



استفاده از داده‌های شبکه‌های اجتماعی در پیش‌بینی نتایج انتخابات

مهشید شهیدی^۱

ادغام علوم اجتماعی با رشته‌های علوم کامپیوتر و مهندسی زمینه جدیدی از مطالعه را ایجاد کرده است: علوم اجتماعی محاسباتی. این حوزه روش‌های محاسباتی را برای منابع جدید داده‌های دیجیتال مانند رسانه‌های اجتماعی برای توسعه نظریه‌های رفتار انسان اعمال می‌کند.

ظهور اینترنت و دیجیتالی شدن انبوه، در سال‌های اخیر حجم بی‌سابقه‌ای از داده‌های دیجیتالی را به بازار عرضه کرده است. برخلاف مجموعه داده‌های متداول که توسط دانشمندان علوم اجتماعی جمع‌آوری شده است، این منابع جدید دیجیتالی غالباً هنگام تکمیل روابط، جزئیات مفصلی درباره تحول روابط اجتماعی در بین جمعیت‌های بزرگ ارائه می‌دهند. در همین حال، انواع مختلفی از تکنیک‌های جدید در حال حاضر برای تجزیه و تحلیل این مجموعه داده‌های بزرگ و پیچیده در دسترس است. این موارد شامل اشکال مختلف تجزیه و تحلیل متن خودکار، آزمایشات میدانی آنلاین، همکاری جمعی و بسیاری دیگر با الهام از یادگیری

^۱ کارشناسی ارشد پژوهشگری علوم اجتماعی از دانشگاه تهران، پژوهشگر تحلیل داده شرکت لایفوب

ماشین است. افزایش سریع داده‌های دیجیتالی - در کنار روش‌های جدید برای تجزیه و تحلیل آنها - زمینه‌ی میان رشته‌ای جدید به نام علوم اجتماعی محاسباتی به وجود آورده است.

علوم اجتماعی محاسباتی یک رشته میان رشته‌ای است که با استفاده از تکنیک‌های محاسباتی به داده‌های سایت‌های رسانه‌های اجتماعی، نظریه‌های رفتار انسان را پیش می‌برد.

یکی از بسترهای مهم در استفاده از ظرفیت علوم اجتماعی محاسباتی، استفاده از داده‌های فضای مجازی در راستای شناخت افکار، ذائقه‌ها و نگرش‌های اجتماعی و سیاسی است. بسیاری از مطالعات اخیر از داده‌های رسانه‌های اجتماعی به عنوان "حسگر اجتماعی" با هدف پیش‌بینی متفاوت استفاده کرده است. پدیده‌های اقتصادی، اجتماعی و سیاسی، از نتایج انتخابات گرفته تا موفقیت در گیشه برای فروش فیلم‌ها.

همان گونه که ذکر شد پیش‌بینی انتخابات نیز از مهم‌ترین فضاها در کاربرد این علم می‌باشد که در سال‌های گذشته در سطح دنیا رواج یافته است.

به طور مثال گروه رسانه و ارتباطات دانشگاه هنگ‌کنگ در سال ۲۰۲۰، فراتحلیلی از مطالعات انجام شده پیرامون قدرت پیش‌بینی کننده‌ی



شبکه‌های اجتماعی برای تخمین نتایج آراء انتخاباتی صورت داده‌اند. آنها ۹ مطالعه که از داده‌های شبکه‌های مجازی همچون "توییتر"، "فروم"، "وبلاگ‌ها"، "یوتیوب"، و... برای پیش‌بینی آراء استفاده کرده بودند را مورد بررسی قرار دادند.

به زعم اسکوریچ، لیو و جیدکا (۲۰۲۰)، پژوهش‌های انجام شده عموماً از دو روش برای پیش‌بینی نتایج بهره برده‌اند: ۱. تحلیل احساسات ۲. تحلیل ویژگی‌های ساختاری.

در روش تحلیل احساسات، عقاید مثبت و منفی با تحلیل لغات سنجیده می‌شوند. مثلاً اوکتر و همکاران با استفاده از ابزار واژه‌نامه‌ی احساسات‌یاب^۲، به این نتیجه رسیده‌اند که احساسات ابراز شده در توئیت‌ها، با میزان حمایتی که از اوباما در سال ۲۰۰۸ صورت گرفت، همخوانی دارد. به زعم نویسندگان این مقاله، روش فراگیری ماشین^۳، بهتر از تحلیل لغات برای پیش‌بینی آراء عمل می‌کند. چرا که علائم ضمنی را که در سخنان مثبت و منفی توئیت‌کنندگان هستند را نیز ثبت می‌کنند.

اما شیوه‌ی دوم که شیوه‌ی تحلیل (ویژگی‌های) ساختاری است، بر ساختار اجتماعی مکالمات درون یک پلتفرم اجتماعی مثل توییتر، فیسبوک و... تمرکز می‌کند. این روش نه بر احساسات مثبت و منفی که بر میزان

مشارکت و تعامل افراد در "پست‌ها"، "کامنت گذاشتن"، "ری‌توئیت‌کردن‌ها"، "منشن‌کردن"، "فالوو کردن" و... تمرکز دارد.

جیدکا و همکاران (۲۰۱۹)، نیز میزان قدرت پیش‌بینی‌کننده‌ی شبکه‌های اجتماعی برای انتخابات را با بررسی پست‌های توییتر، در سه کشور آسیایی - مالزی، هند و پاکستان - سنجیده‌اند. در ادبیات پژوهشی این مطالعه مسئله‌ی وثاقت توئیت برای پیش‌بینی نتایج انتخابات مطرح شده است. مطالعاتی که توئیت را ابزار مناسبی برای پیش‌بینی انتخابات دانسته‌اند، از یک یا چند روش زیر برای پیش‌بینی استفاده کرده‌اند:

۱. تجزیه و تحلیل حجمی^۴ (Vol): شمارش ساده‌ی تعداد دفعات منشن کردن، بازتوئیت‌کردن، میزان طرفداران و لایک کردن و غیره

۲. تجزیه و تحلیل احساس^۵ (Sen): تراکم احساسات مثبت یا منفی در پست‌ها، ابراز احساسات نسبت به نامزدهای خاص یا احزاب سیاسی و...

۳. تحلیل شبکه^۶ (Net): ویژگی‌های شبکه‌ی کاربران رسانه‌های اجتماعی که از نامزدهای خاص یا احزاب سیاسی حمایت یا بحث می‌کنند.

پژوهشگران بر سر استفاده‌ی این شیوه‌ها اتفاق نظر ندارند؛ اسکوریک و همکاران^۷ (۲۰۱۵)، با انجام فراتحلیل مطالعات انجام شده در این زمینه به

^۵ . Sentiment analyses

^۶ . Network analyses

^۷ . Skoric et al.

^۲ . OpinionFinder sentiment lexicon

^۳ . machine learning

^۴ . Volumetric analyses



این نتیجه رسیده‌اند که بیش از نیمی از پژوهش‌ها، از شیوه‌ی تحلیل حجمی بهره برده‌اند. در مطالعه‌ی پیش رو جیدکا و همکاران (۲۰۱۹)، از هر سه شیوه تجزیه و تحلیل حجمی، تجزیه و تحلیل احساس و تحلیل شبکه، در سه کشور مذکور، (مالزی، هند و پاکستان) در دوره‌ای که انتخابات در این کشورها برگزار شده است، بهره برده‌اند.

اطلاعات این پژوهش با استفاده از توئیتر استریمینگ API^۸ و ابزار توئیتر آرشیویست^۹، بدست آمده که بواسطه‌ی آنها نام احزاب سیاسی و نام دو تن از مهمترین رهبران آنها، ردیابی شد. حدود ۳.۴ میلیون توئیت در خلال تاریخ معرفی نامزدهای انتخابات و روز رای‌گیری، گردآوری شد. (۱۴ حزب در مالزی (۱.۱ میلیون توئیت)، ۱۵ حزب در هند (۱.۲ میلیون توئیت) و ۱۱ حزب در پاکستان (۱.۱ میلیون توئیت)). به منظور تحلیل ابتدا، توئیتهای هر حزب به چهار شیوه پیش‌پردازش^{۱۰} شد: ۱. توئیتهای عمومی بدون فیلتر ۲. توئیتهای خاص و حفظ توئیتهایی که از حزب مشخصی نام برده‌اند ۳. توئیتهای مثبت و توئیتهای خاصی که احساسات مثبتی را به نمایش گذارده بودند ۴. وزن‌دهی موقت به توئیتهایی که به زمان انتخابات نزدیک‌ترند. در این مطالعه هر یک از این فرایندهای پیش‌پردازش در رابطه با میزان قدرت آن در تبیین نتایج انتخابات مورد بررسی قرار گرفتند.

نتایج این مطالعه نشان می‌دهد که در مورد مالزی، روش تحلیل حجمی، بهتر است و خطاهای کمتری دارد، در مورد هند و پاکستان، پیش‌بینی‌ها با استفاده از رسانه‌های اجتماعی با نظرسنجی‌های سنتی، نتایج همسانی را به بار می‌آورد. به طور کلی، رسانه‌های اجتماعی ممکن است مزیت کمتری نسبت به نظرسنجی‌های سنتی داشته باشند.

در تحلیل داده‌های مرتبط با این سه کشور، انجام مراحل پیش‌پردازش مانند وزن دادن به توئیتهایی که نزدیک انتخابات گذاشته می‌شوند و یا وزن دادن به نویسندگان موثر، می‌تواند به همراه روش تحلیل احساسات، چشم‌انداز روشنی از نتایج انتخابات پیش روی ما بگذارد. با وجود این، در پاکستان که اکثر احزاب در توئیتر حضور کم‌رنگ‌تری دارند، اعمال همه‌ی مراحل پیش‌پردازش به طور همزمان، منجر به از دست رفتن اطلاعات ارزشمندی می‌شود. در مورد نمونه‌ی هند، مشاهده شد که وزن‌دهی موقتی می‌تواند قدرت پیش‌بینی را دقیق‌تر کند؛ البته ممکن است توئیتهای قدیمی‌تر نیز تأثیر زیادی بر کاربران داشته باشند، اما نتایج نشان می‌دهد که توئیتهای نزدیک‌تر به انتخابات، تأثیرگذارترند و لذا بهتر است که به آنها وزن داده شود. برای پژوهش‌های بعدی پیشنهاد می‌شود که در بازه‌های زمانی متفاوت (نزدیک‌تر به انتخابات یا دورتر از آن) به توئیتهای وزن داده

^{۱۰} . pre-process

^۸ . Twitter's streaming API

^۹ . Tweet Archivist



شود تا میزان قدرت پیش‌بینی‌کنندگی توئیت‌های وزن‌داده شده، در هر یک از بازه‌های زمانی سنجیده شود.

تحلیل شبکه‌ی اجتماعی، ارتباطات و الگوی ساختاری مباحثات مربوط به احزاب سیاسی را نشان می‌دهد و با تمرکز بر الگوی تعاملات احزاب، بیش از میزان برآورد کردن یا کمتر از میزان برآورد کردن تاثیر(توئیت‌ها را)، که نقص روش تحلیل حجمی و تحلیل احساسات است را رفع می‌کند. یافته‌ها نشان می‌دهد که در توئیت‌های موفقیت انتخاباتی به گستردگی وابسته است تا عمق، یعنی حزب سیاسی که شبکه کاربران وسیع‌تری دارد، نسبت به حزب سیاسی که شبکه متمرکزتر اما کاربران فعال‌تری دارد، موفقیت بیشتری در کسب آراء بدست می‌آورد. و مسلماً هر چه توئیت‌های نزدیک انتخابات احساسات مثبت بیشتری را نسبت به یک حزب به اشتراک بگذارند، نتایج انتخابات بیشتر به نفع آن حزب خواهد بود.

اسکوربیچ، لیو و جیدکا (۲۰۲۰)، چهار متغیر زمینه‌ای را که ممکن است بر قدرت تبیینی شبکه‌های اجتماعی اثرگذار باشد، را وارد تحلیل کرده‌اند. ۱. نمره‌ی دموکراسی ۲. سیستم انتخاباتی ۳. آزادی مطبوعات ۴. میزان نفوذ اینترنت

۱۱ . Economist Intelligence Unit

۱. نمره دموکراسی: این نمره بر اساس شاخص دموکراسی که توسط واحد اطلاعات اقتصاددان^{۱۱} تهیه شده است، محاسبه گردیده. بر اساس آن کشورها به چهار دسته تقسیم می‌شوند:

- دموکراسی کامل: کشورهایی که نمره دموکراسی‌شان ۸ تا ۱۰ است؛ مثل آمریکا، انگلیس و آلمان

- دموکراسی ناقص: آن دسته از کشورهایی که نمره دموکراسی‌شان ۶ تا ۸ است؛ مثل فرانسه و هنگ‌کنگ.

- نظام‌های ترکیبی: آنهایی که نمره‌ی دموکراسی‌شان ۴ تا ۶ است، مثل سنگاپور و ونزئلا و اکوادور

- رژیم‌های توتالیتر: که از دموکراسی نمره‌ی ۰ تا ۴ را دریافت می‌کنند مثل نیجریه

۲. سیستم انتخاباتی: سیستم انتخاباتی به این امر که نظام انتخاباتی یک کشور نماینده و بیانگر عقاید رای‌دهندگان هست یا نه ارجاع دارد. مقوله‌بندی بر اساس شبکه دانش انتخاباتی^{۱۲} صورت گرفته است.

۳. آزادی مطبوعات: درجه آزادی مطبوعات از گزارش سالانه Freedom House گرفته شده است. این گزارش به عنوان جامع‌ترین مجموعه

۱۲ . ACE Electoral Knowledge Network



اطلاعات موجود در مورد آزادی رسانه های جهانی ، استقلال رسانه را در ۱۹۹ کشور ارزیابی کرده است.

۴. نفوذ اینترنتی: میزان نفوذ اینترنت با استفاده از داده های اتحادیه بین المللی ارتباطات از راه دور^{۱۳} (ITU) ، درصد افراد استفاده کننده از اینترنت در یک کشور را اندازه گیری می کند.

در نهایت نتایج این فراتحلیل نشان می دهد، در روش تحلیل احساسات، ابزار فراگیری ماشین بهتر از ابزار تحلیل لغات، نتایج را پیش بینی می کند. چرا که ابزار تحلیل لغات، مبتنی بر انگلیسی رسمی است اما بیشتر صحبت های کاربران در شبکه های اجتماعی به صورت غیر رسمی است. گرچه احساسات برای پیش بینی مهم اند اما نتایج این فراتحلیل نشان می دهد که فهم تعاملات درون شبکه های اجتماعی اهمیت بیشتری دارد. در پژوهش هایی که هم از روش تحلیل ویژگی های ساختاری (یعنی تحلیل مناسبات درون شبکه ها که در غالب فالوو کردن، ری توئیت کردن، لایک کردن و... حاصل می شود) و هم از روش تحلیل فراگیری ماشین برای تحلیل احساسات استفاده شده، پیش بینی های دقیق تری بدست آمده است. اگرچه بیشتر مطالعات تنها میزان فالوورها و دوستان، و ری توئیت ها و لایک را مدنظر داشته اند اما برای پیش بینی دقیق تر می توان کاربران سلبریتی، اینفلوئنسر ها و تبلیغ کنندگان را نیز وارد تحلیل کرد. اسکوریچ، لیو و

جیدکا(۲۰۲۰)، پیشنهاد می دهند که می توان از ابزارهایی که اهمیت کاربر را می سنجند، همچون ابزارهای اندازه گیری پیج رنک^{۱۴}، برای فهم ساختار تعاملی شبکه های اجتماعی بهره برد. ویژگی های ساختاری تعاملات می تواند میزان تراکم بحث های آنلاین را نیز نشان دهد. شبکه های متراکم تر، کاربران فعال تری دارند و لذا پتانسیل بیشتری برای تأثیر بر رأی دهندگان دارند. طبق این فراتحلیل اکثر مطالعات انجام شده در این حوزه، به لحاظ جمعیت شناختی جامعه ی هدف خود را متمایز نکرده اند. متمایز کردن گروه های سنی امری مهم و قابل انجام است که این سه نویسنده آنرا پیشنهاد می دهند.

به طرز جالب توجهی وبلاگ ها، پیش بینی کننده های وثیق تری از دیگر شبکه های اجتماعی هستند چرا که نویسندگان آنها اطلاعات مشخصا سیاسی در اختیار کاربران قرار می دهند و عموما چهره های سیاسی هستند که در توئیتر و فیسبوک نیز فعالیت دارند. این فضاها مانند تالارهای گفتگو هستند که در آن شهروندان علاقه مند به سیاست گرد هم می آیند.

نتایج تحلیل R۲ نشان می دهد، در کشورهایی که نظام ترکیبی، بر آنها حاکم است، (یعنی کشورهایی که خصایص کشورهای دموکرات و توتالیتار را همزمان دارند)، به منظور پیش بینی نتایج انتخابات بیشتر می توان بر داده های شبکه های مجازی اتکا کرد. در بیشتر کشورهایی که نظام ترکیبی

PageRank . ۱۴

International Telecommunication Union (ITU)/World Telecommunication/ICT Indicators Database, ۲۰۱۵. ۱۳



بر آنها حاکم است، رسانه های اجتماعی کمتر از رسانه های سنتی، که معمولاً تحت کنترل شدید دولت هستند، سانسور و کنترل می شوند. در این نظام ها دانش و مشارکت شهروندان در فعالیت های سیاسی، آزادی مطبوعات و آزادی بیان ممکن است به شدت محدود شود. مسئله مهمی که باید به آن توجه داشت این است که به دلیل مسائل حریم خصوصی و فضای ناامن سیاسی، شهروندان اغلب تمایل به ابراز عقایدشان در فضاهای خصوصی دارند و بدین دلیل عقاید شهروندان نامشهودتر است، و عموماً صفحه هایی که کاربران آنها عقاید خود را به وضوح ایراد می کنند، به دلایل استراتژیک دست به چنین کاری می زنند که این نشان می دهد عقاید ابراز شده، نمی تواند نماینده میانگین جمعیت شهروندان باشد. این چالشی است که پیش روی تحقیقات اینچنینی قرار دارد. همچنین چون این پژوهش از محدود پژوهش هایی است که دست به مقایسه ی مطالعات انجام شده در این زمینه زده، با مشکلاتی همچون ناهمخوانی مطالعات به لحاظ منبع داده، روش بررسی، جمعیت مورد مطالعه و... روبرو بوده است. لذا به زعم نویسندگان، تحلیل شبکه های اجتماعی نباید جایگزین بررسی افکار عمومی با روش پیمایش شود، چون هر یک از این دو روش با استفاده از منابع متفاوت، چشم انداز متفاوتی بر فهم رفتار سیاسی مردم می گشایند.

منابع:

- Jaidka, K., Ahmed, S., Skoric, M., & Hilbert, M. (۲۰۱۹). Predicting elections from social media: a three-country,

three-method comparative study. *Asian Journal of Communication*, ۲۹(۳), ۲۵۲-۲۷۳.

- Skoric, M. M., Liu, J., & Jaidka, K. (۲۰۲۰). Electoral and Public Opinion Forecasts with Social Media Data: A Meta-Analysis. *Information*, ۱۱(۴), ۱۸۷.



تحلیل و بررسی نتایج به دست آمده در خصوص

افکارسنجی داده‌های شبکه‌های اجتماعی در انتخابات ۱۴۰۰

شرکت لایفوب در زمره نخستین شرکت‌ها در زمینه استفاده از داده‌های فضای مجازی برای پیش‌بینی نتایج انتخابات بوده که دو انتخابات گذشته (ریاست جمهوری سال ۹۶ و مجلس سال ۹۸) را با درصد تشابه بالا به آرای واقعی، پیش‌بینی کرده است.

در انتخابات ریاست جمهوری ۱۴۰۰ نیز لایفوب از ابتدای آغاز تبلیغات ریاست جمهوری به داده‌کاوی محتواهای منتشرشده در شبکه‌های اجتماعی پرداخت.

تجزیه و تحلیل حجمی در خصوص کاندیداها یکی از خروجی‌های مستمر بود که از آغاز تبلیغات کاندیداها تا روز انتخابات ادامه داشت.

قبل از مناظره اول و پس از پایان هر مناظره نیز دو خروجی مربوط به تجزیه و تحلیل احساسات مثبت و منفی نسبت به کاندیداها ارائه شد. تحلیل رویکرد محتواها از یک سو و تحلیل رویکرد کاربران در کامنت‌ها از سوی دیگر با عنوان افکارسنجی کاربران شبکه‌های اجتماعی منتشر گردید. در یک روز مانده به انتخابات نیز این افکارسنجی مجدداً انجام گرفت و به دلیل نزدیکی به روز انتخابات این خروجی به عنوان پیش‌بینی از سوی مخاطبان در نظر گرفته شد.

اما نتایج این افکارسنجی و به نوعی پیش‌بینی، با اختلافاتی در رتبه‌بندی نامزدها و درصد آرای واقعی همراه بود. همچنین در مقایسه‌ی نتایج پیش‌بینی سایر شرکت‌های داده‌کاوی با نتایج واقعی انتخابات نیز، این اختلاف‌ها مشاهده شد.

اولین تفاوت و به نظر کارشناسان لایفوب، مهمترین اختلاف، در خصوص رتبه‌بندی کاندیداهاست. اهمیت رتبه‌بندی بسیار مهمتر از تطابق درصدها است. اغلب مجموعه‌های پیش‌بینی کننده پس از انتخابات به ارائه تطابق درصدهای کاندیداها به نتایج خودشان استناد می‌کنند که این باعث گمراهی مخاطبان خواهد شد. ابتدا رتبه‌بندی مهم است و در درجه بعد تطابق درصد آرا. تمام مجموعه‌های پیش‌بینی کننده بر اساس افکارسنجی کاربران شبکه‌های اجتماعی نتوانستند رتبه بندی صحیح را پیش‌بینی کنند و همگی آقای همتی را به عنوان نفر دوم اعلام نمودند. اما چرا رتبه‌بندی مهمتر است، چون اگر انتخابات به فرض قرار بود به مرحله دوم برسد، پیش‌بینی این موضوع که چه کسی به عنوان رقیب نفر اول به دور دوم راه پیدا خواهد کرد، از اهمیت فراوانی برخوردار است.

حال دلایل به وجود آمدن این اختلافات چیست؟

یکم، بحث فراوان بودن آرای باطله در این انتخابات است، همانطور که ایام انتخابات توضیحاتی در خصوص پیش‌بینی‌ها بیان کردیم، دستیابی به میزان آرای باطله با استفاده از افکارسنجی محتواهای شبکه‌های اجتماعی، دشوار است. به عنوان نمونه در بررسی کامنت‌های اینستاگرام بسیاری از کاربران، «عباس بوعدار» را به عنوان رأی خود عنوان کردند. اما آیا می‌توان



گفت این فرد در پای صندوق نیز به عباس بوعدار رأی می‌دهد؟ از کجا معلوم که برای شوخی این کامنت را درج نکرده باشد و اصلا از کجا معلوم که در انتخابات شرکت کند. اما همانطور که در گزارش آرای باطله بیرون آمده از صندوق مشاهده شد، بسیاری از افراد همین نام را در برگه رأی نوشته بودند.

دوم، دو قطبی نبودن انتخابات، فقدان ساختار حزبی در ایران و همچنین تعدد نامزدها تا روزهای پایانی بود که موجب سردرگمی بسیاری از افراد در خصوص کاندیدای مورد نظر خود و در نتیجه عدم حمایت مشخص از آن شخص در شبکه‌های اجتماعی شد. در بسیاری از نظرسنجی‌های میدانی حدود ۲۰ درصد از افرادی که قصد شرکت در انتخابات را داشتند، پس از مناظره سوم هنوز نامزد خود را انتخاب نکرده بودند.

اما مهمترین دلیل اختلاف، ناشی از زمینه‌های ساختاری، نظیر میزان دسترسی و ضریب نفوذ اینترنت است که با توجه به میزان آرای محسن رضایی در مناطق قومیت‌گرا که در ادامه به آن اشاره می‌شود، می‌توان دو فرضیه را مطرح کرد. یکی عدم امکان نمونه‌گیری درست و قابل تعمیم به ویژه با تأکید بر این مناطق است و دیگری عدم دسترسی و فعالیت کاربران در شبکه‌های اجتماعی در برخی نقاط.

با توجه به اینکه آمار دقیقی از میزان کاربران شبکه‌های اجتماعی به تفکیک استانی موجود نیست، ناچاریم برای راستی‌آزمایی این فرضیه به مقایسه آمار ضریب نفوذ اینترنت در برخی استان‌ها با پیش‌بینی انجام شده و همچنین نتیجه انتخابات در استان‌ها بپردازیم.

آنچه در خصوص نتایج انتخابات به دست آمده، در برخی استان‌ها آرای همتی نسبت به رضایی بالاتر بوده و این یعنی مطابقت بیشتر با رتبه‌بندی که از طریق افکارسنجی فضای مجازی به دست آمده است. منتها با بررسی دقیق‌تر، این نکته به دست آمد که این استان‌ها ضریب نفوذ اینترنت بالایی دارند. بر اساس گزارش سازمان مقررات و ارتباطات رادیویی از بین ۳۱ استان کشور در شهریور ۹۹، بیشترین ضریب نفوذ کلی اینترنت به تهران با ۱۲۶.۰۸ درصد اختصاص یافته است. سمنان نیز با ۱۱۳.۶ درصد در جایگاه سوم قرار دارد. استان‌های گیلان با ۱۰۵/۲۱ درصد و یزد با ۱۰۵/۰۸ نیز جزو استان‌های با ضریب نفوذ اینترنت بالا هستند.

نام کاندیدا	تهران	سمنان	گلستان	یزد
رییسی	۲۲%	۷۳%	۵۶%	۷۱%
همتی	۴%	۶%	۱۷%	۸%
رضایی	۳%	۵%	۱۰%	۵%
قاضی‌زاده	۲%	۴%	۳%	۴%

همانطور که در جدول فوق مشخص شده است، آرای همتی در استان‌های با ضریب نفوذ اینترنت بالا، از رضایی بیشتر بوده و این نتایج به پیش‌بینی به دست آمده از فضای مجازی نزدیک‌تر است.

از آن طرف بررسی‌ها نشان داد که اغلب استان‌هایی که درصد آرای رضایی در آن‌ها بسیار بالا بود و اختلاف کمتری نسبت به سایر مناطق با



آرای ریسی داشتند، استان‌هایی بودند که ضریب نفوذ اینترنت کمتری دارند.

طبق گزارش ذکر شده، استان سیستان و بلوچستان با ۶۳.۷۰ درصد از کمترین میزان ضریب نفوذ اینترنت برخوردار است. استان‌های چهارمحال و بختیاری، کهگیلویه و بویر احمد و لرستان نیز نسبت به سایر استان‌ها ضریب نفوذ کمتری دارند. جالب آنجاست که قومیت‌گرایی نیز در این استان‌ها بیشتر از سایر استان‌هاست.

نام کاندیدا	سیستان و بلوچستان	چهارمحال و بختیاری	کهگیلویه و بویر احمد	لرستان
ریسی	۵۲%	۴۸%	۴۴%	۵۹%
رضایی	۲۱%	۳۹%	۴۳%	۲۳%
همتی	۱۰%	۴%	۴%	۶%
قاضی‌زاده	۲%	۲%	۱%	۲%

بر اساس جدول فوق میزان آرای ریسی و رضایی در این استان‌ها بیشترین میزان نزدیکی را به یکدیگر داشته‌اند که به احتمال فراوان در افکارسنجی‌های صورت گرفته، میزان دسترسی کاربران و استفاده از شبکه‌های مجازی در این مناطق مؤثر واقع شده و باعث خطا در پیش‌بینی شده است.

بخشی از این تفاوت در سبب آرای میان دو نامزد انتخابات (همتی و رضایی) را می‌توان ناشی از «صدای سیاسی قدرت» دانست، به این معنا که

میزان دسترسی به فضای مجازی و میزان مشارکت و فعالیت در بین طرفداران همتی بیشتر از رضایی بوده است و حامیان همتی، صدای سیاسی بلندتری در شبکه‌های اجتماعی و رسانه‌ها داشته‌اند.

اما با تمامی اختلافات موجود در پیش‌بینی از طریق افکارسنجی با نتایج انتخابات، یکی از مهم‌ترین مزیت‌های افکارسنجی نسبت به نظرسنجی کم‌هزینه بودن و کم‌زمان‌بر بودن آن است. شاید به خاطر دلایل ذکر شده نتوان از طریق افکارسنجی شبکه‌های اجتماعی به پیش‌بینی دقیق آرای کلی کشور دست یافت اما می‌توان با پیش‌بینی قابل قبولی پیروز انتخابات را مشخص کرد. اینکه بتوان با افکارسنجی داده‌های شبکه‌های اجتماعی، با هزینه و انرژی چندین برابر کمتر از نظرسنجی میدانی و تلفنی، قبل از برگزاری انتخابات، پیروز انتخابات را با درصد اختلاف کمی مشخص کرد، دستاورد بزرگی است.

به طور مشخص اگر نمونه‌گیری‌های دقیق استانی و منطقه‌ای دقیق به حوزه داده‌کاوی فضای مجازی افزوده شود، پیش‌بینی در خصوص رتبه‌بندی و درصد آرای کاندیداها نیز دقیق‌تر خواهد شد.