### شیوه نامه برچسب‌گذاری مجموعه‌داده

در این پایان‌نامه، یک مجموعه‌داده فارسی با استفاده از رابط برنامه‌نویسی کاربردی[[1]](#footnote-1) شبکه‌اجتماعی توییتر در بستر سامانه ذکاوت پژوهشگاه ارتباطات و فناوری اطلاعات[[2]](#footnote-2) جمع‌آوری شده‌است که شامل 27000 توییت فارسی در بازه زمانی 1/3/1400 تا 30/9/1400 می‌باشد. در این مجموعه‌داده در کنار متن کامل هر توییت، نام کاربر منتشر کننده و تاریخ دقیق انتشار آن نیز تعیین شده‌است. برای گردآوری این مجموعه‌داده فارسی در مورد واکسن کرونا، از کلیدواژه‌های مختلفی استفاده کرده‌ایم که چند نمونه از آن‌ها در جدول 3-1 نشان داده شده‌است.

9 جدول 3-1 کلیدواژه‌های مجموعه‌داده جمع‌آوری شده

|  |  |  |  |
| --- | --- | --- | --- |
| کلیدواژه‌ها | | | |
| واکسن کرونا | واکسن کوید-19 | واکسیناسیون | علائم واکسن کرونا |
| عوارض واکسن کرونا | ترس از واکسن | تردید واکسن | اعتماد به واکسن |
| واکسن ایرانی | واکسن خارجی | واکسن نمیزنم | نه به واکسن اجباری |
| نه به واکسن | امید به واکسن | واکسن بخرید | ضد واکسیناسیون |
| علاقه به واکسن | مخالف واکسن | واکسن میسازیم | واکسن نزنید |

پس از بررسی مجموعه‌داده جمع‌آوری شده، موضوع مهمی به چشم می‌خورد که می‌توانست در نتایج تحلیل میزان مقاومت مردم نسبت به واکسیناسیون کرونا (که هدف اصلی این پایان‌نامه است) تاثیر بسزایی داشته باشد. با وجود این‌که تمامی توییت‌ها بیان‌گر احساسات مردم نسبت به واکسن کرونا بودند اما جامعه هدف آنان با یکدیگر متفاوت بود. به عنوان مثال، برخی از نبود واکسن و کند بودن فرآیند واکسیناسیون گلایه داشتند، برخی از ساخت واکسن ایرانی و انتظار برای استفاده از آن ابراز امیدواری و رضایت می‌کردند، برخی دیگر از عوارض بعد واکسن هراسان بودند و بعضی نیز از دریافت واکسن خشنود به نظر می‌رسیدند، عده‌ای نیز کسانی که توطعه‌های علیه واکسن را باور کرده و در مقابل واکسینه شدن مقاومت می‌کردند را به سخره گرفته بودند و تعدادی نیز آن‌هایی که به واکسن اعتماد کرده و حاضر به دریافت آن شده‌بودند را به عدم آگاهی و سهل‌انگاری محکوم می‌کردند. لذا به منظور ارائه یک تحلیل درست از میزان مقاومت نسبت به واکسیناسیون کرونا و علل آن، باید علاوه بر برچسب‌گذاری توییت‌ها بر اساس نوع هیجان، مخاطب یا جامعه هدف آن‌ها نیز دسته‌بندی بشود. جدول 3-2 به خوبی نشان می‌دهد که چند توییت می‌توانند یک نوع هیجان رو نشان دهند در حالی که جامعه هدف آن‌ها با یکدیگر متفاوت است.

10 جدول 3-2 نمونه‌هایی از مجموعه‌داده که نشان‌دهنده‌ی یک نوع هیجان یکسان با مخاطب متفاوت هستند

|  |  |  |
| --- | --- | --- |
| توییت | هیجان | مخاطب |
| بابا خسته شدیم دیگه.کرونا تو ایران خیلی هوشمندانه داره عمل میکنه،مواقع خاص میره خونشون،بعدش یهو میاد بیرون،عدد و امارا خیلی هوشمندانه بالا و پایین میشن.همه دنیا اروم شده واکسن زدن.ما هرروز نقشه کرونامون تغییر رنگ میده که تازه اونم واقعی نیست.اه.به کجا باید گله و شکایت کرد؟ | Anger | **Government** |
| در امارات و بحرین کسانیکه سینوفارم زدن بخاطر ایمنی پایین این واکسن میتونن بوستر فایزر بزنن. بعد ایران هنوز اندر خم یک کوچه برای همین سینوفارم بنجله ! | Anger | **Vaccine** |
| محمدمعتمدی خودش بچه‌شو تو هلند به دنیا آورده دست زن و بچه‌شو گرفته برده اونجا زندگی کنن، و هی میاد میره چهارتا خالتور میزنه و میخونه،بعد در نهایت میره بالا منبر و از واکسن برکت دفاع میکنه و بر دستان محققان و دانشمندان ایرانی بوسه میزنه! | Anger | **People** |

در نتیجه به کمک یک گروه متشکل از دانشجویان و اساتید مجرب در زمینه هوش مصنوعی، 11000 توییت از این مجموعه‌داده در 4 دسته هیجان (ترس یا تردید[[3]](#footnote-3)، تحقیر یا تمسخر[[4]](#footnote-4)، نفرت یا عصبانیت[[5]](#footnote-5)، شادی یا رضایت[[6]](#footnote-6)) و 3 دسته مخاطبان (دولت و عوامل اجرایی در واکسیناسیون[[7]](#footnote-7)، واکسن[[8]](#footnote-8)، افراد متفرقه[[9]](#footnote-9)) به صورت دستی برچسب‌گذاری شدند. در ادامه نکاتی که در فرآیند برچسب‌گذاری توییت‌ها مورد توجه قرار گرفته شده‌است توضیح داده می‌شود.

**1-1-2-3- نکته اول**

توییت‌هایی که بیان‌گر طنز و کنایه هستند و یا مضمون خبری دارند، در صورتی که بطور صریح بیان‌گر یک نوع هیجان از دسته‌بندی منتخب نباشند، حذف می‌شوند (برچسب del برای آن‌ها در نظر گرفته شده‌است). نمونه‌هایی از این نوع توییت‌ها در جدول 3-3 نشان داده شده‌است.

11جدول 3-3 نمونه‌هایی از مجموعه‌داده که باید حذف شوند

|  |  |
| --- | --- |
| توییت | هیجان |
| پیرمرده‌بهم میگه: دخترم تو که جوونی چرا اومدی واکسن بزنی؟ گفتم: دلم پیره پدر جان، نگاه به ظاهرم نکن اینا همش بوتاکسه 😂 | del |
| رئیس جمهور فیلیپین گفته هرکسی می‌خواد واکسن نزنه اینجا رو ترک کنه! لعنتی تو چرا رئیس جمهور ما نیستی. | del |
| با دیپ استید آشنا بشید. جیم سی. اسمیت؛ مدیرعامل اجرایی تامسون رویترز از اعضای مجمع جهانی اقتصاد و دوست خوبمون کلاوس شواب  وووو عضو هیئت مدیره فایزر https://t.co/6IyDtt2Q9v | del |

**2-1-2-3- نکته دوم**

توییت‌هایی که نشان‌دهنده لحن تمسخرآمیز و تحقیر هستند و یا مضمون توهین آمیز غیر مستقیم دارند، به سبب ماهیت موضوع بیماری کرونا و واکسیناسیون آن، در دسته احساس تحقیر یا تمسخر (انزجار[[10]](#footnote-10)) قرار میگیرند. نمونه‌هایی از این نوع توییت‌ها در جدول 3-4 نشان داده شده‌است.

12جدول 3-4 نمونه‌هایی از مجموعه‌داده که نشان‌دهنده‌ی احساس تحقیر یا تمسخر هستند

|  |  |
| --- | --- |
| توییت | نوع هیجان |
| در جدول درصد واکسیناسیون کامل ایران با یک درصد کل جمعیت در رده ۱۱۵ام دنیا قرار داره!همه دنیا دارن واکسینه میشن ایران منتظر برکته! | Disgust |
| دختره متولد 68 نوشته که از ایران متنفرم و کلی ادامه داده، بعدش نوشته دیگه جور شد و منم با آشنابازی واکسن زدم. لامصب بذار جوهر توییت قبلیت خشک شه بعد بیا بگو. تو خودت یکی از آدمایی هستی که باعث میشه منم از اينجا متنفر شم. | Disgust |

**3-1-2-3- نکته سوم**

توییت‌هایی که حاوی کلمات توهین آمیز (فحش و ناسزا)، خشم و عصبانیت با لحن جدی هستند و به طور کلی بار منفی صریحی را به مخاطب منتقل می‌کند، همگی در دسته احساس خشم قرار میگیرند. نمونه‌هایی از این نوع توییت‌ها در جدول 3-5 نشان داده شده‌است.

13جدول 3-5 نمونه‌هایی از مجموعه‌داده که نشان‌دهنده عصبانیت هستند

|  |  |
| --- | --- |
| توییت | نوع هیجان |
| بابا خسته شدیم دیگه.کرونا تو ایران خیلی هوشمندانه داره عمل میکنه،مواقع خاص میره خونشون،بعدش یهو میاد بیرون،عدد و امارا خیلی هوشمندانه بالا و پایین میشن.همه دنیا اروم شده واکسن زدن.ما هرروز نقشه کرونامون تغییر رنگ میده که تازه اونم واقعی نیست.اه.به کجا باید گله و شکایت کرد؟ | Anger |
| دلالان کسانی هستند که به جای واکسن، داروی رمدسیویر وارد کردند.  چون نتوانستند واکسن مورد نظر چون فایزر رو وارد کنند و روسیه اجازه خرید به این دلالان ندادبرای ایران بادارویی مافیای عوارض جانبی زیاد باعث کشتار مردم در این روزها شدند.قوه قضاییه مجلس دولت جدید حساب این دلالان رو برسید | Anger |

**4-1-2-3- نکته چهارم**

برخی از توییت‌ها ممکن است همزمان بیان‌گر چندین هیجان مختلف باشند و یا مخاطب یک توییت ممکن است به بیش از ۱ دسته (از دسته‌بندی ارائه‌شده) اشاره داشته باشد؛ لذا برای این توییت‌ها فقط دسته‌ای که بیشترین ارتباط را دارد در نظر گرفته می‌شود و در صورتی که مضمون توییت نشان‌دهنده دو نوع کلاس با احتمال ۵۰٪ باشد، برای آن توییت نیز برچسب "حذف" در نظر گرفته شده‌است. نمونه‌هایی از این نوع توییت‌ها در جدول 3-6 نشان داده شده‌است.

14جدول 3-6 نمونه‌هایی از مجموعه‌داده که نشان‌دهنده‌ی دو نوع هیجان و یا دو نوع مخاطب هستند

|  |  |  |
| --- | --- | --- |
| توییت | هیجان | مخاطب |
| میگه مرداد کارت دانشجوییتو بیار تا برات واکسن بزنیم. اون پرستاری که توییت داده بود واکسن زده چی بود ایدیش؟ زنده هست؟ اگه زنده مونده تا منم واکسن بزنم. | ~~disgust, fear~~  Fear | Vaccine |
| ما باید از این واکسن داخلی حمایت کنیم تا با خیال راحت تو کشور خودمون واکسن داشته باشیم نه این‌که منتظر کشورای دیگه باشیم که بهمون بفروشن #قیمت\_واکسن | Happiness | ~~government, vaccine~~  Government |

چند نمونه از دسته‌بندی درست توییت‌ها در انواع هیجان و مخاطب در جدول 3-7 نشان داده شده‌است.

15جدول 3-7 نمونه‌هایی از برچسب‌گذاری درست در مجموعه‌داده

|  |  |  |
| --- | --- | --- |
| توییت | هیجان | مخاطب |
| واکسن زدم :) | Happiness | Vaccine |
| عُرضه‌ی یه واکسیناسیون هم ندارید. هر چی بی عرضگی و بی کفایته تو وجود شماها جمع شده. فقط ادعا!! این همه آدم مسن باید تو هم بلولن بخاطر یه دوز واکسن؟! طرف رو‌روتخت آمبولانس آوردن تو صف! از شرم برید بمیرید! #واکسن\_بخرید | Anger | Government |
| بگم از واکسن میترسم میخندین بهم😐😭؟ چنروزه یادم اومده باس واکسن کزاز بزنم بخاطر دبیرستان و هی یادم میاد میخام برم گمشم اصن نمیخاممممم:)این ترس چیه من دارم چرا میترسم دقیقا؟ خوشبحالتون اگه ترس این چیزارو ندارین | Fear | Vaccine |
| من هنوز نتونستم به ملت بفهمونم که تب و فلان بعد از واکسن عوارض نیست .دیوانه کردن مارو | Disgust | People |

1. Application Programming Interface [↑](#footnote-ref-1)
2. ICT Research Institute [↑](#footnote-ref-2)
3. Fear [↑](#footnote-ref-3)
4. Disgust [↑](#footnote-ref-4)
5. Anger [↑](#footnote-ref-5)
6. Happiness [↑](#footnote-ref-6)
7. Government [↑](#footnote-ref-7)
8. Vaccine [↑](#footnote-ref-8)
9. People [↑](#footnote-ref-9)
10. Disgust [↑](#footnote-ref-10)