

## تشخیص اخبار جعلی منتشر شده در شبکه اجتماعی اینستاگرام بر اساس تارنمای خبری ایسنا

## مبنتی بر الگوریتم متن کاوی و معنی‌شناسی در زبان فارسی

مرتضی دستلان<sup>\*۱</sup>، مهشید ده بزرگی<sup>۲</sup>، بهزاد مریدی<sup>۳</sup>، محسن امامی<sup>۴</sup>

۱. استادیار زبان‌شناسی همگانی، دانشگاه پیام نور، تهران، ایران

۲. کارشناسی ارشد زبان‌شناسی همگانی، دانشگاه پیام نور، تهران، ایران

۳. استادیار زبان‌شناسی همگانی، دانشگاه پیام نور، تهران، ایران

۴. استادیار مهندسی کامپیوتر، دانشگاه پیام نور، تهران، ایران

پذیرش: ۱۴۰۰/۰۲/۲۵

دریافت: ۱۳۹۹/۱۰/۱۷

## The Detection of Fake Sport News Published in Instagram Social Network based on ISNA News Website: A Semantic Study on the Basis of Mining Algorithm in Persian Language

Morteza Dastlan<sup>\*1</sup>, Mahshid Dehbozorgi<sup>2</sup>, Behzad Moridi<sup>3</sup>, Mohsen Emami<sup>4</sup>

1. Assistant Professor of General Linguistics, Payame Noor University, Tehran, Iran

2. M.A. in General Linguistics, Payame Noor University, Tehran, Iran

3. Assistant Professor of General Linguistics, Payame Noor University, Tehran, Iran

4. Assistant Professor of Computer Engineering, Payame Noor University, Tehran, Iran

Received: 2021/01/06

Accepted: 2021/05/15

10.30473/ijl.2022.57037.1429

## Abstract

The distribution of fake sport news is not based on the satisfaction of sport men, sport clubs and sport fans. Correspondingly, the identification of fake news is important and practical. This research has been done in the framework of computational linguistics. The linguistic data are based on a corpus of sports news from ISNA website and Instagram program. In this way, sports news is downloaded from the ISNA website in a period of time, and then in a few pages of the Instagram program, sports news is downloaded and compared in terms of being fake or not. The N-gram method and long and short term memory (LSTM) method have been used to identify fake news from non-fake ones. The method proposed in this paper has been implemented on four valid and existing datasets and has been compared with the previous six methods. The accuracy of this method is acceptable compared to other methods, and the results obtained indicate that this method is suitable and accurate enough to identify fake news among the news published on Instagram.

**Keywords:** Fake news, Computational Linguistics, N-Gram, LSTM, Mining Algorithm, Persian Language, Sociolinguistics.

## چکیده

اخبار جعلی ورزشی که در شبکه‌های اجتماعی پخش می‌شوند برای ورزشکاران، باشگاه‌های ورزشی، ورزش دوستان و ... ناخوشایند هستند، پس تشخیص اخبار جعلی و غیرجعلی امری مهم و کاربردی است. این پژوهش، در چارچوب زبان‌شناسی رایانه‌ای انجام گرفته است. پیکره‌ی زبانی این تحقیق بر اساس بارگیری اخبار ورزشی از تارنمای ایسنا و برنامه اینستاگرام بوده است؛ بدین صورت که اخبار ورزشی از تارنمای ایسنا در یک محدوده زمانی بارگیری شده است و سپس در چند صفحه برنامه اینستاگرام، اخبار ورزشی بارگیری شده و از لحاظ جعلی بودن یا غیرجعلی بودن با هم مقایسه شده‌اند. برای شناسایی اخبار جعلی از غیرجعلی از روش ان-گرم و روش حافظه طولانی کوتاه-مدت (ال اس تی ام) استفاده شده است. روش پیشنهاد شده در این مقاله روی چهار دیتاست معتبر و موجود پیاده سازی شده و با شش روش پیشین مقایسه شده است. دقت این روش نسبت به دیگر روش‌ها قابل قبول می‌باشد و نتایج به دست آمده، بیانگر این است که برای شناسایی اخبار جعلی، در میان خبرهای منتشر شده در اینستاگرام، این روش، مناسب می‌باشد و دقت کافی را دارد.

**کلیدواژه‌ها:** اخبار جعلی، زبان‌شناسی رایانه‌ای، ان-گرم، ال اس تی ام، الگوریتم متن کاوی، زبان فارسی، زبان‌شناسی اجتماعی.

\*Corresponding Author: Morteza Dastlan

Email: mo.dastlan@pnu.ac.ir

\* نویسنده مسئول: مرتضی دستلان

## مقدمه

در فضای مجازی، اخبار غیرواقعی یا شایعه‌پراکنی بخصوص درباره افراد معروف بسیار دیده می‌شود، یکی از این موارد، اخبار جعلی یا غیرواقعی درباره ورزشکاران یا باشگاه‌ها می‌باشد که این موارد می‌تواند آسیب‌های جدی به افراد یا باشگاه‌های مورد نظر که اخبار جعلی در مورد آن‌ها نوشته شده است، وارد کند. دلیل این موضوع می‌تواند تخریب یک باشگاه یا یک ورزشکار یا برخی اوقات تنها برای سرگرمی باشد. به طور طبیعی هیچ ورزشکار یا باشگاهی دوست ندارد مورد تبلیغات جعلی، صرف نظر از نوع رسانه (متنی، به صورت عکس، ویدئو، صوت و ...) قرار گیرد و همواره رقیبان یا افرادی که به هر دلیلی خصومتی با آن باشگاه یا ورزشکار دارند، احتمال دارد به انتشار این نوع اخبار مبادرت ورزند. از این رو، شناسایی اخبار جعلی منتشر شده در شبکه‌های اجتماعی دارای اهمیت فراوانی است. در حالت عادی، برای پیدا کردن و شناسایی اخبار جعلی لازم است که نیروی انسانی و وقت زیادی صرف شود، اما طراحی یک سیستم کامپیوتری هوشمند برای یافتن، مقایسه و درک اخبار صحیح و در انتها شناسایی اخبار جعلی، بسیار به صرفه‌تر، سریع‌تر و معقول‌تر می‌باشد. در این پژوهش، سعی بر آن است که راهی برای تشخیص اخبار ورزشی جعلی یا تحریف شده بیابیم. در این راستا، ابتدا اخبار ورزشی تارنمای ایسنا را در بازه زمانی مشخصی دانلود می‌کنیم. تارنمای ایسنا بدین علت انتخاب شده است که یکی از تارنماهای خبرگزاری اصلی کشور است (خبرگزاری دانشجویان ایران) و تمامی اخبار را پوشش می‌دهد و در ضمن این اخبار قابل دانلود نیز می‌باشند. سپس این اخبار را در پست‌های شبکه اجتماعی اینستاگرام دنبال کرده و آنها را با هم مقایسه می‌کنیم و بدین ترتیب، جعلی یا غیرجعلی بودن خبرها را تعیین می‌کنیم. برای انجام این کار، از روش ان-گرم و روش حافظه طولانی کوتاه-مدت (ال اس تی ام) استفاده می‌کنیم که در ادامه به توضیح آن‌ها خواهیم پرداخت.

## پیشینه تحقیق

مرداوغلو<sup>۱</sup> و گنچ<sup>۲</sup> (۲۰۲۰) عقیده داشتند که پیدا کردن ترندها<sup>۳</sup> در شبکه‌های اجتماعی از روی هشتگ‌ها و به

اشتراک گذاشتن آدرس‌های اینترنتی پست‌ها، یک گزینه مهم برای شناسایی اخبار جعلی می‌باشد، به دلیل این که اشخاصی که اخبار جعلی را تولید می‌کنند قدرت پخش کردن سریع آنها در شبکه‌های اجتماعی را نیز دارا هستند و این کار را با توجه به تعداد پروفایل‌هایی که در اختیارشان است انجام می‌دهند؛ البته این پروفایل‌ها الزاماً توسط ربات‌های نرم‌افزاری کنترل نمی‌شوند و بعضی از آن‌ها توسط افراد حقیقی اداره می‌شوند. اگرچه شناسایی پروفایل‌های متخلف کار آسانی نیست اما در نظر گرفتن ترند شدن یک موضوع می‌تواند برای شناسایی اخبار و مطالب جعلی بسیار مفید باشد. دی بیر<sup>۴</sup> و متی<sup>۵</sup> (۲۰۲۰) معتقدند که انتشار گسترده اطلاعات از طریق بسترهای رسانه دیجیتال و قضاوت در مورد اعتبار آنها، برای افراد و جوامع بسیار مهم است. اخبار جعلی یک رویداد جدید نیست، اما پیامدهای آن در سال‌های اخیر رخ نموده است. آزار دهنده بودن، تأثیر منفی گذاشتن و همراه‌کننده افراد و یا حتی ملت‌ها بودن، می‌تواند از نتایج اخبار جعلی باشد. از نظر آنها رویکردهای متنوعی برای شناسایی اخبار جعلی وجود دارد. آنها در پژوهش خود رویکردهای اصلی که در حال حاضر برای شناسایی اخبار جعلی در دسترس است و نحوه استفاده از این رویکردها در موقعیت‌های مختلف را شناسایی کرده‌اند و به طور نظام‌مند در پژوهش خود آورده‌اند.

دی اولیوریا<sup>۶</sup>، مدیروس<sup>۷</sup> و ماتوس<sup>۸</sup> (۲۰۲۰) معتقدند که اخبار جعلی تهدیدی برای حقیقت است، که باعث بدتر شدن دموکراسی، روزنامه‌نگاری و بی‌اعتباری نهادهای دولتی می‌شود و انسان از تشخیص آن‌ها عاجز است. آنها در پژوهش خود، یک تحلیل محاسباتی مبتنی بر پردازش زبان طبیعی، با استفاده از الگوریتم‌های کارآمد یادگیری ماشین برای کشف اخبار جعلی در متن استخراج شده از رسانه‌های اجتماعی ارائه داده‌اند. در این تحلیل، اخباری که از توویتر در نظر گرفته شده است، تقریباً به ۳۳۰۰۰ مورد بالغ می‌شود. در ارزیابی تشخیص کیفیت، دقت ۰.۸۶٪ و صحت ۰.۹۴٪ به دست آمده است. رویکرد آن‌ها حداقل هزینه را در پی دارد، در حالیکه این پتانسیل را دارد که شاخص اطمینان بالایی را در مورد تمایز اخبار جعلی از واقعی ارائه دهد.

4. De Beer  
5. Matthee  
6. De Oliveira  
7. Medeiros  
8. Mattos

1. Mertoğlu  
2. Genç  
3. Trend

میلیون عضو و حدود ۲ میلیارد و ۲۰۰ میلیون عضو داشته‌اند. با پیشرفت تعداد اعضا، مسأله انتشار مسائل جعلی و تحریف شده از قبیل اخبار و شایعات، رنگ جدی‌تری به خود گرفته است و به تبع آن الگوریتم‌های زیادی برای شناسایی اخبار جعلی در دنیای مجازی عنوان شده است. (چافی<sup>۶</sup>، ۲۰۲۱) در ادامه به توضیح چند روش متداول تشخیص اخبار جعلی می‌پردازیم.

### روش‌های QP<sup>۷</sup>

حل معادله پیدا کردن خط بهینه برای داده‌ها به وسیله روش‌های QP که روش‌های شناخته شده‌ای در حل مسائل محدودیت‌دار هستند صورت می‌گیرد. قبل از تقسیم خطی برای اینکه ماشین بتواند داده‌های با پیچیدگی بالا را دسته‌بندی کند داده‌ها را به وسیله تابع phi به فضا با ابعاد خیلی بالاتر می‌بریم. برای اینکه بتوانیم مسأله ابعاد خیلی بالا را با استفاده از این روش‌ها حل کنیم از قضیه دوگانی لاگرانژ برای تبدیل مسأله مینیمم‌سازی مورد نظر به فرم دوگانی آن که در آن به جای تابع پیچیده phi که ما را به فضایی با ابعاد بالا می‌برد، تابع ساده‌تری به نام تابع هسته که ضرب برداری تابع phi است ظاهر می‌شود استفاده می‌کنیم. در اینجا از توابع هسته مختلفی از جمله هسته‌های نمایی، چند جمله‌ای و سیگموئید می‌توان استفاده نمود. در اینجا برای تفکیک و تشخیص اخبار جعلی و غیرجعلی از روش ان-گرم که از نظریه مخفی مارکوف اقتباس شده است، استفاده می‌شود. (سالاری و ادیب‌نیا، ۱۳۸۹)

### روش مارکوف

روش زنجیره مارکوف<sup>۸</sup> که به افتخار آندری مارکوف<sup>۹</sup> ریاضی‌دان اهل روسیه این گونه نام‌گذاری شده، یک روش احتمالاتی است که در طول چند سال اخیر، کاربرد آن در همه رشته‌ها به سرعت افزایش یافته است (نیکوگفتار و همکاران، ۱۳۹۱).

در علم برنامه‌ریزی تصادفی، زنجیره‌های مارکوف در دو دیدگاه بسیار مهم مورد بررسی قرار می‌گیرد. اول آنکه دارای ساختار عملی و اثبات‌های منطقی بوده و دوم آنکه در بسیاری از مسائل و پدیده‌های طبیعی صدق می‌کند. یک

گاهیروال و همکاران<sup>۱</sup> (۲۰۱۸) عقیده دارند که پردازش زبان طبیعی<sup>۲</sup> وظیفه پردازش و کاوش روی متون و موارد مربوط به زبان انسان‌ها را نیز دارد. با استفاده از روش‌ها و الگوریتم‌های موجود در این علم و همچنین الگوریتم جنگل تصادفی<sup>۳</sup> می‌توان به هر متن یک نمره تحت عنوان F نسبت داد که بیانگر میزان حقیقی یا جعلی بودن متن است. این نمره به درصد بیان می‌شود و برای کاهش خطا از منطق فازی نیز استفاده می‌شود. متون مورد مطالعه با لیبل‌های: کاملاً جعلی، جعلی، حقیقی و کاملاً حقیقی مشخص می‌شوند.

مین یانگ و همکاران<sup>۴</sup> (۲۰۱۸) اظهار نظر می‌کنند که یکی از ترفندهای ساخت اخبار و محتوای جعلی، ساخت تصاویر ترکیبی است که با پردازش تصویر و سیگنال، برخی از آنها را می‌توان به طور سریع تشخیص داد. برای اینکه بتوانیم دقت کار را بالا ببریم به یک دیتاست بزرگ از تصاویر واقعی نیاز داریم که از طریق آن بتوانیم مقایسه را انجام دهیم. این محاسبات در زمینه پردازش تصویر و برپایه پردازش سیگنال‌ها انجام می‌گیرد، به این صورت که در ابتدا هر تصویر را به یک داده جدید از جنس سیگنال تبدیل کرده و سپس سیگنال‌های هر تصویر با هم مقایسه می‌شوند.

### چارچوب نظری

شبکه اجتماعی، ساختاری اجتماعی است که از گروه‌هایی (عموماً شخصی یا سازمانی) تشکیل شده است که توسط یک یا چند نوع خاص از وابستگی‌ها مانند ایده‌ها و تبادلات مالی، دوستان، خویشاوندان و لینک‌های وب به هم متصل می‌شوند. تحلیل شبکه‌های اجتماعی، روابط اجتماعی را با اصطلاحات رأس و یال<sup>۵</sup> می‌نگرد. به بیان دیگر، در ساده‌ترین شکل، یک شبکه اجتماعی نگاشتی از تمام یال‌های مربوط، میان رأس‌های مورد مطالعه است. رأس‌ها بازیگران فردی درون شبکه‌ها و یال‌ها روابط میان این بازیگران محسوب می‌شوند. انواع زیادی از یال‌ها می‌تواند میان رأس‌ها وجود داشته باشد. با پیشرفت دنیای دیجیتال، شبکه‌های اجتماعی از محبوبیت زیادی برخوردار شدند و انسان‌های زیادی به سمت شبکه‌های اجتماعی گرایش پیدا کردند، برای مثال توئیتر و فیسبوک در سال ۲۰۱۸ به ترتیب، حدود ۳۳۰

6. Chaffey  
7. Quadratic Programming  
8. Markov Approach modeling  
9. Andrey Markov

1. Gahirwal et al.  
2. NLP  
3. Random Forrest  
4. Minyoung et al.  
5. Edge

#### 4-grams

$$P(w_i | w_1 w_2 \dots w_{i-1}) \approx P(w_i | w_{i-3} w_{i-2} w_{i-1})$$

#### 5-grams

$$P(w_i | w_1 w_2 \dots w_{i-1}) \approx P(w_i | w_{i-4} w_{i-3} w_{i-2} w_{i-1})$$

این روش یک نمایش ساده است که در پردازش زبان‌های طبیعی و بازیابی اطلاعات استفاده می‌شود. همچنین به عنوان مدل فضایی برداری شناخته می‌شود. در این مدل، یک متن (مانند یک جمله یا سند) به صورت یک بسته چند مجموعه از کلمات آن، بی‌توجه به دستور زبان و حتی نظم کلمات نمایش داده می‌شود (ویسی و صامتی، ۲۰۱۳). در روش بای-گرم بر اساس متن و پیکره انتخاب شده، دو جمله یا عبارت یا کلمه که به صورت متوالی بعد از هم می‌آیند انتخاب می‌شود و به عنوان لیبیل‌های یک متن شناخته می‌شوند که این روند را می‌توان به تعداد مورد نیاز در پیکره مورد نظر انتخاب کرد و احتمال توالی آنها را در متون انتخابی مورد بررسی قرار داد. در این روش برای انتخاب نمونه‌ها و عبارات قابل قبول باید از نمونه‌هایی که به صورت تکیه کلام هستند یا از عباراتی که مشخصاً مربوط به نویسنده هستند، انتخاب کرد که در واقع از طریق هشتگ‌های انتخابی در اخبار قابل شناسایی هستند و از این طریق متن را لیبیل‌گذاری می‌کنیم؛ به طور مثال بعد از تعیین لیبیل‌های مشخص شده اگر یک جمله یا عبارت از چهار کلمه تشکیل شده باشد ابتدا تعداد تکرارهای دو کلمه‌ای که پشت سرهم تکرار شده‌اند در نظر گرفته می‌شود؛ در مرحله دوم کلمه دوم و سوم و تعداد تکرار آنها را مورد بررسی قرار می‌دهیم و در مراحل بعدی به ترتیب توالی دو کلمه سوم و چهارم را بررسی خواهیم کرد.

#### روش RNNs<sup>۲</sup>

در این روش، با هدف آموزش، تعدادی نمونه مستقل وارد مدل کردیم و سپس آن را آموزش دادیم. فرض کنید نمونه‌ها را نمی‌توان مستقل در نظر گرفت؛ به عنوان مثال در یک جمله توالی لغت‌ها یک مفهوم را می‌رساند و هر لغت به لغت‌های قبل و بعد وابسته است. برای نشان دادن این توالی می‌توان از یک مدل دینامیک استفاده کرد. هر واحد این شبکه یک ورودی و یک خروجی دارد. چیزی که آن را از

زنجیره مارکوف دنباله‌ای از متغیرهای تصادفی  $X_1, X_2, X_3, \dots$  است که دارای خاصیت مارکوف هستند یعنی:

$$P_r(X_{n+1} = X | X_1 = X_1, X_2 = X_2, \dots, X_n = X_n) = P_r(X_{n+1} = X | X_n = X_n)$$

هرگاه در یک سیستم یا فرآیند، ویژگی فوق صدق کند، آن سیستم یا فرآیند را سیستم یا فرآیند مارکوفی گویند. اصطلاح زنجیره نیز بیانگر این موضوع است که هر رویداد تنها به رویداد قبل از خود وابسته است. زنجیره مارکوف به عنوان یک روش حافظه‌دار شناخته می‌شود و طول حافظه مورد استفاده را مرتبه زنجیره گویند (آذر و همکاران، ۱۳۹۷).

#### روش N-gram

در حوزه‌های زبان‌شناسی رایانشی و احتمالات، n-gram دنباله‌ای پیوسته از n قلم در یک دنباله معین از متن یا کلام است. بسته به کاربردشان، اقلام می‌توانند واج، هجا، حرف، واژه، یا جفت‌باز باشند n-gramها معمولاً از یک پیکره متنی یا شفاهی جمع‌آوری می‌شوند. به n-gram با اندازه‌های یک، دو، و سه به ترتیب تک‌گرم، دوگرم، و سه‌گرم گفته می‌شود و نامگذاری به همین منوال ادامه می‌یابد. دو مزیت مهم مدل‌های n-gram (و الگوریتم‌های وابسته) سادگی و میزان‌پذیری است؛ در n های بزرگ، مدل می‌تواند محتوای بیشتری را با بدهستان مشخص زمان-مکان ذخیره کند، و آزمایش‌های کوچک را به‌طور بهینه scale up کند. مدل‌های N-gram نوع خاصی از مدل‌های زبانی هستند که آموزش سریع و استفاده آسان دارند. منشأ آنها را می‌توان در کار شانون<sup>۱</sup> جستجو کرد که وظیفه حدس زدن حرف بعدی را در یک متن ارائه می‌دهد. چنین مدل‌هایی از نظر آماری، احتمال پیروی از یک توکن از یک توالی قبلی را تخمین می‌زنند. بر این اساس، احتمال یک دنباله به عنوان حاصل ضرب احتمال هر توکن برای پیروی از پیشوند خود تعریف می‌شود (جیمز و همکاران<sup>۲</sup>، ۲۰۱۸).

در روش ان-گرم یک مدل مناسب از نحوه پیکره‌بندی نویسنده در استفاده از لیبیل‌ها به دست می‌آید که باید با توجه به عدد (n)، تعداد تکرار آنها سنجیده شود. اما خود روش ان-گرم دارای چند روش مختلف است که شامل موارد زیر می‌باشد:

#### Bigram

$$P(w_i | w_1 w_2 \dots w_{i-1}) \approx P(w_i | w_{i-1})$$

#### Trigrams

$$P(w_i | w_1 w_2 \dots w_{i-1}) \approx P(w_i | w_{i-2} w_{i-1})$$

3. Recurrent neural network

1. Shanon

2. Jimenez et al.

واژه شباهتی به تلگرام و فوری بودن انتقال عکس‌ها دارد. در وب‌گاه اینستاگرام این اپلیکیشن این‌گونه معرفی شده است: اینستاگرام یک راه سریع، زیبا و سرگرم‌کننده برای به اشتراک گذاشتن زندگی‌تان با دوستان و خانواده‌تان است. عکس یا ویدئو بگیرید، فیلتر مورد نظرتان را انتخاب کنید و بعد آن را پست کنید، ساده است. در سال ۲۰۱۵ در بین شبکه‌های اجتماعی مختلف، اینستاگرام با ۴۰۰ میلیون کاربر، مقام هفتم را در جهان به خود اختصاص داده است و تنها ۲۸ درصد کاربران آمریکایی بوده‌اند. اما نکته قابل توجه این است که ۹۸ درصد کاربران با استفاده از تلفن همراه هوشمند از اینستاگرام استفاده کرده‌اند (سامانی و فراهانی، ۱۳۹۵).

### روش تحقیق

در زمینه شناسایی اخبار جعلی تحقیقات چندانی در حوزه زبان فارسی صورت نگرفته و کارهای زیادی انجام نشده است. لذا در این تحقیق، پژوهشگر قصد دارد برای مسأله تشخیص اخبار ورزشی جعلی که در شبکه‌های اجتماعی منتشر می‌شوند راه حلی ارایه کند. در این راستا، ابتدا دیتاست‌های مربوط به تارنمای خبرگزاری ایسنا و موارد مربوط به یک بازه زمانی مشخص را دانلود می‌کنیم، سپس در مرحله پیش‌پردازش، پست‌های شبکه اجتماعی اینستاگرام، که در آن‌ها اخبار ورزشی مورد نظر، مورد بحث قرار گرفته‌اند را در یک انبار داده ذخیره می‌کنیم. با استفاده از روش N-Gram احتمال این که این متن در پیکره متون اخبار اصلی منتشر شده توسط تارنمای ایسنا آمده باشد را محاسبه می‌کنیم بدین معنا که تعیین می‌کنیم آیا متن کپی‌برداری مستقیم است، کپی تکه‌ای از متن است یا یک متن مستقل از اخبار مورد نظر است. روش N-Gram روشی نسبتاً سریع‌تر نسبت به الگوریتم‌های خوشه‌بندی می‌باشد و به همین دلیل از این روش استفاده می‌کنیم. در تحلیل داده‌ها در صورتی که: الف) متن پست، مستقل از اخبار ورزشی باشد، نظر شخصی افراد در مورد اخبار مورد نظر بوده و به عنوان اخبار ورزشی جعلی و یا تحریف شده محسوب نمی‌شود، ب) در غیر این صورت باید مفهوم جمله با مفهوم جملاتی که از طریق الگوریتم‌های تحلیل معنایی<sup>۳</sup> برآورد شده است، مقایسه شود و اگر در مقایسه مفهومی و معنایی تفاوتی وجود نداشت، اخبار جعلی و یا تحریف شده نیست، در غیر این صورت، اخبار جعلی یا

شبکه‌های قبلی متفاوت کرده است، این است که هر واحد از واحد کناری نیز تأثیر می‌پذیرد. ورودی کناری نشان دهنده تأثیر لغت‌های قبلی در این توالی است. بنابراین در هر لحظه هر خروجی به ورودی همان لحظه و کل سابقه پیشین مسأله وابسته می‌شود (گودفلو و همکاران<sup>۱</sup>، ۲۰۱۶).

### روش LSTM<sup>۲</sup>

از آنجا که بخاطر سپاری همه مراحل پیشین مسأله کاری مشکل است بنابراین به جای اینکه همه مراحل پیشین در نظر گرفته شوند می‌توان از یک مارکوف با درجه دلخواه استفاده کرد. به عنوان مثال، اگر در یک جمله برای هر لغت فقط تأثیر لغت پیشین را در نظر بگیریم از یک مارکوف با درجه ۱ استفاده شده است و اگر تا دو لغت قبل مؤثر باشد مارکوف مرتبه ۲ است. اگر بخواهیم این خاصیت مسأله را به گونه‌ای تغییر بدهیم که خود مدل یاد بگیرد که نگهداری و تأثیر تا چند مرحله قبل برایش مفید است از یک حافظه طولانی کوتاه-مدت استفاده می‌کنیم. در واقع از چند دروازه با پارامتر داخلی استفاده می‌کنیم تا بتواند میزان بخاطر سپاری و فراموشی را هم یاد بگیرد (گودفلو و همکاران، ۲۰۱۶). یکی از مزیت‌های روش ال اس تی ام نسبت به سایر روش‌ها این است که نیاز به دریافت ورودی‌های با تعداد ویژگی‌های ثابت ندارد؛ به این معنی که شبکه عصبی مستقل از طول تعداد پارامترها است و نیازی نیست که تمام داده‌ها با یک سایز یکسان به آن داده شود. در این روش، داده‌ها را با هر طولی (هر تعداد حروف و کاراکتر) می‌توان به word2vector تبدیل کرد.

### برنامه اینستاگرام

اینستاگرام، یکی از پرمخاطب‌ترین شبکه‌های اجتماعی در جامعه ایرانی است که کاربران بی‌شماری آن را دنبال می‌کنند. کاربرد به اشتراک‌گذاری عکس، ویدئو و پخش زنده، همچنین دنبال کردن صفحات کاربران مختلف در سراسر دنیا و تعامل با آنها، از جمله خدماتی است که این شبکه اجتماعی به صورت رایگان به کاربران خود ارائه می‌دهد (تاجیک اسماعیلی و همکاران، ۱۳۹۹).

عنوان اپلیکیشن اینستاگرام که کاربران را قادر می‌سازد عکس بگیرند و به سادگی به اشتراک بگذارند، از دوربین‌های کداک "اینستنت" اقتباس شده است و گرام این

1. Goodfellow et al.  
2. Long Short-Term Memory

3. Semantic analysis

تحریف شده، تشخیص داده می‌شود.

### تحلیل داده‌ها

پیکره زبانی این تحقیق بر اساس بارگیری اخبار ورزشی از تارنمای ایسنا و برنامه اینستاگرام در محدوده زمانی ۹۸/۱۱/۱ تا ۹۹/۱/۳۱ بوده است. بدین صورت که اخبار ورزشی از تارنمای ایسنا در این محدوده زمانی بارگیری شده است و سپس در چند صفحه برنامه اینستاگرام که اخبار ورزشی گذاشته می‌شده است، اخبار ورزشی بارگیری شده است و سپس از لحاظ جعلی بودن یا غیرجعلی بودن با هم مقایسه شده‌اند.

در بازه زمانی مورد نظر حدود ۱۵۰۰ خبر ورزشی در تارنمای ایسنا یافت شد، که از این تعداد حدوداً ۵۰۰ خبر در صفحات مختلف برنامه اینستاگرام بازیابی شد. از این تعداد، ۵۰۰ خبر که کاملاً با اخبار موجود در تارنمای ایسنا مورد مطابقت قرار گرفته بود، بطور دستی توسط پژوهشگر جدا گردید. سپس ۴۰۰ خبر از این ۵۰۰ خبر نیز به طور دستی توسط پژوهشگر، جعلی یا غیرواقعی شد، بدین صورت که افعال و یا کلمات اصلی تغییر داده شد. به طور مثال، جمله «وزارت ورزش، دستور رفتن کی‌روش را داد»، تیتراژ یک خبر در اینستاگرام بوده که در تارنمای ایسنا هم وجود داشته و واقعی می‌باشد که پژوهشگر برای جعلی کردن این خبر این جمله را به «وزارت ورزش دستور رفتن کی‌روش را نداد» و یا «وزارت ورزش دستور رفتن کی‌روش را داد» تغییر داده است و برچسب جعلی روی این خبر گذاشته است. به تک تک اخبار برچسب جعلی و غیرجعلی زده شده است، بدین صورت که تمامی خبرها تک به تک با خبرهای تارنمای ایسنا مقایسه شده و خبرهای واقعی جدا گردیده و سپس برچسب واقعی به آنها زده شده است. سپس همان خبرهای واقعی را جعلی کرده و برچسب جعلی به آنها زده‌ایم و در پیکره اخبار جعلی وارد کرده‌ایم.

پس از به دست آمدن اخبار مورد نیاز که در این پژوهش حدوداً ۵۰۰ خبر در اینستاگرام انتخاب شده است که با اخبار ایسنا مطابقت دارد، این اخبار را به دو دسته ۴۰۰ و ۱۰۰ تایی تقسیم کردیم. دسته ۴۰۰ تایی را با حالت دستی (توسط پژوهشگر) به اخبار غیرواقعی تبدیل کردیم (مانند منفی کردن افعال و یا استفاده از کلمات متضاد و یا تغییر دادن اسامی یا تاریخ و ...) و برچسب جعلی به آنها زدیم. سپس این ۴۰۰ خبر برچسب خورده را با خبرهای واقعی که آنها نیز برچسب خورده‌اند وارد برنامه ال اس تی ام کردیم تا نرم‌افزار با استفاده از هوش مصنوعی بتواند یاد بگیرد و بقیه جملات بعدی را خود برچسب‌گذاری کند. پس از اینکه ۴۰۰ خبر جعلی و غیرجعلی را وارد نرم‌افزار ال اس تی ام کردیم، ۱۰۰ خبر مورد نظر را بدون این که برچسب زده شود، وارد نرم‌افزار کردیم تا نرم‌افزار به خبرها برچسب جعلی و غیرجعلی بزند. مراحل پیدا کردن اخبار جعلی از غیرجعلی در این پژوهش به صورت زیر است:

- ابتدا اخبار واقعی را از تارنمای ایسنا استخراج کرده و سپس با انتخاب پست‌های اینستاگرام مربوط به اخبار واقعی (واقع شده در اینستا)، آنها را (توسط N-gram)

در این پژوهش، ابتدا الگوهای واقعی و اصلی را، از منبع اصلی (تارنمای ایسنا) که حاوی اخبار معتبر و اصلی است، از طریق الگوریتم ریشه‌یابی به دست می‌آوریم. بدین صورت که حتی در صورت نقل قول جملات در اخبار، الگوی اصلی و حقیقی آنها باز هم قابل تشخیص باشد، که این امر از طریق ریشه‌یابی انجام می‌شود. الگوریتم ریشه‌یابی در این حوزه به معنای حذف پس‌وندها و پیش‌وندهای کلمات و استخراج ریشه آن‌ها است. در هر زبان، واژه‌ها با توجه به نقش معنایی و نحوی خود در جمله به شکل‌های ظاهری متفاوتی حضور می‌یابند. این شکل ظاهری متفاوت از جهتی نشان‌دهنده معنای متفاوت این واژه‌هاست، اما با توجه به این که تمامی آن‌ها از یک ریشه مشتق شده‌اند، از نظر معنا قرابت نسبتاً زیادی خواهند داشت. از همین رو در بسیاری از کاربردهای پردازش زبان طبیعی و بازیابی اطلاعات، نیاز داریم تا همه مشتقات یک واژه را به ریشه آن، که همان شکل ساده واژه می‌باشد، تبدیل نماییم. ریشه‌یابی کلمات در سیستم‌های بازیابی اطلاعات یک امر ناگزیر است (ذاکری، ۱۳۹۶). سپس اخبار به دست آمده در اینستاگرام را از طریق روش ان-گرم و باهم‌آیی با اخبار ایسنا مقایسه می‌کنیم تا اخبار مربوطه را پیدا کنیم. بدین صورت که به طور مثال اخباری در اینستاگرام پیدا کردیم با مضمون «مینو رایولا: می‌خواهم در تابستان امسال یک بازیکن به رئال مادرید منتقل کنم، فکر نمی‌کنم هالند امسال از دورتمند جدا شود.» برای جستجوی این خبر در پیکره اخبار واقعی، از کلماتی که باهم‌آیی دارند و روش ان-گرم استفاده می‌کنیم، بدین صورت که ابتدا کلماتی که باهم‌آیی دارند را جستجو می‌کنیم، مثلاً «رئال مادرید» یا «مینو رایولا» و سپس از طریق روش ان-گرم کلمات را ابتدا تک تک و سپس دو به دو و .... مقایسه می‌کنیم و با استفاده از فرمول، آن درصدی به دست می‌آید که نشان‌دهنده وجود یا عدم وجود خبر جعلی در پیکره اخبار واقعی است.

مورد آن‌ها تصمیم‌گیری شود.

در مرحله اول ابتدا در بازه زمانی مورد نظر حدود ۱۵۰۰ خبر ورزشی در تارنمای ایسنا انتخاب شد، اما تنها ۶۵۰ خبر در آن بازه زمانی در صفحات مختلف برنامه اینستاگرام یافت شد. در مرحله بعد ۵۰۰ خبر را که با اخبار موجود در تارنمای ایسنا مورد مطابقت قرار گرفته بود توسط برنامه ان-گرم، باهم‌آیی، هم‌معنایی و در برخی موارد توسط شم‌زبانی، جدا کردیم. در زیر چندین مثال برای نشان دادن این موضوع ارائه شده است.

به طور مثال، خبر زیر در تارنمای ایسنا آمده است:

#### ۱. «تمدید قرارداد ژيرو و کابایرو با چلسی

مهاجم و دروازه‌بان تیم فوتبال چلسی قراردادشان را با این تیم تمدید کردند.

به گزارش ایسنا به نقل از ساکرت، بعد از شیوع ویروس کرونا در قاره اروپا خیلی از رویدادهای ورزشی تعطیل شده است و باشگاه‌های لیگ برتری تمرینات خودشان را برای بازی‌های لیگ برتر انگلیس آغاز کردند.

باشگاه چلسی چهارشنبه شب خبر از تمدید قرارداد دو بازیکن خود داد. ابتدا اولیویه ژيرو مهاجم فرانسوی چلسی قرارداد خود را برای یک فصل دیگر تمدید کرد تا مهاجم سابق آرسنال فصل بعد هم در تیم آبی پوش لندن باشد.

همچنین ویلی کابایرو دروازه‌بان آرژانتین و تیم چلسی که توانست کپا آریزابالاگا دروازه‌بان چلسی را نیمکت‌نشین کند نیز قرارداد خود را برای یک فصل دیگر تمدید کرد. این خبر در اینستاگرام به صورت زیر آمده است:



شکل ۱. خبر مورد نظر در اینستاگرام

بررسی می‌کنیم تا ببینیم واقعی هستند یا خیر.

- پیاده‌سازی مرحله ریشه‌یابی بر روی تمام متون انتخابی
- اعمال روش بای-گرم بر روی متون
- محاسبه داده‌های آماری Q1 و Q3 برای بای-گرم‌های به دست آمده
- اگر بای-گرم بدست آمده کوچکتر از Q1 باشد، متن ورودی جعلی می‌باشد.
- اگر بزرگتر از Q3 بود متن ورودی غیر جعلی می‌باشد.
- اگر بین Q1 و Q3 بود، برای تصمیم‌گیری دوباره به ال اس تی ام فرستاده می‌شود.
- ورودی ال اس تی ام باید عدد باشد؛ بنابراین با تکنیک تبدیل، کلمه، به بردارهای عددی تبدیل می‌شود.
- با داده‌های جعلی و غیر جعلی برچسب خورده، ال اس تی ام آموزش داده می‌شود.
- متن‌های بین Q1 و Q3 را کلاسه‌بندی می‌کنیم.
- خطای کار ال اس تی ام را محاسبه می‌کنیم.
- انواع برچسب‌ها نیز به ترتیب زیر است:
- موجود در پیکره (حقیقی): اگر داده‌های به دست آمده، در پیکره موجود باشند که به طور قطع متن "غیر جعلی و واقعی" است و متعلق به همان منبع معتبر و اصلی است.
- احتمالاً جعلی یا حقیقی: در صورتی که متن برچسب «احتمالاً جعلی یا حقیقی» گرفته باشد باید آن را به طبقه‌بندی پیاده‌سازی شده، به وسیله شبکه عصبی ال اس تی ام دهیم، تا در آن‌جا درباره آن تصمیم‌گیری شود.
- ناموجود در پیکره (جعلی): در صورتی که برچسب‌های مورد نظر در متن وجود نداشته باشند و همچنین احتمال وجود آنها از طریق قرار گرفتن در طبقه‌بندی پیاده‌سازی شده توسط شبکه عصبی مورد بررسی قرار گیرد و نتیجه مشابهی به دست آید بنابراین متن، جعلی خواهد بود.

داده‌ها و برچسب‌های انتخاب شده را که قبلاً به صورت الگوهای استاندارد از اخبار منبع اصلی تهیه شده‌اند در جدولی با ۴ چارک<sup>۱</sup> قرار می‌دهیم و چارک‌های دور و نزدیک و چارک Q1 - Q3 را مورد مطالعه قرار می‌دهیم و بدین صورت ۲۴۰ داده انتخابی دسته‌بندی می‌شوند و توالی کلمات را که شامل دو چارک اخبار جعلی و اخبار واقعی می‌شوند را، در شبکه ال اس تی ام قرار می‌دهیم، تا آموزش ببینند و در نهایت Q1 - Q3 را به شبکه ال اس تی ام می‌دهیم تا در

1. Quartile

پشت سرهم تکرار شده‌اند در نظر گرفته می‌شود؛ در مرحله دوم، کلمهٔ دوم و سوم و تعداد تکرار آنها را مورد بررسی قرار می‌دهیم و در مراحل بعدی به ترتیب توالی دو کلمهٔ سوم و چهارم را بررسی خواهیم کرد.

برای پیدا کردن برخی خبرها نیز، از باهم‌آیی استفاده می‌کنیم، بدین صورت که مثلاً واژه‌های «باشگاه» و «بارسلونا» باهم‌آیی دارند و می‌توان واژه‌هایی که با هم باهم‌آیی دارند را در اخبار واقعی (تارنمای ایسنا) پیدا کرد و سپس آن‌ها را در اخبار اینستاگرام (توسط هشتگ) جستجو کرد.

سپس اخبار به دست آمده در اینستاگرام را با اخبار ایسنا مقایسه می‌کنیم تا اخبار مربوطه را پیدا کنیم. بدین صورت که به طور مثال خبری را در اینستاگرام پیدا کردیم با مضمون «خلیل‌زاده: پولم را بدهند می‌مانم و پنج سال دیگر تمدید می‌کنیم». برای جستجوی این خبر در پیکرهٔ اخبار واقعی از کلماتی که باهم‌آیی دارند و روش آن-گرم استفاده می‌کنیم بدین صورت که ابتدا کلماتی که باهم‌آیی دارند را جستجو می‌کنیم، مثلاً «خلیل‌زاده» یا «تمدید کردن» و سپس از طریق روش آن-گرم کلمات را ابتدا تک‌تک و سپس دو به دو و ... مقایسه می‌کنیم و با استفاده از فرمول، آن درصدی به دست می‌آید که نشان‌دهندهٔ وجود یا عدم وجود خبر اینستاگرام در ایسنا است.

صنوفی (۱۳۸۲) باهم‌آیی را این چنین تعریف می‌کند: در یک حوزهٔ معنایی، رابطهٔ معنایی بنیادینی میان واژه‌ها وجود دارد که به باهم‌آیی آنها منجر می‌گردد: مانند دندان / گاز گرفتن، درخت / کهنسال، سگ / هار و غیره. باهم‌آیی دو نوع است باهم‌آیی همنشینی و باهم‌آیی متداعی. وقوع واژه‌هایی با ویژگی‌های بنیادین مشترک بر روی محور همنشینی به نوعی باهم‌آیی منجر می‌شود که آن را باهم‌آیی همنشینی می‌نامند. صفاتی نظیر پیر، قدیمی، کهنه، کهنسال و غیره، با توجه به موصوف همشین‌شان در زبان فارسی جزء باهم‌آیی همنشینی محسوب می‌شوند. باهم‌آیی واژه‌ها بر حسب ویژگی‌ای که آن‌ها را در یک حوزهٔ معنایی قرار می‌دهد باهم‌آیی متداعی نامیده می‌شود مانند ماه، ستاره، خورشید و یا سیب، پرتقال، خیار و غیره.

### جمع‌آوری پیکره

پس از پیدا کردن اخبار مورد نظر به این نکته پی بردیم که تنها ۱۰۰ خبر از این ۵۰۰ خبر جعلی بودند و چون پیکرهٔ اخبار جعلی با غیر جعلی باید برابر باشند، به طور دستی ۴۰۰

«دی‌مارتزیو: چلسی بند تمدید ۱ ساله قرارداد اولیویر را فعال کرده و مهاجم فرانسوی تا تابستان ۲۰۲۱ در لندن ماندگار شد.»

خبری که در تارنمای ایسنا آمده است جزء پیکرهٔ اخبار واقعی و غیر جعلی است و برای پیدا کردن این خبر در پیکرهٔ اخبار اینستاگرام از روش آن-گرم استفاده می‌کنیم بدین صورت که ابتدا واژه «ژیرو» (که کلید واژه در این خبر هست) را در پیکرهٔ اینستاگرام جستجو می‌کنیم. به طور مثال عدد ۳ آمده است که یعنی سه بار این کلمه تکرار شده است. سپس واژهٔ «قرارداد ژيرو» (دو کلید واژه که باید در اخبار موجود باشد) را جستجو می‌کنیم که عدد ۱ آمده است بدین معنا که تنها یک بار این دو کلمه پشت سر هم آمده‌اند. از فرمول بای-گرم:

$$P(w_i | w_1 w_2 \dots w_{i-1}) \approx P(w_i | w_{i-1})$$

برای پیدا کردن واژه‌ها استفاده می‌کنیم. بدین معنا که احتمال اینکه این خبر در پیکرهٔ ما باشد ۱ تقسیم بر ۳ است یعنی ۳۳٪ و این احتمال هنگامی که تری-گرم، فور-گرم و ... را پیدا کنیم، این درصد بالاتر رفته و خبر نهایی یافت می‌شود.

فرمول تری-گرم

$$P(w_i | w_1 w_2 \dots w_{i-1}) \approx P(w_i | w_{i-2} w_{i-1})$$

فرمول فور-گرم

$$P(w_i | w_1 w_2 \dots w_{i-1}) \approx P(w_i | w_{i-3} w_{i-2} w_{i-1})$$

فرمول فایو-گرم

$$P(w_i | w_1 w_2 \dots w_{i-1}) \approx P(w_i | w_{i-4} w_{i-3} w_{i-2} w_{i-1})$$

در روش بای-گرم بر اساس متن و پیکرهٔ انتخاب شده دو جمله یا عبارت یا کلمه که به صورت متوالی بعد از هم می‌آیند انتخاب می‌شوند و به عنوان برجسب‌های یک متن شناخته می‌شوند که این روند را می‌توان به تعداد مورد نیاز در پیکرهٔ مورد نظر انتخاب کرد و احتمال توالی آنها را در متون انتخابی مورد بررسی قرار داد. در این روش برای انتخاب نمونه‌ها و عبارات قابل قبول باید از نمونه‌هایی که به صورت تکیه کلام هستند یا از عباراتی که مشخصاً مربوط به نویسنده است، انتخاب شود که در واقع از طریق هشتگ‌های انتخابی در اخبار قابل شناسایی هستند، و از این طریق متن را برجسب‌گذاری می‌کنیم؛ به طور مثال بعد از تعیین برجسب‌های مشخص شده اگر یک جمله یا عبارت از چهار کلمه تشکیل شده باشد، ابتدا تعداد تکرارهای دو کلمه‌ای که



اگر و فقط اگر این سه جمله از شرایط صدق یکسانی برخوردار باشند، آن وقت می‌توانیم بگوییم که این سه جمله محتوای گزاره‌ای یکسانی دارند.

معنی‌شناسان منطقی سعی بر آن داشته‌اند تا معلوم کنند، مسأله معنی‌داری در کدام حوزه‌های رفتار زبان طبیعی به شکلی بارزتر جلوه می‌کند. برخی از این حوزه‌ها را می‌توان وجه، زمان دستوری، نمود و نیز افعالی دانست که از نگرش گزاره‌ای برخوردارند.

در پیکره، به اخباری که تیتراهای یکسان داشتند اما متن آنها متفاوت بود برخوردیم. در مثال زیر می‌بینیم که تیترا هر دو خبر تا حدودی با هم متفاوت است:

## ۲. ایسنا: «خلیل‌زاده: پولم را بدهند می‌مانم و پنج سال دیگر تمدید می‌کنیم.»

اینستاگرام: «خلیل‌زاده: پول قراردادم را بدهند در پرسپولیس می‌مانم.»

ابتدا از لحاظ معنایی هر دو خبر را با هم مقایسه می‌کنیم که واقعی بودن آن مشخص شود. تیترا متن دوم را از طریق صحت صدق و کذب با متن اول مقایسه می‌کنیم. در تیترا خبر منتشره در اینستاگرام، در جمله اول کلمه «قرارداد» و «در پرسپولیس» اضافه شده و جمله دوم کامل حذف شده است. اما از لحاظ معنایی یکسان هستند و از طریق هم‌معنایی قابل اثبات است. اما با این که تیترا هر دو خبر از لحاظ معنایی با هم یکی هستند، ولی متن خبر دوم با اول متفاوت است و به موضوعی غیر از موضوع تیترا پرداخته است که چندین مورد اینچنینی در پیکره یافت شد و به علت ناهمگون بودن از پیکره حذف شدند.

به علاوه، تیترا خبر اول کاملاً منطبق بر قواعد نحوی زبان فارسی نیست، («تمدید می‌کنیم» از لحاظ شمارش با فاعل هم‌خوانی ندارد) که در تارنماهای خبری به علت عجله داشتن در گذاشتن خبر و به روزرسانی سریع اخبار، بسیار دیده می‌شود و این مشکلات در پیکره نادیده گرفته شده است و هنگام ریشه‌یابی، ریشه هر دو واژه، یک واژه در نظر گرفته شده است. در برخی دیگر از خبرهای موجود در پیکره، به اخباری برخوردیم که تیترا آن‌ها یکسان نبود اما متن آن‌ها یکسان بود. در مثال زیر ابتدا خبر در تارنمای ایسنا منتشر شده و سپس همان خبر در اینستاگرام:

## ۳. «جزییات قرارداد بیرانوند با آنتورپ»

تیم فوتبال آنتورپ با علیرضا بیرانوند دروازه‌بان تیم ملی فوتبال ایران برای ۳ سال به توافق نهایی رسید.

خبر باقیمانده را جعلی کردیم. بدین صورت که افعال و یا کلمات اصلی را تغییر دادیم به طور مثال جمله «وزارت ورزش، دستور رفتن کی‌روش را داد» تیترا یک خبر در اینستاگرام بوده که در تارنمای ایسنا هم وجود داشته و واقعی می‌باشد که برای جعلی کردن این خبر، جمله را به «وزارت ورزش دستور رفتن کی‌روش را نداد» و یا «وزارت ورزش دستور رفتن کی‌روش را داد» تغییر دادیم و برچسب جعلی روی این اخبار گذاشتیم. پس در این مرحله پیکره مطالعه کامل گردید و دیتابیس، متشکل از ۵۰۰ خبر جعلی و ۵۰۰ خبر غیرجعلی، یعنی در مجموع ۱۰۰۰ خبر شد.

## برچسب‌زنی

در مرحله بعد، به تک‌تک خبرها برچسب جعلی و غیرجعلی زدیم، بدین صورت که تمامی خبرها را تک به تک با اخبار تارنمای ایسنا مقایسه و خبرهای واقعی را جدا کردیم و سپس برچسب واقعی به آنها زدیم، سپس همان خبرهای واقعی را جعلی کرده و برچسب جعلی به آنها زدیم و در پیکره اخبار جعلی وارد کردیم. این نکته قابل ذکر است که برای مقایسه اخبار غیرجعلی در تارنمای ایسنا از اخبار موجود در اینستاگرام، از خلاصه‌برداری، باهم‌آیی و محک صدق و کذب استفاده کردیم.

لاینز (۱۳۹۱) معتقد است که برای شناسایی جمله‌های معنی‌دار و بی‌معنی می‌توان از شرایط صدق استفاده کرد. شرایط صدق، نظریه مشروط به صدق معنی و به لحاظ تاریخی برگرفته از اثبات‌گرایی است و همچون اثبات‌گرایی گونه‌های متفاوتی دارد که با یکدیگر تفاوت‌های جزئی دارند.

(الف) «جان اسمیت غیر متأهل است.»

(ب) «جان اسمیت متأهل نیست.»

(ج) «جان اسمیت مجرد است.»

مفاد این جمله‌ها بیانگر مجموعه گزاره‌هایی است، که هرکدام ارزش خاص خود را دارند، آن هم بر حسب این که به هنگامی که سخن می‌گوییم، جان اسمیت به چه کسی ارجاع داده می‌شود و در آن زمان، این فرد متأهل است یا خیر. ما در چنین مواردی، به خوبی می‌دانیم که به لحاظ تجربی، چطور می‌شود صدق یا کذب گزاره‌های بیان شده در مفاد این جمله را معلوم کرد. به علاوه بدون اینکه نیازی به تحقیق تجربی درباره وضعیت تأهل جان اسمیت باشد، می‌توانیم مدعی شویم که بر حسب دانش زبانی‌مان، جمله (الف) از همان شرایط صدق جمله‌های (ب) و یا حتی (ج) برخوردار است.

در برخی موارد نیز، تیتیر خبرها کاملاً مشابه بودند اما خبر اینستاگرام خیلی خلاصه شده بود به صورتی که یک خبر چند صفحه‌ای به چند خط تبدیل شده بود.

#### ۵. «جهانبخش: آژاکس بالاترین رقم تاریخ خودش را به من پیشنهاد داد»

بازیکن ایرانی تیم برایتون فاش کرد آکمار مانع از رفتن او به تیم فوتبال آژاکس شد.

به گزارش ایسنا و به نقل از اسپورت باکس، علیرضا جهانبخش بازیکن پیشین آکمار (AZ) و بهترین گلزن اربودویزه هلند در فصل ۲۰۱۷-۲۰۱۸ بود و به همین ترتیب این بازیکن ایرانی مجبور شد باشگاه جدید خود را از چند تیم بزرگ انتخاب کند. آژاکس حتی می‌خواست برای جذب او تا جایی پیش برود که رکورد نقل و انتقالات خود را بشکند.

جهانبخش سرانجام در تابستان سال ۲۰۱۸ با ۱۹ میلیون یورو به باشگاه لیگ برتری برایتون و هوو آلیون رفت. در همین رابطه بازیکن ایرانی گفت: «من پیشنهادات جالب زیادی داشتم. جدی‌ترین‌ها از آژاکس و PSV بود. آژاکس بالاترین رقم تاریخ خودش را به من پیشنهاد داد، اما اجازه نداد که به باشگاه دیگر هلندی بروم. آنها به من پیشنهاد بهتری دادند، اما من تصمیم گرفتم که لیگ هلند را ترک کنم. در آن زمان میرالم سلیمانی، گران‌ترین بازیکن آژاکس بود و می‌توان متصور بود که آژاکس دست کم ۱۷ میلیون یورو به جهانبخش پیشنهاد داده بود.

با این حال، این امر با علاقه تیم هلندی متوقف نمی‌شود، زیرا او می‌توانست در ایتالیا یا آلمان نیز بازی کند. بازیکن ایرانی در بخش دیگری از صحبت‌های خود گفت: «مذاکراتی با ناپولی انجام شد و پیشنهادهای خوبی از آلمان داشتم، از جمله از باشگاه‌هایی که همیشه در بین ۵ تیم برتر بود. من هنوز یک قرارداد ۲ ساله داشتم و هم‌چنان به افزایش قیمت تأکید داشتم و به همین خاطر تصمیم به رفتن به لیگ برتر انگلیس گرفتم.

در پایان سال ۲۰۱۹، علیرضا اولین گل خود را در برایتون به ثمر رساند. او به زیبایی توانست برای تیمش گلزنی کند و در بازی برابر بورنموث نیز گل دیگری برای تیمش به ثمر رساند.»

این خبر در اینستاگرام به صورت زیر آمده است:  
«جهانبخش: آژاکس بالاترین رقم تاریخ خودش را به من پیشنهاد داد:

علیرضا جهانبخش به اسپرت باکس گفت: من

به گزارش ایسنا به نقل از سایت دالمپورتال، آنتورپ با پرداخت ۶۳۰ هزار یورو که مبلغ قرارداد علیرضا بیرانوند با پرسپولیس بود، توانست با قراردادی ۳ ساله با بند تمدید یک ساله به توافق نهایی برسد.

البته بیرانوند معاینات پزشکی خود را انجام نداده و در تابستان این امر رخ خواهد داد (تارنمای خبرگزاری ایسنا).

متن خبر در اینستاگرام به صورت زیر است:  
«توافق‌نامه سه‌جانبه آنتورپ بلژیک، پرسپولیس و بیرانوند امضا شد. بیرانوند هم با آنتورپ امضا کرد.

مدیر روابط عمومی باشگاه پرسپولیس ضمن تقدیر از وفاداری بهترین گلر آسیا، از توافق سه‌جانبه باشگاه آنتورپ بلژیک، باشگاه پرسپولیس و علیرضا بیرانوند خبر داد. بیرانوند هم دقایقی قبل با آنتورپ قرارداد بست» (خبر مورد نظر در اینستاگرام).

در مثال بالا، عکس‌های هر دو خبر یکی است و انتظار می‌رود که تیتیر و متن خبر هم یکسان باشد اما تیتیر آنها با هم متفاوت است ولی متن آنها هم معنا هستند و یک معنای اصلی دارند. این نوع اخبار را در پیکره نگه داشته و تیتیر خبر واقعی را به عنوان تیتیر درست ثبت کرده‌ایم.

در برخی از خبرها، تیتیر خبرها با هم هم‌معنا هستند، اما باید از طریق ریشه‌شناسی، باهم‌آیی یا هم‌معنایی، آنها را در کنار هم قرار دهیم؛ ولی در برخی خبرها مانند مثال زیر، برای اثبات هم‌معنایی دو خبر باید از شم زبانی استفاده کنیم.

#### ۴. ایسنا: «اسکوچیچ: تمام تلاشم را می‌کنم تا تیم ملی به جام جهانی صعود کند»

اینستاگرام: «قول بزرگ دراگان اسکوچیچ به ۸۰ میلیون ایرانی»

در نگاه اول به نظر می‌رسد که این دو تیتیر کاملاً با هم متفاوت هستند و غیر از واژه اسکوچیچ بقیه با هم متفاوت است که نه می‌توان از طریق ریشه‌شناسی و نه از طریق باهم‌آیی و یا ان-گرم، هم‌معنایی این دو تیتیر را اثبات کرد. در این موارد ما در مرحله آخر از شم زبانی استفاده کردیم، بدین صورت که فردی که به زبان فارسی مسلط بود و فارسی زبان مادری وی بود، این جملات را بررسی و هم‌معنا بودن آنها را تایید کرده است. پس این نوع خبرها در پیکره استفاده شده است. به علاوه، متن این خبرها شبیه به هم است؛ همان طور که در مثال بالا دیده می‌شود، غیر از خلاصه بودن خبر اینستاگرام در بقیه موارد هر دو خبر تقریباً یکسان هستند.

آنالیزهای آماری و داده‌کاوی، پسوندها حذف کردند، که این روش هم می‌تواند راهی برای تشخیص ریشه باشد. معروف‌ترین الگوریتم ریشه‌یابی در انگلیسی پورتر می‌باشد. برای ریشه‌یابی به برچسب کلمات نیازی نیست، بنابراین برای این مؤلفه برچسب‌گذاری صورت نمی‌گیرد.

### ال اس تی ام

در مرحله بعد، از ۱۰۰۰ خبر به دست آمده به عنوان پیکره (۵۰۰ خبر جعلی و ۵۰۰ خبر غیرجعلی)، ۸۰۰ خبر (۴۰۰ خبر جعلی و ۴۰۰ خبر غیرجعلی) را به‌طور تصادفی انتخاب می‌کنیم (اخبار باید یکی باشند یعنی همان خبری که جعلی آن را انتخاب کردیم باید غیر جعلی آن را نیز انتخاب کنیم) و آنها را به عنوان ورودی به شبکه ال اس تی ام (در برنامه متلب) وارد می‌کنیم. ۸۰۰ خبر به عنوان ورودی‌های برنامه ال اس تی ام برای یادگیری جداسازی اخبار جعلی از غیرجعلی این برنامه محسوب می‌شوند. بدین صورت که ما ابتدا ۸۰۰ خبر را، برچسب‌زده (جعلی و غیرجعلی) به عنوان ورودی این برنامه انتخاب می‌کنیم و سپس ۱۰۰ خبر باقیمانده را به عنوان تست وارد این برنامه می‌کنیم که ببینیم برنامه به‌طور صحیح اخبار جعلی را از غیرجعلی تشخیص می‌دهد یا خیر.

اما در ابتدا باید کلمات به وکتور (عدد) تغییر یابند. در این مقاله ما از تبدیل Word2Vec برای این منظور استفاده کرده‌ایم. همانطور که قبلاً نیز گفتیم این روش نسبت به سایر روش‌ها مزیتی دارد که نیاز به دریافت ورودی‌های با تعداد ویژگی‌های ثابت ندارد.

در آخرین مرحله ال اس تی ام، داده‌ها را با هر طولی (هر تعداد حروف و کاراکتر) به word2vector تبدیل می‌کنیم. برای هر کلمه یک بردار ۳۵۰ تایی تهیه می‌کنیم و از روی مجموع داده‌ها ۱۷۰ اپوک<sup>۱</sup> می‌زنیم تا نتیجه مطلوب حاصل شود.

### نتایج حاصله

بر اساس پیکره‌ای که شامل ۱۰۰۰ جمله می‌شود، داده‌ها به دو دسته تقسیم می‌شوند که طبق روند توضیح داده شده، در نهایت ۱۷۰ اپوک زده می‌شود و جدول شماره ۱ به دست می‌آید.

پیشنهادات جالب زیادی داشتیم. جدی‌ترین‌ها از آژاکس و PSV بود. آژاکس حتی بیشتر از آن‌چه که قبلاً برای بازیکنان انجام داده بودم پیشنهاد داد، اما AZ اجازه نداد که باشگاه دیگر هلندی بروم.

پ.ن: در آن زمان میرالم سلیمانی، گران‌ترین بازیکن آژاکس بود و می‌توان متصور بود که آژاکس دست کم ۱۷ میلیون یورو به جهانبخش پیشنهاد داده بود.» (خبر مورد نظر در اینستاگرام)

همان‌طور که در این خبرها دیده می‌شود، تیترو دو خبر کاملاً یکسان هستند، اما خبر اینستاگرام بسیار خلاصه‌تر از خبر تارنمای ایسنا است. به طوری که در چند خط کل خبر را خلاصه کرده است. اما پژوهشگر این امر را امری قابل چشم‌پوشی یافته و این نوع خبرها را در پیکره جای داده است. در اینستاگرام چون خبر مورد نظر باید در یک عکس جا داده شود، پس این مسأله امری طبیعی است و بیشتر اخبار مورد نظر در اینستاگرام خلاصه‌تر از همان خبر در تارنمای ایسنا بوده است.

### ریشه‌یابی

مرحله بعد ریشه‌یابی است، که توسط کتابخانه هضم (HAZM) که در کتابخانه برنامه پایتون موجود است، صورت می‌پذیرد. هضم کتابخانه پردازش زبان فارسی است که در پایتون، جاوا و سی‌شارپ برای توکن‌سازی استفاده می‌شود. برای استفاده از هضم، همه متن فارسی باید تبدیل به یونیکد شود. البته دیتاست هضم بخش زیادی از اصطلاحات زبان فارسی را پوشش نمی‌دهد، کلمات اضافی را حذف نمی‌کند و مفاهیم را هم از کلمات تشخیص نمی‌دهد (امیرخانی و همکاران، ۲۰۲۰).

این ابزار برای ریشه‌یابی لغات و تشخیص نوع کلمه ساخته شده و از آن برای ریشه‌یابی (اسم مکان، اسم زمان، حالت فاعلی، مفعولی و ...) استفاده می‌شود. ریشه‌یابی اصطلاحی است که برای فرآیند کاهش دادن یک کلمه و رسیدن به ریشه آن به کار می‌رود. منظور از ریشه در این تعریف، ریشه زبانی نیست و هدف این است که فرمت‌های گوناگون یک کلمه دارای ریشه‌های یکسان باشند. معمولاً ریشه‌یابی لغات بر اساس قواعد ساختارهای و سپس حذف پسوندها صورت می‌پذیرد. تاکنون روش مؤثری برای حذف پسوندها ارائه نشده است. در تلاشی که در آزمایشگاه فناوری وب انجام شده است، سعی شده تا بر اساس

۱. هر یک باری که کل داده‌ها توسط شبکه عصبی دیده می‌شوند و آموزش می‌بینند یک اپوک (Epoch) گفته می‌شود.

خورده و سپس ۲۰۰ خبر برچسب نخورده باقیمانده را در ال اس تی ام وارد کردیم تا برنامه به آنها برچسب بزند. نتایج حاصله در جدول ۲ نشان داده شده است:

جدول ۲. نتایج حاصل از برچسب‌زنی برنامه ال اس تی ام

| تعداد کل اخبار               | اخبار غیر جعلی | اخبار جعلی |
|------------------------------|----------------|------------|
| ۱۰۰                          | ۱۰۰            | ۱۰۰        |
| تعداد خبرهای تشخیص داده شده  | ۸۱             | ۷۵         |
| تعداد خبرهای تشخیص داده نشده | ۱۹             | ۲۵         |

برای محاسبه معیارهای ارزیابی از فرمول‌های جدول ۳ استفاده کردیم.

جدول ۳. فرمول‌های محاسبه معیارهای ارزیابی

|                                                                  |                        |
|------------------------------------------------------------------|------------------------|
| $Accuracy = \frac{tp + tn}{tp + tn + fp + fn}$                   | فرمول محاسبه Accuracy  |
| $Precision = \frac{tp}{tp + fp}$                                 | فرمول محاسبه Precision |
| $Recall = \frac{tp}{tp + fn}$                                    | فرمول محاسبه Recall    |
| $F1 = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall}$ | فرمول محاسبه F1:       |

متغیر tp: تعداد خبرهای حقیقی درست تشخیص داده شده.  
 متغیر tn: تعداد خبرهای حقیقی اشتباه تشخیص داده شده.  
 متغیر fp: تعداد خبرهای جعلی درست تشخیص داده شده.  
 متغیر fn: تعداد خبرهای جعلی اشتباه تشخیص داده شده.  
 Accuracy = 0.5  
 Precision = 0.51  
 Recall = 0.76  
 F = 0.61

### بحث و نتیجه‌گیری

روش پیشنهاد شده در این مقاله بر روی ۴ دیتاست معتبر و موجود پیاده‌سازی شده و با ۶ روش پیشین مقایسه شده است. همان‌گونه که در جدول ۱ مشاهده می‌شود، نتایجی که در ردیف دقت<sup>۱</sup> آمده است، نمایانگر دقت الگوریتم‌ها می‌باشد که با یکدیگر مقایسه شده‌اند. درنگاهی اجمالی می‌توان میزان بهبود نتایج بدست آمده از روش به کار گرفته شده در این مقاله را مشاهده کرد.

مشخصات ال اس تی ام طراحی شده:

- ۱ " توالی ورودی توالی ورودی با ۳۵۰ بعد
- ۲ " ال اس تی ام ال اس تی ام با ۲۳۰ لایه پنهان
- ۳ " کاملاً متصل ۲ لایه کاملاً متصل
- ۴ " تابع فعالسازی سافت مکس
- ۵ " نوع خروجی طبقه‌بندی ۲ کلاس

جدول ۱. نتایج ال اس تی ام

| مرحله | تعداد تکرار | زمان مصرفی (ثانیه) | میزان خطا | میزان دقت | نرخ یادگیری |
|-------|-------------|--------------------|-----------|-----------|-------------|
| 1     | 1           | 0.16               | 0.7231    | 72.83%    | 0.0100      |
| 6     | 50          | 504.76             | 0.7356    | 25.10%    | 0.0100      |
| 12    | 100         | 1081.63            | 0.6656    | 70.83%    | 0.0100      |
| 17    | 150         | 1626.54            | 0.6834    | 71.53%    | 0.0100      |
| 23    | 200         | 2165.50            | 0.6837    | 71.33%    | 0.0100      |
| 28    | 250         | 2737.00            | 0.6959    | 73.00%    | 0.0100      |
| 34    | 300         | 3295.55            | 0.4836    | 76.63%    | 0.0100      |
| 39    | 350         | 3889.07            | 0.6898    | 77.17%    | 0.0100      |
| 45    | 400         | 4384.69            | 0.5772    | 70.50%    | 0.0100      |
| 50    | 450         | 4960.54            | 0.4019    | 77.73%    | 0.0100      |
| 56    | 500         | 5475.52            | 0.2349    | 78.00%    | 0.0100      |
| 62    | 550         | 6133.67            | 0.2397    | 75.83%    | 0.0100      |
| 67    | 600         | 6577.67            | 0.3364    | 78.33%    | 0.0100      |
| 73    | 650         | 7166.54            | 0.1278    | 78.53%    | 0.0100      |
| 78    | 700         | 7659.77            | 0.3235    | 80.00%    | 0.0100      |
| 84    | 750         | 8201.73            | 0.2299    | 79.33%    | 0.0100      |
| 89    | 800         | 8752.51            | 0.0123    | 81.27%    | 0.0100      |
| 95    | 850         | 9286.05            | 0.1153    | 81.50%    | 0.0100      |
| 100   | 900         | 9864.73            | 0.0134    | 81.83%    | 0.0100      |
| 106   | 950         | 10399.92           | 0.0129    | 85.30%    | 0.0100      |
| 112   | 1000        | 10937.98           | 0.0025    | 82.83%    | 0.0100      |
| 117   | 1050        | 11472.65           | 0.0025    | 88.43%    | 0.0100      |
| 123   | 1100        | 12013.98           | 0.1056    | 89.33%    | 0.0100      |
| 128   | 1150        | 12569.59           | 0.1020    | 90.00%    | 0.0100      |
| 134   | 1200        | 13095.69           | 0.0008    | 100.00%   | 0.0100      |
| 139   | 1250        | 13644.28           | 0.0115    | 95.17%    | 0.0100      |
| 145   | 1300        | 14178.39           | 0.0007    | 100.00%   | 0.0100      |
| 150   | 1350        | 14737.49           | 0.0008    | 100.00%   | 0.0100      |

در قیاس با روش‌های پیشین، دقت نسبتاً قابل قبولی در نتایج بدست آمده، حاصل شده است؛ که البته در صورتی که از روش‌های بهتری برای تبدیل کلمات به اعداد قابل ورود به ال اس تی ام استفاده شود انتظار می‌رود دقت بالاتری بدست آید. همانطور که گفتیم، برنامه ال اس تی ام یک نوع هوش مصنوعی است که با وارد کردن داده‌های آموزشی، می‌تواند فرایند را یاد بگیرد و به داده‌های بدون برچسب، برچسب جعلی یا غیر جعلی بزند. برای این کار، ابتدا ۸۰۰ خبر برچسب

1. Accuracy

انتخابات و ... افراد منتشرکننده اخبار غیرواقعی را سریع‌تر پیدا نمایند و با آن‌ها برخورد کنند.

با توجه به اینکه در توازن بین دقت و سرعت با بالا بردن سرعت، دقت الگوریتم پایین می‌آید برای جلوگیری از این امر در مرحله دوم و با استفاده از روشی با دقت بالا سعی کردیم دقت الگوریتم پیشنهادی را نیز بالا ببریم.

همان‌طور که مشاهده می‌شود دقت این روش نسبت به دیگر روش‌ها قابل قبول است. میانگین زمانی روش پیشنهادی در مقایسه با سه روش RST+Castillo، LIWC+Castillo و TriFN بهتر می‌باشد اما در مقایسه با سه روش RST، LIWC و Castillo زمان نسبتاً بیشتری صرف می‌کند.

نتایج به دست آمده، بیانگر این است که برای شناسایی اخبار جعلی، در میان اخبار منتشر شده در اینستاگرام، روش پیشنهادی ارائه شده در این مقاله روش مناسبی می‌باشد دقت کافی را دارد.

این تحقیق می‌تواند برای بیشتر اقسام جامعه بالاخص ورزشکاران، باشگاه‌های ورزشی، مربیان ورزشی، ورزش دوستان، خبرنگاران ورزشی، محققان، علاقمندان به تشخیص اخبار جعلی از غیر جعلی و ... مفید باشد.

از آنجا که الگوریتم‌های شناسایی اخبار جعلی، عموماً بر پایه دقت یا سرعت، کار خود را انجام می‌دهند در این مقاله ما سعی داشتیم با استفاده از الگوریتم‌های ترکیبی و در جایگاه درست، یک توازن منطقی بین سرعت و دقت را برقرار کنیم که البته با توجه به نتایج به دست آمده در این امر نیز موفق بوده‌ایم. بر این اساس:

دقت نتایج بدست آمده از طریق روش ال اس تی ام و با توجه به آنکه داده‌های ورودی باید به صورت اعداد (بردار) باشند، افزایش یافته است.

با در نظر گرفتن این موضوع که روش ما یک روش ترکیبی است، سرعت رسیدن به نتیجه مطلوب در روش بای-گرام (روش اول) از روش ال اس تی ام (روش دوم) بیشتر است چرا که روش دوم زمان بر است. اما روش اول برای داده‌های ساده‌تر کاربرد دارد و روش دوم برای گرفتن نتایج کامل و دقیق ارجحیت دارد.

اخبار منتشر شده غیرواقعی در صورتی که از پایگاه خبری اخبار حقیقی بسیار دور یا بسیار نزدیک باشد سریعاً قابل شناسایی است و اینستاگرام (یا دیگر شبکه‌های اجتماعی) با استفاده از این روش‌ها می‌توانند در زمان‌های مختلف (مخصوصاً در زمان‌های دارای اهمیت ویژه از قبیل:

## منابع

امیری، علی، عادل آذر و میثم شهبازی. (۱۳۹۹). ارائه رویکردی مبتنی بر برنامه ریزی تصادفی و زنجیره مارکوف به منظور بهینه سازی بازتولید و برون سپاری در زنجیره تامین. *مطالعات مدیریت صنعتی*. ۱۸ (۵۷). ۱-۴۲.

تاجیک اسماعیلی، سمیه، ابوالفضل علیپور و سروناز تربتی. (۱۳۹۹). نقش اینستاگرام در توسعه برند شخصی (مورد مطالعه: کاربران اینترنتی اینستاگرام، ۱۳۹۸).

پژوهش‌های ارتباطی. ۲۷ (۳). ۳۵-۵۷.

<http://ensani.ir/file/download/article/1614585842-9616-103-2.pdf>

ذاکری، مرتضی. (۱۳۹۶). مشخص کردن برچسب اجزای سخن.

<https://elmnet.ir/Article/11185539-25131/%D8%A8%D8%B1%DA%86%D8%B3%D8%A8-%D8%B2%D9%86%DB%8C-%D8%A7%D8%AC%D8%B2%D8%A7%DB%8C-%D8%B3%D8%AE%D9%86-%D8%AF%D8%B1-%D9%86%D9%88%D8%B4%D8%AA%D9%87-%D9%87%D8%A7%DB%8C->

[%D9%81%D8%A7%D8%B1%D8%B3%DB%8C-%D8%A8%D8%A7-%D8%A7%D8%B3%D8%AA%D9%81%D8%A7%D8%AF%D9%87-%D8%A7%D8%B2-%D8%B4%D8%A8%DA%A9%D9%87-%D9%87%D8%A7%DB%8C-%D8%B9%D8%B5%D8%A8%DB%8C-%D9%88-](#)

[%D8%A8%D8%A7%D8%B2%D9%86%D9%85%D8%A7%DB%8C%DB%8C-%DA%A9%D9%84%D9%85%D8%A7%D8%AA](#)

سالاری، سیدمجتبی، و فضل الله ادیب‌نیا. (۱۳۸۹). ۱۰ الگوریتم از برترین‌های داده کاوی. سیزدهمین کنفرانس دانشجویی مهندسی برق ایران تربیت مدرس. ۲۴-۲۶.

سامانی، سمانه، و اعظم فراهانی. (۱۳۹۵). هویت آنلاین و اینستاگرام (مطالعه چگونگی نمایش هویت جوانان در اینستاگرام). رسانه. ۲۷ (۲). ۸۵-۱۰۴.

<http://ensani.ir/file/download/article/20170201142336-9616-399.pdf>

صفوی، کورش. (۱۳۸۲). پژوهش درباره با هم‌آیی واژگان در زبان فارسی. *متن پژوهی ادبی*. ۷ (۱۸). ۱-۱۳.

- مخازن نفتی جنوب باختر ایران. فصلنامه علمی- پژوهشی علوم زمین. ۹، (۱). ۶۵-۷۷.
- <https://iranjournals.nlai.ir/bitstream/handle/123456789/90400/4CF4E2C9954C0BDCED92BA969E000F4F.pdf?sequence=-1&isAllowed=y>
- Amirkhani, H., Jafari, M. A., Amirak, A., Pourjafari, Z., Jahromi, S. F., & Kouhkan, Z. (2020). Farstail: A Persian natural language inference dataset. arXiv preprint arXiv:2009.08820.
- Chaffey, D. (2021). Global social media research summary 2016. Smart Insights: Social Media Marketing. <https://www.smartinsights.com/social-media-marketing/social-media-strategy/new-global-social-media-research/>
- De Beer, D., & Matthee, M. (2020). Approaches to identify fake news: A systematic literature review. In International Conference on Integrated Science (pp. 13-22). Springer, Cham.
- De Oliveira, N. R., Medeiros, D. S., & Mattos, D. M. (2020). A sensitive stylistic approach to identify fake news on social networking. IEEE Signal Processing Letters, 27, 1250-1254.
- Gahirwal, M., Moghe, S., Kulkarni, T., Khakhar, D., & Bhatia, J. (2018). Fake news detection. International Journal of Advance Research, Ideas and Innovations in Technology, 4(1), 817-819.
- Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep learning. Book in preparation for MIT Press. URL: <http://www.deeplearningbook.org>, 1.
- Jimenez, M., Maxime, C., Le Traon, Y., & Papadakis, M. (2018). On the impact of tokenizer and parameters on n-gram based code analysis. In 2018 IEEE International Conference on Software Maintenance and Evolution (ICSME) (pp. 437-448). <https://orbilu.uni.lu/bitstream/10993/36135/1/icsme3.pdf>
- Mertoğlu, U., & Genç, B. (2020). Automated fake news detection in the age of digital libraries. Information Technology and Libraries, 39(4). <https://doi.org/10.6017/ital.v39i4.12483>
- Minyoung Huh, Andrew Liu, Andrew Owens, Alexei A. (2018). Fighting fake news: Image splice detection via learned self-consistency. Efron; Proceedings of the European Conference on Computer Vision (ECCV), pp. 101-117.
- Veisi, H., & Sameti, H. (2013). Speech enhancement using hidden Markov models in Mel-frequency domain. Speech Communication, 55(2), 205-220.
- لاینز، جان. (۱۳۹۱). درآمدی بر معنی‌شناسی زبان، ترجمه کورش صفوی، تهران: علمی.
- نیکوگفتار، هانیبه، عباس بحرودی، بهزاد تخم‌چی، غلامحسین نوروزی، بهزاد مهرگینی. (۱۳۹۲). مدل‌سازی با زنجیره مارکوف، مطالعه موردی: رخساره‌های سنگی یکی از