

فصل ۱ – آشنایی با مفهوم اخبار جعلی و روش‌های تشخیص آن

مقدمه

همه‌گیری جدید ویروس کرونا (Covid-19) از سال 2019 توجه بسیاری را به خود جلب کرده و باعث بحران عمیق بهداشتی در سرتاسر جهان شده است. از بدو ورود این ویروس در جوامع انسانی اطلاعات بسیاری تاکنون در خصوص آن منتشر شده است. شبکه‌های اجتماعی که امروزه بخشی جدایی ناپذیر در زندگی روزمره انسان‌ها هستند، سهم بسیار بالایی در تولید و انتقال داده‌های مرتبط با ویروس کرونا دارند. ویروس کرونا با ایجاد ترس و وحشت در میان مردم باعث بروز مشکلات بسیاری شده است. در میان انبوه داده‌های تولید شده در بستر شبکه‌های اجتماعی، اطلاعات جعلی و گمراه‌کننده‌ای وجود دارد که منجر به تشدید هراس جمعی در خصوص ویروس کرونا

شده است. سرعت انتشار شایعات مانند سرعت انتشار سایر اطلاعات در شبکه‌های اجتماعی بسیار بالاست. به جهت جلوگیری از انتشار طوفان افکار منفی و نادرست در زمان همه‌گیری‌ها، مبارزه با شایعات و جلوگیری درست و به موقع از انتشار آن‌ها نقش به‌سزایی در مبارزه با بیماری دارد و اگر نتوان مانع انتشار به موقع شایعات شد، تهدیدی برای ثبات اجتماعی و امنیت شهروندان خواهند بود. بنابراین شبکه‌های اجتماعی این توانایی را دارند که باعث انتشار گسترده اطلاعات درست و نادرست شوند و میلیون‌ها نفر را تحت تاثیر آن قرار دهند. در نتیجه لازم است قبل از اینکه شایعات مرتبط با کووید-19 تاثیرات جدی و مخرب بر زندگی مردم بگذارند، شناسایی و مهار گردند.

از زمان آغاز این همه‌گیری تحقیقات زیادی در حوزه تشخیص شایعات در جهان صورت گرفته است. به زبان فارسی نیز اقداماتی برای تشخیص شایعات با موضوعات متنوع انجام شده است؛ اما موضوع اصلی پژوهش ما تشخیص شایعات پیرامون همه‌گیری کووید-19 می‌باشد. هدف اصلی ما از انجام پروژه پیش‌رو در گام اول، گردآوری یک مجموعه داده با برچسب شایعه و غیر شایعه در خصوص بیماری کرونا به زبان فارسی است. در گام دوم پروژه، با بهره‌گیری از الگوریتم‌های یادگیری عمیق به تشخیص شایعه خواهیم پرداخت.

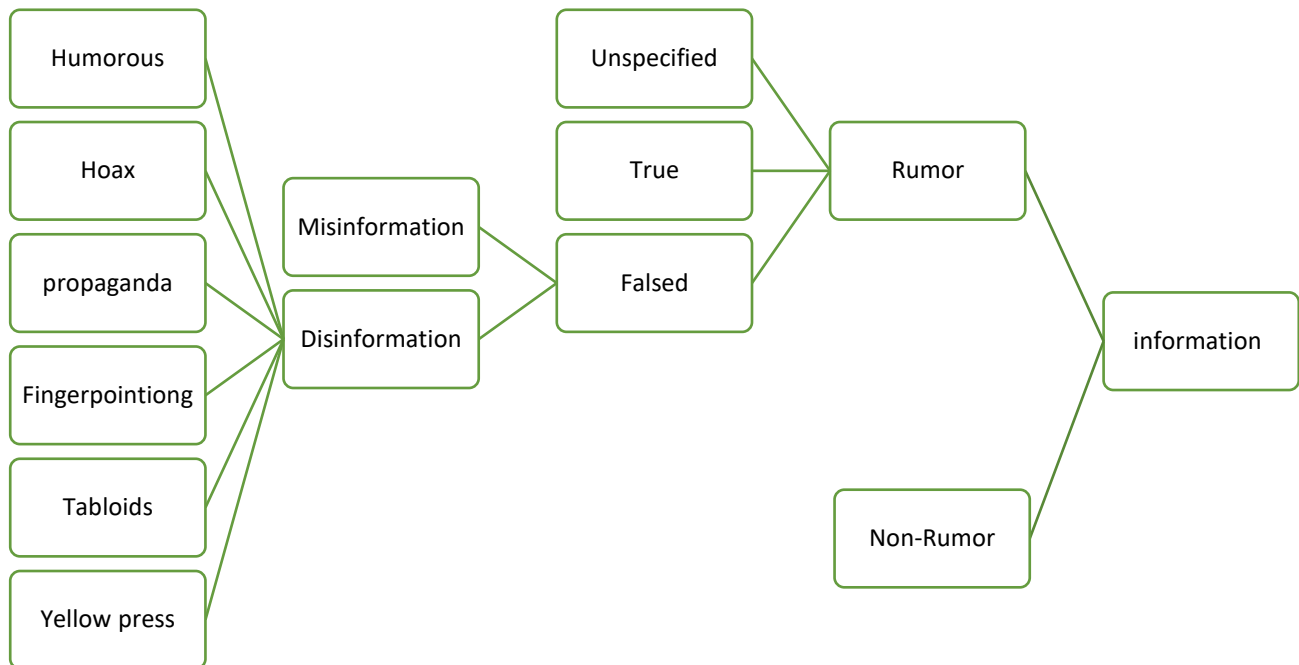
1-1 تعریف شایعه و انواع آن

شایعه به اطلاعاتی گفته می‌شود که صحت آن مشکوک است. برخی شایعات ممکن است درست، برخی نادرست و برخی تائید نشده باشند.

در تعریفی دیگر از شورای عالی فضای مجازی شایعه عبارت است از هر گونه اطلاعات، اخبار و محتوای خبری منتشر شده در فضای مجازی شامل انواع صور متنی، تصویری یا چندرسانه‌ای که مابه‌ازایی در واقعیت نداشته یا شکل تحریف شده‌ای از یک واقعیت است که موجب آسیب و اخلال در نظم عمومی، تشویش اذهان عمومی،

کاهش ثبات، انسجام، اعتماد، سلامت و امنیت فرد و جامعه در ابعاد اقتصادی، اجتماعی، سیاسی و یا کسب منافع نامشروع می‌شود.

رسانه‌های اجتماعی این قدرت را دارند که هر اطلاعاتی را خواه درست یا نادرست، تبدیل به ویروس کنند و میلیون‌ها نفر را تحت تاثیر آن قرار دهند. از این رو لازم است این شایعات را قبل از اینکه تأثیر جدی بر زندگی مردم بگذارد، شناسایی و مهار کرد. به طور کلی شایعات را می‌توان به صورت زیر طبقه بندی نمود:



با توجه به طبقه‌بندی فوق، اطلاعات به دو دسته شایعه و غیر شایعه تقسیم می‌شوند. اطلاعات با برچسب شایعه ممکن است درست، نادرست و یا تأیید نشده باشند. شایعات درست حاوی اطلاعات صحیح در مورد یک خبر، متن یا ... هستند. و اما شایعات نادرست، عامل اصلی گمراهی و تشویش، اخلاق و ... هستند. عموماً شایعات نادرست به دو دسته misinformation و disinformation تقسیم می‌شوند. جهت آشنایی با تعاریف هر یک از این دو دسته و زیر مجموعه‌های مربوط به آن به تعاریف زیر توجه نمائید:

➤ **Misinformation:** این دسته از شایعات صرفاً اشتباهات صادقانه مردم هستند که بدون نیت و قصد قبلی منتشر می‌شوند.

➤ **Disinformation:** شایعات عمدی هستند که برای گمراه کردن مردم استفاده می‌شوند. این اطلاعات تحت عنوان اطلاعات نادرست برچسب‌گذاری می‌شوند که براساس قصد ایجاد کننده آن به شرح زیر طبقه‌بندی می‌شوند:

- طنز (Humorous): بی‌ضررترین نوع شایعات است. منابعی که این نوع اطلاعات را منتشر می‌کنند اخبار و داستان‌هایی را جعل می‌کنند تا جنبه سرگرم کننده به آن بدهند. انگیزه معمولاً سرگرم کردن مردم می‌باشد. این دسته از اطلاعات از قبل نادرست اعلام شده و فقط برای اهداف طنز در نظر گرفته می‌شوند. مانند طنزهای خبری.
- فریب (Hoax): جعل عمدی اخبار است. هدف آن ایجاد وحشت در بین مردم و ایجاد دردسر برای افرادی است مورد هدف آن هستند. مانند اخبار نادرست مبنی بر افزایش قیمت محصولات غذایی و ... برای ایجاد رعب و وحشت.
- مبالغه در تبلیغ (propaganda): نوعی از اخبار و اطلاعات غلط است که با سیاه‌نمایی سعی در تغییر خبر به نفع خود دارد. به عنوان مثال شلوغ کاری‌های رسانه‌ای، مبالغه در محتوای خبر منتقل شده و ...
- Finger printing: این دسته از شایعات نادرست به دنبال منافع شخصی، سازمانی و ... است. هدف این نوع شایعات خدشه‌دار کردن وجهه فرد، سازمان، حزب، گروه و ... است.

- **Tabloid**: نوعی شایعه است که بر داستان‌های پرشور و شایعات در مورد افراد مشهور تاکید می‌کند.
- **Yellow press**: حالت روزنامه‌نگاری زرد و اخبار زرد را دارد. نوعی از اطلاعات است، که با تحقیق کم و یا بدون تحقیق گزارش می‌شوند. تنها هدف خبرنگار جلب توجه با استفاده از تیترهای جذاب و بدون توجه به صحت اخبار است. آنها به خود زحمت نمی‌دهند که در یک خبر و داستان عمیق شوند و فقط آن را منتشر می‌کنند.

1-2 مراحل تشخیص شایعات

فرآیند بررسی ادعای یک خبر به جهت تشخیص صحت آن یک فرآیند چند مرحله‌ای چالش برانگیز است. معمولاً برای انجام این فرآیند مراحل زیر باید طی شود [20]:

- 1- بازیابی اسنادی که به طور بالقوه مدرکی برای اثبات یک ادعا در مورد خبر هستند.
 - 2- پیش بینی موضع هر سند در رابطه با ادعا
 - 3- تخمین و برآورد میزان قابل اعتماد بودن اسناد
 - 4- تصمیم گیری بر اساس تجمیع حداکثریت آراء برای کلیه استادی که از مرحله (1) بدست آمده‌اند.
- با توجه به فرآیند 4 مرحله‌ای فوق دو گام ضروری برای تشخیص شایعات در شبکه‌های اجتماعی وجود دارند که عبارتند از:

- جمع آوری مجموعه داده برچسب دار
- استفاده از الگوریتم‌های مطرح برای تشخیص شایعات

• مجموعه داده:

به علت متفاوت بودن ماهیت شبکه‌های اجتماعی، محققان وجود مجموعه داده را برای تشخیص شایعات یک نیاز ضروری می‌دانند. تاکنون تحقیقات زیادی با هدف گردآوری مجموعه داده برای شایعات انجام شده است که تعدادی از معروف‌ترین آنها در بخش پیشینه تحقیق ذکر گردید.

اما جدای از مجموعه داده‌های موجود، جمع‌آوری مجموعه داده برچسب‌دار متناسب با زبان فارسی کاری بسیار ارزشمند و ضروری است. در این مرحله از کار باید عملیات fact checking انجام شود. روش‌های متنوعی برای این عملیات وجود دارد. مانند سیستم‌های محاسباتی که برخی محققان ایجاد کرده‌اند که ادعاهایی را که ممکن است نیاز به تأیید داشته باشند را شناسایی میکنند. از دیگر اشکال آن سایت‌های راستی‌آزمایی است که با استفاده از آنها می‌توان ادعای فرد یا افراد را مورد بررسی قرار داد. اما نوع دیگری از fact checking ارزیابی خود متن مورد بررسی است. یعنی نشانه‌هایی را بیابیم که در منابع خبری رایج غیر معمول است. این یک مشکل طبقه‌بندی متون است که معمولاً توسط تکنیک‌های NLP به آن پرداخته می‌شود. بنابراین چالش اصلی ما در این مرحله گردآوری مجموعه داده برچسب‌دار شایعات، به زبان فارسی است.

فصل 2 – برچسب گذاری مجموعه داده

در این فصل، مراحل اجرایی برچسب‌گذاری مجموعه داده جمع‌آوری شده با هدف تشخیص شایعات در شبکه‌های اجتماعی با متن فارسی در خصوص بیماری کرونا به ترتیب شرح داده می‌شود. لازم به ذکر است مجموعه داده مورد استفاده در این پژوهش مبتنی بر داده‌های جمع‌آوری شده فارسی توئیت در بستر سامانه ذکاوت پژوهشگاه ارتباطات و فناوری اطلاعات بوده و همچنین به جهت پردازش زبان فارسی نیز از ابزارهای توسعه داده شده در این سامانه استفاده خواهد شد.

توجه نمایید که مجموعه داده جمع‌آوری شده در این پژوهش، تمامی اطلاعات منتشر شده در خصوص بیماری کرونا به زبان فارسی می‌باشد. و هدف ما برچسب زدن داده‌های مرتبط با کرونا و تشخیص شایعاتی است که در طول دوران این همه‌گیری در بستر توئیت منتشر شده است.

2-1 هدف برچسب زنی مجموعه داده

برچسب‌گذاری داده، بخشی از مرحله پیش پردازش در هنگام توسعه یک مدل یادگیری ماشینی (ML) است. این نیاز به شناسایی داده‌های خام (یعنی تصاویر، فایل‌های متنی، ویدئوها) و سپس افزودن یک یا چند برچسب به آن داده‌ها دارد تا زمینه آن را برای مدل‌ها مشخص کند و به مدل یادگیری ماشینی امکان پیش‌بینی دقیق را بدهد.

برچسب‌گذاری داده‌ها زیربنای موارد مختلف یادگیری ماشینی و یادگیری عمیق، از جمله بینایی رایانه و پردازش زبان طبیعی (NLP) است. فواید برچسب‌گذاری عبارتند از:

- پیش‌بینی‌های دقیق‌تر:

برچسب‌گذاری دقیق داده‌ها تضمین کیفیت بهتری را در الگوریتم‌های یادگیری ماشین تضمین می‌کند و به مدل اجازه می‌دهد تا خروجی مورد انتظار را آموزش دهد و به دست آورد.

- قابلیت استفاده بهتر از داده‌ها:

برچسب گذاری داده ها همچنین می تواند قابلیت استفاده از متغیرهای داده را در یک مدل بهبود بخشد. برای مثال، ممکن است یک متغیر طبقه بندی را به عنوان یک متغیر باینری مجدداً طبقه بندی کنید تا برای یک مدل قابل مصرف تر شود. تجمیع داده ها به این روش می تواند مدل را با کاهش تعداد متغیرهای مدل بهینه کند یا امکان گنجاندن متغیرهای کنترلی را فراهم کند.

در نتیجه برای رسیدن به یک مدل یادگیری ماشین دقیق و همچنین استفاده بهینه و درست از مجموعه داده، باید پس از جمع آوری مجموعه داده های خود در خصوص بیماری کرونا آن ها را برچسب گذاری نمائیم تا به بهترین نتیجه مورد نظر برسیم.

2-2 گام های تعیین برچسب های منتخب برای تشخیص شایعات

در این مسیر به جهت تایید برچسب های متناسب با داده های جمع آوری شده از بستر توئیت، مراحل زیر طی شده است:

- کاربران توئیت با بیشترین میزان تاثیرگذاری در خصوص موضوع کرونا شناسایی شدند و سمت، تعداد پست های منتشر شده از آن ها مورد بررسی قرار گرفت. در ادامه نقش این افراد برای هر نوع اقدامی در خصوص بیماری کرونا اعم از اطلاع رسانی، افزایش آگاهی و ... بررسی گردید. جدول زیر لیستی از کاربران تاثیرگذار در شبکه اجتماعی توئیت با محوریت کرونا است:

| نام | نام کاربری | تعداد پست | تعداد دنبال کننده | سمت |
|-----------|-------------|-----------|-------------------|-------------------------|
| سعید نمکی | Saeednamaki | 129 | 33.6 k | وزیر بهداشت دولت یازدهم |

| | | | | |
|---|---------|---------|-----------------|--------------------------|
| مدیرکل روابط عمومی وزارت بهداشت دولت یازدهم | 30.5 k | 7493 | Drjahanpur | دکتر کیانوش جهانپور |
| خبرنگار | 44.1 k | 9397 | Vahabzade_ali | علیرضا وهاب زاده |
| مجری برنامه و تلویزیون فعال اصولگرا | 196.2 k | 4097 | Sabeti_twt | امیرحسین ثابتی |
| هفتمین رئیس جمهور | 847 k | 3442 | Rouhani_ir | حسن روحانی |
| رئیس مجلس شورای اسلامی | 240.8 k | 551 | Mb_ghalibaf | محمدباقر قالیباف |
| گوینده شبکه خبر | 178.6 k | 5401 | Elmira_sharifi | المیرا شریفی |
| خبرگذاری دانشجویان ایران | 247.3k | 79.7k | Isna_farsi | خبرگذاری ایسنا |
| شبکه‌ی تلویزیون ماهواره‌ای، کانال رادیویی و وبسایت خبری به زبان فارسی مستقر در لندن | 609.5k | 129.6 k | IranIntl | ایران اینترنشنال |
| کاربر توئیتر | 5388 | 53.7 k | Abrahimsoltani2 | ابراهیم سلطانیان |
| روزمره نویس | 80.1 k | 92.2 k | inlygif770 | مرمی مهریون فالو فالو |

- با استفاده از خبرگذاری‌های معتبر و همچنین سایت‌های صحت‌سنج مانند factnameh و factYar، یک فکت نامه معتبر ایجاد نمودیم، که شامل 175 خبر صحیح و تأییدشده 100 خبر نادرست و جعلی درخصوص بیماری کووید-19 می‌باشد.

2-3 نمونه برچسب برای مجموعه داده توئیتر

در این بخش نمونه‌هایی از اخبار برچسب خورده را به همراه دلایل انتخاب برچسب شرح داده می‌شود:

الف) توئیت‌ها با محتوای طنز و کنایه

در این مطالعه توئیت‌هایی که محتوای طنز و کنایه را در خود داشته باشند، به گونه‌ای که به طور صحیح محتوای درست یا نادرست بودن را منتقل نکنند، از مجموعه داده حذف خواهند شد:

| توئیت | برچسب | نشانه |
|--|-------|-------|
| هر کس #واکسن گیرش نماید، بیاد سرکوچه ما هر دوز دو دلار . فراوون موجوده | حذف | del |
| عمر ما به سر رسید کلاغ هم به خونه‌اش رسید واکسن بیصاحب هنوز نرسیده :// | حذف | del |

ب) توئیت‌های درست

این دسته از توئیت‌ها اطلاعاتی می‌باشند که کاملاً صحیح هستند و با مراجع معتبر جهان در زمینه کووید-

19 مطابقت دارند:

| توئیت | برچسب | نشانه |
|--|-------|-------|
| تصویر بیلبوردی در شیکاگو حاوی توصیه های بهداشتی منسوب به پیامبر اسلام | درست | — |
| باور غلط نوشیدن الکل برای جلوگیری از کرونا باعث مسمومیت و مرگ شماری از ایرانیان شده است. | درست | — |
| ویدئو پرستاری در ایتالیا که با دریافت رشوه واکسن را به جای تزریق در دستمال خالی میکند. | درست | — |

د) توئیت‌های نادرست

گفته یا آمار این دسته از اطلاعات نادرست است یا دست کم سندی معتبر آن را رد می‌کند:

| توئیت | برچسب | نشانه |
|-------|-------|-------|
|-------|-------|-------|

| | | |
|---|--------|--|
| * | نادرست | طلاب می‌توانند بدون نوبت واکسن بزنند. |
| * | نادرست | ۲ هزار ایرانی دو هفته پس از تزریق واکسن کرونا فوت کرده‌اند. |
| * | نادرست | 82 درصد از زنان بارداری که در سه ماهه اول و دوم بارداری واکسن کرونا زدند دچار سقط جنین شدند. |

ه) نمونه توئیت‌ها با برجسب تحریفات

و دسته آخر نوعی از اخبار و اطلاعات غلط است که با سیاه‌نمایی و تحریف اطلاعات، سعی در تغییر ماجرا به نفع خود دارد.

| نشانه | برجسب | توئیت |
|-------|---------|--|
| p | تحریفات | در کشورهای موفق در کنترل کرونا "مرکز بحران" یا جایی که بیشترین شیوع ویروس گزارش شده رو به مردم گزارش می‌کنن که هم امکانات بیشتری اونجا باشه، هم مردم بیشتر مراقب باشن هم سفرها به اونجا کمتر باشه، اینجا اومدن آمار استان به استان رو لغو کردن، فقط در ایران اشغالی با مدیریت اقتصادش. |
| p | تحریفات | وحشتناک وحشتناک وحشتناک، آمار مبتلایان به کرونا در ایران به شدت رو به افزایش است. |

برای دریافت اطلاعات بیشتر از پروژه با ایمیل zekavat@itrc.ac.ir تماس بگیرید.