

## تشخیص زود هنگام شایعات در رسانه‌های اجتماعی با استفاده از سیستم استنتاج فازی عصبی تطبیقی

سید ابوالفضل حمزه پور<sup>۱</sup>، پویا پارسا<sup>۲</sup>، علیرضا یوسف پور<sup>۳</sup>، سید سعید حمیدی<sup>۴</sup>

<sup>۱</sup> دانشجوی کارشناسی ارشد شبکه های کامپیوتری، دانشکده فنی مهندسی، دانشگاه آزاد، قائم شهر

<sup>۲</sup> دانشجوی کارشناسی ارشد نرم افزار، دانشکده فنی مهندسی، دانشگاه آزاد، قائم شهر (نویسنده مسئول)

<sup>۳</sup> استادیار، دانشکده فنی مهندسی، دانشگاه آزاد، قائم شهر

<sup>۴</sup> مربی، دانشکده فنی مهندسی، دانشگاه آزاد، قائم شهر

### چکیده

شایعات، اخبار تأیید نشده و اغلب اشتباهی هستند که به صورت وسیع در سطح جامعه منتشر و موجب سلب اعتماد یا افزایش کاذب اعتماد گره‌های یک شبکه به یک نهاد یا موضوع می‌شوند. با فراگیر شدن شبکه‌های اجتماعی در سال‌های اخیر، به رغم کاربردهای مثبت آن‌ها، انتشار شایعات ساده‌تر شده است. شایعات یک چالش امنیتی در شبکه‌های اجتماعی محسوب می‌شوند، از این رو تشخیص زود هنگام شایعات چالش مهمی در سازوکارهای امنیت است. در این پژوهش، برای تشخیص شایعات رویکردی مبتنی بر سامانه استنتاج فازی تطبیقی عصبی (ANFIS) ارائه شده است. این رویکرد برای تشخیص شایعه بر پایه استنتاج پرسش‌گری‌های کاربران طراحی شده و در ارزیابی‌های انجام شده توانسته بهبود قابل توجهی در مقابل سایر روش‌های پیشین بدست آورد.

**واژه‌های کلیدی:** شبکه اجتماعی، تشخیص شایعه، توئیت‌های پاسخ، سیستم استنتاج فازی عصبی تطبیقی

## ۱. مقدمه

اهمیت رسانه‌های اجتماعی، به عنوان منبع اصلی اطلاعات به روز از این واقعیت ناشی می‌شود که هر کسی می‌تواند فوراً اطلاعات مربوط به اخبار فوری را پست کند، به اشتراک بگذارد و یا جمع‌آوری کند. این انعطاف‌پذیری در به اشتراک‌گذاری و تبادل اطلاعات میان کاربران این شبکه‌ها قریب به اتفاق با حجم عظیمی از اطلاعات جدید در هرثانیه همراه است. متأسفانه، این اطلاعات همیشه قابل اعتماد نیستند، زیرا عدم میانجی‌گری یک خط سردبیری امکان انتشار اطلاعات تایید نشده را فراهم می‌کند. این سناریو به نفع گسترش اطلاعات نادرست و مغرضانه است که به دنبال تأثیرگذاری بر افکار عمومی با اطلاعات نادرست است. به طور خلاصه، آنچه که فضیلت شبکه‌های اجتماعی به نظر می‌رسید، به تهدید اصلی آن نیز تبدیل شده است. تا جایی که شناسایی شایعات در این شبکه‌ها و کنترل آن در اولین فرصت برای به حداقل رساندن آسیب از اهداف دولت‌ها و صاحبان این شبکه‌ها شده است. بنابراین طراحی و پیاده‌سازی سیستم‌هایی که به صورت خودکار شایعات را شناسایی کنند، از اهمیت بسیار زیادی برخوردار است.

امرتشخیص شایعه گاه نیازمند دانشی ضمنی است تا صحت و سقم یک ادعا با ادله عقلی سنجیده شود که این امر برای انسان قابل انجام است و چون این کار برای کامپیوتر دشوار است، برای داشتن این اطلاعات، نیازمند برخورداری از یک دانش‌نامه می‌باشد که توسط آن، اطلاعات سنجیده شود و استدلال اضافه شود. تا به امروز روش‌های زیادی برای تشخیص و شکست شایعات ارائه شده است.

مطالعات نشان داده شبکه‌های اجتماعی در تحلیل صحت شایعات ارزش زیادی دارند. نشان داده شده است که کاربران شبکه‌های اجتماعی تمایل دارند شایعات نادرست را زیر سوال ببرند یا انکار کنند [۱]، یکی از اولین روش‌هایی که از این دیدگاه پیروی می‌کرد توسط کاستیو<sup>۱</sup> و همکاران ارائه شد [۲] که ویژگی‌های زیادی را از تئوئتر استخراج کرد تا اخبار را بر اساس سطح صحت آن‌ها طبقه‌بندی کنند. در ادامه، پیشرفت‌ها و توسعه‌های زیادی برای روش‌های معرفی شده توسط کاستیو و همکاران ارائه شد. برای مثال، ژائو<sup>۲</sup> و همکاران [۳] نشان داد با استفاده از تئوئیت‌هایی که اعتبار پست قبلی را پرس‌وجو می‌کنند یا به چالش می‌کشند می‌توان مشخص شود که آیا شایعه است یا خیر. آن‌ها برای اولین بار از واژه تئوئیت‌های پرس‌وجو استفاده کردند، یک تئوئیت در صورتی که با یک دسته عبارات منظم و تنظیم شده به صورت دستی مطابقت داشته باشد، پرس‌وجو تلقی می‌شود. اخیراً وگنو<sup>۳</sup> و همکاران نشان داد که تشخیص موضع خودکار بر اساس ویژگی‌های واژگانی یک مشکل چالش‌برانگیز است که برای حل موفقیت‌آمیز به معماری‌های یادگیری عمیق پیچیده نیاز دارد [۴].

در این تحقیق ما قصد داریم یک سیستم تشخیص شایعه مبتنی بر رویکردهای شبکه‌های عصبی برای تشخیص زود هنگام شایعات در شبکه‌های اجتماعی ایجاد کنیم. جهت رسیدن به این مهم از کنترل ویژگی‌های به دست آمده از پست‌های پاسخ و ساختار ارتباطی کاربری که پست اصلی را تولید کرده به کمک سیستم استنتاج عصبی فازی تطبیقی بهره خواهیم برد. پیش‌تر آن‌چه ژائو و همکاران [۳] روی آن تمرکز داشته‌اند وجود یک سری قطعه عبارات تنظیم شده برای شناسایی توییتهای پرس‌وجو بوده؛ این توییتهای پرس‌وجو بر اساس شباهت‌ها خوشه‌بندی می‌شدند، و هر خوشه در نهایت یک شایعه نامزد تلقی می‌شود. ما معتقدیم که عموم مردم همیشه صحت شایعات را به شکل مشخص و عبارات منظم زیر سوال نخواهند برد و حتی امکان دارد کاربرانی به دلایلی مختلف اقدام به تایید اخبار تایید نشده بزنن. برای مهار بهتر زمینه پیرامون یک تویییت، ما معتقدیم که طبقه‌بندی‌کننده باید از ترکیب ویژگی‌های محتوای پاسخ‌ها با ویژگی‌های اجتماعی از خود تئوئیت اصلی، به تجزیه و

Castillo

Zhao

Buguño

تحلیل شایعات نوظهور اقدام کند. با تشخیص زود هنگام شایعات می‌توان هم برای محدود کردن انتشار اطلاعاتی که ممکن است متعاقباً نادرست باشد و هم برای کاهش خطر آسیب به افراد و جامعه قدمی موثر برداشت. مطالبی که در ادامه به آن خواهیم پرداخت به این شرح است: ابتدا در بخش بعدی پیشینه‌ای از شایعات در تحقیقات گذشته بیان خواهیم کرد، در بخش سوم به بیان روش پیشنهادی این تحقیق و شرح کامل آن می‌پردازیم، در بخش چهارم نتایج حاصل از تحقیق و نمودارهای عملکرد و مقایسه آورده‌ایم و در پایان به بیان نتیجه عملکرد و تحقیق خود پرداخته‌ایم.

## ۲. پیشینه و ادبیات تحقیق

در ادبیات تحقیقاتی از تعاریفی مختلفی برای شایعه استفاده شده است که با یکدیگر متفاوت هستند. برای مثال، برخی از تحقیقات، شایعه را به عنوان اطلاعاتی که نادرست تلقی می‌شوند بیان شده (به عنوان مثال، کای<sup>۴</sup> و همکاران [۵] و لیانگ<sup>۵</sup> همکاران [۶])، در حالی که اکثر متون، شایعات را به عنوان «گزاره‌های اطلاعاتی تایید نشده مرتبط به یک واقعه» [۷] تعریف می‌کنند. تعریف ارائه شده توسط فرهنگ لغت‌های بزرگ، مانند فرهنگ لغت انگلیسی آکسفورد<sup>۶</sup> که شایعه را به عنوان «یک داستان یا گزارش از حقیقت نامشخص یا مشکوک در حال انتشار» تعریف می‌کند. این اطلاعات تایید نشده ممکن است درست یا حدوداً یا کاملاً نادرست باشد. امکان دارد هیچ‌گاه صحت درستی یا نادرستی آن مشخص نشود. دانوان<sup>۸</sup> این گونه بیان می‌کند که «یک شایعه را می‌توان به عنوان یک مورد از اطلاعاتی که هنوز تأیید نشده است قبول کرد در حالی که ارزش حقیقت آن در زمان انتشار حل نشده باقی می‌ماند». یک شایعه زمانی به عنوان تایید نشده تعریف می‌شود که هیچ مدرکی یا هیچ تاییده رسمی از منابع معتبر برای تایید آن وجود نداشته باشد. طبقه‌بندی‌های مختلفی برای شایعه وجود دارد، وقتی صحبت از توسعه سیستم طبقه‌بندی شایعات می‌شود، عاملی که تا حد زیادی رویکردهای مورد استفاده را تعیین می‌کند، ویژگی‌های زمانی آن‌ها است: شایعات نوظهور: شایعاتی که در زمینه اخبار فوری به وجود می‌آیند عموماً شایعاتی هستند که قبلاً مشاهده نشده بودند. شایعات دیرینه: شایعاتی که مدت‌ها در محافل دست‌به‌دست می‌شوند بدون اینکه صحت و سقم آن‌ها با قطعیت ثابت شود. در پژوهش‌های گذشته [۹] ویژگی‌های مختلفی برای تشخیص شایعات استفاده شده که در هر روش به سطوح مختلفی از مهندسی ویژگی نیاز است که برخی روش‌ها نیازمند استخراج دستی کلیه بردارهای ویژگی و برخی دیگر تنها نیازمند سطحی از استخراج دستی این بردارها هستند؛ با توجه به این که بسیاری از ویژگی‌ها به هم مرتبط هستند بسیار مهم و ضروری است که این ویژگی‌ها به درستی استخراج شوند و این امر یک گام اساسی در تشخیص شایعات است. براساس شکل ۱ ویژگی‌ها را از چهار بعد می‌توان مورد بررسی قرارداد:

ویژگی‌های مبتنی بر محتوا

ویژگی‌های مبتنی بر کاربر

ویژگی‌های مبتنی بر شبکه انتشار

ویژگی‌های موقت زمانی و مکانی

Cai

Liang

Oxford

Donovan



شکل ۱-ویژگی‌های مورد استفاده در تشخیص شایعات [۹]

### ۳. روش تحقیق

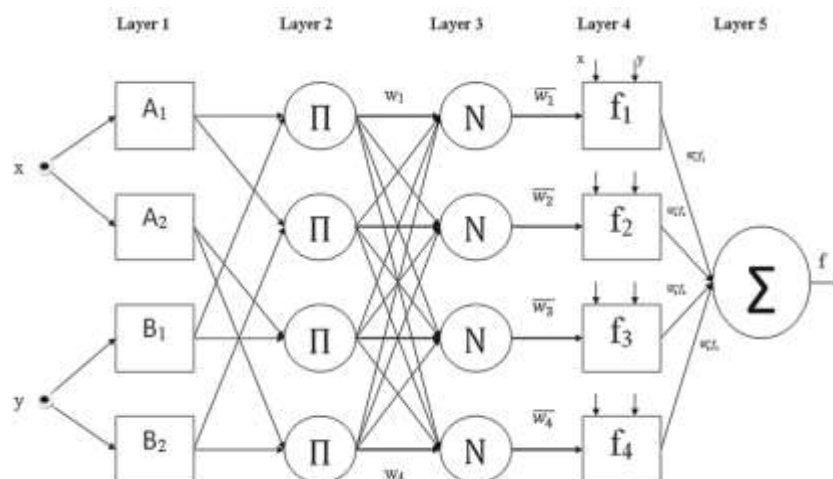
در این تحقیق قصد داریم با استفاده از سیستم استنتاج عصبی فازی تطبیقی<sup>۸</sup> (ANFIS) و به کمک توئیت‌های پاسخ جهت تشخیص شایعه یا واقعیت توئیت‌ها در زمان نشر آن‌ها اقدام کنیم. از آنجایی که این سیستم، شبکه‌های عصبی و مفاهیم منطق فازی را یکی می‌کند، می‌توان از امکانات هر دو آن‌ها به صورت هم‌زمان بهره برد. سیستم استنتاج آن مطابق با مجموعه قوانین فازی اگر-آنگاه می‌باشد که قابلیت یادگیری برای تقریب زدن توابع غیرخطی را دارد و در نتیجه این سیستم به خوبی می‌تواند با کمک تجمیع ویژگی‌های محتوایی استخراج شده از توئیت‌های پاسخ با ویژگی‌های اجتماعی خود توئیت اصلی پست‌ها را به دو دسته شایعه و واقعیت کلاس‌بندی کند. به این ترتیب تصمیم‌گیری و طبقه‌بندی نهایی می‌تواند با دقت بیشتری انجام گیرد. سیستم فازی استنتاجی عصبی تطبیقی<sup>۱۰</sup> (ANFIS) [۱۰] از دو بخش به نام‌های مقدمه و پیامد تشکیل شده است. آموزش ANFIS به معنای تعیین پارامترهای متعلق به این قسمت‌ها با استفاده از الگوریتم بهینه‌سازی است. ANFIS از جفت داده‌های ورودی-خروجی موجود در طول آموزش استفاده می‌کند. سپس قوانین فازی اگر-آنگاه مبنی بر اتصال این قطعات به یکدیگر به دست می‌آید.

Adaptive fuzzy neural inference system

Inference

If-then

Adaptive Network Fuzzy Inference System



شکل ۲- ساختار ANFIS که دارای دو ورودی و یک خروجی است [۱۵]

ساختار ANFIS همانطور که در شکل ۲ مشاهده می‌شود از پنج لایه تشکیل شده است. در این شکل ساختار ANFIS با ۲ ورودی و یک خروجی آورده شده است که از چهار تابع عضویت و چهار قانون تشکیل شده است. در ادامه لایه‌های ساختار ANFIS با توجه به شکل ۳ آمده است.

### لایه اول

این لایه فازسازی<sup>۲</sup> نامیده می‌شود. این لایه از توابع عضویت برای به‌دست آوردن خوشه‌های فازی از مقادیر ورودی استفاده می‌کند. پارامترهایی تابع عضویت را تعیین می‌کنند و این پارامترها به‌عنوان پارامتر مقدمه نامیده می‌شوند.  $\{a, b, c\}$  مجموعه پارامترهای مقدماتی هستند. درجات عضویت هر تابع عضویت با توجه به پارامترها مقدماتی و به‌کمک رابطه‌های ۱ و ۲ محاسبه می‌شوند.

$$\mu_{A_i}(x) = \text{gbellmf}(x; a, b, c) = \frac{1}{1 + \left| \frac{x-c}{a} \right|^{2b}} \quad (1)$$

$$O_i^1 = \mu_{A_i}(x) \quad (2)$$

### لایه دوم

این لایه به‌عنوان لایه قانون<sup>۳</sup> نامیده می‌شود. وزن انتقال ( $w_i$ ) برای قوانین با استفاده از درجات عضویت گرفته شده از لایه قبلی تولید می‌شود. مقادیر  $w_i$  با ضرب مقادیر عضویت به‌صورت زیر بدست می‌آید (رابطه ۳).

$$O_i^2 = \omega_i = \mu_{A_i}(x) \cdot \mu_{B_i}(y) \quad i = 1, 2. \quad (3)$$

### لایه سوم

این لایه به‌عنوان لایه نرمال‌سازی<sup>۴</sup> نامیده می‌شود و وزن انتقال نرمال شده متعلق به هر قانون را محاسبه می‌کند (رابطه ۴).

Fuzzification layer

Rule layer

Normalization layer

$$O_i^3 = \bar{\omega}_i = \frac{\omega_i}{\omega_1 + \omega_2 + \omega_3 + \omega_4} \quad i \in \{1,2,3,4\} \quad (4)$$

#### لایه چهارم

این لایه به عنوان لایه فازی زدای نامیده می شود. مقادیر وزنی قوانین در هر گره از این لایه مطابق با رابطه ۵ محاسبه می شود. این مقدار با استفاده از چند جمله ای مرتبه اول تعیین می شود.

$$O_i^4 = \bar{\omega}_i f_i = \bar{\omega}_i (p_i x + q_i y + r_i) \quad (5)$$

خروجی لایه نرمال سازی و  $\{p_i, q_i, r_i\}$  مجموعه پارامتر است. این ها هستند پارامترهای پیامد نامیده می شود. تعداد پارامترهای پیامد هر قانون یکی بیشتر از تعداد ورودی به سیستم است.

#### لایه پنجم

این لایه به عنوان لایه جمع نامیده می شود. خروجی واقعی ANFIS با جمع کردن خروجی های به دست آمده برای هر قانون در لایه فازی سازی به دست می آید (رابطه ۶).

$$O_i^5 = \sum_i \bar{\omega}_i f_i = \frac{\sum_i \omega_i f_i}{\sum_i \omega_i} \quad (6)$$

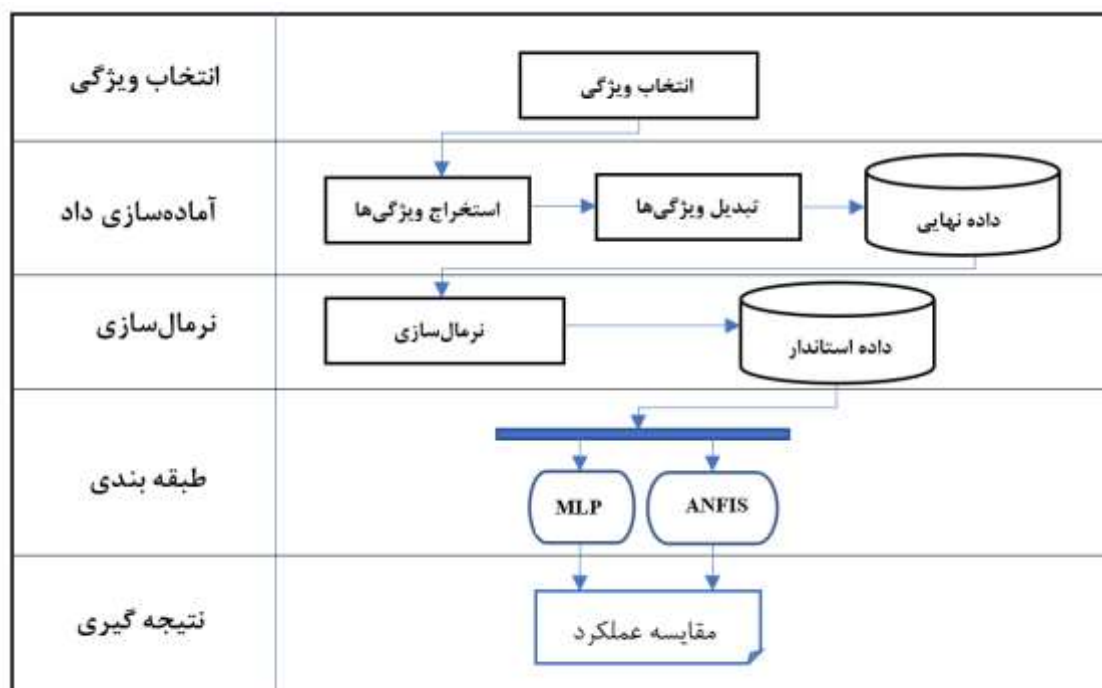
### ۱.۳. چهارچوب و معماری پیشنهادی

به طور کلی سیستم تشخیص یک مجموعه تئیت در طول زمان را به عنوان خط زمانی  $TL = \{t_1, t_2, \dots, t_n\}$  به عنوان ورودی خود می پذیرد و طبقه بندی کننده یک برچسب  $Y = \{NR, R\}$  را اختصاص می دهد. اگر طبقه بندی کننده یک تئیت را به عنوان شایعه گزارش کند، برچسب R را به آن اختصاص می دهد، در غیر این صورت برچسب NR به آن اختصاص می یابد. همانطور که در شکل ۳ نشان داده شده است، مراحل کار به انتخاب ویژگی، آماده سازی داده، نرمال سازی، طبقه بندی و نتیجه گیری قسمت بندی می شود که در ادامه به طور کامل توضیح داده شده است.

---

Defuzzification layer

Summation layer



شکل ۳- معماری مدل پیشنهادی

### ۲.۳. مجموعه داده

طبق آنچه که قبل تر بیان کرده ایم یک شایعه واقعی مربوط به اخباری است که بدون تایید منتشر می شوند و پس از مدتی شواهد تاییدی برای آن موجود است. شایعه ای نادرست بدون تایید منتشر می شود که پس از مدتی شواهد کافی برای رد صحت آن وجود دارد. غیر شایعه مربوط به یک خبر قابل تایید از لحظه انتشار است. در نهایت، خبر تایید نشده شایعه ای است که هیچ مدرکی برای تایید یا رد صحت آن وجود ندارد.

وظیفه تشخیص خودکار شایعات مستلزم مجموعه داده های برچسب گذاری شده است که از طریق آن می توان آموزش یک ماشین یادگیری را انجام داد. ادغام این مجموعه داده ها فرآیندی پرهزینه است که نیازمند تلاش کارشناسانی است که اخبار را برای این اهداف برچسب گذاری می کنند. در این تحقیق، برای ارزیابی تجربی، ما از مجموعه داده های عمومی توئیت با نام PHEME استفاده کرده ایم که شامل دو مجموعه داده شایعه واقعی و غیر شایعه است.

نتایج بررسی های انجام شده در روند شناسایی و طبقه بندی اخبار منتشر شده در این مجموعه داده توسط یک گروه خبرنگار حرفه ای مطابق با جدول ۱ بوده است.

جدول ۱- درصد توئیت های شایعه و غیر شایعه در مجموعه داده های PHEME

موضوع	شایعه	غیر شایعه
تیراندازی در شارلی ابدو	۴۵۸ (۲۲٪)	۱۶۲۱ (۷۸٪)

<sup>۱</sup>[https://figshare.com/articles/dataset/PHEME\\_dataset\\_of\\_rumours\\_and\\_non-rumours/4010619/1](https://figshare.com/articles/dataset/PHEME_dataset_of_rumours_and_non-rumours/4010619/1)

ناآرامی فرگوسن	۲۸۴ (۸٪/۲۴)	۸۵۹ (۲٪/۷۵)
سقوط هواپیمای جرمن‌وینز	۲۳۸ (۷٪/۵۰)	۲۳۱ (۳٪/۴۹)
تیراندازی در اتاوا	۴۷۰ (۸٪/۵۲)	۴۲۰ (۲٪/۴۷)
گروگان‌گیری سیدنی	۵۲۲ (۸٪/۴۲)	۶۹۹ (۲٪/۵۷)

### ۳.۳. انتخاب ویژگی

ما از دو نوع ویژگی طبقه‌بندی کننده استفاده کرده‌ایم: ویژگی‌های مبتنی بر محتوا و ویژگی‌های اجتماعی. از ترکیب ویژگی‌های اجتماعی به صورت مجزا و ترکیب آن با ویژگی محتوایی در آزمایش استفاده شده است.

#### ویژگی‌های مبتنی بر محتوا

ما از پنج ویژگی محتوایی استفاده کرده‌ایم که همه آن‌ها به‌عنوان بخشی از داده‌های استخراجی API توئیتر می‌باشند که به صورت متاداده و مرتبط به پاسخ‌های یک توئیت هستند.

**استفاده از علامت سوال:** یک ویژگی باینری که نشان می‌دهد آیا توئیت حداقل یک علامت سوال در آن وجود دارد یا خیر. وجود علامت سوال می‌تواند و ممکن است نشان دهنده عدم قطعیت در توئیت باشد.

**استفاده از علامت تعجب:** یک ویژگی باینری که می‌تواند نشان دهد آیا توئیت حداقل یک علامت تعجب در خود دارد یا خیر. علامت تعجب ممکن است نشان دهنده تاکید یا تعجب باشد.

**استفاده از هشتک:** یک ویژگی باینری که نشان می‌دهد آیا توئیت حداقل یک هشتک در خود دارد یا خیر. وجود هشتک نشان دهنده تاکید بر موضوع و جلب توجه دارد.

**استفاده از آدرس اینترنتی:**<sup>۸</sup> یک ویژگی باینری که نشان می‌دهد آیا توئیت حداقل یک آدرس اینترنتی در خود دارد یا خیر. وجود آدرس اینترنتی در توئیت به منظور ارجاع به یک متن یا عکس بوده و می‌تواند نشان دهنده عدم قطعیت و رد خبر باشد.

#### ویژگی‌های اجتماعی

ما از هشت ویژگی اجتماعی استفاده می‌کنیم که همه آن‌ها به‌عنوان بخشی داده‌های API توئیتر می‌باشند که در متاداده مرتبط با نویسنده توئیت بازبایی شده هستند که از هفت عدد آن‌ها به‌عنوان ویژگی اصلی و یکی به عنوان ویژگی فرعی استفاده می‌شود.

**تعداد توئیت:** تعداد توئیت پست شده توسط نویسنده اصلی. از آنجایی که این عدد می‌تواند به طور قابل توجهی در بین کاربران متفاوت باشد، ما آن‌ها را از گرد کردن لگاریتم مبنای ۱۰ تعداد توئیت‌ها بدست می‌آوریم:  $\log_{10}(\text{statuses\_count})$

**تعداد فهرست شده:** این ویژگی از گرد کردن لگاریتم مبنای ۱۰ تعداد لیست‌هایی که یک کاربر به آن‌ها تعلق دارد محاسبه می‌شود:  $\log_{10}(\text{listed\_count})$

**نسبت فالو:** در این ویژگی ما به تعداد فالوورها نگاه می‌کنیم. با این حال، ممکن است گاهی اوقات تعداد فالوورها دروغین هستند، به عنوان مثال، کاربرانی هستند که بی‌جهت بسیاری دیگر را دنبال می‌کنند تا فالوورهای بیشتری جذب کنند.

برای کنترل این اثر، نسبت فالو را از لگاریتم فالوورها نسبت به دوستان تعریف می‌کنیم:

$$\log_{10}(\text{followers\_count}/\text{favorites\_count})$$

**تعداد لایک:** این ویژگی از گرد کردن لگاریتم مبنا ۱۰ تعداد لایک توییت نویسنده اصلی دریافت کرده می‌باشد:

$$\log_{10}(\text{favorite\_count})$$

**تعداد باز توییت:** این ویژگی از گرد کردن لگاریتم مبنا ۱۰ تعداد باز توییت‌های توییت نویسنده اصلی بدست می‌آید:

$$\log_{10}(\text{retweet\_count})$$

**سن:** مدت فعالیت کاربر است. از مدت فعالیت کاربر بر حسب روز که در توییت سپری شده، محاسبه می‌کنیم، یعنی از روزی که حساب کاربری ایجاد شده تا روز توییت فعلی.

**تأیید شده:** یک ویژگی باینری است که نشان می‌دهد کاربر توسط توییت تأیید شده است یا خیر. کاربران تأیید شده کسانی هستند که توییت هویت آن‌ها را تأیید کرده است و معمولاً افراد معتبری هستند.

**تعداد پاسخ:** این ویژگی از تعداد توییت‌های پاسخ به جهت واکنش به توییت اصلی زده شده بدست می‌آید. از این ویژگی به عنوان ویژگی فرعی برای محاسبه نسبت ویژگی‌های محتوایی استفاده می‌شود.

در جدول‌های ۲ و ۳ ویژگی‌های مبتنی بر محتوا و اجتماعی استفاده شده نشان داده شده است.

جدول ۲- ویژگی‌های مبتنی بر محتوا

ویژگی	نوع ویژگی
استفاده از علامت سوال	باینری
استفاده از علامت تعجب	باینری
استفاده از هشتک	باینری
استفاده از آدرس اینترنتی	باینری

جدول ۳- ویژگی‌های اجتماعی

ویژگی	نوع ویژگی
تعداد توییت	عدد
تعداد فهرست شده	عدد
نسبت فالو	عدد
تعداد لایک	عدد
تعداد باز توییت	عدد
سن	عدد
تأیید شده	باینری
تعداد پاسخ	عدد

Like

Retweet

Reply

## ۴.۳. آماده سازی داده

داده های مورد استفاده در این تحقیق از مجموعه توئیت های موضوع شارلی ابدو مطابق با ویژگی های بیان شده در جدول ۲ و ۳ استخراج گردیده اند. با توجه به اینکه ویژگی های مبتنی بر محتوا از توئیت های پاسخ به توئیت اصلی استخراج گردیده اند و هر توئیت طبیعتاً بیش از یک پاسخ می تواند داشته باشد و به منظور برداری کردن داده ها، بعد از استخراج ویژگی های محتوایی از پاسخ ها هر توئیت حاصل جمع تعداد هر کدام از آن ویژگی ها مقابل ویژگی های اجتماعی توئیت اصلی در یک سطر نوشته شده. از آنجایی که اعداد ویژگی های اجتماعی استخراج شده می تواند به طور قابل توجهی در تمام توئیت های موجود متفاوت باشد، نسبت تعداد ویژگی مورد نظر در پاسخ ها بر کل تعداد پاسخ در جدول نهایی ثبت شده، به طور مثال ابتدا تمام پاسخ هایی مربوط به یک توئیت که دارای ویژگی «استفاده از علامت سوال» بودن تعیین می شود، سپس نسبت عدد مجموع تعداد پاسخ های دارای علامت سوال بر کل پاسخ های توئیت اصلی به عنوان داده نهایی در جدول قرار داده شده. چنان که در پایان هر سطر شامل نسبت ویژگی های محتوایی پاسخ ها با ویژگی های اجتماعی استخراج شده از توئیت اصلی مطابق با جدول ۴ خواهد شد، با توجه به ماهیت پژوهش مبنی بر اثر توئیت های پرس و جو بر امر تشخیص شایعات در روند آموزش سامانه از بررسی توئیت های که پاسخی با خود به همراه نداشته اند خودداری شده و به جهت بهبود کیفیت آموزش از توئیت های با پاسخ های بالا ۲۰ عدد استفاده شد، به طوری که تعداد کل داده های بدست آمده ۳۷۳ عدد بوده، که از این داده ها به صورت تصادفی ۷۰٪ آن برای مرحله آموزش مدل و ۳۰٪ آن برای مرحله واریسی آنتخاب شده است.

جدول ۴- داده های نهایی بعد از تبدیل

ویژگی	نوع ویژگی	تبدیل
تعداد توئیت	عدد	$\log_{10}(\text{statuses\_count})$
تعداد لیست	عدد	$\log_{10}(\text{listed\_count})$
نسبت فالو	عدد	$\log_{10}(\text{followers\_count}/\text{favorites\_count})$
تعداد لایک	عدد	$\log_{10}(\text{favorite\_count})$
تعداد باز توئیت	عدد	$\log_{10}(\text{retweet\_count})$
سن	عدد	-
تایید شده	باینری	true = 1 & false = 0
نسبت علامت سوال	عدد	$\frac{\text{Question\_Count}}{\text{Reaction\_Count}}$
نسبت علامت تعجب	عدد	$\frac{\text{Exclamation\_Count}}{\text{Reaction\_Count}}$

Charlie Hebdo

Training

Testing

$\frac{\text{Hashtac\_Count}}{\text{Reaction\_Count}}$	عدد	نسبت هشتک
$\frac{\text{URL\_Countt}}{\text{Reaction\_Count}}$	عدد	نسبت آدرس اینترنتی

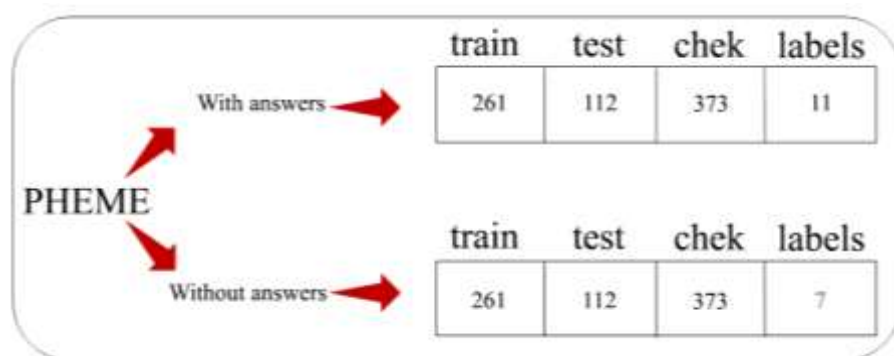
### ۴.۳. نرمال سازی

برای افزایش دقت و سرعت عملکرد مدل، داده‌های نهایی را در رنج ۰ تا ۱ طبق رابطه ۷ نرمال سازی شده است تا یک مجموعه داده استاندارد داشته باشیم.

$$X_n = L + \frac{(X - X_{min})(H - L)}{X_{max} - X_{min}} \quad (7)$$

### ۵.۳. طبقه بندی

پس از آماده سازی داده‌ها، جهت آموزش سیستم طبقه بندی پیشنهادی و نشان دادن تاثیر پست‌های پاسخ، داده‌ها را به دو ترکیب، اولی، داده‌های ویژگی‌های توثیت اصلی و دیگری داده‌های توثیت اصلی به همراه ویژگی‌های پاسخ‌ها مورد بررسی قرار گرفته‌اند.



شکل ۴- نمایش تقسیمات دو حالت داده

جهت مقایسه عملکرد مدل پیشنهادی با داده طبقه بندی کننده یکسان با یک سامانه شبکه عصبی MLP مورد ارزیابی قرار می گیرد.

### ۴. نتایج تحقیق و مقایسه

برای درک عملکرد مدل‌ها در هر کلاس، ماتریس‌های درهم ریختگی برای دو حالت تقسیمات داده در شکل ۵ و ۶ آمده است. آنچه ماتریس‌های درهم ریختگی در هر دو مدل نشان می‌دهد دقت حالت ترکیبی عملکرد بهتری را به همراه داشته.

		با پاسخ ها			بدون پاسخ ها		
Output Class	Target Class	0	1		0	1	
		184 49.3%	4 1.1%	97.9% 2.1%	174 46.6%	12 3.2%	93.5% 6.5%
0	1	8 2.1%	177 47.5%	95.7% 4.3%	18 4.8%	169 45.3%	90.4% 9.6%
1	0	95.8% 4.2%	97.8% 2.2%	96.8% 3.2%	90.6% 9.4%	93.4% 6.6%	92.0% 8.0%

شکل ۵- ماتریس درهم ریختگی داده‌های آزمایش در شبکه عصبی فازی پیشنهادی و مقایسه

		با پاسخ ها			بدون پاسخ ها		
Output Class	Target Class	0	1		0	1	
		172 46.1%	26 7.0%	86.9% 13.1%	166 44.5%	32 8.6%	83.8% 16.2%
0	1	20 5.4%	155 41.6%	88.6% 11.4%	26 7.0%	149 39.9%	85.1% 14.9%
1	0	89.6% 10.4%	85.6% 14.4%	87.7% 12.3%	86.5% 13.5%	82.3% 17.7%	84.5% 15.5%

شکل ۶- ماتریس درهم ریختگی داده‌های آزمایش در شبکه عصبی MLP

#### ۱.۴. مقایسه عملکرد

ارزیابی عملکرد روش پیشنهادی در این پژوهش با استفاده از معیارهای دقت،<sup>۵</sup> صحت،<sup>۶</sup> پوشش<sup>۷</sup> و F1<sup>۸</sup> مطابق با رابطه‌های ۸ و ۹ و ۱۰ و ۱۱ بوده است.

$$Accuracy = \frac{TP + TN}{TP + FN + TN + FP} \quad (۸)$$

$$Precision = \frac{TP}{TP + FP} \quad (۹)$$

Accuracy

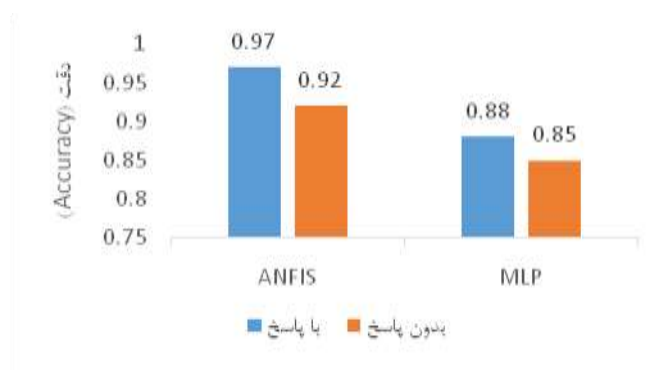
Precision

Recall

$$Recall = \frac{TP}{TP + FN} \quad (10)$$

$$F\_Score = \frac{2 \times Recall \times Precision}{Recall + Precision} \quad (11)$$

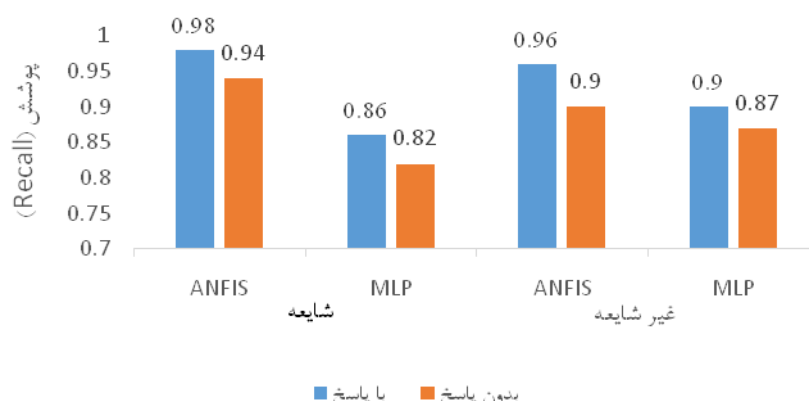
شکل ۷ معیار دقت ANFIS پیشنهادی همراه با روش MLP را در هر دو حالت پیش‌تر بیان شده نشان می‌دهد. ANFIS پیشنهادی از نظر دقت در مقایسه با ANFIS مقایسه و روش MLP در هر دو حالت عملکرد بهتری دارد. روش پیشنهادی نسبت به مدل‌های مقایسه همان‌طور که در شکل ۸ و ۹ نشان داده شده است، در معیار صحت و پوشش بهتر عمل کرده.



شکل ۷- معیار دقت سیستم طبقه بندی ANFIS و MLP



شکل ۸- معیار صحت سیستم طبقه بندی ANFIS و MLP در دو کلاس شایعه و غیر شایعه



شکل ۹- معیار پوشش سیستم طبقه بندی ANFIS و MLP در دو کلاس شایعه و غیر شایعه

در شکل ۱۰ معیار F1 برای شایعات و غیر شایعات برای روش پیشنهادی و مدل‌های مقایسه نشان داده شده است. نمودارها موجود در شکل نشان از برتری مدل ANFIS پیشنهادی نسبت به مدل‌های مقایسه در هر دو کلاس شایعه و غیر شایعه دارد.



شکل ۱۰- معیار F1 سیستم طبقه بندی ANFIS و MLP در دو کلاس شایعه و غیر شایعه

معیارهای دقت، صحت، پوشش و F1 نشان می‌دهد که ANFIS نسبت به روش مقایسه عملکرد بهتری را داشته و همچنین نمودار دو حالت ویژگی‌ها نشان از تاثیر بهبود طبقه‌بندی در حالتی که از توثیقات پاسخ استفاده شده دارد. آنچه نمودارها نشان می‌دهند ANFS روش پیشنهادی از روش مقایسه عملکرد بهتری را داشته و قادر است شایعات را بهتر از مدل مقایسه در این پژوهش طبقه‌بندی کند.

برای مقایسه بین رویکرد پیشنهادی و نتایج حاصل از سایر پژوهش‌ها، در جدول ۵ دقت روش پیشنهادی که توسط الخدایر و همکاران [۱۱] و همچنین وانگ و همکاران [۱۲] گزارش شده‌اند، در کنار بهترین عملکرد روش پیشنهادی ما آورده شده است. این پژوهش‌ها نیز از مجموعه داده PHEME استفاده کرده‌اند.

## جدول ۵-مقایسه عملکرد روش پیشنهادی با سایر پژوهش‌ها

نام مدل	P	R	F1
[۱۱]LSTM-RNN	۷۶٪	۷۴٪	۷۴٪
[۱۲]KMGCN	۸۸٪	۸۸٪	۸۸٪
رویکرد پیشنهادی	۹۶٪	۹۸٪	۹۷٪

## ۵. نتیجه‌گیری

در این پژوهش تلاش شد با استفاده از ANFIS و به کمک توئیت‌های پاسخ جهت تشخیص شایعه یا واقعیت توئیت‌ها در زمان نشر آن‌ها اقدام کنیم. این سامانه برای مقابله با چالش پراکندگی داده‌ها از قابلیت یادگیری خودکار شبکه‌های عصبی و همچنین جهت برخورد با عدم قطعیت در داده‌ها به فازی سازی آن‌ها می‌پردازد. این ویژگی‌ها باعث شده ANFIS قابلیت یادگیری برای تقریب زدن توابع غیرخطی را دارا باشد و در نتیجه این سیستم به خوبی توانسته با مسائلی خیر خطی همچون مسئله شایعات نوظهور نسبت به دیگر سیستم‌های یادگیری ماشین برخورد کند.

بر اساس نتایج به دست آمده، روش پیشنهادی یعنی ANFIS با کمک از توئیت‌های پاسخ نسبت به روش‌های طبقه‌بندی MLP و ANFIS بدون کمک توئیت‌های پاسخ همچنین روش MLP بدون کمک توئیت‌های پاسخ به دقت ۹۷٪ در طبقه‌بندی شایعات رسیده که این دقت برای روش مقایسه ذکر شده فوق به ترتیب برابر با ۸۸٪، ۹۲٪، ۸۵٪ می‌باشد، این نتیجه نشان از بهبود تشخیص و طبقه‌بندی شایعات توسط این سامانه پیشنهادی را دارد.

## منابع و مراجع

۱. M. Mendoza, B. Poblete, and C. Castillo, "Twitter under crisis: Can we trust what we RT?", SOMA 2010 - Proceedings of the 1st Workshop on Social Media Analytics, pp. 71–79, 2010.
۲. C. Castillo, M. Mendoza, and B. Poblete, "Information credibility on Twitter", Proceedings of the 20th International Conference Companion on World Wide Web, WWW 2011, pp. 675–684, 2011, doi: 10.1145/1963405.1963500.
۳. Z. Zhao, P. Resnick, and Q. Mei, "Enquiring minds: Early detection of rumors in social media from enquiry posts", WWW 2015 - Proceedings of the 24th International Conference on World Wide Web, pp. 1395–1405, May 2015, doi: 10.1145/2736277.2741637.
۴. M. Bugueño and M. Mendoza, "Applying Self-attention for Stance Classification", Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), vol. 11896 LNCS, pp. 51–61, 2019.
۵. G. Cai, H. Wu, and R. Lv, "Rumors detection in Chinese via crowd responses", ASONAM 2014 - Proceedings of the 2014 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, pp. 912–917, Oct. 2014, doi: 10.1109/ASONAM.2014.6921694.
۶. G. Liang, W. He, C. Xu, L. Chen, and J. Zeng, "Rumor Identification in Microblogging Systems Based on Users' Behavior", IEEE Trans Comput Soc Syst, vol. 2, no. 3, pp. 99–108, Sep. 2015.

۷. N. DiFonzo and P. Bordia, "Rumor, Gossip and Urban Legends", *Diogenes*, vol. 54, no. 1, pp. 19–35, 2016.
۸. P. Donovan, "How Idle is Idle Talk? One Hundred Years of Rumor Research", *Diogenes*, vol. 54, no. 1, pp. 59–82, 2016.
۹. ف. صادقی و ا. جلالی بیدگلی, "مروری بر روش‌های تشخیص شایعات در شبکه‌های اجتماعی", *فصلنامه علمی ترویجی منادی امنیت فضای تولید و تبادل اطلاعات (افتا)*, جلد ۸, شماره ۱, صفحه ۳–۱۴, ۱۳۹۸.
۱۰. D. Karaboga and E. Kaya, "Adaptive network based fuzzy inference system (ANFIS) training approaches: a comprehensive survey", *Artificial Intelligence Review 2018 52:4*, vol. 52, no. 4, pp. 2263–2293, Jan. 2018, doi: 10.1007/S10462-017-9610-2.
۱۱. S. A. Alkhodair, S. H. H. Ding, B. C. M. Fung, and J. Liu, "Detecting breaking news rumors of emerging topics in social media", *Inf Process Manag*, vol. 57, no. 2, p. 102018, Mar. 2020.
۱۲. Y. Wang, S. Qian, J. Hu, Q. Fang, and C. Xu, "Fake news detection via knowledge-driven multimodal graph convolutional networks", in *ICMR 2020 - Proceedings of the 2020 International Conference on Multimedia Retrieval*, 2020. doi: 10.1145/3372278.3390713.