویژه در رابطه با رسانه های اجتماعی می باشد(Mostafa et al., 2019) . بر این اساس چارچوب کلی الگوی پیشنهادی، چارچوب مدل یادگیری نظارت شده می باشد که به بخش های جمع آوری داده، پیش پردازش داده ها[۹۵] و برچسب گذاری[۹۶]، مهندسی ویژگی ها شامل استخراج ویژگی[۹۷] و انتخاب ویژگی[۹۸]، تقسیم مجموعه داده ها به بخش آموزش[۹۹] و آزمون[۱۰۰]، انتخاب و پیاده سازی الگوریتم یادگیری ماشین و ارزیابی مدل تقسیم گردید(Chen et al, 2016). فرایند کار الگوریتم یادگیری نظارت شده بدین صورت است که پس از پیاده سازی الگوریتم مورد نظر بر روی داده های آموزشی، مدل الگوهای موجود را یاد می گیرد و سپس بر روی مجموعه داده های آزمون پیاده سازی کرده و خروجی را پیش بینی می کند. در نهایت دقت مدل با توجه به معیارهای ارزیابی مورد سنجش قرار می گیرد. با توجه به توضیحات بالا الگوی پیشنهادی با استفاده از یادگیری نظارت شده با تکیه بر تشخیص شواهد از غیر شواهد در میان توئیت ها و نحوه تاثیرگذاری آنها بر فرایند سیاستگذاری علم و فناوری در قالب سه خروجی ارائه شده است که بر اساس نمودار ۱، بخش های الگوی پیشنهادی آن به شرح زیر می باشد.

در گام اول که جمع آوری داده ها از رسانه های اجتماعی توئیتر مورد نظر است کلید واژه های مرتبط با سیاستگذاری علم و فناوری در قالب کلید واژه هایی مربوط به موضوعات سیاستگذاری علم و فناوری تهیه و بر اساس آنها تمامی توییتهای فارسی زبان در بازه زمانی ۶ ماهه دوم سال ۱۳۹۸ (معادل ۲۳ آگوست ۲۰۱۹ تا ۱۹ مارس ۲۰۲۰) که شامل این کلمات کلیدی یا هشتگ آنها بوده است با استفاده از امکانات توئیتر API دانلود می گردد.

پیش پردازش مجموعه داده یکی از اصلی ترین مراحل فرایند یادگیری ماشین می باشد که در این بخش داده های خام برای پیاده سازی الگوریتم های یادگیری ماشین آماده می شود. عملیات پیش پردازش داده ها شامل حذف توییت های تکراری و ریتوییت ها، تبلیغات و توییتهای غیر مرتبط با حوزه سیاستگذاری علم و فناوری می باشد.

---

[94] Supervised Learning
[95] Preprocessing
[96] Labeling
[97] Feature Extraction
[98] Feature Selection
[99] Train
[100] Test

همچنین در این مرحله داده های پرت[۱۰۱] شـــامل داده هایی که مقادیر کاملاً متفاوت نســـبت به مجموعه داده را دارا بودند و مقادیر گمشـــده یعنی داده هایی که ناقص بودند یا به طور ناقص دانلود شـــده بودند نیز حذف می گردد. به علاوه عملیات نرمال سازی[۱۰۲] و تبدیل داده ها[۱۰۳] بر روی نمونه ها صورت می گیرد.

در مرحله بعدی برچسب گذاری انجام می شود، برچسب گذاری بر روی داده ها یکی از کلیدی ترین و دقیق ترین بخش های آماده سازی مجموعه داده ها است که توسط متخصصان و کارشناسان مربوطه صورت می گیرد. بدین صورت که پس از بررسی هر یک از نمونه ها و تشخیص توئیت مورد نظر به عنوان شواهد یا غیرشواهد برچسب های مورد نظر به نمونه ها تخصیص داده می شود. در نهایت، در این روش اتخاذ شده، سه نوع گزارش به سیاستگذار علم و فناوری ارائه می گردد که دو گزارش پس از برچسب گذاری تهیه می گردد. این دو گزارش صرفا در رابطه با توییتهایی با برچسب شواهد می باشد. در گزارش اول تعداد توییت های حاوی کلمات کلیدی است که به صورت بصری نمایش داده میشود که تعداد بالای توییت ها در هر کلمه کلیدی نشان دهنده اهمیت حوزه موضوعی مورد نظر از دید کاربران و میزان پرداختن به آن موضوع است.

خروجی دوم(گزارش دوم)، بر اســـاس تحلیل میزان احســـاس موجود در متن ارائه می گردد. از آنجایی که تحلیل احساس[۱۰۴] در حوزه متن کاوی و پردازش زبان طبیعی[۱۰۵] کمک بسزایی به تشخیص احساس درون متن می نماید و با توجه به کاربرد وسیع و موفق این تکنیک در حوزه های مختلف از جمله تحلیل نظرات مشتریان و پردازش متون، در این بخش به کار گیری می شود(Jianqiang & Xiaolin, 2017). آماده سازی متن در این روش شامل حذف فاصله ها، لینک ها، کاراکتر های خاص، اعداد، علائم نگارشی، ضمایر، بازگردانی شکل اصلی کلمه (به عنوان مثال اگر کلمه حریم های خصوصی را در متن داشته باشیم پس از بازگردانی کلمه حریم خصوصی را به ما برمی گرداند) می باشد. پس از آماده سازی با استفاده از تحلیل احساس که یکی از روش های پردازش زبان طبیعی است میزان قطبیت موجود در متن بر اســـاس اندازه گیری وزن احســـاس کلمات تعیین می گردد. در این روش به هر کلمه یک وزن از احساس اختصاص داده میشود و در نهایت با توجه به جمع بندی کلمات موجود در متن خروجی در سه دسته مثبت، منفی و خنثی به دست میآید. خروجی این بخش نیز می تواند به سیاست گذار در رابطه با دریافت یک بینش نسبی نسبت به احساسات و نظرات کاربران در رابطه با موضوعات مختلف حوزه سیاستگذاری علم و فناوری کمک نماید.

---

[101] Outliers
[102] Normalization
[103] Data transformation
[104] Sentiment analysis
[105] Natural language processing

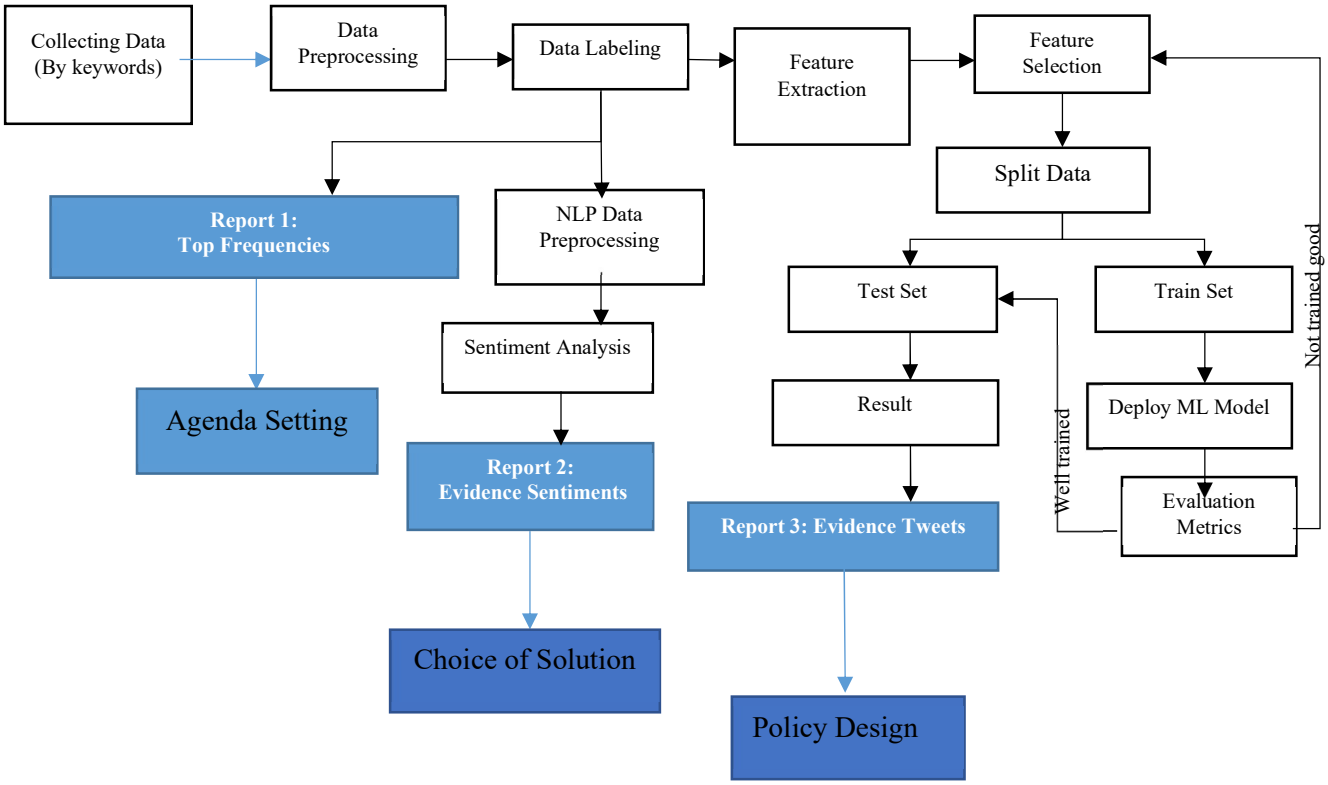
مهندسی ویژگیها شامل استخراج و انتخاب ویژگی هایی است که در استخراج ویژگی و تولید ویژگی های جدید مد نظر است به نحوی که بتوانند در یادگیری مدل کمک قابل توجهی بکند. ویژگی هایی که خوب انتخاب شده باشند، دقت مدل یادگیری را بالا میبرند و ویژگیهای کم اهمیت دقت مدل را پایین می آورد. با توجه به موفقیت بسیاری از این ویژگیها در پژوهش های مختلف، از جمله تشخیص اسپم در رسانه های اجتماعی (Hijawi et al, 2016 ; Sedhi & Sun , 2018) در این پژوهش مجموعه وسیعی از ویژگیهای مبتنی بر متن از قبیل تعداد کاراکترها، ایموجی ها، اعداد، علائم نگارشی، لینک ها، هشتگها، منشن ها و غیره و ویژگی های مبتنی بر حساب کاربری از قبیل سن حساب کاربری، تعداد توئیتها، دنبال کنندگان و دنبال شوندگان، ریتوئیتها، لایکها و غیره را ارائه و مورد بررسی قرار خواهد گرفت. در بخش انتخاب ویژگی از مجموعه ویژگیها، زیر مجموعه ای از بهترین ها انتخاب میشوند که مدل با این زیرمجموعه بالاترین دقت را دارا خواهد بود.

در گام بعدی الگوریتم های یادگیری ماشین متعددی روی مجموعه داده(توئیت ها) با زیرمجموعه مناسب از ویژگیها پیاده سازی می شود(Abayomi et al, 2019). بدین منظور ابتدا تمامی توئیت ها با هر نوع دسته بندی از کلید واژه ها با هم تجمیع شده و به صورت یک مجموعه داده تجمیع شده در می آید در این حالت تمامی نمونه ها با برچسب دسته مربوطه خود شناخته خواهند شد. برای یادگیری و ارزیابی الگوریتم ها، مجموعه داده می بایست به دو بخش آزمون و آموزش تقسیم گردد، در این کار به میزان درصد بالایی از داده ها به مجموعه آموزش و مابقی به مجموعه آزمون تخصیص داده می شود. عمل یادگیری الگوریتم بر روی مجموعه آموزش صورت می پذیرد و پس از ارزیابی مدل بر اساس معیارهای ارزیابی، درصورتیکه معیارهای ارزیابی نشان دهنده یادگیری ضعیف الگوریتم با زیر مجموعه ویژگی های انتخاب شده باشند می بایست مجددا مجموعه ویژگی های جدید انتخاب گردیده و الگوریتم دوباره ارزیابی گردد و اگر معیار ها قابل قبول بودند، الگوریتم روی مجموعه آزمون پیاده سازی می شود و خروجی مدل با توجه به معیارهای ارزیابی سنجیده می شود و پس از ارزیابی هر یک از الگوریتم ها، الگوریتم با بالاترین دقت انتخاب می گردد.

معیارهای[۱۰۶] اصلی ارزیابی الگوریتم یادگیری ماشین نظارت شده شامل دقت[۱۰۷]، صحت[۱۰۸] ،فراخوانی[۱۰۹] و میانگین هارمونیک[۱۱۰] می باشند که عملکرد الگوریتم بر اساس آنها سنجیده می شود(Thatwat, 2018). در این بخش نتایج میتوانند به عنوان بررسی شواهد در رسانه های اجتماعی، در اختیار سیاستگذاران علم و فناوری قرار گرفته تا به

[106] Metrics
[107] Accuracy
[108] Precision
[109] Recall
[110] F-Measure

صورت دقیق تر در یک حوزه خاص در سیاست گذاری مورد استفاده قرار گیرد. نمودار ۱.۲، نمایانگر الگوی یاد شده با استفاده از این الگوریتم هاست.



نمودار ۱.۲. الگوی مفهومی کاربست داده های رسانه های اجتماعی در سیاستگذاری علم و فناوری

همانگونه که در توضیحات شکل گیری الگوی ارائه شده، بیان شد، این الگو قرار است با بکارگیری روش ها و تکنیک های تحلیل داده های رسانه های اجتماعی به سیاستگذار علم و فناری در خصوص ارائه شواهد معتبر از طریق رسانه های اجتماعی کمک کند. بر اساس مدل چرخه ای سیاستگذاری از یانگ و کوئین(۲۰۰۲)، سیاستگذاری در فرایندی شش مرحله ای(تنظیم دستور کار، تدوین گزینه های سیاستی، انتخاب گزینه های سیاستی، تدوین سند سیاستی، اجرای سیاست ها، ارزیابی سیاست ها) صورت می گیرد. بر اساس این مدل که به فعالیت های سیاستگذار در هر مرحله از فرایند سیاستگذاری پرداخته است، سیاستگذاری متشکل از اقدامات و فعالیت هایی است که بایستی در مراحل مختلف انجام شود. اولین مرحله که در این مدل و از جمله بسیاری از مدل های سیاستگذاری وجود دارد، مرحله تعیین دستور کار است و اینکه چه موضوعات و مسائلی برای مداخله سیاستگذارانه مد نظر قرار گیرد. با توجه به روش های تحلیل داده های توئیتر که در بالا ارائه شده است، نتایجی که پس از استخراج داده ها، تمیز سازی داده ها و برچسب گذاری موضوعی داده ها، انجام می شود می تواند در قالب خروجی 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 تولید شده در توئیتر در رابطه با موضوعات مختلف سیاستگذاری علم و فناوری را تدارک می بیند.

دومین خروجی که این الگوی تحلیل داده های توئیتر تدارک می بیند، تحلیل احساس داده های استخراج شده و برچسب گذاری شده در رابطه با کلید واژه های مختلف علم و فناوری است. تحلیل احساس، میزان قطبیت محتوای تولید شده(توئیت های تولید شده) در رابطه با کلید واژه های سیاستگذاری علم و فناوری را مشخص می سازد. حجم و قطبیت محتوای تولید شده در رابطه با کلید واژه های مختلف می تواند به سیاستگذار در انتخاب راه حل های پیشنهادی که در دست دارد، کمک کند. سیاستگذار در این مرحله نیازمند کسب اطلاعات در مورد نظرات ذی نفعان مختلف سیاستگذاری در رابطه با راه حل های مختلف سیاستی است که یکی از کانال های اطلاع از نظرات ذی نفعان در رابطه با گزینه های سیاستی مطرح، داده های توئیتر و استفاده از تحلیل احساس این داده هاست. این خروجی می تواند به عنوان دومین خروجی فرعی، مورد استفاده سیاستگذار علم و فناوری قرار گیرد.

سومین خروجی و خروجی اصلی که این الگو با بکار گیری الگوریتم های یاد شده پس از تحلیل داده های توئیتر ارائه می کند، تشخیص و جداسازی توئیت هایی است که می تواند به عنوان شواهد در تدوین سیاست ها مورد استفاده قرار گیرد. از آنجا که سیاستگذار برای استفاده از داده های توئیتر، مشکل تشخیص شواهد از غیر شواهد را در میان

انبوهی از توئیت های مرتبط با حوزه علم و فناوری دارد، توسعه الگوریتمی که بتواند به او در این رابطه کمک کند از جمله اهداف این پژوهش بوده است. در این راستا یک الگوریتم یادگیرنده طراحی شد که در بالا قابل مشاهده است. در این الگوریتم که بخشی از الگوی اصلی است، داده ها را پس از پاک سازی و انتخاب بر اساس ویژگی های مشخص، به دو قسمت داده های آزمون و داده های آموزش تقسیم می کند و با استفاده از فرایند یادگیری و پیاده سازی آن روی حجمی از توئیت ها، الگوریتم، توئیت هایی که می تواند به عنوان شواهد مورد استفاده قرار گیرد را از غیر شواهد تشخیص خواهد داد. پس از این مرحله، خروجی بدست آمده که بر اساس تحلیل شواهد موجود در توئیتر در مورد سیاستگذاری علم و فناوری است، می تواند مورد استفاده سیاستگذار جهت تدوین سیاست ها قرار گیرد.

بدین منظور پژوهشگر بر اساس مطالعه پیشینه و استفاده از نظرات خبرگانی، ویژگی هایی که می تواند یک محتوا مانند توئیت را تبدیل به شواهد مناسب در زمینه سیاستگذاری علم و فناوری گرداند پیشنهاد داده است. مواردی همچون مرتبط بودن با کلید واژه های سیاست علم و فناوری، موضوعات علم و فناوری موضوع اصلی توئیت باشد و نه موضوع فرعی آن، توئیت ها اظهار نظر شخصی افراد باشد یا اظهار نظرهای شخصی دیگر کاربران را ریتوئیت کرده باشد، عدم وجود لینک های آموزشی و تبلیغاتی در توئیت و ... از جمله این موارد است.

استخراج داده ها، پاک سازی داده ها، تحلیل داده ها، تست مدل و بررسی دقت آن و همچنین تعیین قطبیت و بررسی رفتار کاربران، در بخش بعدی و بر اساس داده های ۶ ماهه دوم سال ۹۸ توئیتر فارسی، صورت گرفته است که گزارش آن در فصل بعدی آمده است.

نتیجه گیری

این فصل بر اساس منطق های ارائه شده در فصل پیشین در خصوص کاربست داده های رسانه های اجتماعی در سیاستگذاری شکل گرفت. الگوی ارائه شده در این فصل بر اساس الگوریتم های یادگیری ماشین و سه عنصر از الگوی چرخه ای سیاستگذاری از یانگ و کوئین(۲۰۰۲)، است. سه عنصر از الگوی یانگ و کوئین(۲۰۰۲)، یعنی تنظیم دستور کار، انتخاب گزینه های سیاستی و تدوین سند سیاستی، با الگوریتم یادگیری ماشین، ترکیب شده است و در نتیجه یک الگو با سه خروجی(تعداد کلمات کلیدی و هشتگ های موضوعی، تحلیل احساس و رفتار کاربر، جداسازی شواهد از غیر شواهد) ارائه گردیده است. هر کدام از این خروجی ها قابلیت کاربست در یکی از عناصر الگوی یانگ و کوئین(۲۰۰۲)، را دارد. این خروجی های سه گانه می تواند این سه عنصر از مدل سیاستگذاری را جهت انجام بهترین عملکرد یاری دهد. در واقع کانالی جدید از داده ها که متمرکز بر خواست ذی نفعان است در اختیار سیاستگذار قرار می دهد.

فصل سوم

آزمون مدل پیشنهادی

مقدمه:

در این فصل مدل ارائه شده در فصل پیشین با استفاده از داده های شش ماهه توئیتر آزمون شده است. از آنجا که سیاستگذار برای استفاده از داده های رسانه های اجتماعی از جمله توئیتر، مشکل تشخیص شواهد از غیر شواهد را در میان انبوهی از توئیت های مرتبط با حوزه علم و فناوری دارد، توسعه الگوریتمی که بتواند به او در این رابطه کمک کند، هدف اصلی این پژوهش بود. در این راستا یک الگوریتم یادگیرنده طراحی شد که در فصل دوم توضیح داده شد. در این الگوریتم که بخشی از الگوی اصلی است، داده ها را پس از پاک سازی و انتخاب بر اساس ویژگی های مشخص، به دو قسمت داده های آزمون و داده های آموزش تقسیم کرده و با استفاده از فرایند یادگیری و پیاده سازی آن روی حجمی از توئیت ها، الگوریتم، توئیت هایی که می تواند به عنوان شواهد مورد استفاده قرار گیرد را از غیر شواهد تشخیص داد و با تقریب خوبی این کار صورت گرفت. در این فصل فرایند ازمون مدل تشریح شده است.

۱.۳.منابع جدید داده ها

بر اساس بررسی پیشینه موضوع، مشخص شد که تاکنون مطالعات محدودی در ارتباط با تحلیل شبکه های اجتماعی مرتبط با حوزه سیاست گذاری انجام شده است. اکثر این پژوهش ها بر روی تشخیص عنوان[۱۱۱] پست ها، تحلیل گراف شبکه و استخراج احساس پست های شبکه های اجتماعی متمرکز شده است(Driss, Mellouli, Trabelsi, 2019; Gintova, 2019; Panagiotopoulosa, Bowena, Brooker, 2017). اما این موضوع که در میان تمامی این پیام ها(پست ها)، چه پیامی می‌تواند به عنوان شواهد قلمداد شود و در فرایند سیاستگذاری مبتنی بر شواهد، مورد استفاده قرار گیرد، مورد مطالعه قرار نگرفته است. از این رو تفکیک و تشخیص پیام های شواهد از غیر شواهد می تواند به عنوان یکی از موضوعات مهم در حوزه تحلیل شبکه های اجتماعی و سیاستگذاری به شمار رود.

پاناگاپلوس، باون و بروکر(۲۰۱۷)، در قالب پژوهشی که در خصوص تعیین ارزش داده های رسانه های اجتماعی به عنوان شواهد سیاستی از منظر سیاستگذاران انجام دادند، ویژگی های داده های رسانه های اجتماعی در مقابل منابع سنتی شواهد را از منظر چارچوب قابلیت های جمعی، معرفی کردند. در زیر، جدول۳ این ویژگی ها را بر اساس چارچوب یاد شده نشان می دهد.

جدول ۱.۳. منابع سنتی داده در مقابل داده های رسانه های اجتماعی در سیاستگذاری مبتنی بر شواهد

	منابع سنتی شواهد	داده های رسانه های اجتماعی
کسب نظرات مخاطبان	روشهای جمع آوری داده مانند نظرسنجی ها، گروه های کانونی، مصاحبه با خبرگان و غیره	رویکردهای مبتنی بر استفاده از کلمات کلیدی برای جمع آوری داده و تحلیل شبکه برای آماده سازی محتوا استفاده می شود.
	نمونه گیری و فرایند انتخاب مشارکت کنندگان مشخص است	نمونه گیری مبتنی بر مشارکت و درگیرشدن خود کاربران صورت می گیرد.
	اطلاعات مشارکت کنندگان به طور مستقیم از آنها دریافت می شود.	مخاطب با توجه به فراداده موجود هر پلتفرم رسانه اجتماعی در مورد کاربران، ساخته می شود.

[111] Topic Detection

	شرکت کنندگان موافقت و مشارکت فعال در تولید داده ها دارند.	شرکت کنندگان منفعلانه و با رضایت ضمنی[۱۱۲] مشارکت می کنند.
	ورودی داده ها بسته به روش جمع آوری آن، می تواند خصوصی یا عمومی باشد.	داده ها از فضاهای عمومی دیجیتال، تأمین می شود.
مقیاس پذیری ورودی	محدودیت های هزینه و زمان در جمع آوری بازخوردهای مشارکت کنندگان وجود دارد.	داده ها در لحظه در دسترس است.
	تأمین مجموعه های اطلاعاتی بزرگتر، معمولاً با صرف منابع زیادی امکان پذیر است.	تامین داده ها صرف نظر از تعداد مشارکت ها تا حد زیادی از ویژگی صرفه به مقیاس[۱۱۳] سود می برد.
قابلیت های ارتباطی	روابط پایدار و طولانی مدت با ذی نفعان سیاست ها وجود دارد.	روابط میان تولید کنندگان شواهد با پویایی بیشتری در شبکه های دیجیتال ایجاد می شود.
	تأثیرگذاران سنتی بر یک حوزه سیاستی، شناخته شده و رفتارهای آنها قابل پیش بینی است.	ایجاد تأثیرگذاران و نقاط کانونی، بیشتر به روند ها و موضوعات مختلف در شبکه، وابسته اند.
روش های تحلیل داده ها	از طریق تجزیه و تحلیل آماری، تجزیه و تحلیل های اقتصاد سنجی و روشهای علوم اجتماعی، شواهد «سخت» جمع آوری می شود.	با استفاده از تجزیه و تحلیل درگیری کاربر، متن کاوی و سایر تکنیک های تحلیل مبتنی بر شبکه، «شواهد نرم» جمع آوری می شود.

جدول بالا به وضوح نشان می دهد که در مقایسه با روشهای ثابت و سنتی پیشی، پیچیدگی های فنی تجزیه و تحلیل داده های رسانه های اجتماعی به انواع جدیدی از قابلیت های علوم داده(به عنوان مثال ابزارها و مهارت های

[112] Implicit Consent
[113] Scalable

ادغام/دستکاری مجموعه های کلان داده) نیاز دارد. به عنوان مثال، روش هایی مانند تجزیه و تحلیل هشتگ ها، به طور کلی در تحقیقات علوم اجتماعی ایجاد شده اند(به عنوان مثال برنت و همکاران[۱۱۴]، ۲۰۱۲ ؛ هیفیلد و همکاران[۱۱۵] ، ۲۰۱۳) اما برای فعالیت های تجاری یا سیاستی که تمرکز آنها روی خلاصه آمار ها و اندازه گیری های وایرال شده ها است، اعمال نشده است. این موضوع پیچیدگی استفاده از این داده ها در سیاستگذاری ها را نشان می دهد که نیازمند تحلیل های متفاوتی نسبت به دیگر زمینه های استفاده از داده های رسانه های اجتماعی است.

از آنجا که سیاستگذار برای استفاده از داده های رسانه های اجتماعی و از جمله توئیتر، مشکل تشخیص شواهد از غیر شواهد را در میان انبوهی از توئیت های مرتبط با حوزه علم و فناوری دارد، توسعه الگوریتمی که بتواند به او در این رابطه کمک کند، هدف اصلی این پژوهش بود. در این راستا یک الگوریتم یادگیرنده طراحی شد که در فصل دوم قابل مشاهده است. در این الگوریتم که بخشی از الگوی اصلی است، داده ها را پس از پاک سازی و انتخاب بر اساس ویژگی های مشخص، به دو قسمت داده های آزمون و داده های آموزش تقسیم می کند و با استفاده از فرایند یادگیری و پیاده سازی آن روی حجمی از توئیت ها، الگوریتم، توئیت هایی که می تواند به عنوان شواهد مورد استفاده قرار گیرد را از غیر شواهد تشخیص خواهد داد. پس از این مرحله، خروجی بدست آمده که بر اساس تحلیل شواهد موجود در توئیتر در مورد سیاستگذاری علم و فناوری است، می تواند مورد استفاده سیاستگذار جهت تدوین سیاست ها قرار گیرد. همانگونه که در پیش از این نیز در شرح الگوی پیشنهادی رفت، تلاش شده چارچوبی ارائه شوده که بتوان از میان تمامی توئیت های منتشر شده، توئیت های شواهد را از غیر شواهد با توجه به معیارهایی(معیارهای چندگانه شواهد) تشخیص داد.

در چارچوب پیشنهادی، نحوه تاثیر تحلیل های داده های شبکه اجتماعی توئیتر در فرایند تصمیم گیری سیاست گذاران ارائه شده است. همان گونه که در الگوی ارائه شده مشاهده گردید، این الگو بر مبنای توسعه سه بخش تاثیرگذار در فرایند سیاستگذاری طراحی شده است. بخش اول، گزارش آماری موضوعات پرتکرار در پیام های کاربران ارائه می شود؛ ۲- خروجی بخش دوم، میزان قطبیت و احساس درون متنها را با استفاده از پردازش زبان طبیعی، استخراج کرده و به عنوان گرایش نظرات کاربران در موضوع های سیاستی در حوزه علم و فناوری ارائه می دهد؛ ۳- در این بخش یک الگورتیم پیشنهادی با استفاده از تکنیک های یادگیری ماشین که پیامهای شواهد را از پیام های غیر شواهد تشخیص می دهد، ارائه شده است.

114

115

۲.۳. جمع آوری داده[۱۱۶]

امروزه توئیتر با بیش از ۵۰۰ میلیون کاربر فعال روزانه تبدیل به یکی از پرطرفدارترین شبکه های اجتماعی میکروبلاگینگ شده است که کاربران در آن با محدودیت ۲۸۰ کاراکتری در موضوعات متنوع نظرات، پیام ها، فیلم ها و عکس های خود را به اشتراک می گذارند. در راستای ارزیابی الگوی پیشنهادی، اولین مرحله ای که به منظور تشخیص توئیت های شواهد از غیر شواهد انجام شد، اقدام به جمع آوری داده از شبکه اجتماعی توئیتر با استفاده از ابزار توئیتر API[117] بود. از اینرو ۳۹ کلمه کلیدی در حوزه ی «سیاست گذاری علم و فناوری» با بررسی ادبیات موضوع در زبان فارسی(قانعی راد و همکاران، ۱۳۹۰؛ قاضی نوری و قاضی نوری، ۱۳۹۱؛ قاضی نوری و همکاران، ۱۳۹۳؛ احمدیان و همکاران، ۱۳۹۷؛ نامداریان، ۱۳۹۸؛ کلانتری و همکاران، ۱۳۹۸؛ قدیمی و حجازی، ۱۳۹۸)، انتخاب و بر اساس این کلمات کلیدی، استخراج توئیت ها آغاز شد. پژوهش گر در انتخاب این کلمات کلیدی از روش های نظام مند تحلیل محتوا، استفاده نکرده است بلکه بر اساس موضوع و اهداف این مقالات، حوزه هایی که در آن ها بحث و بررسی شده بود به عنوان کلمات کلیدی انتخاب شد. البته از آنجایی که برای تعدادی از کلمات کلیدی به حد کافی توئیت وجود نداشت، آن کلمات کلیدی در همان ابتدا حذف گردیدند و برای بقیه موارد داده ها استخراج گردید. آنگونه که در جدول۲، برخی از کلمات کلیدی بر اساس اهمیت شان و یا تعداد توئیت های مرتبط، هم با هشتگ و هم بدون هشتگ در توییت ها جستجو شده اند. در این فرایند بر اساس کلمات کلیدی، اقدام به جمع آوری توییت های حاوی حداقل یکی از این ۳۹ کلمه کلیدی در بازه زمانی ۶ ماهه دوم سال ۱۳۹۸ مطابق(۲۳ آگوست تا ۱۸ مارس سال ۲۰۱۸) شد. بر اساس کلمات کلیّدی انتخاب شده، تعداد ۲۸۲۷۷ توییت جمع آوری گردید. پس از حذف بخشی از توییت ها در حین پاکسازی اولیه آنها، تعداد ۱۴۰۲۹ توئیت باقی ماند. در مجموع در مرحله اول حدود ۱۴ هزار توییت(پیش از پاک سازی داده ها) در این بازه زمانی استخراج و جمع آوری گردید.

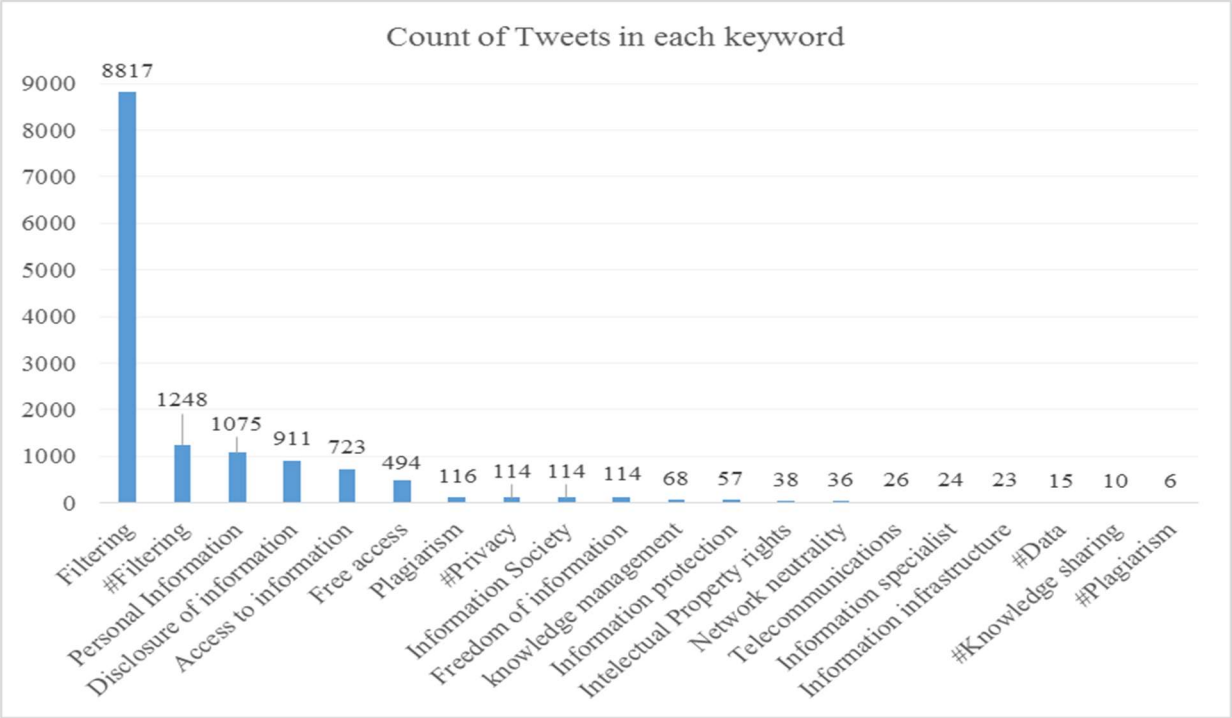
جدول۲.۳. کلمات کلیدی استخراج شده از ادبیات موضوع

سواد اطلاعاتی(#)	دسترسی به اطلاعات	آزادی اطلاعات	جامعه اطلاعاتی(#)
حریم خصوصی(#)	بازار های اطلاعاتی	فیلترینگ(#)	افشای اطلاعات
#داده	کالاهای اطلاعاتی	اشتراک دانش(#)	شکاف دیجیتال
متخصص اطلاعات	قیمت گذاری اطلاعات	مدیریت دانش	توسعه ICT
حقوق مالکیت فکری	سیاست های اطلاعات	دسترسی آزاد(#)	پردازش اطلاعات(#)
حفاظت از اطلاعات(#)	سیاست های ارتباطات	کیفیت داده	سیاستگذاری علم(#)

[116] Data Collection
[117] https://developer.twitter.com

اطلاعات علمی فناورانه	سیاستگذاری علم و فناوری	محرمانگی داده	سیاستگذاری فناوری
زیر ساخت اطلاعاتی	اطلاعات شخصی	ارتباطات راه دور	همانند جویی اطلاعات
زیر ساخت ارتباطی	سرقت ادبی(#)	صنعت اطلاعات	دسترسی آزاد
شبکه سازی علم	بیطرفی شبکه	زیر ساخت اطلاعات	

نمودار زیر تعداد توییت های دانلود شده را نشان می دهد 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 کلیدواژه های حوزه سیاستگذاری علم و فناوری در میان کاربران فارسی زبان بوده است. در این توئیت ها، قریب به اتفاق کاربران، رویکرد اعتراضی نسبت به فیلتر شدن سرویس های فناورانه داشتند. خشم کاربران از سیاستگذاران، در این توئیت ها قابل مشاهده بود. پس از آن توییت هایی که کلمات کلیدی اطلاعات شخصی، افشای اطلاعات، دسترسی به اطلاعات و دسترسی آزاد داشتند، به ترتیب پر اهمیت ترین موضوعات از دید کاربران بوده است. کاربران به موضوع افشای اطلاعات آنها از سوی سامانه های دولتی و خصوصی به شدت معترض بودند و بعضا سرویس های یکپارچه ای که اطلاعات شخصی آنها را در دسترس دیگر سازمان ها قرار می داد، مورد اعتراض بودند. موضوعاتی مانند سرقت ادبی، و نیاز به سیاستگذاری این حوزه، جزء خواست های کاربران توئیتر بود که در رابطه با این موضوع توئیت زده بودند و از بی سیاستی و عدم اجرای سیاست های این حوزه، شکایت می کردند. #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 کم اهمیت بوده و به برخی از موضوعات مانند زیرساخت ارتباطاتی، شبکه سازی علم، بی طرفی شبکه، همانندجویی اطلاعات و مدارک و غیره در توئیت ها از سوی کاربران پرداخته نشده است.



نمودار۳.۱. تعداد توئیت های مرتبط با هر کدام از کلمات کلیدی

نمایش موضوعات پرتکرار از دید کاربران در قالب ابر کلمات در شکل زیر آورده شده است. این ابر کلمات اهمیت موضوعات مختلف در حوزه علم و فناوری را در میان کاربران ایرانی شبکه اجتماعی توئیتر نشان می دهد که چه موضوعاتی در این دوره زمانی بیشتر مورد توجه کاربران بوده است. همانگونه که در ابر کلمات هم مشخص است، کاربران فارسی زبان، از میان کلید واژه های استخراج شده، فیلترینگ و رفع آن را مهمترین خواسته خود مطرح می کردند. دسترسی آزاد به اطلاعات و مدارک نیز از موضوعات مهمی است که مورد نظر کاربران بوده است و بر اساس ابرکلمات نمایش داده شده، حجم زیادی از توئیت ها مرتبط با این کلید واژه می باشد.

نمودار۳.۲. ابر کلمات مرتبط با هر کدام از کلمات کلیدی

۳.۳. پیش پردازش داده[۱۱۸]

پیش پردازش داده ها یکی از مهمترین بخش های فرایند یادگیری ماشین می باشد که به منظور افزایش دقت یادگیری الگوریتم های یادگیری ماشین انجام می گیرد. پیش پردازش داده ها، عمده زمان فرآیند یادگیری و پیاده سازی الگوریتم یادگیری ماشین را به خود اختصاص می‌دهد. بخشی از فرایند پیش پردازش در طی عمل برچسب گذاری صورت گرفته است و پیامهای اسپم، تبلیغاتی و غیرمرتبط، حذف شدند. سپس پیام های تکراری از مجموعه داده به همراه داده های پرت حذف گردید. نمونه های دارای مقادیر گمشده[۱۱۹] موجود در مجموعه داده نیز از مجموعه داده حذف شدند. مقادیر گمشده، مقادیری از توئیت هاست که یا کامل نبودند و یا بصورت ناقصی دانلود شده بودند. ضمنا برخی از حساب های کاربری پس از بازه جمع آوری داده ها حذف شده بودند که داده های آن هم در بروزرسانی مجموعه داده ها، حذف گردید.

به منظور پردازش داده ها، ابتدا تمام پیام ها می بایست به عنوان شواهد و غیر شواهد برچسب دریافت می کرد. این مرحله توسط بررسی ادبیات موضوع، بر اساس دارا بودن معیارهایی(معیارهای شواهد) انجام شد که به تمامی نمونه ها

[118] Data Preprocessing
[119] Missing value

برچسب شواهد و غیر شواهد تعلق گرفت. منطق برچسب زنی بر اساس پیشینه پژوهش در حوزه سیاستگذاری مبتنی بر شواهد، شناسایی کلید واژه های سیاستگذاری علم و فناوری استوار بود. بر اساس پیش‌فرض‌های روش سیاست‌گذاری مبتنی بر شواهد، چالش هایی وجود دارد که باعث می‌شود میان محققین و سیاست‌گذاران در خصوص استفاده از انواع شواهد، اختلاف نظر بوجود بیاید. اول اینکه، تعریف آنچه که به عنوان «شواهد» از آن یاد می‌شود، تعریفی ذهنی است و معیار های عینی برای آن تعریف نمی شود. پژوهشگرانی که از روش سیاست‌گذاری مبتنی بر شواهد دفاع می‌کنند، معمولاً تعریف نسبتاً دقیقی از شواهد دارند که صرفاً تحقیقات آکادمیک دقیق را دربر می‌گیرد(Head, 2008). در مقابل هنگامی که سیاست‌گذاران از روش سیاست‌گذاری مبتنی بر شواهد سخن می‌گویند معمولاً دایره تعریفشان گسترده‌تر است و تجربیات حرفه‌ای، دانش سیاسی، نظر ذینفعان، و ارزیابی‌های دولتی را نیز در بر می‌گیرد (Head, 2008). این تمایز نشان می‌دهد که علت اصلی اختلاف‌نظر میان پژوهشگران و سیاست‌مداران در رابطه با روش سیاست‌گذاری مبتنی بر شواهد از کجا نشات می‌گیرد. از اینرو در این پژوهش، تلاش شد تا معیارهای مناسبی در حین عمل برچسب زنی توئیت ها و تقسیم آنها به شواهد و غیر شواهد، مد نظر قرار گیرد.

البته هر یک از این دیدگاه‌ها(پژوهشگران و سیاستگذاران) مزایا و معایب خاص خود را دارد. از یک طرف، سیاست‌گذاری مبتنی بر شواهد مد نظر پژوهشگران ممکن است منابع مفید، قانونی و به هنگام (Clarke & Margetts, 2014)، باشد ولی واقعیات عملی تصمیم‌گیری را نادیده بگیرد. از طرف دیگر، شواهد مد نظر سیاستگذار نیز ممکن است مفهومی بی‌معنا از شواهد تلقی شود. منفعت سیاست‌گذاران در این است که خودشان را مدافع این روش معرفی کنند، و به صورتی انتخابی از شواهد استفاده می‌کنند (Janssen & Helbig, 2016) که تفاوت چندانی با سیاست‌گذاری مبتنی بر عقاید نخواهد داشت.

چالش دوم آنکه سیاست‌گذاری مبتنی بر شواهد ممکن است واقعیات دیگری را در تصمیم‌گیری نادیده بگیرد. حامیان آکادمیک سیاستگذاری مبتنی بر شواهد، از یک هنجار سیاسی سخن می‌گویند که در آن، «شواهد» مهمترین عامل تعیین‌کننده تصمیمات سیاسی است. این دیدگاه می‌تواند جنبه‌هایی دیگر نظیر مشروعیت، افکار عمومی، فرهنگ رایج جوامع، مدیریت ریسک، روابط سیاسی، تفسیر رسانه‌ها و گروه‌های ذینفع را نادیده بگیرد(Kum et al., 2015). بنابراین «شواهد» به تنهایی برای تعریف اقدامات و تصمیمات کافی نیست حتی اگر به درستی درک شده باشد.

چالش سوم آنکه، منافع خاص سیاست‌گذاران معمولاً نادیده انگاشته می‌شود. طرفداران رویکرد آکادمیک به تصمیم‌گیری به سادگی تصور می‌کنند که این روش می‌تواند به نفع جامعه تمام شود (Clarke & Margetts, 2014)بدون اینکه سازوکار علّی و منافع سیاست‌گذاران را مورد توجه قرار دهند. بر اساس اعتبار داده شده به انواع

شواهد از سوی پژوهشگران، سلسله مراتب شواهد را به شرح زیر می توان تعریف کرد. در زیر سلسله مراتب ارزشِ شواهد، از منظر تاکید پژوهشگران پیشین آورده شده است.

جدول۳.۳. سلسله مراتب شواهد

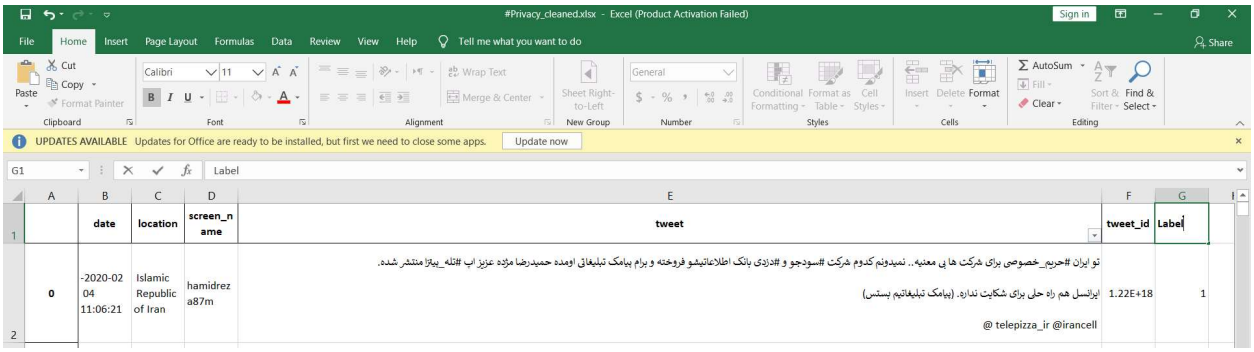
سطح شواهد	نوع شواهد
+++++	نتایج آزمایش های کنترل شده، آمار و ارقام، مرور سیستماتیک
++++	آزمایش های کنترل نشده، مطالعات موردی، پیمایش
+++	فراتحلیل، مرور سیستماتیک
++	مطالعات مروری غیر تحلیلی
+	عقاید خبرگانی

باید در نظر گرفت که این نوع از سیاستگذاری درواقع یک شیوه نسبی است و بسته به شرایط، برخی شواهد معتبرتر خواهند بود. از اینرو در موضوعات مختلف، سیاستگذار نیازمند شواهد مختلفی با درجاتی متفاوتی از رسمیت است. در حقیقت، بسیاری از شکست‌ها در امر سیاست‌گذاری ناشی از ضعف در مدیریت اطلاعات و بی‌توجهی به شواهد متنوع در میدان مورد نظر است. سیاست‌گذاری مبتنی بر شواهد لزوماً به معنی تکنوکراسی صرف نیست، بلکه در واقع ضرورت تلفیق دیدگاه‌های تکنیکی و سیاسی را مورد تاکید قرار می‌دهد (Howlett, 2009). در واقع این روش اگر به درستی اجرا شود، می‌تواند قدرت نظارتی عمومی را نیز افزایش دهد و شهروندان را در سیاستگذاری دخیل سازد. در رابطه با سیاستگذاری حوزه علم و فناوری نیز، با توجه به پیچیده و تغییر پذیر بودن حوزه، سیاستگذار نیازمند استفاده از دانش عمومی در سطحی گسترده است. شبکه های اجتماعی می تواند این بستر را به راحتی و با هزینه بسیار کمی در اختیار او قرار دهد. از اینرو ایده استفاده از این داده ها برخاسته از این منطق است که می بایست شواهد متنوعی در سیاستگذاری ها مورد استفاده قرار گیرد.

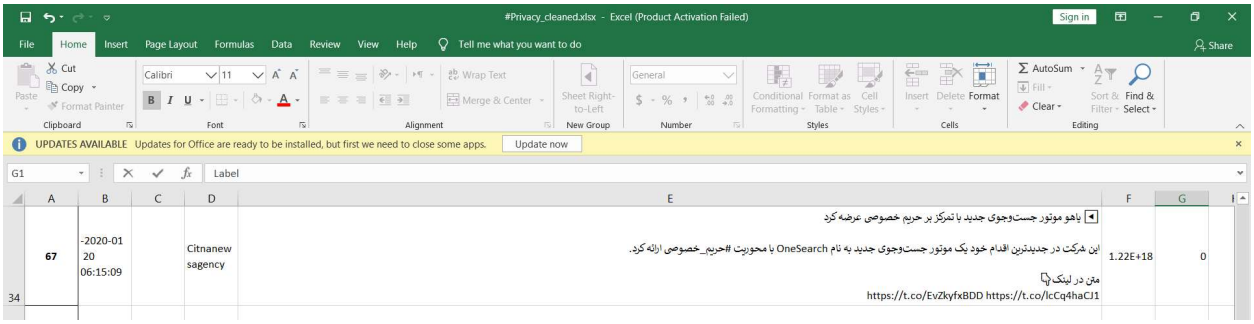
بر اساس مواردی که در بالا بیان شد، در این مرحله پژوهشگر تمامی توئیت های استخراج شده را تک تک مطالعه و بر حسب محتوای آن، برچسب شواهد و غیر شواهد به آن زده است. پژوهشگر بر اساس مطالعه پیشینه و استفاده از نظرات خبرگانی، ویژگی هایی که می تواند یک محتوا مانند توئیت را تبدیل به شواهد مناسب در زمینه سیاستگذاری علم و فناوری گرداند پیشنهاد داده است. مواردی همچون مرتبط بودن با حوزه های موضوعی سیاست علم و فناوری، موضوعات علم و فناوری موضوع اصلی توئیت باشد و نه موضوع فرعی آن، توئیت ها اظهار نظر شخصی افراد باشد یا اظهار نظرهای شخصی دیگر کاربران را ریتوئیت کرده باشد، عدم وجود لینک های آموزشی و تبلیغاتی در توئیت و ... از جمله این موارد است. شکل های ۱ و ۲ یک نمونه از پیام با برچسب شواهد و غیر شواهد را نمایش می دهند

که از مجموعه داده انتخاب شده است. این دو شکل از میان انبوهی از داده ها در مجموعه داده[۱۲۰] پژوهش در نظر گرفته شده است تا نشان دهد چگونه توئیت ها بررسی و هر کدام بر اساس معیارهای پیشنهادی، برچسب شواهد و غیر شواهد دریافت کرده است. همانگونه که مشاهده می شود، توئیت های شواهد برچسب (۱) و توئیت های غیر شواهد برچسب (۰) دریافت کرده اند. برای اینکه عمل یادگیری الگوریتم در مراحل بعدی به درستی اتفاق بیافتد و داده های پرت از مجموعه داده های آموزش و آزمون حذف شود، برچسب زنی به توئیت هایی که کلمات ناسزا داشتند(برچسب ۲ دریافت کردند) نیز گسترش پیدا کرد.

شکل۳.۱. یک نمونه توئیت با برچسب شواهد(۱) با موضوع حریم خصوصی در فضای دیجیتال



شکل۳.۲. یک نمونه توئیت با برچسب غیر شواهد(0) با موضوع حریم خصوصی در فضای دیجیتال



در زیر نمونه بیشتری از توئیت ها(توئیت های شواهد) که در این پژوهش مورد بررسی قرار گرفته است، در قالب جدول آورده شده است.

جدول ۳.٤. نمونه ای از توئیت های شواهد

ردیف	متن توئیت با برچسب شواهد	نام کاربری

[120] Data Set

۱	با گونه‌ی جالبی از موجودات مواجه هستیم که با فیلترشکن میاد توئیتر و از فیلترینگ حمایت می‌کنه	IraniZorro
۲	جناب @azarijahromi لطفا در مورد رفع فیلتر توییتر اقدام جدی کنید و بعنوان عیدی مردم قرار دهید.کلیه مسئولین با اینترنت بدون فیلترینگ, فعال هستند و برای مردم عادی فیلتر شده است.یا برای همه اینترنت بدون فیلترینگ در نظر بگیرید یا به این تناقض پایان دهید. پیشاپیش سال نو مبارک! https://t.co/1wf0fRTwmO	vahid_kz1
۳	تا چند سال پیش، صنف ما از دو طرف بدجور کوبیده می شد. تحریم های IT از خارج از کشور به بهانه ایرانی بودن، فیلترینگ احمقانه از داخل کشور. جدیدا از یه سری همکارا هم باید ضربه بخوریم که بهانه های مختلف، شدن ابزار ایجاد تبعیض و تسهیل خفقان و... به قولی از ماست که برماست..	iman1024s
۴	عشاق فیلترینگ از همین شبکه فیلتر شده برند و در سروش آی گپ پست بگذارند اینجا توئیتر است .	FashiMohsen
۵	واقعا وقتی با یه سرچ ساده همه اطلاعات شخصی رو میشه پیدا کرد نمیتونن کسی رو برای انتشار این اطلاعات محکوم کنن.	monagili
۶	این حق کاربر هست که بتونه از امکانت وب و نت بدون ترس از دست دادن اطلاعات شخصی یا بدافزارها استفاده کنه. من حتی از ترس وی پی ان ها رو از گوشیم پاک کردم و دیگه با گوشیم نمیتونم از اپلیکیشن های یوتیوب، پینترست، تلگرام و ... استفاده کنم و این واقعا ترسناکه @Saeed0f @Amir_s_f	Miichkaaaa
۷	یادش بخیر قدیما سرقت ادبی میکردیم واو جا نیوفته، مثل الان نبود قانون کپی‌رایت و حقوق محفوظ شاعر و نویسنده، اینترنت درست حسابی هم نبود کسی چک کنه بگه خودت ننوشتی مجله پیک‌عدالت ۱۳۸۸ https://t.co/LqLDudAJIA	HassanTaherii
۸	@omidamramagiiir تو مقاله اگه بیش از ۱۰ تا ۱۵ درصد مطالب کپی شده باشه به عنوان سرقت ادبی محسوب میشه و فقط باید برداشت خودتو بنویسی و رفرنس بدی. به هر حال به نقطه‌ی مشترک نرسیدیم	Arashlouck
۹	البته این رتبه چاپ مقاله هست. یادمون هم باشه که ایران در زمینه سرقت ادبی چه داخل و چه خارج پرونده سیاهی داره. دیگه خودمون که میدونیم بازار انقلاب چه خبره. و انتشار مقاله ملاک نیست. استفاده از اون دانش هست، و اکثر نویسنده ها این مقاله ها جذب خارج میشن	Sisyphus34
۱۰	اهمیت #حریم_خصوصی در دنیای امروز! آزادی اینترنت با سه شاخص رتبه بندی می شود: توسعه و دسترسی، محدودیت ها (#فیلترینگ)، نقض حقوق کاربران. در سالیان اخیر #ایران تنها در شاخص دسترسی پیشرفت کرده .	born_to_live_

	در این توئیت مثالی می زنم از اهمیت حریم خصوصی کاربران.	
۱۱	#ویدئوکال راه ارتباطی چندان محترمانه ای نیست و مطمئن هستم که طی سالهای آینده آداب استفاده از آن تغییر خواهد کرد. شاید نوعی تجاوز به #حریم_خصوصی و خلوت دیگران محسوب شود،کما اینکه همین حالا نیز در اکثر مواقع موجب رنجش گیرنده تماس می شود و او از سر ناچاری به دوستش پاسخ میدهد.	syzyf_thesecond
۱۲	#حریم_خصوصی توی ایران یه مفهوم کاملا تجملاتی محسوب میشه ..... https://t.co/HNzNuvNnsG	Irantruthseeker
۱۳	داده جمع‌آوری می‌شه؛ خوش‌مون بیاد یا نه؛ اما حتما باید ناشناس بشه تا #حریم_شخصی افراد رعایت بشه. اما وقتی شرکت‌های کوچیک، که زورشون کمتره، می‌تونن اینقدر راحت از زیر مسئولیت‌شون شونه خالی کنن، وضعیت شرکت‌های بزرگ که دیگه گفتن نداره... یکی‌ش افتضاح #فیسبوک	hmdprs
۱۴	هر روز حجم زیادی از اخبار و اطلاعات به سمت ما روانه می شود.اینکه در این وانفسا چه واکنشی به آنها داشته باشیم"مهم"است.نباید به همه آن ها پرداخت!گویا"عمدی"در ارسال این حجم از اخبار است!گویا با"سبک جدیدی"از #سانسور روبرو هستیم!بنظرم دیگر سانسور بمعنی عدم دسترسی به اطلاعات نیست.شاید	K_Y_4

پس از مرحله برچسب زنی داده ها، نوبت به استخراج ویژگی های مناسب برای تحلیل داده ها رسید که تلاش شد بر اساس بررسی ادبیات موجود، مشورت های خبرگانی و پیشنهادات پژوهشگر، ویژگی های استخراج شود. با توجه به تفاوت در ارزش برخی از ویژگیها(به عنوان مثال مقادیر ویژگی مانند تعداد جملات، در بازه صفر تا ۱۲ جمله است در حالی که ارزش ویژگی ای مانند تعداد فالوئرها از صفر تا حدود چند میلیون نیز می رسد)، به منظور جلوگیری از حساسیت الگوریتم یادگیری ماشین و نادیده گرفته شدن برخی از ویژگیهایی که ارزش آنها خیلی کوچکتر از ارزش ویژگیهای دیگر است، عمل نرمال سازی صورت گرفت. از اینرو، با توجه به تفاوت در ارزش برخی از ویژگیها، به منظور جلوگیری از نادیده گرفتن مقادیر کوچک توسط الگوریتم یادگیری ماشین، مقادیر ویژگی ها با استفاده از فرمول زیر نرمال سازی شده که سبب افزایش دقت الگوریتم می گردد. در واقع با استفاده از این تکنیک، مقیاس اندازه گیری ارزشهای تمام ویژگیها تغییر می کند و با مقدار جدید همگی در یک بازه ثابتی قرار می گیرند که سبب افزایش دقت الگوریتم می گردد. تکنیک استانداردسازی یکی از تکنیکهای نرمال سازی است که برای تبدیل مقادیر ویژگیها در یک بازه مشخص استفاده می شود. برخلاف نرمالسازی این تکنیک اطلاعات مفید در رابطه با مقادیر داده های پرت را نگه می دارد. به عبارتی مقادیر تمام ویژگیها را برای هر نمونه در یک بازه معین محدود می نماید. اگر

σ را به عنوان انحراف از معیار برای نمونه x و M را میانگین داده ها در نظر بگیریم، ارزش جدید V در هر ویژگی و برای هر نمونه با استفاده از فرمول زیر محاسبه می گردد.

$$V = \frac{(x - M)}{\sigma}$$

٤.٣. استخراج ویژگی ها[۱۲۱]

همانگونه که در بالا گفته شد، اکنون زمان استخراج ویژگی هایی است که در این رابطه بتواند کمک کننده باشد. در پژوهش های پیشین مرتبط با حوزه تحلیل رفتار کاربران شبکه های اجتماعی، به منظور تحلیل داده های کاربران از ویژگی های مبتنی بر حساب کاربری و ویژگی های مبتنی بر متن استفاده گردیده است که نشان از موفقیت این ویژگیها دارد. بعنوان مثال برای تشخیص اسپم در بسیاری از مطالعات پیشین(Chen et al, 2016; Sedhai & Sun, 2017)، از ویژگی های مبتنی بر حساب کاربری برای بررسی خصوصیات پروفایل کاربران استفاده گردیده و در برخی دیگر از ویژگی‌های مبتنی بر متن استفاده شده است(Hijawi et al, 2017). با توجه به موفقیت این ویژگی‌ها در مطالعات مختلف، این پژوهش نیز به منظور تشخیص شواهد از غیر شواهد اقدام به استفاده از برخی از این ویژگیها و همچنین استخراج ویژگیهای جدید در حوزه تشخیص پیام های شواهد از غیر شواهد نمود. در این پژوهش هر دو دسته از ویژگی‌ها ۱- مبتنی بر نام کاربری و ۲- مبتنی بر متن مورد استفاده قرار گرفته است تا بوسیله آنها الگو و رفتار کاربرانی که پیام‌های شواهد ارسال می‌کنند نسبت به کسانی که پیام هایشان شواهد محسوب نمی شود مورد بررسی قرار گیرد. جدول زیر ویژگی های مبتنی بر متن استخراج شده از ادبیات موضوع را به منظور بررسی الگوهای موجود در متن پیام های شواهد نشان می دهد(Hijawi et al., 2017; Chen et al., 2017; Sedhai & Sun, 2017). از ویژگی های استفاده شده مربوط به متن، توییت هایی که حاوی کلمات ناسزا و رکیک هستند، تعداد جملات هر توئیت، تعداد خط ها، تعداد فاصله ها، طول توییت، تعداد ایموجی ها، تعداد لینکها، تعداد علائم نگارشی، تعداد کاراکترها، طول کلمات، زمان ارسال توییت، تعداد هشتگ، تعداد URLs، تعداد منشن های موجود در متن و غیره مورد استفاده قرار گرفته است.

جدول ۵.۳. ویژگی های مبتنی بر متن

[121] Feature Extraction

No	Feature Name	Description
1	Swear Word	Having swear words in tweet
2	Tweet Time	Time of a tweet sent
3	No_Sentences	The number of sentences in a tweet
4	No_Lines	The number of lines that a tweet has
5	No_Mentions	The number of mentions included in a tweet
6	No_Urls	The number of URLs included in a tweet
7	No_Hashtags	The number of hashtags included in a tweet
8	No_Digits	The total number of digits in a tweet
9	No_Emojis	The number of emojis included in a tweet
10	No_Spaces	The number of spaces included in a tweet
11	Length Of Tweet	The length of a tweet
12	Max Length Of Words	Maximum length of words that a tweet has
13	Mean Length Of Words	The mean length of words that a tweet has
14	No_Exclamation Marks	The number of exclamation marks included in a tweet
15	No_Question Marks	The number of question marks included in a tweet
16	No_ Punctuations	The number of punctuations included in a tweet except question and exclamation marks
17	No_Words	Total number of words that a tweet has
18	No_Characters	Total number of characters have been used in a tweet
19	Digits To Chars Ratio	The number of digits to the number of characters ratio in a tweet
20	Lines To Sentences Ratio	The number of lines to the number of sentences ratio in a tweet
21	Words To Sentences Ratio	The number of words to the number of sentences ratio in a tweet
22	Hashtags More Than 2	Having more than 2 hashtags in a tweet
23	No_Words Less Than 3 Chars	Total number of words with less than 3 characters that a tweet has
24	No_Words More Than 5 Chars	Total number of words with more than 5 characters that a tweet has
25	Video	Tweet contains video
26	Image	Tweet contains image

ویژگی های مبتنی بر فعالیت حساب کاربری به منظور بررسی خصوصیات حساب کاربری افراد ارائه شده است. از قبیل تعداد دنبال کنندگان، تعداد دنبال شونده ها، تعداد توییت های ارسال شده، تعداد لایک ها، تعداد لیست ها(لیست های دوستان نزدیک و ...) و برخی از ویژگی های مربوط به ایجاد حساب کاربری مانند دارا بودن عکس بک گراند، دارا بودن عکس پروفایل، داشتن توصیف، وجود URLs در توصیف و غیره ارائه شده است. تمام ویژگی های فوق برای تمام توئیت های جمع آوری شده استخراج گردیده است.جدول ۴ این ویژگی ها را نشان می دهد.

جدول ۶.۳. ویژگی های مبتنی بر حساب کاربری

No	Feature Name	Description
1	No_Followings	The number of followers of this twitter user
2	No_Followers	The number of followings/friends of this twitter user
3	Ff_Ratio	The number of followers to the number of followings ratio
4	Description	Having description in the profile
5	No_Likes	The number of user favorites by this twitter user
6	Url In Description	Having URL in description of twitter user
7	No_Lists	The number of lists that this twitter user added
8	No_Tweets	The number of tweets this twitter user sent
9	Profile Image	Having profile image in the twitter profile account
10	Background Image	Having background image in the twitter profile account
11	Profile Background Image	Having profile background image in the twitter profile account

۵.۳. انتخاب ویژگی[۱۲۲]

این مرحله، مرحله ای است بایستی ویژگی های استخراج شده، ارزیابی و مناسب ترین آنها انتخاب شود. ویژگی هایی مناسب ترند که بتوانند یادگیری بیشتری برای الگوریتم ایجاد کنند. از مجموعه ویژگی های ارائه شده در جداول ۳ و ۴ برخی از ویژگی ها برای مدل یادگیری ماشین از اهمیت بالاتری برخوردار بودند و دارای تاثیر بیشتری و برخی دارای تأثیری کمتر بر روی یادگیری مدل بودند. ویژگی های دارای اهمیت کمتر می بایست از مجموعه ویژگی‌ها حذف می گردید و به منظور افزایش دقت و کارایی مدل پیشنهادی زیر مجموعه ای از بهترین ویژگی ها از مجموعه کامل ویژگی های ارائه شده می بایست انتخاب می شد. محاسبه بهره اطلاعاتی[۱۲۳] یکی از روش های موفقیت آمیز در حوزه انتخاب ویژگی در بسیاری از مطالعات پیشین مورد استفاده قرار گرفته است که این پژوهش نیز از این شاخص برای انتخاب ویژگی های مناسب استفاده کرده است. این روش مبتنی بر محاسبه بی نظمی[۱۲۴] است و میزان دارا بودن اطلاعات موجود در هر ویژگی را با استفاده از این معیار محاسبه می نماید.

به منظور پیاده سازی مدل پیشنهادی ابتدا تمامی مجموعه توییت های دانلود شده در هر موضوع، بدون در نظر گرفتن دسته بندی موضوعات آنها، به صورت یکپارچه به یک مجموعه داده تبدیل شد و از میان آنها تعداد ۵۹۰۰ توییت به صورت اتفاقی انتخاب گردید. در حین فرایند تعیین برچسب دسته ها، پس از حذف توئیت هایی که کلمات کلیدی در آنها در معنای دیگری به کار رفته یا متن توییت مرتبط با سیاست های داخلی کشور نبود، تعداد ۴۵۶۰ توئیت باقی

[122] Feature Selection
[123]
[124] Entropy

ماند. در میان ویژگی های استخراج شده در جدول زیر، در مدل تشخیص شواهد، برخی از ویژگی ها از اهمیت بیشتر و برخی از اهمیت کمتری برخوردار هستند. در جدول۵ مقادیر بهره اطلاعات[۱۲۵] مربوط به هر ویژگی نشان داده شده است. بهره اطلاعات نشان می دهد که کدام ویژگی بیشترین بار اطلاعاتی را با خودش دارد و در واقع به چه میزان اطلاعات راجع به خروجی در خودش دارد. میزان بهره اطلاعات بالاتر به این معناست که ویژگی مزبور تاثیر بیشتری در خروجی الگوریتم دارد.

یکی از معیارهای نشان‌دهنده میزان تاثیر و کارایی هر ویژگی در روش فیلتر بهره اطلاعات است. این مقدار بر اساس میزان آنتروپی هر ویژگی به دست می‌آید. آنتروپی در واقع نشان دهنده میزان کم اطلاعاتی است که یک ویژگی خاص در اختیار قرار می دهد که با استفاده از معادله اول به دست می‌آید. یعنی در مجموعه‌ی داده از روی یک ویژگی چقدر می‌توان تشخیص داد که دسته بندی نهایی چیست. معادله دوم نیز نحوه محاسبه مقدار بهره را نشان می‌دهد که از آن با عنوان بهره اطلاعات یاد می شود، که مقدار آن با استفاده از آنتروپی هر یک از ویژگی‌ها محاسبه شده و از این طریق میزانِ اطلاعاتی که از یک ویژگی را می‌تواند به دست آورد محاسبه می نماید. به عبارتی دیگر این معیار به ما می‌گوید که یک ویژگی خاص چقدر می‌تواند اطلاعات بیشتری راجع به خروجی به ما بدهد که بر اساس فرمول زیر محاسبه می شود.

$$Entropy(x) = -\sum_{x} P(x)\, log_2\, P(x)$$

$$Gain(x.a) = Entropy(x) - \sum_{i \in a} \frac{|x_i|}{|x|} Entropy(x_i)$$

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 ویژگی ها را همه حساب های کاربری دارا بودند و در نتیجه بهره اطلاعات آن صفر گردید. به عنوان مثال در مجموعه داده مطالعه شده، تمامی حساب های کاربری موجود در مجموعه داده، دارای Description و عکس پروفایل در حساب کاربری خود بودند، از اینرو بهره اطلاعات این ویژگی ها برابر با صفر است چون نمی تواند الگوریتم را برای تشخیص شواهد از غیر شواهد آموزش دهد. همچنین مشخص گردید که زمان ارسال توئیت ها در تمامی ساعات شبانه روز بین دسته های شواهد و غیر شواهد بطور تقریبا یکسان توزیع شده بود، از اینرو ویژگی ای مانند زمان انتشار توئیت ها ویژگی خوبی برای یادگیری الگوریتم محسوب نمی شد و حذف گردد. همچنین هیچ

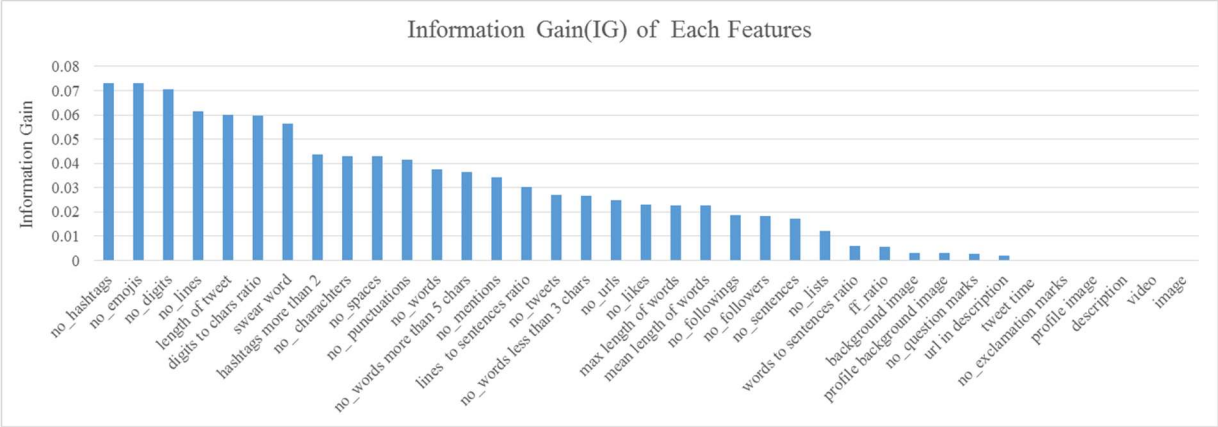
[125] Information Gain (IG)

یک از پستهای کاربرانی که متنهای شواهد و غیر شواهد توئیت کرده اند حاوی عکس و ویدئو نبوده است، لذا این باعث شد که ارزش این ویژگی نیز در این کار برابر با صفر باشد.

جدول ۷.۳. مقادیر بهره اطلاعات هر یک از ویژگی ها

No	Feature Name	IG	No	Feature Name	IG
1	No_Hashtags	0.07309	20	Max Length of Words	0.02279
2	No_Emojis	0.07305	21	Mean Length of Words	0.02258
3	No_Digits	0.07054	22	No_Followings	0.01878
4	No_Lines	0.06164	23	No_Followers	0.01825
5	Length of Tweet	0.06003	24	No_Sentences	0.01737
6	Digits to Chars Ratio	0.05954	25	No_Lists	0.01227
7	Swear Words	0.05654	26	Words To Sentences Ratio	0.00597
8	Hashtags More Than 2	0.04358	27	Ff_Ratio	0.00544
9	No_Characters	0.04293	28	Background Image	0.00313
10	No_Spaces	0.04287	29	Profile Background Image	0.00313
11	No_ Punctuations	0.04163	30	No_Question Marks	0.00258
12	No_Words	0.03741	31	URL In Description	0.00199
13	No_Word More Than 5 Chars	0.03626	32	Tweet Time	0
14	No_Mentions	0.03428	33	No_Exclamation Marks	0
15	Lines To Sentences Ratio	0.03036	34	Profile Image	0
16	No_Tweets	0.02697	35	Description	0
17	No_Words Less Than 3 Chars	0.02653	36	Video	0
18	No_Urls	0.02469	37	Image	0
19	no_likes	0.02287			

با در نظر گرفتن مقادیر بهره اطلاعاتِ ویژگی‌ها، زیر مجموعه های مختلفی از داده ها برای پیاده سازی مدل دسته بندی[۱۲۶] انتخاب گردیدند و مشخص گردید که برخی از ویژگی ها در کنار ویژگی های دیگر برای مدل یادگیری ماشین مفید هستند. بر این اساس ویژگیهای شماره ۱ تا ۲۵ به عنوان زیرمجموعه ویژگیهای مناسب، در این پژوهش انتخاب شد که می تواند الگوریتم را جهت تشخیص توئیت شواهد از غیر شواهد، آموزش دهد. نمودار۴، مقادیر مربوط به بهره اطلاعات هر ویژگی را نشان می دهد که ویژگیهای با بهره اطلاعات کمتر از ۰.۰۱ از زیر مجموعه ویژگی ها حذف گردیدند. از میان زیر مجموعه انتخاب شده تعداد ۲۰ ویژگی مبتنی بر متن[۱۲۷] و ۵ ویژگی مبتنی بر حساب کاربری[۱۲۸] می باشد.



نمودار۴.۳. مقادیر مربوط به بهره اطلاعات هر یک از ویژگی ها

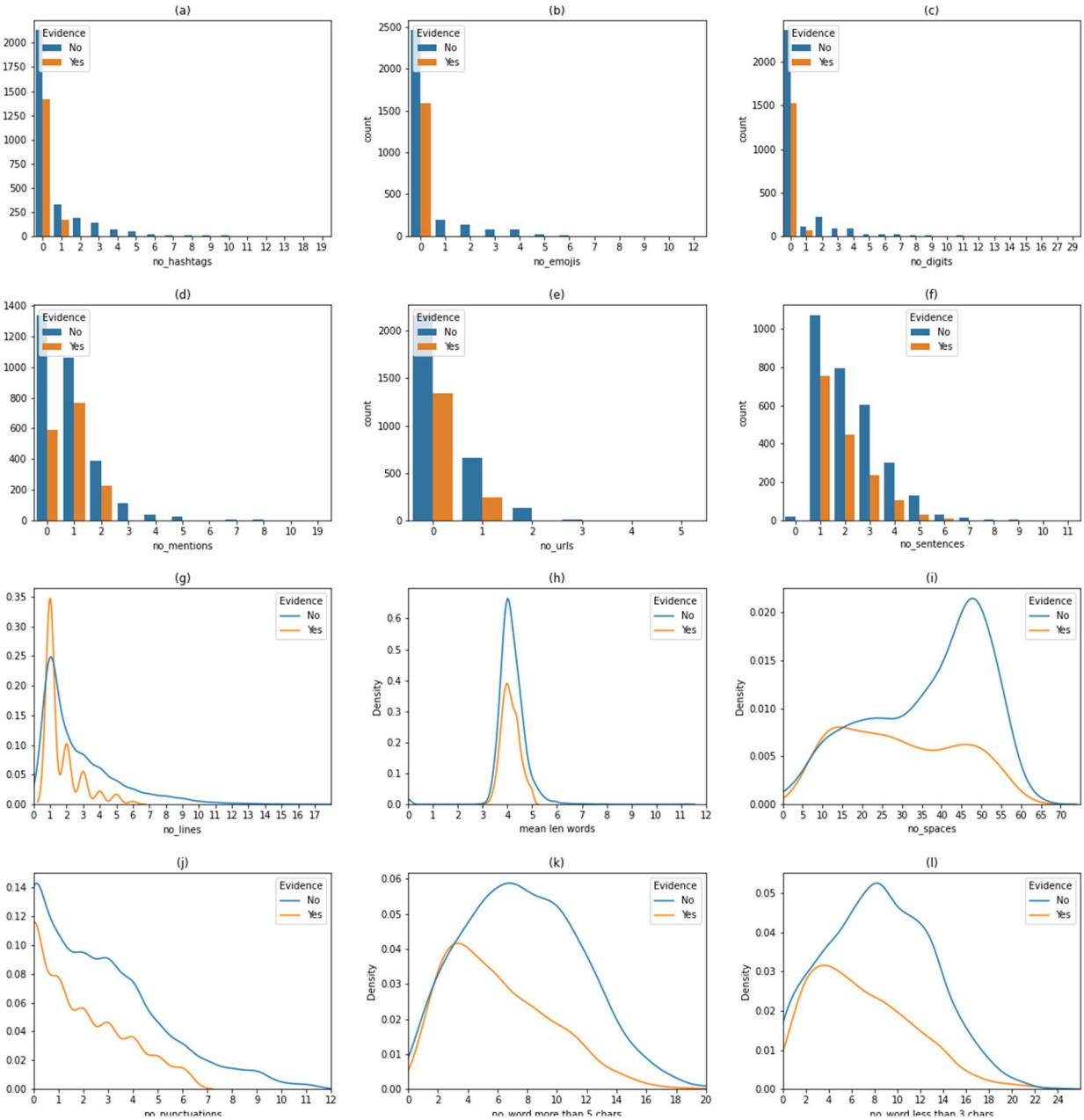
۶.۳. شناسایی رفتار کاربران منتشر کننده توئیت شواهد

به دنبال یکی از اهداف پژوهش، یعنی کشف الگو ها و رفتارهای کاربرانی که پیام های شواهد ارسال می کنند، تلاش شد تا رفتارهایی که این کاربران در تولید شواهد از خود نشان می دهند، مشخص گردد. همانگونه که در نمودارهای زیر قابل مشاهده است، شواهدی که کاربران توئیتر تولید می کنند دارای ویژگی هایی است. این ویژگی ها در قالب نمودارهای زیر نمایش داده شده است. بر اساس نمودار (a) اکثر توئیت های شواهد حاوی هشتگ نبوده و تعداد بسیار کمی صرفا دارای یک هشتگ هستند. توئیت هایی که می توان آنها را به عنوان شواهد قلمداد کرد، دارای هشتگ نبودند یا هشتگ های کمی داشتند. این موضوع نشان می دهد، که متن هایی که می توان آنها را به عنوان

[126] Classification Model
[127] Content Based
[128] Account Based

شواهد در این حوزه قلمداد کرد، غالبا توسط کاربر هشتگ گذاری نمی شود. نمودار (b) نشان می دهد در اکثر توئیت های شواهد، هیچ کاراکتر ایموجی وجود نداشته است. یعنی کاربرانی که می توان توئیت آنها را به عنوان شواهد قلمداد کرد، در توئیت خود در اغلب موارد ایموجی استفاده نمی کنند. عدم استفاده از ایموجی در اغلب توئیت های شواهد، نشان دهنده فضای رسمی انتشار این توئیت ها می تواند قلمداد شود. با توجه به نمودار (c) می توان دریافت که در ۹۶ درصد از توییت های شواهد، عدد بکار نرفته است. در واقع می توان گفت، کاربران منتشر کننده شواهد، بسیار کم از عددها و ارقام استفاده کرده اند. توئیت های شواهد بجای ارائه اعداد و ارقام، در پی توضیح، تبیین و یا انتقاد از موضوعات بوده اند. همچنین نمودار (d) نشان میدهد که اکثر کاربران تولید کننده شواهد، تمایلی به استفاده از بیش از یک خطاب[۱۲۹] در توئیتهایشان ندارند که به نظر می رسد به این دلیل باشد که کاربر به دنبال انتشار نظر شخصی خود یا پاسخ به توییت کاربر مشخص دیگری باشند که نظر خود را به اشتراک گذاشته است و به دنبال خطاب و جلب توجه دیگر کاربران نیست. نمودار (e) بیانگر عدم وجود لینک URL در بیش از ۸۷ درصد از توئیتهای شواهد است و ۱۳ درصد دیگر فقط حاوی یک لینک URL هستند. این موضوع نشان می دهد استناد به متن های خارج از پلتفرم توئیتر در تولید شواهد، کمتر اتفاق افتاده است و شواهد درون پلتفرمی تولید و توزیع شده اند. همچنین در نمودار (f) می توان مشاهده نمود که کاربران شواهد غالباً تمایل به نوشتن نظر خود در یک جمله یا کمتر از ۳ جمله دارند. که نشان از اختصار گویی توئیت های شواهد می باشد. به همین ترتیب بر اساس نمودار (g) کاربران بیشتر تمایل به استفاده از یک خط نوشته دارند و تمایلی به انتشار توئیت های طولانی و در چند خط پیاپی در این رابطه وجود ندارد. و مطابق نمودار (i) نزدیک به ۷۰ درصد از کاربران از تعداد ۵ تا ۴۰ فاصله در متن های شواهد برای بیان نظرات خود استفاده کرده اند و ۳۰ درصد کاربرانی که شواهد تولید کرده اند، خارج ازاین بازه از فاصله در متن استفاده کرده اند. در ۳۶ درصد از پیامهای غیر شواهد، ۵ تا ۴۰ فاصله در متن استفاده شده، و در ۶۴ درصد دیگر از پیام های غیر شواهد، خارج از این بازه از فاصله در متن های خود استفاده کرده اند. به علاوه از الگوهای به دست آمده در رفتار کاربران می توان به استفاده کمتر از تعداد علائم نگارشی، استفاده کمتر از کلمات با طول بیشتر از پنج حرف همچنین بکارگیری محدود تعداد کلمات با طول کمتر از ۳ حرف اشاره کرد. همچنین الگوهای رفتاری که از ویژگی های مرتبط با حساب کاربری(دنبال کننده و دنبال شونده و لیست ها) استخراج می شود، در نمودار های زیر نمایش داده شده است. در استخراج الگوهای رفتاری مبتنی بر حساب کاربری، مشخص گردید که ۷۵ درصد از کاربرانی که پیام شواهد ارسال کرده اند کمتر از ۱۷ هزار توئیت ارسالی داشتند و ۲۵ درصد از این افراد که شواهد تولید کرده اند بیش از ۱۷ هزار توئیت داشته اند. این در حالی است که ۳۵ درصد از کاربرانی که پیام های غیر شواهد ارسال کرده اند زیر ۱۷ هزار توئیت داشته اند و ۶۵ درصد از این افراد، بالای ۱۷ هزار توئیت داشته اند.

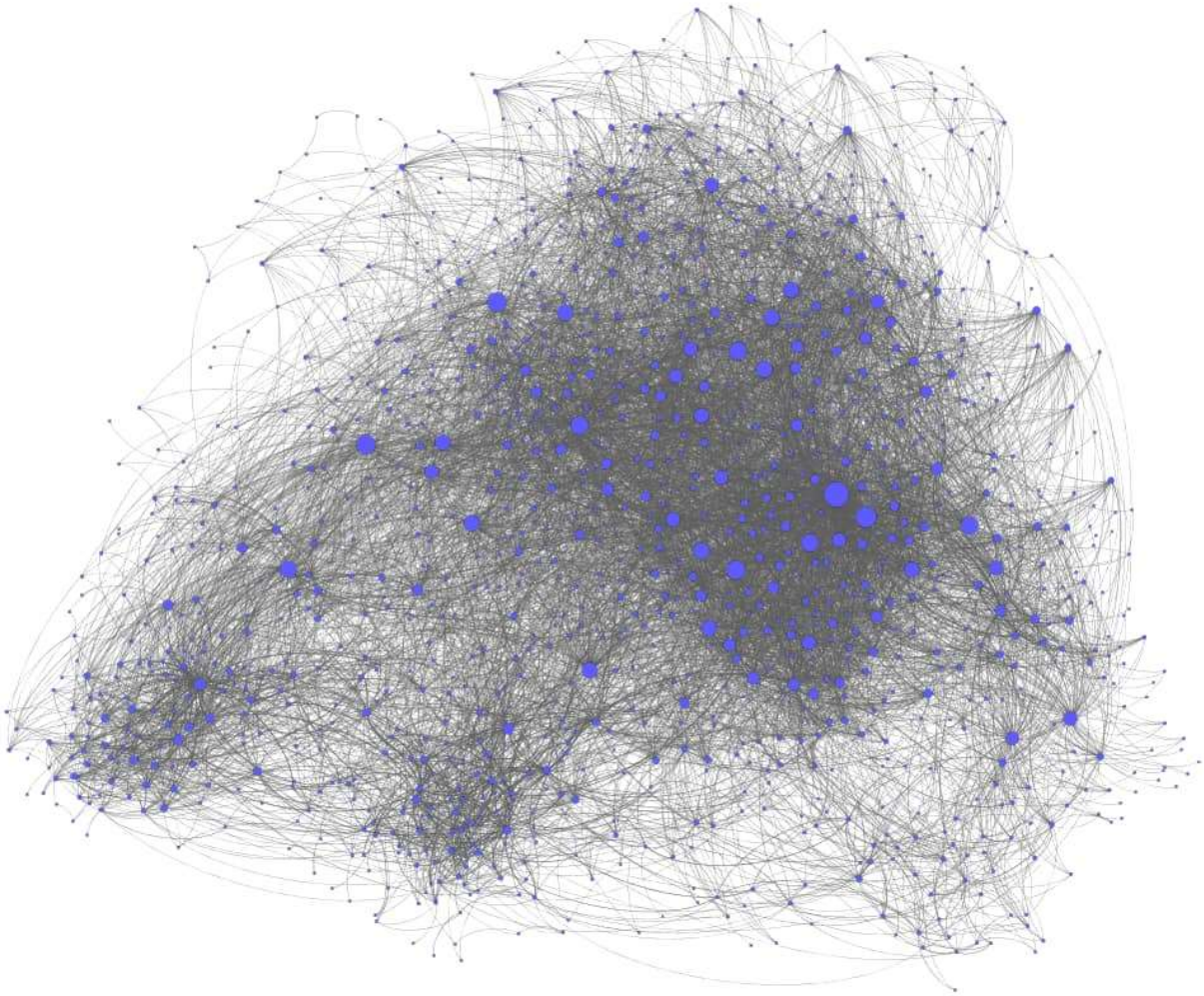
[129] Mention (@)



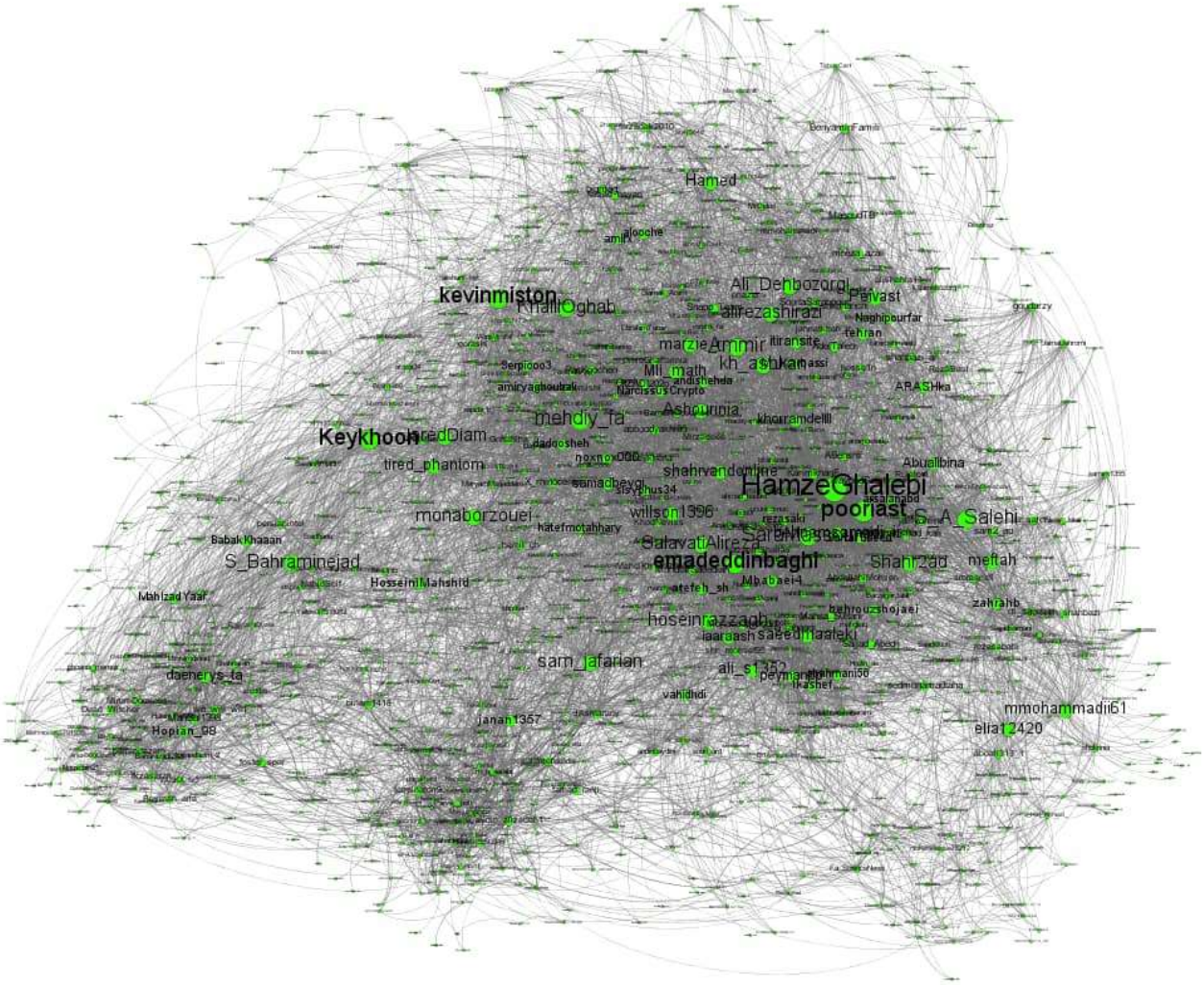
نمودار۵.۳. مجموعه اطلاعات مرتبط با ویژگی ها

یکی از فعالیت های تحلیلی که در رابطه با شناسایی رفتار کاربران و اجتماعات ایجاد شده توسط آنها در شبکه های اجتماعی انجام می شود، استفاده از اطلاعات مربوط به گراف شبکه ی ساخته شده از کاربران و نوع فعالیت های آنها است. با استفاده از ابزار API توئیتر اقدام به جمع آوری نام کاربری دنبال کنندگان(فالوئینگ های) هر یک از کاربرانی که پیام شواهد ارسال کرده بودند گردید. پس از استخراج روابط بین آنها گراف شبکه ایجاد شد. در این

شبکه کاربران گره های گراف و ارتباط بین آنها بعنوان یال در نظر گرفته شده است. می توان با استفاده از یال های یک گره یا یال های موجود بین هر گره میزان ارتباط بین آنها را بدست آورد. در این شبکه کاربرانی هستند که تعداد یال های آنها از دیگر گره ها بیشتر است این افراد بعنوان کاربران مهم و تاثیر گذار در شبکه شناخته می شوند. تاثیرگذارترین 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 ارسال کنندگان توئیت شواهد، گراف آنها ترسیم شد. در این گراف، تاثیرگذاران شبکه و میزان ارتباطات آنها مشخص شده است.



شکل۳.۳. گراف شبکه کاربران منتشر کننده توئیت شواهد



شکل۴.۳. گراف شبکه کاربران منتشر کننده توئیت شواهد بهمراه نام کاربری آنها

همانگونه که در گراف بالا مشخص است، تاثیرگذاران شبکه توئیت هایی ارسال کرده اند که باعث برانگیخته شدن دیگر کاربران و در نتیجه اشتراک زیاد آن متن در یک شبکه بزرگ از توئیتر شده است. در زیر متن توئیت های ۱۱ نفر از نفوذدارترین کاربران توئیتر که در کلید واژه های استخراج شده توئیت ارسال کرده اند به همراه تعداد ارتباطاتی که توئیت آنها برانگیخته است، آورده شده است. معیار انتخاب این کاربران، تعداد ارتباطاتی بوده که ایجاد کرده بودند(بیش از ۳۰۰ ارتباط) همانگونه که در جدول ۶ دیده می شود، برخی از این کاربران نیز بیش از یک توئیت منتشر کرده اند که همگی توانسته شبکه خوبی از ارتباطات را ایجاد کند. همانگونه که از متن پرنفوذترین توئیت ها مشخص است، این کاربران بیش از همه موضوع فیلترینگ اطلاعات، اپلیکیشن ها و شبکه های اجتماعی را مد نظر داشته اند. موضوع بعدی دسترسی به اطلاعات و حریم خصوصی اطلاعات بوده است که مد نظر کاربران قرار گرفته است.

جدول ۸.۳. پرنفوذترین کاربران در کلید واژه های استخراج شده

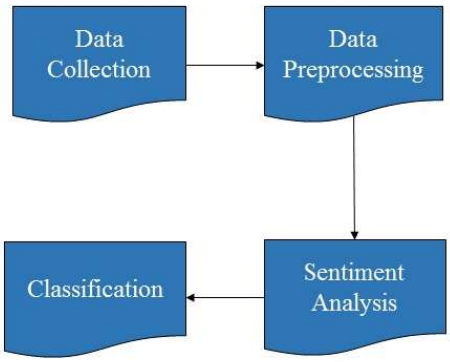
ردیف	تعداد یال ها	نام کاربری	متن توئیت
۱	۱۰۶۰	HamzeGhalebi	مهم ترین چالش امنیتی فیلترینگ گسترده، انتقال بخشی از مرجعیت اطلاع رسانی به خارج از کشور است. دلیل مهمش هم این است که اساسا سیاستگذار به یک سری از سوالات و موضوعات که برای مردم مساله است نمی پردازد. مثل سیاست فیلترینگ اینترنت. وقتی کلا همه چیز فیلتر است اثر فیلترینگ کم می شود.
۲	۵۸۵	daenerys_ta	@ShahramBahrami8 حضور پر رنگ و دلسوزانه شما در توئیتر فارسی باعث دلگرمی ماست. یا باید رفع فیلترینگ کنن یا تمام حساب‌های مقامات ایران در تمام شبکه‌های اجتماعی فیلتر مسدود بشه. فکر می‌کنم آمریکا این قدرت رو داشته باشه. بهتره به سیاست‌مدار ها اعلام کنیم مردم.
۳	۵۸۲	willson1396	@azarijahromi میشه شمام فیلترینگ رو اول خودت تحمل کنی؟
۴	۴۱۰	pooriast	@azarijahromi @barari_ir @mrahimi واقعا منم میخوام آتیش بزنم، 🙃 حالا اگر 🔥 نه ولی خوب دسترسی به اطلاعات و ایجاد یک شبکه ارتباطی برای علاقمندان نجوم و برنامه فضایی باید و میتونه عملی باشد.
			پس فیلترینگ یا نظارت یا دسترسی به اطلاعات مردم خوب است. #رمزارز
			@azarijahromi با این فیلترینگ شدید، قطع دم به دم و نقشه‌های شوم برای اینترنت، توسعه اقتصاد دیجیتال چیز آزاردهنده اگر نه نفرت‌انگیزی است.

			شما- نه فقط شخص شما همه با کمک هم- مبنای اینترنت را لگدمال کرده‌اید الان نوید اقتصاد دیجیتال و به اشتراک گذای اطلاعات نوعی دهن‌کجی است.
			امروز هرکس را دیدم شلیک موشک به هواپیما را باور کرده بود. خواستم به سرباز خبرنگارها تبریک بگم عملیات موفق بوده است. با تشکر از جمهوری اسلامی و تدبیر کنترل افکارعمومی با فیلترینگ.
			در قضاوتی حیاتی، دیوان عالی هند آزادی بیان را ارجح دانست و #قطع_اینترنت در این کشور را ممنوع کرد. چندی پیش نیز دیوان عالی ترکیه #فیلترینگ ویکی‌پدیا را ممنوع کرد و این پلتفرم آزاد شد.
			@aboali313_ عزاداری و انتقام و وحدت را ول کردید افتادید دنبال جهرمی و اینستاگرام، آقا جهرمی را استیضاح کنید، اعدام کنید، اما ملت را خلع‌سلاح نکنید و وحدت را مخدوش نکنید. فیلترینگ شبکه‌های اجتماعی خیانت بوده است.
			این لاشخوری برای #فیلترینگ در شرایطی که بیش از همیشه نیاز به اطلاعات داریم #خائنانه است الان باید توئیتر، فیسبوک و یوتیوب نیز باز شود تا وحدت ملی در سطح جهانی طنین‌انداز شود نه اینکه ابزارهای محدود را هم از خود خلع کنیم. با همین اینستاگرام موج حمایت طنین‌انداز شده و غضب دشمن طبیعه https://t.co/icW3OSCGVB
۵	۳۷۸	peymanpp	در زمانهای دور، قدرت داشتن به معنی دسترسی به اطلاعات بود. اکنون قدرت یعنی آگاهی بر اینکه از چه اطلاعاتی باید چشم پوشید
			انسانها واقعا نمی‌دانند به چه اطلاعاتی توجه نشان دهند، و اغلبْ وقت خود را صرف بحث و بررسی درباره‌ی موضوعات فرعی می کنند.
			@s7az2mm اینها اطلاعات شخصی مردم هست جناب جرجاندی ما حق انتشار نداریم . نقض حریم خصوصی بسیار کار سخیفی است.
۶	۳۶۶	dadoosheh	@ANGREEOMAN خود کرومکست. من با ایفون هرچی تو یوتیوب و نتفلیکس ببینم می‌اندازم رو تلویزیون. میخوام ببینم در ایران ملت همینجوری می‌تونن کار کنن باهاش یا فیلترینگ نمی‌گذاره
۷	۳۶۰	mehdiy_fa	@Ardeshir دقیقا. مورد شرکت آپارات گفته‌های دیگری هم دارد از جهت رانت حکومتی که بواسطه فیلترینگ یوتیوب دریافت کرده و احتمالا کمکهای مالی حکومت را دریافت کرده.

۸	۳۳۰	M_Karbassi	@majidhn371 میشه فراتر از کلیات حجم افشای اطلاعات کاربران رو بگید ؟ چه اطلاعاتی از مردم دست هکرها افتاده تا حداقل بدونن مراقب سوء استفاده‌های احتمالی باشن ؟ به کاربرهایی که اطلاعات‌شون لو رفته شخصا اطلاع‌رسانی شده ؟
۹	۳۲۷	kh_ashkan	@hamshahrinews نوشته ، دسترسی به اطلاعات تازه سخت و عملا غیرممکن بوده چون گوگل در دسترس نیست و یوز –موتور جستجوی ایرانی که ادعا شده در شبکه ملی اطلاعات جایگزین گوگل میشود، کارایی ندارد. قبلا هم کارشناسان به یوز ایراد میگرفتند که بخاطر حجم بالای کاربرانش
۱۰	۳۰۹	Ali_Dehbozorgi	@alirezashirazi بخش دیگری از سود غیر ملموس دسترسی به اطلاعات کاربران است که شهرداری خود رقیب محسوب می شود. دقت کنیم مسئله رابه تاکسی کوچک نکنیم این یک فرایند دولتی است که برای سهیم شدن درهمه درآمد کسب و کارهای نوآور طراحی شده تا به نام مجوز مرجع شوند.
			@alirezashirazi منظورم را از بخش دوم توییت درمورد سود را اصلاح میکنم: شهرداری در تدارک یک فرایند درآمد زایی درقالب یک عوارض جدید به نام نظارت و مجوز دهی است. در صورتی که تنوعی از سهم عوارض و مالیات ها را اکنون دریافت می کند. همچنین دسترسی به اطلاعات کاربران.
			@AliFiroozi19 لطفا موضوع فیلترینگ کسب و کارهای اینترنتی مجاز را هم بدون دعوت و صحبت با مدیران را هم پیگیری بفرمایید.
۱۱	۳۰۸	ali_s1352	@azarijahromi@ pooriast جناب استراکی به نظرم وزیر جوان مدافع سرسخت آزادی اینترنت بین المللی است، هر چند اینترنت داخلی هم برای امنیت کشور ضروری است! اما متوجه این نگرانی شما از بابت فیلترینگ هم هستم، ولی بعید میدانم در این مملکت، وزیر چندان قدرتی در این رابطه داشته باشد.
			به نظرم #فیلترینگ_توئیتر توسط #جمهوری_اسلامی و #ساسپند کردن #اکانت_توئیتر از طرف #دولت_امریکا در یک راستا است! چرا همه میخوان #ملت_ایران را خفه کنند؟؟؟ اونوقت خودشون میان #توئیت_فارسی میزنن!!! این چه وضعشه به حضرت عباس؟؟؟

۷.۳. تحلیل احساس[۱۳۰]

در این بخش به جهت 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 کرده ایم. تحلیل احساس یکی از روش های پردازش زبان طبیعی است که با نام های دیگری از قبیل Review Mining, Sentiment Mining, Subjectivity Mining و غیره شناخته می شود و روش های محاسبه گوناگونی دارد(Mittal & Patidar, 2019). تحلیل احساس در توئیتر به دلایل زیر با برخی چالش ها همراه است: ۱. محدودیت طول متن توییت که دارای محدودیت ۲۸۰ کاراکتری است منجر به تولید متن های فشرده[۱۳۱] می شود. ۲. استفاده از کلمات ناسزا، باعث ایجاد دشواری در تشخیص معنای آنها برای ماشین می شود ۳.در توئیتر استفاده از URL ها، هشتگ ها یا خطاب ها[۱۳۲] مجاز است که در فرایند تحلیل احساس مشکل ایجاد می کند. نمودار زیر فرایند تحلیل احساس را نشان می دهد.



نمودار۶.۳. فرایند تحلیل احساس توئیت ها

پس از استخراج داده ها، فرایند تحلیل احساس شامل ۳ بخش پیش پردازش، تشخیص احساس و دسته بندی انجام می شود. در پیش پردازش گام هایی صورت گرفت: ۱.حذف توییت های تکراری و ریتوییت ها. در این رابطه تمامی توئیت ها بررسی و تمامی توئیت های تکراری و ریتوئیت های تکراری از مجموعه داده ها حذف شدند. ۲.حذف کاراکترهای خاص مانند URL ها، اعداد و علائم نگارشی یکی از مراحل بود که تلاش شد روی مجموعه داده های شواهد انجام شود. ۳.حذف استاپ ورد ها[۱۳۳] که شامل حروف ربط، ضمایر و غیره می شود نیز از جمله مراحلی بود که

[130] Sentiment Analysis

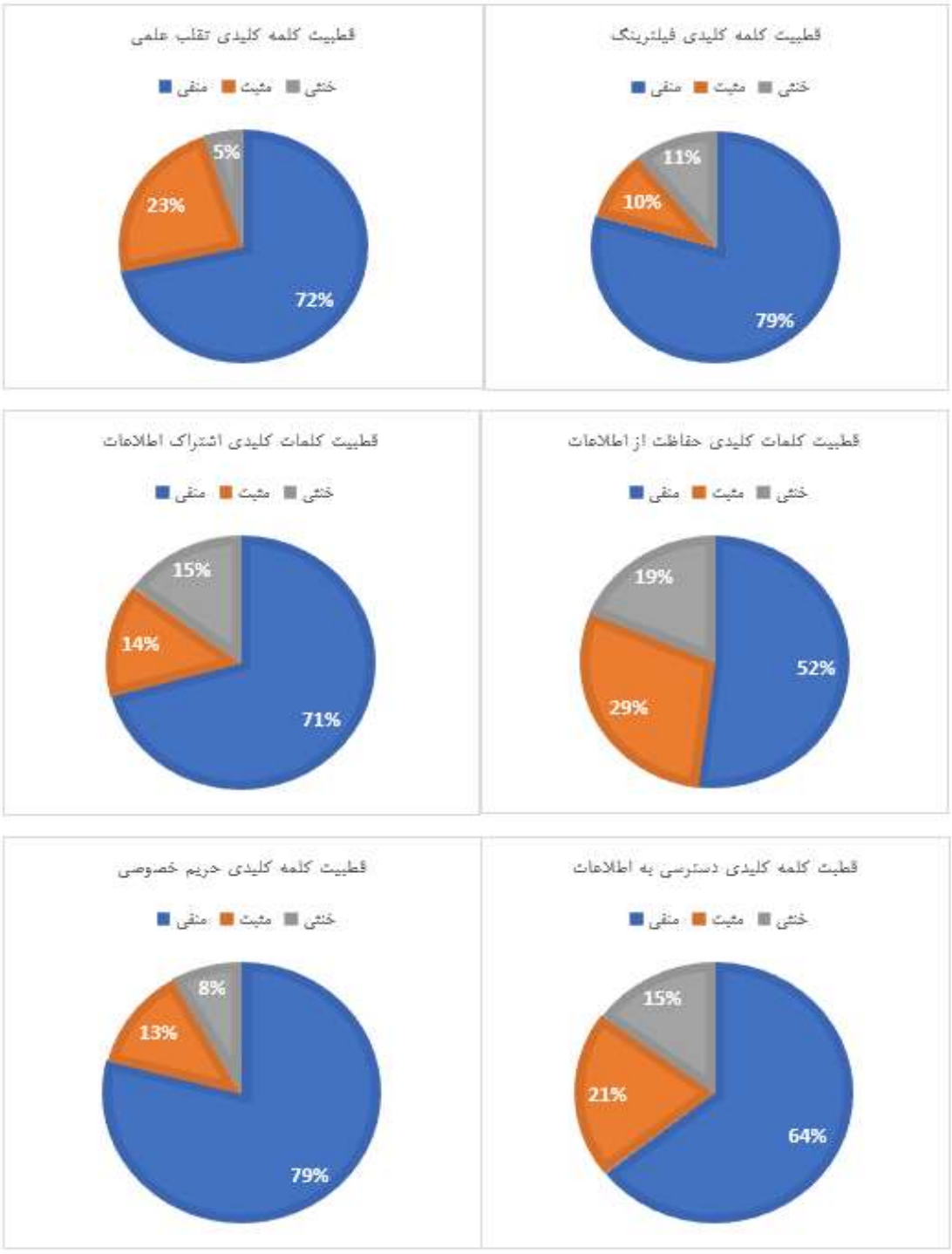
[131] Intensive Statement
[132] Mention
[133] Stop Words

انجام شد. چرا که این حروف ارزش تحلیلی در خصوص تحلیل احساس را ندارند. ۴.حذف کلمات غیر فارسیدر مرحله بعد انجام شد. ۵. بازگرداندن هر کلمه به شکل اصلی آن کلمه[۱۳۴]. به عنوان مثال کلمه کتاب‌ها در این مرحله به کلمه کتاب تبدیل شود. ۶. متن بصورت توکن های کوچکی از کلمات در آمد[۱۳۵].

پس از مراحل پیش پردازش، 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 امتیاز دهی شده است تعیین می‌گردد. بعنوان 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 توییت باید اینکه کلمات یک توییت تا چه حد عقیده و نظر شخصی کاربر است[۱۳۶] و اینکه چقدر بر اساس یک حقیقت و واقعیت بیرونی و بر اساس مشاهدات یا اندازه گیری ها[۱۳۷] است، مشخص گردد. با محاسبه این دو میزان احساس یک توییت محاسبه می گردید و در گام سوم توییت ها در قالب سه دسته مثبت، منفی و خنثی دسته بندی شدند. در زیر قطبیت برخی از کلمات کلیدی پر تکرار، آورده شده است. همانگونه که مشاهده می شود فیلترینگ و حریم خصوصی بیشترین قطبیت منفی را داشته اند و توئیت های مرتبط با حفاظت از اطلاعات بیشترین قطبیت مثبت را دارا می باشد.

مشاهده و بررسی مجموعه توئیت های استخراج شده و همچنین توئیت هایی که به عنوان شواهد جدا شدند، نشان دهنده حساسیت بالای کاربرانی است که در رابطه با کلید واژه های استخراج شده، توئیت منتشر کرده اند. میزان قطبیت استخراج شده در بیشتر کلید واژه ها منفی بوده که نیازمند مداخله فوری سیاستگذار در این حوزه ها می باشد.

134 Stemming
135 Tokenization
136 Subjective
137 Objective



نمودار۷.۳. نمودار قطبیت کلمات کلیدی

۸.۳. دسته بندی[۱۳۸]

پس از مرحله انتخاب ویژگی های مناسب که بر اساس بهره اطلاعات، بتواند پیش بینی خوبی برای الگوریتم ایجاد کند، فرایند تشخیص پیام های شواهد به وسیله الگوریتم های یادگیری ماشین، آغاز شد. از آنجایی که در تشخیص شواهد به دنبال خروجی مشخص و برچسب هر نمونه هستیم، رویکرد پیشنهادی از نوع یادگیری با نظارت و دسته بندی[۱۳۹] بود. همانگونه که ملاحظه شد، مدل اصلی ارائه شده(نمودار۱)، شامل ساختاری با یک دسته بند نظارت شده که پیام‌های شواهد را تشخیص می‌دهد، طراحی شده است. بدین منظور نیازمند یادگیری دسته‌بندها[۱۴۰] هستیم. فرایند یادگیری دسته بندها ابتدا با بخشی از داده های برچسب زده شده، اتفاق می افتد و سپس با دریافت دانش، الگوریتم عمل پیش‌بینی برچسب را برای نمونه های ورودی جدید انجام خواهد داد. برای طی کردن این دو گام که شامل بخش یادگیری و آزمون است، در ابتدا ویژگی های استخراج شده مربوط به نمونه S به صورت 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 می شود. بنابراین هر نمونه در مجموعه داده به شکل $Data=\{(S_1,L_1),(S_2,L_2),(S_3,L_3),\ldots,(S_n,L_n)\}$ در می آید. برای مرحله آموزش، بخش عمده‌ای از مجموعه داده ها انتخاب شده و به منظور یادگیری دسته بندها، داده های 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 دسته بند آن نمونه ها را ندیده است، بدون برچسب هر نمونه به شکل $Test=\{S_1,S_2,\ldots,S_n\}$ برای پیش بینی برچسب آنها به دسته بند داده می‌شود. سپس توسط دسته بند آموزش دیده شده به هر نمونه برچسب دسته ای مشخص تعلق می گیرد. در نهایت دسته های پیش بینی شده هر نمونه با برچسب اصلی آن نمونه مقایسه شده و عملکرد دسته بند توسط معیارهای ارزیابی مورد سنجش قرار می گیرد.

۱.۸.۳. معیارهای ارزیابی[۱۴۱]

در راستای ارزیابی عملکرد مدل دسته بندی جهت تشخیص شواهد، از معیارهایی که به صورت گسترده ای در حوزه ارزیابی دسته بندها توسط محققان استفاده شده، استفاده گردید(Chen et al., 2016). این معیارها بر اساس دو تعریف اصلی محاسبه می گردند:

---

[138] Classification

[139]

[140]

[141] Evaluation Metrics

۱. مثبت و منفی: اگر دسته شواهد را به عنوان دسته(E) در نظر گرفته شود و دسته بند C یک نمونه را در دسته (E) پیش بینی کند، نتیجه پیش بینی برای آن نمونه به صورت مثبت ثبت می شود. اگر نمونه را به عنوان غیر E پیش بینی کند، نتیجه پیش بینی آن نمونه به صورت منفی در نظر گرفته می شود.

۲.درست و غلط: اگر مقدار پیش بینی شده توسط دسته بند C برای یک نمونه به درستی دسته E تشخیص داده شده باشد مقدار درست(T) و اگر به اشتباه تشخیص داده شده باشد مقدار اشتباه(F) تعلق میگیرد.

True Positive (TP): توییت هایی که متعلق به دسته E هستند و به درستی در دسته E دسته‌بندی شده‌اند.

False Positive (FP): توییت هایی که متعلق به دسته E نیستند به اشتباه در دسته E دسته‌بندی شده‌اند.

True Negative (TN): توییت هایی که متعلق به دسته E نیستند و به درستی در دسته غیر E دسته‌بندی شده‌اند.

False Positive (FN): توییت هایی که متعلق به دسته E هستند و به اشتباه در دسته غیر E دسته بندی شده‌اند.

مقادیر بالا برای یک دسته بند از طریق ماتریس در هم ریختگی[۱۴۲] نشان داده شده در جدول زیر قابل محاسبه می باشد.

		Prediction	
		Evidence	Non-Evidence
Label	Evidence	True-Positive (TP)	False-Negative (FN)
	Non-Evidence	False-Positive (FP)	True-Negative (TN)

نمودار۳.۸. ماتریس درهم ریختگی

142

با استفاده از مقادیر بالا عملکرد یک دسته بند توسط معیارهای دقت[۱۴۳]، صحت[۱۴۴]، فراخوانی[۱۴۵] و میانگین هارمونیک[۱۴۶] قابل ارزیابی هستند.

معیار دقت (Accuracy)، یکی از معیارهای مهم و متداول دسته بندی میزان درستی پیش بینی ها است. این معیار میزان پیشگویی های درست دسته بند نسبت به کل نمونه از مجموعه داده را نشان می دهد. در واقع میزان دقت یک الگوریتم به صورت زیر محاسبه میگردد.

$$Accuracy = \frac{TP + TN}{S}$$

معیار صحت(Precision)، نسبت نمونه‌های به درستی پیش‌بینی شده در دسته E نسبت به کل نمونه‌هایی که به عنوان دسته E توسط دسته بند پیش بینی شده اند که از معادله زیر قابل محاسبه می باشد.

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

معیار فراخوانی(Recall)، نسبت نمونه‌های به درستی پیش‌بینی شده در دسته E نسبت به تمام نمونه های متعلق به دسته E است.

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

میانگین هارمونیک f(F-Measure)، در واقع میانگین هارمونیک معیارهای صحت و فراخوانی است که برای ارزیابی عملکرد دسته بندی در هر دسته استفاده می شود و از طریق فرمول زیر به دست می آید.

$$F - measure = \frac{2TP}{2TP + FP + FN}$$

همه این چهار معیار ارزیابی مدل دسته بندی انتخاب شده، جهت ارزیابی این مدل از اهمیت زیادی برخوردار بودند که در زیر نحوه محاسبه آن آمده است.

[143] Accuracy
[144] Precision
[145] Recall
[146] F-Measure

۲.۸.۳. ارزیابی مدل مناسب دسته بندی جهت تشخیص شواهد از غیر شواهد

مجموعه داده های مورد استفاده در این پژوهش جهت پیاده سازی مدل دسته بندی پیشنهادی، شامل ۴۵۶۰ نمونه بود که به دو دسته تقسیم شدند. نمونه ای که برچسب شواهد(۲۹۷۸) داشته و نمونه ای که برچسب غیر شواهد(۱۵۸۳) داشتند. به منظور یادگیری مدل پیشنهادی بود که مجموعه داده به دو بخشِ مجموعه آموزش[۱۴۷] (شامل تعداد ۴۱۰۴ نمونه) و مجموعه آزمون[۱۴۸] (شامل تعداد ۴۵۶ نمونه) تقسیم شده اند. مجموعه داده های آزمون شامل ۲۵۴ نمونه با برچسب غیرشواهد و ۲۰۲ نمونه با برچسب شواهد بوده است. در این بخش در راستای تشخیص شواهد[۱۴۹]، عملکرد ۶ دسته بند[150] شامل ماشین بردار پشتیبان[۱۵۱]، درخت تصمیم[۱۵۲]، X-GBoost ، K نزدیکترین همسایه[۱۵۳]، آنالیز تشخیصی خطی[۱۵۴] و رگرسیون منطقی[۱۵۵] با هم مقایسه گردید. ابتدا مدل های دسته بند بر روی مجموعه داده آموزش پیاده سازی شدند و پس از فرایند یادگیری، دسته بندهای یادگیری شده[۱۵۶] روی مجموعه آزمون پیاده سازی شد و نتایج عملکرد آنها مطابق با جدول۶ بر اساس معیارهای ارزیابی[۱۵۷] برای هر کدام از الگوریتم ها نشان داده شده است.

جدول۹.۳. معیارهای ارزیابی الگوریتم های 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

[147] Train Set
[148] Test Set
[149] Evidence Detection
[150] Classifiers
[151] Support Vector Machine (SVM)
[152] Decision Tree (DT)
[153] K Nearest Neighbor (KNN)
[154] Linear Discriminant Analysis (LDA)
[155] Logistic Regression (LR)
[156] Trained Classifiers
[157] Evaluation Metrics

همانگونه که آمد، به منظور ارزیابی مدل تشخیص شواهد از معیارهای دقت، صحت، فراخوانی و میانگین هارمونیک استفاده شد. این معیارها با استفاده از مقادیر تعیین شده در جدول ماتریس درهم ریختگی[۱۵۸] قابل محاسبه است. برای دسته بند درخت تصمیم(الگوریتم درخت تصمیم) مشخص شد که از میان ۲۰۲ نمونه شواهد تعداد ۱۸۲ پیام را با عنوان TP به درستی و ۲۰ پیام را با عنوان FN به اشتباه برچسب غیرشواهد زده است. همچنین از میان ۲۵۴ نمونه غیر شواهد تعداد ۲۰۶ نمونه را با عنوان TN به درستی و تعداد ۴۸ نمونه را با عنوان FP به اشتباه برچسب گذاری نموده است. این مقادیر در ماتریس زیر قابل مشاهده است.

		Prediction	
		Evidence	Non-Evidence
Actual Label	Evidence	182 (TP)	20 (FN)
	Non-Evidence	48 (FP)	206 (TN)

نمودار۳.۷. ماتریس درهم ریختگی

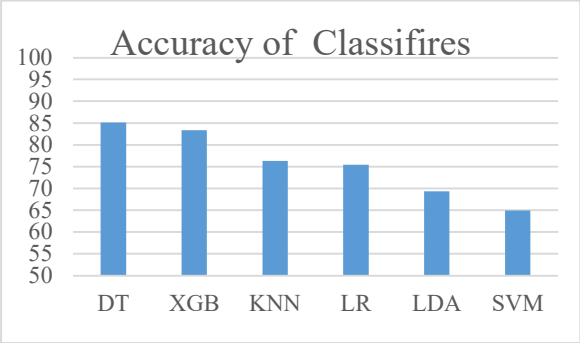
معیار میانگین هارمونیک F در میان دسته بندها برای درخت تصمیم و XGBoost از دیگر دسته بندها بالاتر بود. همانطور که مشاهده می شود، درخت تصمیم و XGBoost توانسته اند به ترتیب مقادیر ۸۴.۲۶ و ۸۱.۵۵ درصد را کسب کنند. پس از آنها الگوریتم های دسته بند K نزدیکترین همسایه، رگرسیون منطقی، LDA و ماشین بردار پشتیبان قرار گرفتند. از جدول زیر می توان نتیجه پیاده سازی دسته بندها را بر روی مجموعه داده های آزمون مشاهده کرد

[158] Confusion Matrix

که درخت تصمیم با ۸۵.۰۹ درصد عملکرد بهتری در مقایسه با دیگر الگوریتم های دسته بند داشته است. جدول۷، میزان دقت عملکرد دسته بندها را نسبت به یکدیگر نشان می‌دهد.

جدول۳.۱۰. میزان دقت عملکرد دسته بندها

Algorithm	Accuracy
Decision Tree (DT)	85.09
XGBoost	83.33
K-Nearest Neigbor(KNN)	76.32
Logistic Regression(LR)	75.44
Linear Discriminate Analasys(LDA)	69.30
Support Vector Machine(SVM)	64.91



نمودار۳.۸. میزان دقت عملکرد دسته بندها

در راستای تشخیص توئیت های شواهد از غیر شواهد، مدل پیشنهادی این پژوهش توانست با دقت ۸۵.۰۹ درصد توئیت های شواهد را از غیر شواهد تشخیص دهد. این میزان از دقت، نسبت به مطالعات پیشین در این حوزه، قابل قبول محسوب می شود. همچنین با توجه به اینکه مدل ارائه شده، جدید بوده و ویژگی ها و دسته بندی های ارائه شده، نیز جدید بوده اند، این میزان از دقت کاملا مطلوب می باشد. تحلیل بیشتر راجع به روش ارائه شده نشان می دهد که ویژگی های مبتنی بر متن بهتر از ویژگی های مبتنی بر حساب کاربری در این کار تاثیر دارد. این موضوع را می توان

اینگونه توضیح داد که کاربرانی که توئیت هایی از جنس شواهد ارسال می کنند، می توانند در مقطعی پیام شواهد ارسال کنند و در مقطعی پیام غیر شواهد ارسال کنند. در مورد کاربرانی که توئیت های غیرشواهد ارسال کرده اند نیز به همین صورت می باشد. به همین دلیل ویژگی های مبتنی بر متن در این کار از اهمیت بالاتری برخوردار بودند که در نتیجه می توان ادعا کرد، ترکیب این دو دسته از ویژگی ها می تواند دقت مدل را افزایش دهد.

الگوریتم انتخاب شده برای مدل پیشنهادی که اکنون یادگیری در مورد آن اتفاق افتاده است، اکنون می تواند با هر مجموعه داده ای از شبکه اجتماعی توئیتر با دقت حدود ۸۵ درصد کار کند و توئیت های شواهد را از غیر شواهد جداسازد. این مدل در واقع کاربست داده های یکی از مهمترین و تاثیرگذارترین شبکه های اجتماعی فارسی زبان را در فرایند سیاستگذاری علم و فناوری، با رویکرد سیاستگذاری مبتنی بر شواهد، مدل کرده است. سیاستگذار با کاربست الگوی سیاستگذاری پیشنهادی و همچنین مدل پیشنهادی جهت تشخیص شواهد، می تواند نظرات، علایق، نیازها، ظرفیت ها و ... در رابطه با کاربران را دریابد و با استفاده از این کانال از شواهد ایجاد شده، تصمیمات سیاستی مناسبی اتخاذ کند.

نتیجه گیری

اولین مانع توسعه گفتمان «سیاست‌گذاری مبتنی بر شواهد» این است که این گزاره بیشتر یک شعار و برای محکوم کردن طرف مخالف به کار می‌رود. به عبارت دیگر از این طریق، دیدگاه‌های مخالف به این متهم می‌شوند که شواهد را نادیده گرفته و تصمیماتشان مبتنی بر واقعیات نیست. اینکه باید در سیاست‌گذاری از داده‌ها و شواهد استفاده شود امر جدیدی نیست. در واقع اصلاً نمی‌توان یک فرآیند تصمیم‌گیری را تصور کرد که در آن به نحوی از اشکالی از شواهد موجود استفاده نشود، البته مخالفان جنبش سیاستگذاری مبتنی بر شواهد نیز به شدت معتقدند که تصمیمات نباید مبتنی بر تعصب، دلبخواه و بر اساس کلیشه‌های موجود اتخاذ شوند. به همین ترتیب، اختلاف در میان سیاست‌گذاران نیز در اصل مفهوم لزوم به کارگیری شواهد نیست بلکه موضوع بر اساس وزن‌دهی و ارزیابی منابع مختلف و بعضاً متعارض اطلاعات در دسترس یا تولیدی است(Syer, 2019). بنابراین یکی از جنبه‌های مهم و قابل بحث این نوع 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 نوآوری ها در این خصوص خواهد بود. از اینرو هدف این پژوهش نیز در راستای چنین نیازی یعنی ارائه چارچوبی برای تشخیص توئیت های شواهد و شناسایی رفتار کاربرانی که چنین توئیت هایی منتشر می کنند بوده است و در برچسب زنی داده ها نیز این هدف مد نظر قرار گرفت. مونتسچی[۱۵۹] (۲۰۰۹) معتقد است که شواهد عبارتند از هر آن چیزی که قابل گزارش‌دهی باشد. قدرت شواهد به اعتبار روش استخراج آن از میدان مورد نظر، بستگی دارد آنچه که به عنوان شواهد توسط پژوهشگران توصیه می‌شود، هر آن چیزی است که قابلیت استخراج داشته و بصورت ملموس قابل بررسی باشد.

از آنجا که رسانه های اجتماعی از جمله توئیتر، فرصت شناخت و درک افکار عمومی را برای سیاستگذاران ایجاد می کنند، استفاده از این رسانه ها یک فرصت بزرگ برای سیاستگذاری ها محسوب می شوند. بر اساس رویکرد سیاستگذاری مبتنی بر شواهد که در سال های اخیر از اقبال خوبی در میان سیاستگذاران برخوردار شده است، سیاستگذار نیازمند شواهد مناسب از کانال های مختلف است تا با کاربست آنها در فرایند سیاستگذاری بتواند سیاست های موثر جهت حل مشکلات عمومی را طراحی کند. یکی از زمینه های مهم سیاستگذاری عمومی، سیاستگذاری علم و فناوری است که در سال های اخیر به دلیل فرو ریختن مرزهای فناوری های مختلف با چالش های زیادی مواجه شده است. سیاستگذار جهت فائق آمدن بر این چالش ها نیازمند استفاده از شواهد متنوع در فرایند سیاستگذاری است.

159

در کنار شواهد رسمی، شواهد غیر رسمی نیز امروزه معتبر شناخته شده است. این پژوهش جهت پاسخ گویی به این پرسش که آیا می توان بر اساس پیش فرض ها و استاندارد هایی که در رابطه با شواهد وجود دارد، از محتوای کاربر ساخته در شبکه اجتماعی توئیتر، به عنوان شواهد مناسب برای سیاستگذاری استفاده کرد. اگر چنین کاری به لحاظ پیش فرض های نظری امکان پذیر است، چه تحلیل هایی اکنون در رابطه با داده های رسانه های اجتماعی صورت گیرد و مناسب ترین آنها برای بکارگیری در توئیتر کدام است که بتوان محتوای این رسانه اجتماعی را به شواهد مناسب برای سیاستگذار علم و فناوری تبدیل کرد.

بدین منظور، پژوهشگر جهت فراهم سازی پیش فرض های نظری، کار را از مفهوم سیاستگذاری مبتنی بر شواهد آغاز و سپس کارهای انجام شده در این خصوص و در رابطه با ویژگی های شواهد را مرور کرده است. سپس به ویژگی داده های رسانه های اجتماعی و بررسی استدلال هایی که سیاستگذار جهت تامین منافع عمومی حق استفاده از این منبع عمومی(داده های رسانه های اجتماعی) را دارد، پرداخته شده است. پس از آن، منطق استفاده از قابلیت های جمعی در سیاستگذاری، شرح داده شده است. پس از آن نیز تحلیل های پر تکراری که در رابطه با داده های رسانه های اجتماعی مورد استفاده قرار می گیرد بررسی شده است. پس از آن این پژوهش به سیاستگذاری علم و فناوری و چالش های آن پرداخته است. در ادامه الگوی مفهومی کاربست داده های رسانه اجتماعی توئیتر در سیاستگذاری علم و فناوری بر اساس حوزه های موضوعی این حوزه از سیاستگذاری، ارائه شده است. پژوهشگر با بررسی الگوریتم های تحلیل داده های رسانه های اجتماعی و استفاده از مشاوره های خبرگانی، الگوی مناسب جهت تحلیل داده های رسانه های اجتماعی را انتخاب و جهت کاربرد در فرایند سیاستگذاری سه خروجی برای آن در نظر گرفته است. دو خروجی فرعی و یک خروجی اصلی که همگی در سه بخش از مدل شش عنصری سیاستگذاری از یانگ و کوئین(۲۰۰۲) می تواند مورد استفاده قرار گیرد.

در خروجی اولِ الگوی ارائه شده، نظر بر آن بوده است که فراهم آوری اطلاعاتی راجع به کمیت های(توصیف های کمی) ایجاد شده در خصوص کلید واژه های اصلی در شبکه اجتماعی توئیتر می تواند به سیاستگذار در خصوص انتخاب دستورکار مناسب برای 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 است می تواند در انتخاب این مسئله درست، کمک کننده باشد. در خروجی دوم که تحلیل احساس و در واقع تحلیل رفتارِ کاربرانی که در رابطه با کلید واژه های مورد نظر توئیت شواهد منتشر می کنند، ارائه می شود، قصد کمک به اتخاذ راه حل های مناسب جهت حل مسئله، در کنار دیگر کانال های اطلاعات، بوده است.

در خروجی سومِ الگو، پژوهشگر در صدد آزمون خروجی های تعریف شده برای الگوی پیشنهادی بر آمده است. مهمترین خروجی این الگوی پیشنهادی یعنی تشخیص شواهد از غیر شواهد، به منظور در اختیار قرار دادن کانالی از شواهد در اختیار سیاستگذار علم و فناوری، است. بدین منظور الگوریتم انتخاب شده با داده های ۶ ماهه توئیتر فارسی آزمون شد و با دقت خوب و قابل قبولی(۸۵.۹)، مورد تایید قرار گرفت. هدف ارائه این خروجی از مدل پیشنهادی، از میان بردن سردرگمی سیاستگذار علم و فناوری در میان انبوهی از توئیت هایی است که در هر لحظه در این پلتفرم منتشر می شود. در واقع این الگوریتم، توئیت هایی از جنس شواهد را جدا ساخته و تنها آنها را در اختیار سیاستگذار قرار می دهد. پس از آن این سیاستگذار است که با در دست داشتن شواهد معتبری از کانالی مانند شبکه اجتماعی توئیتر، می تواند دست به تدوین سیاست ها آنها بزند. با استفاده از این الگوی پیشنهادی، سیاستگذاران می توانند به طور سیستماتیک از داده های 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 می تواند فرایند تدوین، اجرا و ارزیابی سیاست ها را تسهیل کند. چرا که دیدگاهها، نیازها و ظرفیت های این کاربران برخاسته از دانش زمینه ای آنها نسبت به مشکلات حوزه علم و فناوری است. پژوهشگر در تمام فرایند پژوهش خود این نکته را مد نظر داشته است که شبکه های اجتماعی و از جمله توئیتر تنها می توانند بخشی از شواهد مورد نیاز سیاستگذار را فراهم آورد. و هیچ گاه ادعا نشده است که تنها استفاده از این گونه شواهد و چشم پوشی از دیگر شواهد، می تواند سیاستگذار را به اتخاذ سیاست های درست رهنمون کند. اما بایستی تاکید کرد از آنجا که امکانات سیاستگذار برای پیمایش های گسترده و با روایی بالا، جهت دستیابی به نظرات شهروندان،کم است، استفاده از داده های رسانه های اجتماعی که در حجمی انبوه تولید می شود و تحلیل آنها بسیار ارزان تر و سریع از دیگر داده ها است، نباید در فرایند سیاستگذاری مورد غفلت واقع شود.

منابع:

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

۲. اصنافی، امیررضا (۱۳۹۵). تأملی بر آرشیوسازی رسانه‌های اجتماعی. *گنجینه اسناد*، ۱۰۲، ۱۱۳-۹۶

۳. ایروانچی، عارفه.(۱۳۹۴) روشی مستقل از معنا برای استخراج کلمه کلیدی داده‌های متنی کاربران در یک شبکه‌اجتماعی. پایان‌نامه کارشناسی ارشد مهندسی کامپیوتر گرایش نرم‌افزار. دانشکده فنی و مهندسی، دانشگاه علم و فرهنگ

۴. قاضی نوری، سپهر؛ قاضی نوری، سروش(۱۳۹۱). سیاستگذاری علم و فناوری در قالب سیاست های عام و خاص، *رهیافت*، ۵۰، ۴۹- ۵۵

۵. قاضی نوری، سروش، رضایی نیک، نفیسه(۱۳۹۳).بررسی الزامات، چالشها و قابلیتهای شبکه اجتماعی کنشگران مدیریت فناوری و نوآوری ایران. *تحقیقات فرهنگی ایران*، ۷(۲) ، ۷۳- ۴

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

۷. قانعی راد، 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

۸. کلانتری اسماعیل، منتظر غلامعلی، قاضی نوری، سید سپهر(۱۳۹۸). تدوین سناریوهای گذار به وضعیت بهبودیافتهٔ ساختار سیاستگذاری علم و فنّاوری در ایران، *پژوهش های مدیریت راهبردی*، ۷۴، ۷۵- ۱۰۲

۹. خواجه ئیان، داتیس(۱۳۹۸). درآمدی بر مدیریت شبکه های اجتماعی و کسب و کارهای پلتفرمی، تهران: ادبیان روز

۱۰. نامداریان، لیلا(۱۳۹۸). شناسایی حوزه‌های موضوعی سیاست اطلاعات علمی و فناورانه و بررسی نقش پژوهشگاه علوم و فناوری اطلاعات ایران (ایرانداک) در حوزه‌های موضوعی مرتبط. تهران: پژوهشگاه علوم و فناوری اطلاعات ایران(ایرانداک).

11. Abayomi-alli, S. Misra, A. Abayomi-alli, and M. Odusami. (2019), "Engineering Applications of Artificial Intelligence A review of soft techniques for SMS spam classification : Methods , approaches," *Eng. Appl. Artif. Intell.*, vol. 86, pp. 197–212
12. Aggarwal, C. C. (2011). *An introduction to social network data analytics*. In C. C. Aggarwal (Ed.), Social network data analytics (pp. 1–15). United States: Springer.
13. Barnett, J., Burningham, K., Walker, G., & Cass, N. (2012). Imagined publics and engagement around renewable energy technologies in the UK. Public Understanding of Science, 21(1), 36–50.

14. Bekkers, V., Edwards, A., & de Kool, D. (2013). Social media monitoring: Responsive governance in the shadow of surveillance? Government Information Quarterly, 30(4), 335–342.
15. Boyd, D., & Crawford, K. (2012). Critical questions for big data. Information, Communication & Society, 15(5), 662–679.
16. Cairney, P. (2016). *The Politics of Evidence-Based Policy Making*. London: Palgrave Macmillan.
17. Cartwright, N., Hardie, j. (2012). *Evidence-Based Policy A Practical Guide to Doing It Better*. New York: Oxford.
18. Castelló, I., Morsing, M., & Schultz, F. (2013). Communicative dynamics and the polyphony of corporate social responsibility in the network society. Journal of Business Ethics, 118(4), 683–694.
19. Chen, C; Wang, Y; Zhang, J; Xiang, Y; Zhou, W; Min, G. (2016). Statistical Features-Based Real-Time Detection of Drifted Twitter Spam. *IEEE Transactions on Computational Social Systems.* 12(4). 914 – 925.
20. Chen, C; Wang, Y; Zhang, J; Xiang, Y; Zhou, W; Min, G. (2017). Statistical Features-Based Real-Time Detection of Drifted Twitter Spam. *IEEE Transactions on Information Forensics and Security*, 12(4). 914-925.
21. Clarke, A., & Margetts, M. (2014) Governments and Citizens Getting to Know Each Other? Open, Closed, and Big Data in Public Management Reform, *Policy & Internet*. 6 (2), 393–417. doi:10.1002/1944-2866.POI377.
22. Criado, J. I., Sandoval-Almazan, R., & Gil-Garcia, J. R. (2013). Government innovation through social media. *Government Information Quarterly, 30*, 319–326.
23. Culnan, M. J., McHugh, P. J., & Zubillaga, J. I. (2010). How large U.S. companies can use twitter and other social media to gain business value. MIS Quarterly Executive, 9(4), 243–259.
24. Davies, H., Nutley, S., & Smith, P. (2001). What works?: Evidence-based policy and practice in public services. Bristol: Policy Press.
25. Driss, O. B., Mellouli, S., & Trabelsi, Z. (2019). From citizens to government policy-makers: Social media data analysis. *Government Information Quarterly,* 36, 560–570
26. Eastin, M., Brinson, NH., Doorey, A., Wilcox, G. (2016). Living in a big data world: Predicting mobile commerce activity through privacy concerns. *Computers in Human Behavior*. 58: 214- 220
27. Enroth, Henrik. (2014). Governance: the art of governing after governmentality. *European journal of social theory*, 17(1): 60-76.
28. Eom, S. J., Hwang, H., & Kim, J. H. (2018). Can social media increase government responsiveness? A case study of Seoul, Korea. *Government Information Quarterly, 35*,109–122.
29. Evens, T.,& Donders, K. (2018). *Platform Power and Policy in Transforming Television Markets*. London: Palgrave Macmillan.
30. Fern´andez, M., Wandhoefer, T., Allen, B., Cano Basave, A., & Alani, H. (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.

31. Fernando, S., Díaz López, J. A., Şerban, O., Gómez-Romero, J., Molina-Solana, M., Guo, Y. (2019). Towards a large-scale twitter observatory for political events. *Future Generation Computer Systems*, 14(2): 1-8.

32. Gandomi, A., Haider, M. (2015). Beyond the hype: Big data concepts, methods, and analytics. *International Journal of Information Management*. 35. 137-144
33. Giachanou, A., & Crestani, F. (2016). Like it or not: A survey of twitter sentiment analysis methods. ACM Computing Surveys (CSUR), 49(2), 28.
34. 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.
35. Hasan, M. A., & Zaki, M. J. (2011). *A survey of link prediction in social networks*. In C. C. Aggarwal (Ed.), Social network data analytics (pp. 243–275). United States: Springer
36. He, W., Zha, S., & Li, L. (2013). Social media competitive analysis and text mining: A case study in the pizza industry. *International Journal of Information Management*, 33(3), 464–472.
37. Head, B (2010). Reconsidering evidence-based policy: Key issues and challenges, *Policy and Society*. 29: 77–94.
38. Head, B. (2008). Three Lenses of Evidence-Based Policy, *The Australian Journal of Public Administration*, 67(1), 1–11.
39. Head, B. W. (2008). Three lenses of evidence-based policy. *Australian Journal of Public Administration*, 67(1), 1–11.

40. Heidemann, J., Klier, M., & Probst, F. (2012). Online social networks: A survey of a global phenomenon. *Computer Networks*, 56(18), 3866–3878.
41. Helen, D. (2016). Social media for professional development and networking opportunities in academia. *Further and Higher Education*,(5)40, 706-729.

42. Highfield, T., Harrington, S., & Bruns, A. (2013). Twitter as a technology for audiencing and fandom. Information, Communication & Society, 16(3), 315–339.
43. Hijawi, H. Faris, J. Alqatawna, A. M. Al-zoubi, and I. Aljarah, "Improving Email Spam Detection Using Content Based Feature Engineering Approach,"2016
44. Hijawi, W; Faris, H; Alqatawna, J; Al-Zoubi, A; Aljarah, I. (2017). Improving Email Spam Detection Using Content Based Feature Engineering Approach. Jordan Conference on Applied Electrical Engineering and Computing Technologies (AEECT). Jordan
45. Howlett, M. (2009). Policy analytical capacity and evidence-based policy-making: Lessons from Canada. *Canadian Public Administration*, 52(2), 153–175.
46. Ingold, J., & Monaghan, M. (2016). Evidence translation: An exploration of policy makers' use of evidence. Policy & Politics, 44(2), 171–190.
47. Janssen, M., & Helbig, N. (2016). Innovating and changing the policy-cycle: Policy-makers be prepared! *Government Information Quarterly*. 12(3), 120- 132

48. Jianqiang and G. Xiaolin. (2017). "Comparison Research on Text Pre-processing Methods on Twitter Sentiment Analysis," vol. 3536,
49. Kaolan, A., Mazurek, G. (2018). Social media. Chapter book In "media management and economic" edited by Albarran, A. (2018). London: Routledge.
50. Kaplan, A. M., & Haenlein, M. (2010). Users of the world, unite! The challenges and opportunities of social media. *Business Horizons*, 53(1), 59–68.

51. Kasabov, E. (2008). The challenge of devising public policy for high-tech, science-based, and knowledge-based communities: Evidence from a life science and biotechnology community. Environment and Planning C: Government and Policy, 26(1), 210–228.
52. Kietzmann, J. H., Hermkens, K., McCarthy, I. P., & Silvestre, B. S. (2011). Social media? Get serious! Understanding the functional building blocks of social media. Business Horizons, 54(3), 241–251.
53. Kum, H.-C., Joy Stewart, C., Rose, R. A., & Duncan, D. F. (2015). Using big data for evidence-based governance in child welfare. *Children and Youth Services Review*, 58, 127–136.
54. Latham, R. (2014). Information Management Advice 57 Managing Social Media Records Part 4: How to Capture Social Media Records, Tasmanian Archive+ Heritage Office, Issued: June 2014. Frome: https://www.informationstrategy.tas.gov.au
55. Lee, E. J., Lee, H. Y., & Choi, S. (2020). Is the message the medium? How politicians' Twitter blunders affect perceived authenticity of Twitter communication. *Computers in Human Behavior*, 104: 106- 188.
56. Liu, Yili. YING, Xiangxiang (2014). A Review of Social Network Sites: Definition, Experience and Applications. The Conference on Web Based Business Management. NewYork.
57. Lodge, M., & Wegrich, K. (2014). Crowdsourcing and regulatory reviews: A new way of challenging red tape in British government? Regulation & Governance, 9(1), 30–46.
58. Lundberg, J., Laitinen, M. (2020). Twitter trolls: a linguistic profile of anti-democratic discourse. *Language Sciences*, 8(3): 1-14.
59. McCay-Peet, L., & Quan-Haase, A. (2016). 'A model of social media engagement: User profiles, gratifications, and experiences'. In H. O'Brien & M. Lalmas (Eds.), Why engagement matters: Cross-disciplinary perspectives and innovations on user engagement with digital media. Heidelberg: Springer Verlag.
60. Missaoui, R. Sarr, I. (2014). *Social Network Analysis – Community Detection and Evolution*. Canada: springer.
61. Misuraca, G., Codagnone, C., & Rossel, P. (2013). From practice to theory and back to practice: Reflexivity in measurement and evaluation for evidence-based policy making in the information society. Government Information Quarterly, 30, S68–S82.
62. Mittal, A., & Patidar, S. (2019). Sentiment Analysis on Twitter Data: A Survey. Presented in "Association for Computing Machinery". Bangkok, Thailand. July 27–29
63. Mostafa *et al*. (2019). "ScienceDirect Examining multiple feature evaluation and classification methods for improving the diagnosis of Parkinson ' s disease," *Cogn. Syst. Res.*, no. xxx
64. Napoli, P. N. (2019). User Data as Public Resource: Implications for Social Media Regulation, *Policy and Internet*. doi: 10.1002/poi3.216.
65. Newman, J., Cherney, A., & Head, B. W. (2017). Policy capacity and evidence-based policy in the public service. Public Management Review, 19(2), 157–174.
66. Nutley , B., Isabel, W., Huw, D.(2007). *Using evidence: how research can inform public services*. Bristol: policy press.
67. Nutley, S., H. Davies and I. Walter. (2002). Evidence-Based Policy and Practice: Cross Sector Lessons from the UK. Working Paper 9, Research Unit for Research Utilization, University of St Andrews.

68. Oh, S., & Syn, S. (2018). Motivations for sharing information and social support in social media: A comparative analysis of Facebook, Twitter, Delicious, YouTube, and Flickr: Motivations for Sharing Information. *Association for Information Science and Technology*,(10)66, 1-29.

69. Ooms, W., Bell, J., & Kok, R. A. W. (2015). Use of social media in inbound open innovation: Building capabilities for absorptive capacity. Creativity and Innovation Management, 24(1), 136–150.
70. Panagiotopoulos, P., Shan, L. C., Barnett, J., Regan, Á., & McConnon, Á. (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.
71. Panayiotopoulos, P., Bowen, F., & Brooker, P. (2017). The value of social media data: Integrating crowd capabilities in evidence-based policy. *Government Information Quarterly,*34: 601–612.
72. Pang, M.-S., Lee, G., & DeLone, W. H. (2014). In public sector organisations: a publicvalue management perspective. Journal of Information Technology, 29(3), 187–205.
73. Park, C. S., & Kaye, B. K. (2017). The tweet goes on: Interconnection of Twitter opinion leadership, network size, and civic engagement. *Computers in Human Behavior*, 69: 174-180.
74. Park, M., & Lee, T. S. (2013). Understanding science and technology information users through transaction log analysis. *Library hi tech*.
75. Parkhurst, J. (2017). *The politics of evidence; from evidence-based policy to the good governance of evidence*. Routledge. London
76. 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.

77. Parthasarathy, S., Ruan, Y., & Satuluri, V. (2011). *Community discovery in social networks: Applications, methods and emerging trends*. In C. C. Aggarwal (Ed.), Social network data analytics (pp. 79–113). United States: Springer
78. Picazo-Vela, S., Gutiérrez-Martínez, I., & Luna-Reyes, L. F. (2012). Understanding risks, benefits, and strategic alternatives of social media applications in the public sector. Government Information Quarterly, 29(4), 504–511.
79. Prpic, J., & Shukla, P. (2014). The contours of crowd capability. 2014 47th Hawaii International Conference on System Sciences (pp. 3461–3470). IEEE.
80. Prpić, J., Shukla, P. P., Kietzmann, J. H., & McCarthy, I. P. (2015). How to work a crowd: Developing crowd capital through crowdsourcing. Business Horizons, 58(1), 77–85.
81. Prpić, J., Taeihagh, A., & Melton, J. (2015). The fundamentals of policy crowdsourcing. Policy & Internet, 7(3), 340–361.
82. Purtova, N. 2017. “Do Property Rights in Personal Data Make Sense After the Big Data Turn? Individual Control and Transparency.” Journal of Law and Economic Regulation 10 (2): 64–78.
83. Roberts, N., Galluch, P., Dinger, M., & Grover, V. (2012). Absorptive capacity and information systems research: Review, synthesis, and directions for future research. MIS Quartely, 36(2), 625–648.

84. Roy, Rishiraj Saha. Padmakumar, Aishwarya. Prasaad Jeganathan ,Guna. Kumaraguru ,Ponnurangam (2015).Automated Linguistic Personalization of Targeted Marketing Messages Mining User-Generated Text on Social Media. *Springer International Publishing Switzerland*,2015, CICLing 2015, Part II, LNCS 9042, pp203-224.
85. Saif, H., He, Y., & Alani, H. (2012). *Semantic sentiment analysis of twitter. In International semantic web conference* (pp. 508-524). Springer, Berlin, Heidelberg.
86. Sanderson, I. (2002). Evaluation, policy learning and evidence-based policy making. Public Administration, 80(1), 1–22.
87. Schintler, L. A., & Kulkarni, R. (2014). Big data for policy analysis: The good, the bad, and the ugly. Review of Policy Research, 31(4), 343–348.
88. Schwartz, P., & Solove, D. (2011). The PII problem: privacy and new concept of personally identifiable information. *NYU Law Review*. 86: 111- 122
89. Sedhai and A. Sun. (2018). "Semi-Supervised Spam Detection in Twitter Stream," *IEEE Trans. Comput. Soc. Syst.*, vol. 5, no. 1, pp. 169–175
90. Sedhai, S; Sun, A. (2017). Semi-Supervised Spam Detection in Twitter Stream. *Computational Social Systems*. 5(1). 169 - 175
91. Shanahan, E. A., Mcbeth, M.K., & Hathaway, P. L. (2011). Narrative Policy Framework: The Influence of Media Policy Narratives on Public Opinion. *Politics and Policy*, 39, 3, 373-400.
92. Stamatelatos, G., Gyftopoulos, S., Drosatos, G., & Efraimidis, P. S. (2020). Revealing the political affinity of online entities through their Twitter followers. *Information Processing and Management,* 57: 102-172.
93. Sun, J., & Tang, J. (2011). *A survey of models and algorithms for social influence analysis*. In C. C. Aggarwal (Ed.), Social network data analytics (pp. 177–214). United States: Springer.
94. Suzor, N., M. Dragiewicz, B. Harris, R. Gillett, J. Burgess, and T. Van Geelen. (2019). "Human Rights by Design: The Responsibilities of Social Media Platforms to Address Gender-Based Violence Online." *Policy & Internet*. 11 (1): 84–103.
95. Tang, L., & Liu, H. (2010). Community detection and mining in social media. *Synthesis Lectures on Data Mining and Knowledge Discovery*, 2(1), 1–137.
96. Tharwat, "Classification assessment methods," *Appl. Comput. Informatics, 2018*
97. Turow, J., and N. Couldry. (2018). "Media as Data Extraction: Towards a New Map of a Transformed Communications Field." *Journal of Communication*. 68, 415–23.
98. Walker, G., Cass, N., Burningham, K., & Barnett, J. (2010). Renewable energy and sociotechnical change: Imagined subjectivities of "the public" and their implications. Environment and Planning A, 42(4), 931–947.
99. Wang, X., Wei, F., Liu, X., Zhou, M., & Zhang, M. (2011). Topic sentiment analysis in twitter: a graph-based hashtag sentiment classification approach. In Proceedings of the 20th ACM international conference on Information and knowledge management. 1031-1040.
100. Wang, Y., Chen, Q., Kang, C., & Xia, Q. (2016). Clustering of electricity consumption behavior dynamics toward big data applications. *IEEE transactions on smart grid*, 7(5): 2437-2447.
101. Wessels, B. (2012). Identification and practices of identity and privacy in everyday digital communication. *New Media & Society*, 14(8): 1251-1268

102. Williams, M. L., Edwards, A., Housley, W., Burnap, P., Rana, O., Avis, N., et al. (2013). Policing cyber neighbourhoods: Tension monitoring and social media networks. Policing and Society, 23(4), 461–481.
103. Williamson, B & Piattoeva, N. (2018). Objectivity as standardization in data-scientific education policy, technology and governance, *Learning, Media and Technology*. 44(1): 64- 76.
104. Yan, Z,. (2018). Big Data and Government Governance. *International Conference on Information Management and Processing*. IEEE. 110- 114
105. Young, K., Ashby, D., Boaz, A., Grayson, L (2002). Social Science and the Evidence-based Policy Movement, *Social Policy & Society*. 1(3):215 -224.
106. Zahra, S. A., & George, G. (2002). Absorptive capacity: A review, reconceptualization, and extension. *Academy of Management Review*, 27(2), 185–203.

Abstract: One of the best ways to understand about public opinion in order to use in policy process, is through social media. So far, social media data has not been systematically used in the science and technology policy process. The purpose of this study was to use the data of social media users in the science and technology policy process. The research question was: How can social media data be used in science and technology policy? First, the use of social media data as part of credible evidence in the science and technology policy process has been explained. Then, the model of using social media data in science and technology policy has been presented. The presented model was tested with 6-month data of all Persian users of Twitter who published tweets on the subject of science and technology policy.

Keywords: Evidence Base Policy; Science and Technology Policy; Social Media; User Data; Twitter

Iranian Research Institute
for Information Science and Technology (IranDoc)

Provide A Model for Using Social Media Data in Science and Technology Policy

Somayeh Labafi

23/ May/ 2021