



دانشگاه صنعتی امیرکبیر
(پلی تکنیک تهران)



دانشکده مهندسی کامپیوتر

پیاده سازی ابزار جمع آوری اخبار جعلی فارسی و دسته بندی آن

ارائه دهنده:
محمد رضا صمدی

استاد راهنما:
دکتر سعیده ممتازی



بهمن ۱۳۹۹

فهرست مطالب

۱. مقدمه

۲. پیشینه تشخیص اخبار جعلی

۳. تهیه دادگان «تاج»

۴. توضیح و ارزیابی مدل‌های استفاده شده

۵. بهبود دقت تشخیص اخبار جعلی

۶. جمع‌بندی و نتیجه‌گیری



۱. مقدمه

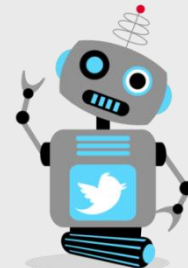
- خبر جعلی چیست؟
- اهمیت تشخیص اخبار جعلی چیست؟
- چه ابزارهای هوشمندی برای تشخیص اخبار جعلی موجود است؟
- تفاوت مسئله تشخیص اخبار جعلی با دیگر مسائل چیست؟
- راه‌حل‌های ارائه شده چیست؟

خبر جعلی چیست؟

خبر جعلی، اطلاعات ساختگی است که از ساختار محتوای اخبار رسانه‌ها تقلید می‌کند اما از لحاظ فرایند انتشار و اهداف با آن‌ها متفاوت است (لیزر و همکاران، ۲۰۱۸).

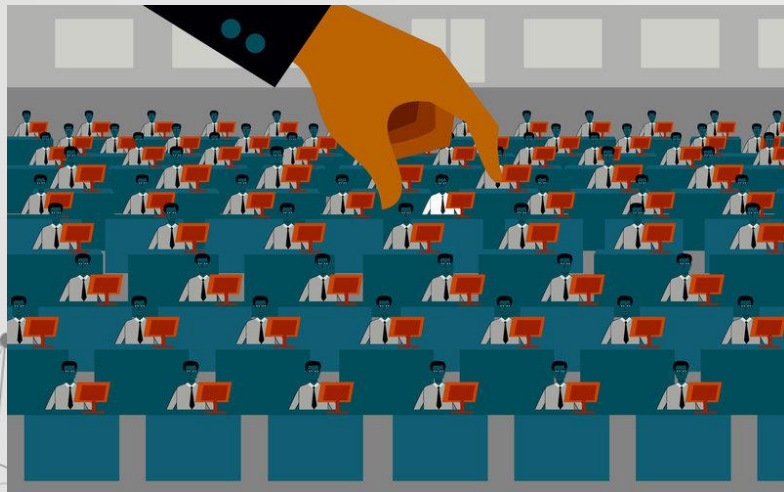
اطلاعات غلط

اطلاعات نادرستی که به اشتباه نقل شده‌اند و قصد فریب افراد را ندارند.



اطلاعات دروغ

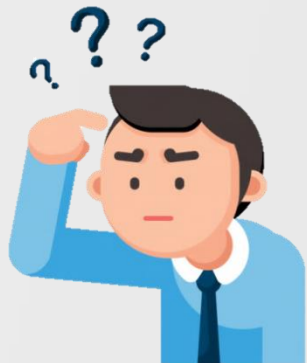
اطلاعاتی که از روی عمد و با هدف فریب افراد منتشر می‌شوند.





تأثيرات عمده انتشار اخبار جعلی

- گمراه کردن افکار عمومی
- سردرگم و مردد شدن افراد جامعه
- آثار و عواقب سیاسی، اجتماعی، فرهنگی



TRUST



- تبدیل صوت سخنرانی شخصیت‌های سیاسی به متن
- ذخیره دائمی جریان داده‌های خبری
- استفاده از سامانه تشخیص ادعا به منظور تسهیل فرایند جستجو
- استفاده از سامانه‌ای برای تصمیم‌گیری نهایی توسط روزنامه‌نگاران



بررسی و تحلیل
موارد شایع در
فضای مجازی



- بررسی شایعات در فیسبوک و دیگر شبکه‌های اجتماعی و ارائه مستندات
- مقابله با رواج **شایعه و دروغ** در فضای مجازی
- گمانه در دی ماه سال ۱۳۹۴ فیلتر شد.
- آخرین فعالیت سایت **شایعات** در ۲۷ خرداد ۱۳۹۷ انجام شده است.



مزایا و معایب سامانه‌های هوشمند

- ✗ غیر قابل استفاده در زبان‌هایی با منابع زبانی اندک (فارسی، آلمانی و ...)
- ✗ توانایی کمتر در تشخیص ادعاهای پیچیده خبری

- ✓ افزایش سرعت تشخیص و جلوگیری از انتشار اخبار جعلی
- ✓ قابلیت تحلیل و بررسی سریع گستره وسیعی از داده‌های خبری

Democratic Debate, September 2019



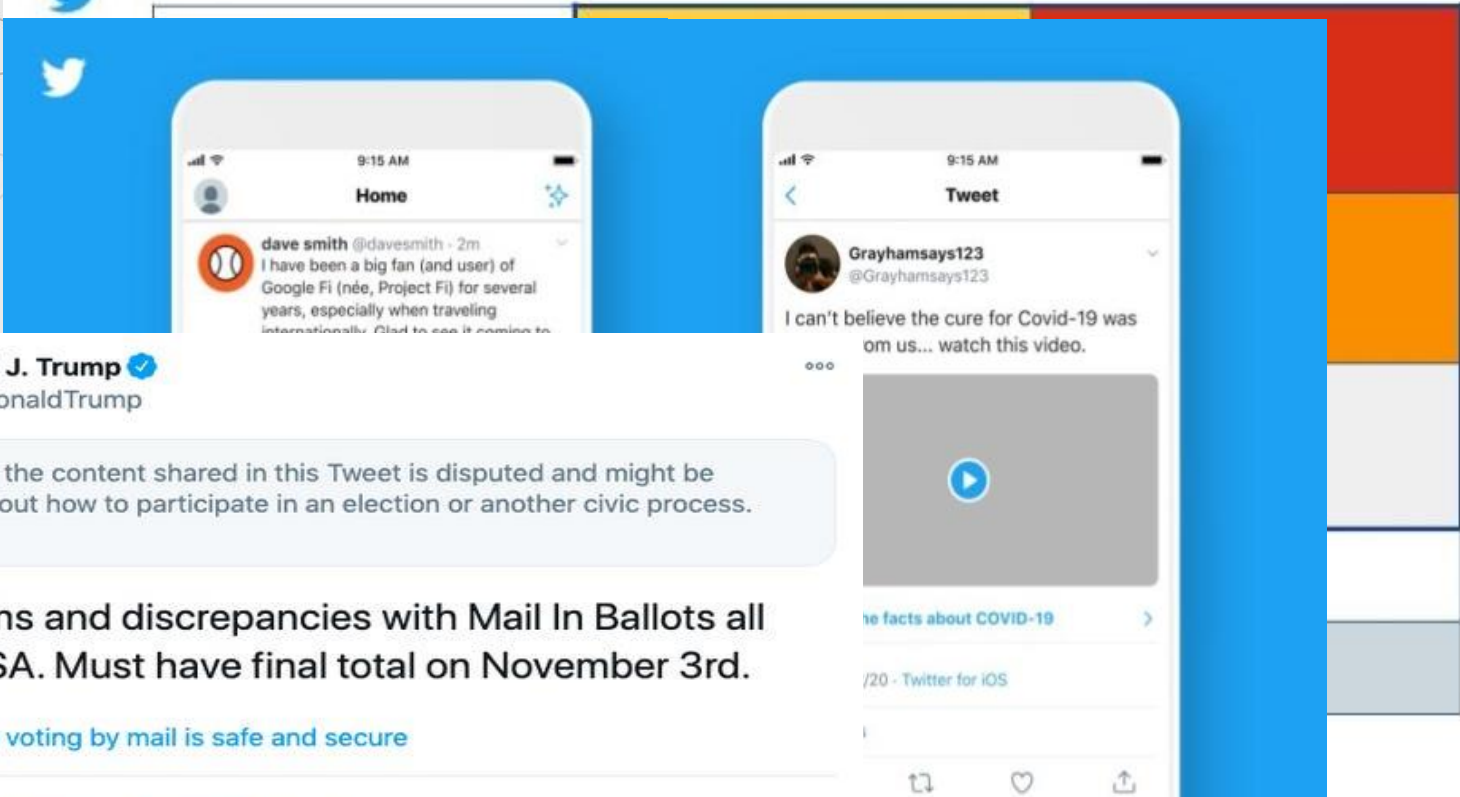
Related Fact-Check from PolitiFact — June 2019

PREVIOUS CLAIM FROM ELIZABETH WARREN
"The insurance companies last year alone sucked \$23 billion in profits out of the health care system. \$23 billion. And that doesn't count the money that was paid to executives, the money that was spent lobbying Washington."

CONCLUSION
Accurate

LIVE — NEWS | THE DEMOCRATIC DEBATE

A lesson in automated journalism:
Bring back the humans



Donald J. Trump ✓
@realDonaldTrump

Some or all of the content shared in this Tweet is disputed and might be misleading about how to participate in an election or another civic process.
[Learn more](#)

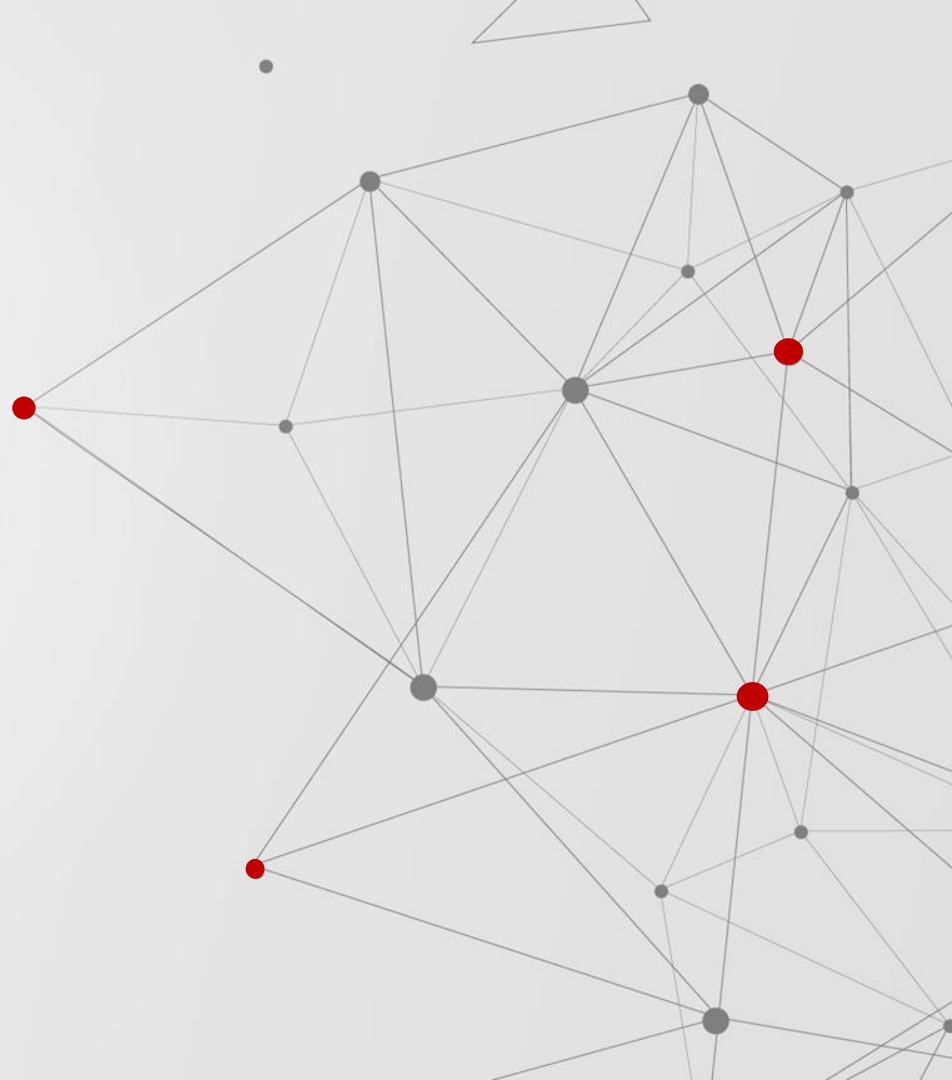
Big problems and discrepancies with Mail In Ballots all over the USA. Must have final total on November 3rd.

⚠ [Learn how voting by mail is safe and secure](#)

11:43 PM · Oct 26, 2020 · Twitter for iPhone

۲. پیشینه تشخیص اخبار جعلی

- چه کارهای مرتبطی در زبان فارسی انجام شده؟
- جزئیات دادگان اخبار جعلی فارسی چیست؟
- ضعف کارهای انجام شده چیست؟



- ارائه یک رویکرد تحلیلی برای تاثیر **ویژگی‌های ساختاری** در کنار ویژگی‌های مبتنی بر محتوا توئیت
- ارائه دادگان مربوط به **شایعات** منتشر شده در شبکه اجتماعی توئیتر شامل ۷۸۳ توئیت جعلی و ۷۸۳

توئیت اصیل از دو سایت شایعات و گمانه

- استفاده از مدل‌های سنتی یادگیری ماشین مانند: بیز ساده، ماشین بردار پشتیبان، کا-نزدیک‌ترین همسایه

و درخت تصمیم.

Feature ID	Value	definition
FR	Binary	'1' represents retweeted post and '0' represents main tweet
NU	Integer	URLs number in tweet
NH	Integer	hashtags number in tweet
NM	Integer	number of accounts mentioned in tweet
FV	Binary	'1' represents verified account, '0' represents not verified
FPP	Binary	'1' represents profile with photo, '0' represents without photo
FH	Binary	'1' represents profile with header, '0' represents without header
FL	Binary	'1' represents profile with location, '0' represents without location
FG	Binary	'1' represents geo-enabled setting, '0' represents not enabled geo setting
FP	Binary	'1' represents protected account, '0' not protected
FU	Binary	'1', account has bio, '0' does not
CP	Real	account calculated page rank
CC	Real	account calculated clustering coefficient
CM	Integer	account calculated modularity class size
NR	Integer	retweets number
NL	Integer	likes number
NS	Integer	stop words number in tweet
NT	Integer	total number of account's tweets
NFG	Integer	following number
NFR	Integer	followers number
NLS	Integer	number of account being listed
NG	Integer	account age

• بررسی توئیتهای ۱۱,۹۸۱ کاربر فارسی زبان در شبکه اجتماعی توئیتر.

• بررسی پیکره بزرگی شامل ۳,۵۹۳,۷۰۴ توئیت و ۴,۳۴۵ توئیت **شایعه فارسی** که عمدتاً درمورد زلزله

◀ کرمانشاه بوده است.

Type	Range	Description	Name
Structure feature	Binary	is this tweet a reply to another?	is_reply
Structure feature	Binary	is this a retweet or not?	is_retweet
Structure feature	Real value	the number of retweets of this tweet divide by the number of the user followers	retweet_per_follower
Structure feature	Real value	number of tweet likes divide by the number of follower users	like_per_follower
Structure feature	Real value	number of likes	number_of_likes
Structure feature	Real value	number of retweets	number_of_retweets
User feature	Real value	number of the tweet sender followers	user_followers_count
User feature	Real value	number of the tweet sender friends	user_friend_count
User feature	Real value	lifetime of the tweet sender (by day)	user_profile_age
User feature	Real value	number of all the statuses of the user who tweets	user_total_statuses_count
User feature	Binary	is this user verified?	is_verified
User feature	Real value	number of all lists which are owned by user	user_listed_count
User feature	Real value	social influence	user_si
Content feature	Binary	is this tweet contains a URL?	has_url
Content feature	Binary	is this tweet contains a mention?	has_mention
Content feature	Binary	is this tweet contains a hashtag?	has_hashtag
Content feature	Binary	is this tweet contains media files?	has_media
Content feature	[1-280]	number of tweet characters. rumor tweets usually are longer than 30 characters	tweet_length

- استفاده از رویکرد **دسته‌بندی موضع** برای تشخیص اخبار جعلی
- جمع‌آوری ۵۳۴ ادعا به همراه ۲۰۱۲۴ متن خبر درمورد آن‌ها
- استفاده از معماری حافظه کوتاه مدت طولانی پشته‌ای برای دسته‌بندی در چهار دسته: موافق، ناموافق، نامرتب، بحث‌شده
- به طور مستقیم به بررسی جعلی یا اصیل بودن اخبار نپرداخته است.

- استخراج دادگانی مرتبط با **شایعات** تلگرام شامل ۸۸۲ شایعه و ۸۸۲ پست اصیل
- استفاده از کانال خبرگزاری‌های فارس، ایرنا، ایسنا، تسنیم و مهر برای اخبار اصیل و سایت‌های شایعات، گمانه و ویکی‌هواکس برای استخراج شایعات
- تحلیل و بررسی یک تئوری به نام **قدرت شایعه** که در سال ۱۹۴۷ توسط آلپورت و پوستمن ارائه شد
- استخراج ۴۲ ویژگی مبتنی بر متن که نشان‌دهنده ارزش خبری، اهمیت، احساسات و ابهامات یک توئیت است.

$$\text{اهمیت} \times \text{ابهام} \approx \text{قدرت شایعه}$$

آمار دادگان موجود فارسی

مقاله	تعداد داده جعلی	تعداد کل داده	منبع اخبار جعلی	منبع اخبار
جهان بخش و دیگران (۲۰۲۰)	۸۸۲	۱,۷۶۴	ویکی‌هواکس/شایعات /گمانه	تلگرام
زرهران و دیگران (۲۰۱۹)	۵۳۴	۲,۱۲۴	شایعات / فیک‌نیوز	وبسایت‌ها
محمودآباد و دیگران (۲۰۱۸)	۴,۳۴۵	۳,۵۸۹,۰۴۹	شایعات	توییتر
زمانی و دیگران (۲۰۱۷)	۷۸۳	۱,۵۶۶	شایعات/گمانه	توییتر

- عدم استفاده از مدل‌های به روز پردازش زبان طبیعی مانند تعبیه‌های مبتنی بر بافت
- وابستگی زیاد به استخراج ویژگی‌های متنی
- ارائه دادگانی با حجم محدود و منحصر به شبکه‌های اجتماعی مانند توئیتر
- محدود بودن منبع استخراج دادگان شایعات موجود به ۳ سایت فارسی



۳. تهیه دادگان «تاج»

- چالش‌های اصلی برای استخراج اخبار جعلی چیست؟
- روش ارائه شده برای تهیه دادگان چیست؟
- برچسب‌گذاری اخبار به چه صورت انجام شده است؟
- جزئیات آماره‌های دادگان «تاج» چیست؟

چالش‌های جمع‌آوری اخبار جعلی



- عدم وجود منبع مستقیم برای جمع‌آوری اخبار جعلی
- نیاز به آگاهی دقیق در مورد وقایع خبری
- نیاز به دادگان متنوع خبری برای تهیه مجموعه دادگان

یک فرض ساده:
اخبار جعلی عمدتاً پس از انتشار توسط
نهادهای رسمی و یا رسانه‌های آگاه **تکذیب**
می‌شوند.

رویکرد معکوس:
با استفاده از اخبار **تکذیب‌شده** به اصل اخبار جعلی برسیم.

روش جمع‌آوری معکوس



جعلی ❌

سریال نون خ با کمک ۳ میلیارد تومانی استانداری در کرمانشاه ساخته می‌شود.

جعلی چند رسانه‌ای 🎥

(ویدئو) جشن و پایکوبی پرستاران بیمارستان رازی رشت - برخط نیوز

پرسشی ❓

ماجرای کمک مالی استانداری کرمانشاه برای ساخت «نون.خ» چیست؟ - شهرخبر

تکذیبی ⚡

عدم صحت خبر ملاقات ابتکار با زنان شاغل در سفارتخانه های خارجی

اطلاعات بیشتر ⓘ

تازه ترین جزئیات از فصل دوم سریال «نون.خ»

نامربوط ⚡

تمجید بهروز شعبی از طراحی و اجرای شوخی های «نون.خ ۲»

جزئیات دادگان «تاج»



- ۵۹۲ وبسایت متمایز
- میانگین طول اخبار : ۳۰۳ توکن
- کوتاه‌ترین خبر: ۹ توکن
- طولانی‌ترین خبر: ۷,۱۷۲ توکن
- استخراج اخبار اصیل از ۵ خبرگزاری معتبر: ایرنا، ایسنا، همشهری‌آنلاین، فارس و مهر.

دی ۱۳۸۸



مهر ۱۳۹۹



دو نمونه عنوان خبر جعلی و خبر تکذیبیه آن‌ها

۱۳۹۷/۱۲/۰۷

**کریمی قدوسی به بولتن نیوز خبر داد: استعفای
زنگنه بعد از نامه ۷۲ نماینده**

مجلس نماینده مشهد در مجلس شورای اسلامی
در گفتگویی اختصاصی با بولتن نیوز گفت؛ ظاهراً
وزیر نفت، استعفاء داده است.

۱۳۹۷/۱۲/۰۸

**وزیر نفت درباره استعفای قریب‌الوقوع خود در سخنانی،
تلویحا این شایعات را تکذیب کرد.**

به گزارش خبرگزاری موج، بیژن زنگنه، وزیر نفت صبح
امروز در حاشیه جلسه هیات دولت در جمع خبرنگاران
در پاسخ به پرسشی درباره شایعه مربوط به کناره‌گیری
اش از کابینه دولت اظهار کرد: بگردند تا برگردیم.


۱۳۹۶/۰۱/۲۰

بدهی بانکی دلیل اصلی تعطیلی «داروگر»
تعطیلی‌های دیگر درپیش است و وزارت صنعت
هشدارهای پیاپی را نشنید.

۱۳۹۶/۰۱/۲۳

خبر تعطیلی داروگر تکذیب شد

در پی انتشار اخباری مبنی بر تعطیلی کارخانه قدیمی و
پرآوازه داروگر رییس سازمان صنعت، معدن و تجارت
استان تهران با تکذیب این خبر گفت: داروگر برند
قدیمی کشور در تولید محصولات بهداشتی است و نه
تنها تعطیل نشده بلکه تولید آن افزایش نیز داشته
است.



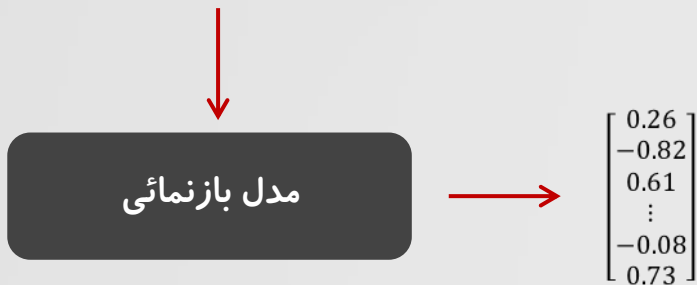
۴. توضیح و ارزیابی مدل‌های استفاده شده

- برای بازنمایی اخبار از چه رویکردی استفاده شده است؟
- چه مدل‌های پایه‌ای برای دسته‌بندی استفاده می‌شوند؟
- جزئیات معماری‌های ارائه شده چیست؟
- چه پیشرفتی روی دادگان قبلی فارسی حاصل شده است؟

بازنمائی اخبار ورودی

- بازنمایی مبتنی بر واژه
- بازنمایی مبتنی بر تعبیه ایستا (ورد۲وک)
- مدل های بازنمایی متن مبتنی بر بافت
 - برت چند زبانه
 - آلبرت-فارسی
 - ایکس.ال.ام-روبرتا
 - پارس برت

محمود احمدی نژاد کاندید انتخابات
ریاست جمهوری ۱۴۰۰ شد.



ورودی مدل برت

تعبیه واژه

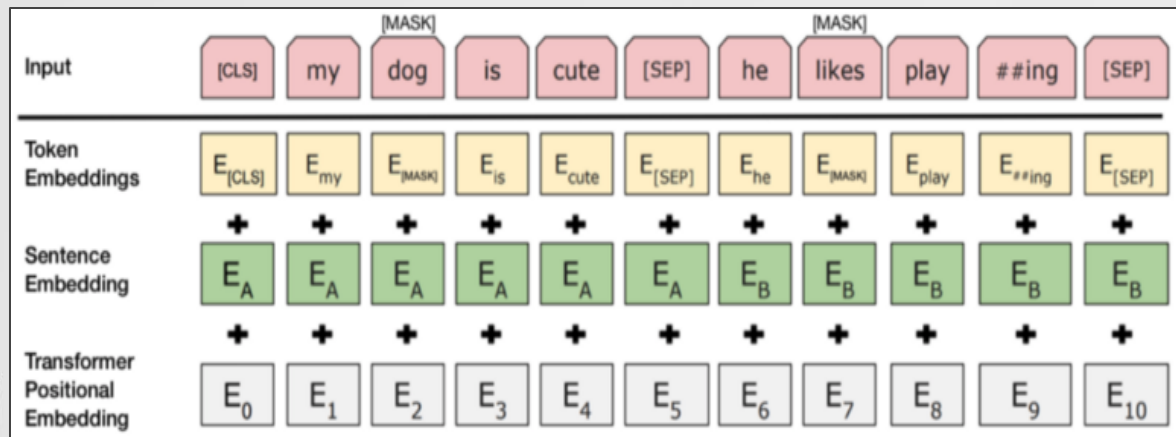
تبدیل هر کلمه به فضای برداری یکتا

تعبیه موقعیت انتقال دهنده‌ها

نشان‌دهنده موقعیت یک کلمه در متن ورودی

تعبیه جمله

هر داده ورودی مدل برت از دو جمله "آ" و "ب" تشکیل شده است که برای پیش بینی جمله بعدی کاربرد دارد.



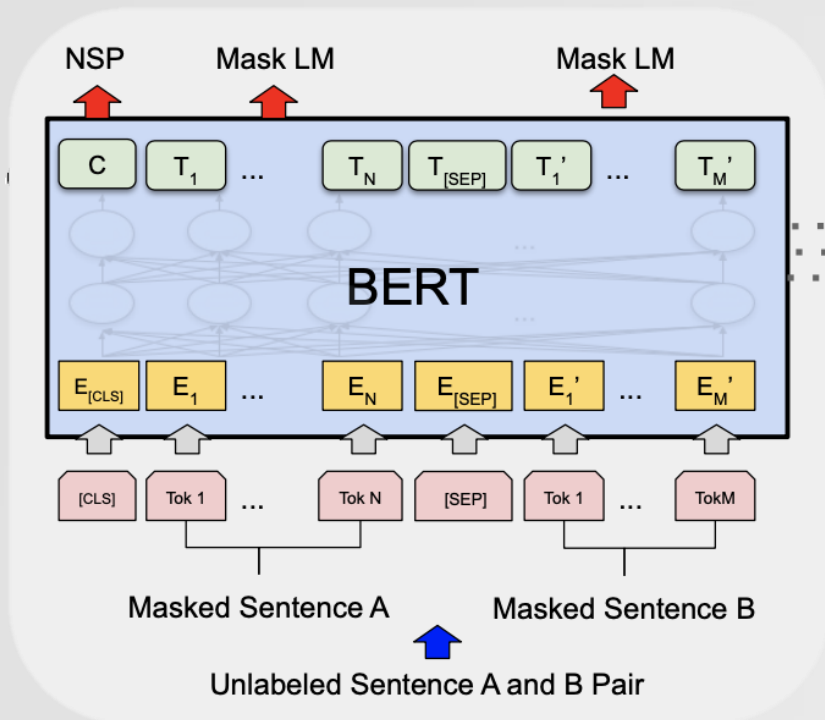
آموزش بدون نظارت

پوشش واژه‌ها

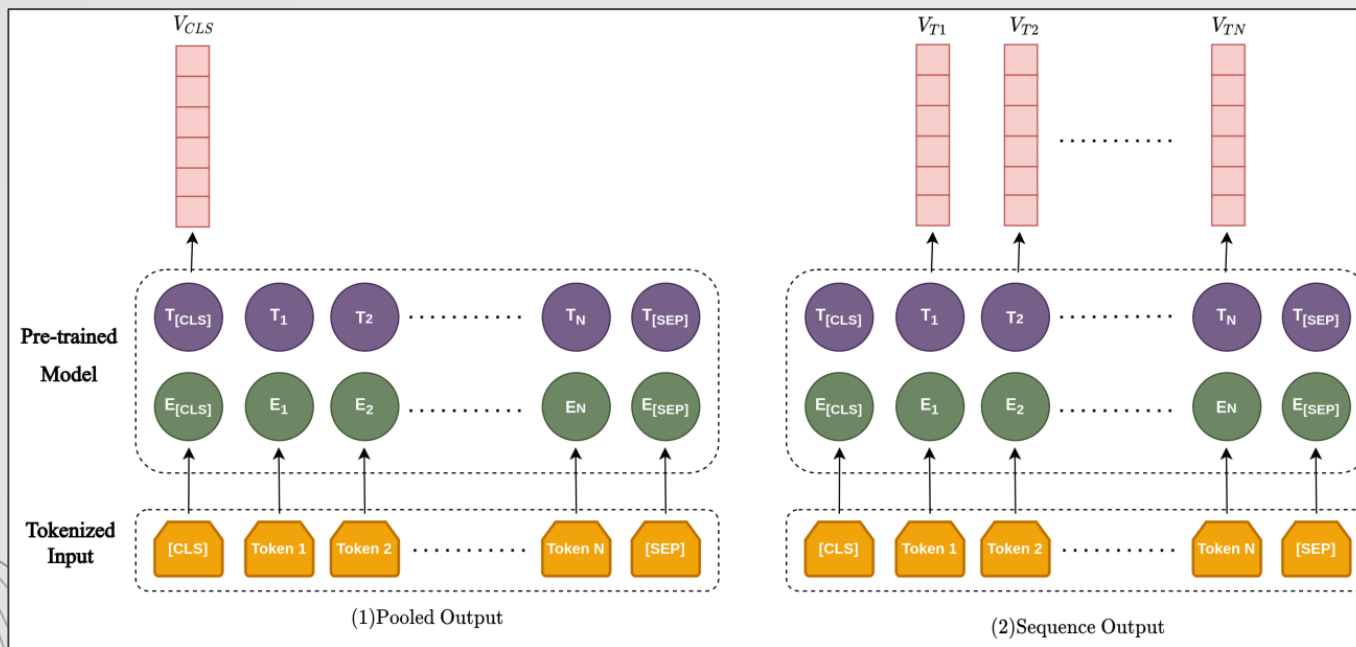
حدس کلمه پوشیده شده با استفاده از اطلاعات زمینه‌ای براساس کلمات مجاور.

پیش‌بینی جمله بعدی

پیش‌بینی اینکه آیا جمله «ب» می‌تواند جمله بعد از «آ» باشد یا نه با استفاده از اطلاعات زمینه‌ای میان عبارات و تعداد بیشتری از کلمات.



بازنمایی تجمعی

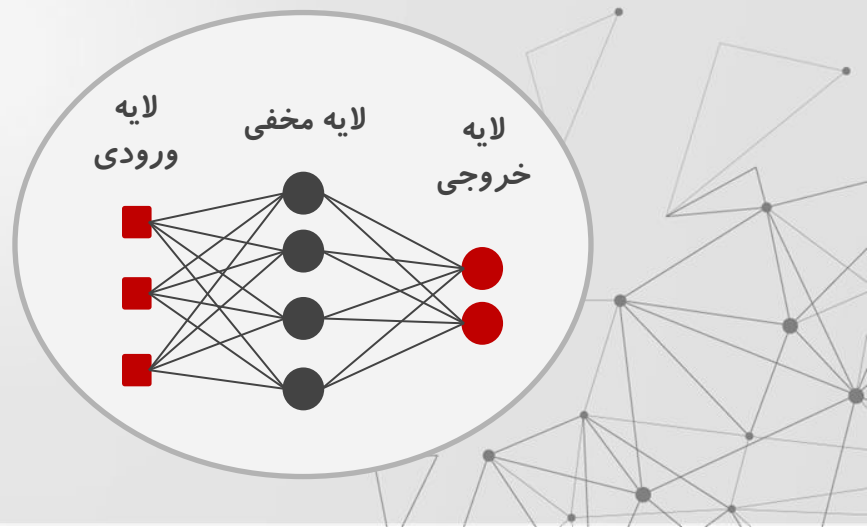
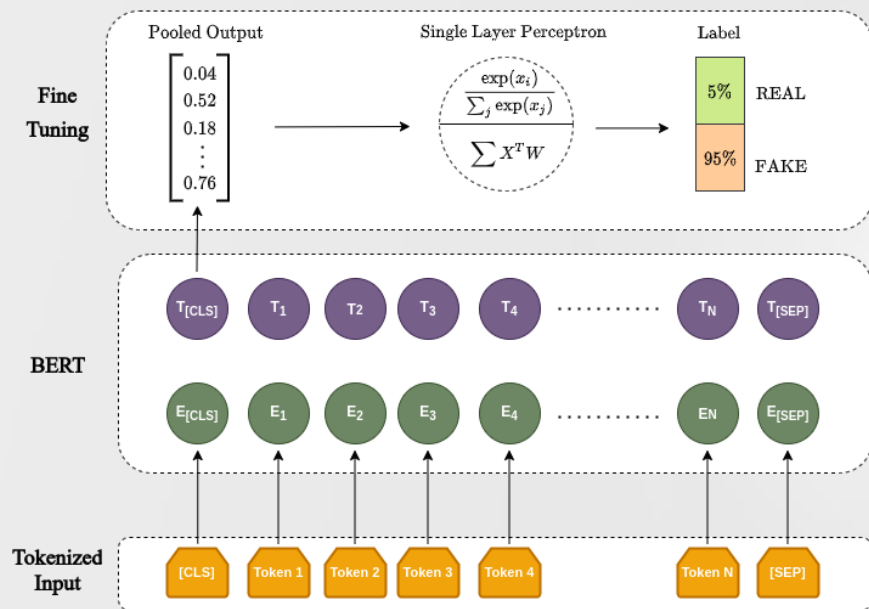


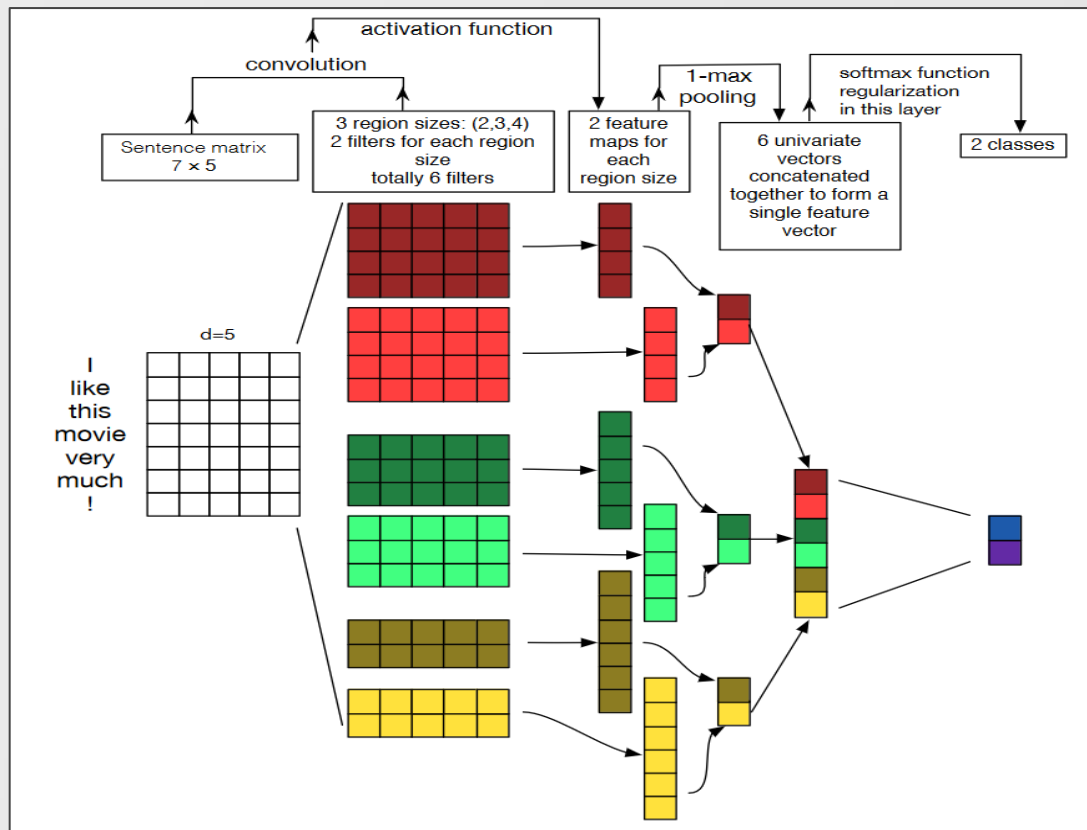
دسته‌بند پرسپترون تک لایه

- حذف تصادفی نورون‌ها در هنگام آموزش مدل به

منظور جلوگیری از «پیش‌برازش» مدل

- استفاده از تابع «بیشینه نرم» برای لایه خروجی

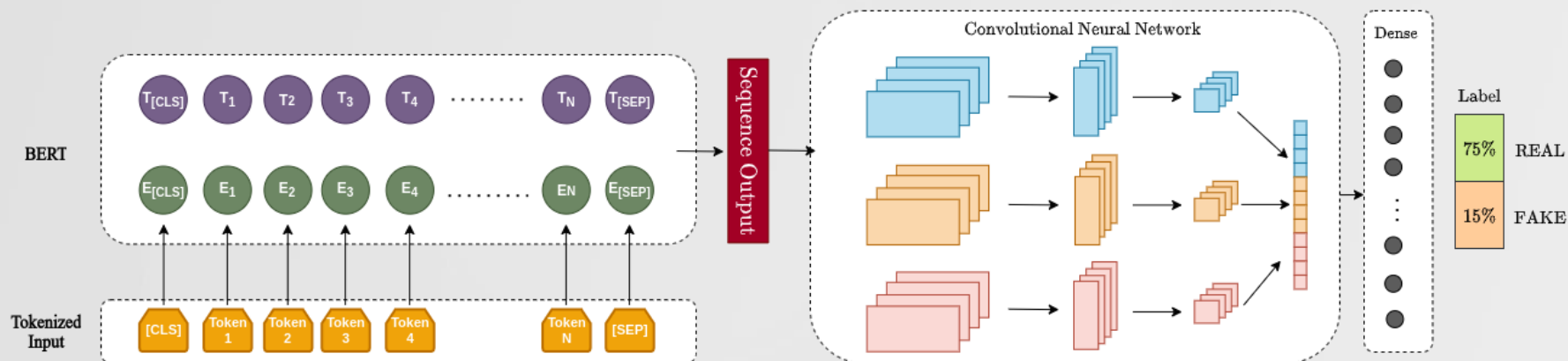




• استفاده از ۳ لایه پیچشی موازی به منظور استخراج ویژگی‌های سطح بالا براساس هم‌نشینی

واژگان متن ورودی

• در هر لایه پیچشی، ۳۰ فیلتر با اندازه هسته ۳، ۴ و ۵ استفاده شده است.



نتایج تشخیص اخبار جعلی فارسی با استفاده از بازنمایی های مختلف و دسته بندی با شبکه عصبی پیچشی

مدل	فراخوانی	صحت	معیار اف ۱	دقت
برت چند زبانه	۸۸.۵۵	۸۴	۸۶.۲۱	۸۷.۳۳
آلبرت-فارسی	۸۷.۰۲	۹۲	۸۹.۴۴	۸۹.۷۵
پارس برت	۸۹.۱۳	۹۳.۷۱	۹۱.۳۶	۹۱.۶۴
ایکس.ال.ام-روبرتا	۸۷.۸۴	۹۰.۸۵	۸۹.۳۲	۸۹.۷۵
ورد ۲وک	۷۷.۳۴	۸۰	۷۸.۶۵	۷۹.۵۱

نتایج تشخیص اخبار جعلی فارسی با استفاده از بازنمایی پارس برت و مقایسه دسته بندها

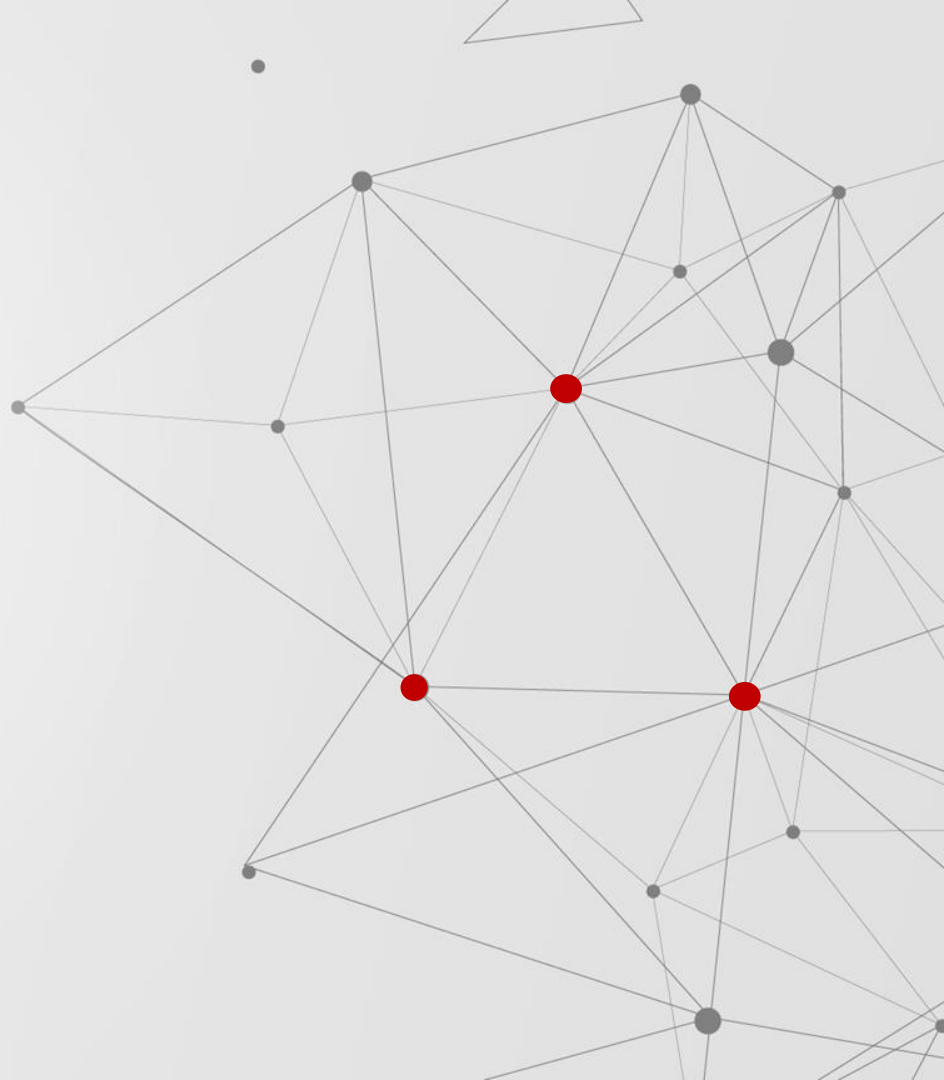
دسته بند	فراخوانی	صحت	معیار اف ۱	دقت
پیچشی	۸۹,۱۳	۹۳,۷۱	۹۱,۳۶	۹۱,۶۴
پرسپترون تک لایه	۹۲,۴۴	۹۰,۸۵	۹۱,۶۴	۹۲,۱۸

ارزیابی و مقایسه مدل های ارائه شده با مدل های پیشین بر روی دادگان فارسی شبکه های اجتماعی

مدل	دادگان	فراخوانی	صحت	معیار اف ۱	دقت
ماشین بردار پشتیبان (زمانی و دیگران، ۲۰۱۷)	توئیتر	۷۱،۴۴	۸۱،۲۴	۷۶،۰۲	۷۴،۲۶ ^{+۳۱،۵۶}
ماشین بردار پشتیبان (جهان بخش و دیگران، ۲۰۲۰)		۷۶،۳	۷۶،۳	۷۶،۳ ^{+۲۰،۱۰}	–
پارس برت-پیچشی		۹۵،۷۱	۹۷،۱۰	۹۶،۴۰	۹۶،۱۲
پارس برت-پرسپترون		۹۴،۲۰	۹۴،۲۰	۹۴،۲۰	۹۳،۷۹
ماشین بردار پشتیبان (جهان بخش و دیگران، ۲۰۲۰)	تلگرام	۸۲،۸	۸۲،۹	۸۲،۸ ^{+۱۱،۳۶}	–
پارس برت-پیچشی		۹۳،۴۹	۹۵،۰۴	۹۴،۲۶	۹۲،۵۱
پارس برت-پرسپترون		۹۲،۰۶	۹۵،۸۶	۹۳،۹۲	۹۱،۹۷

۵. بهبود مدل تشخیص اخبار جعلی

- چرا فراداده‌ها اهمیت زیادی دارند؟
- چه فراداده‌هایی در این پروژه استفاده شده است؟
- بهبود نتایج با استفاده از این فراداده‌ها چه مقدار بوده است؟



فراداده‌های مبتنی بر متن خبر



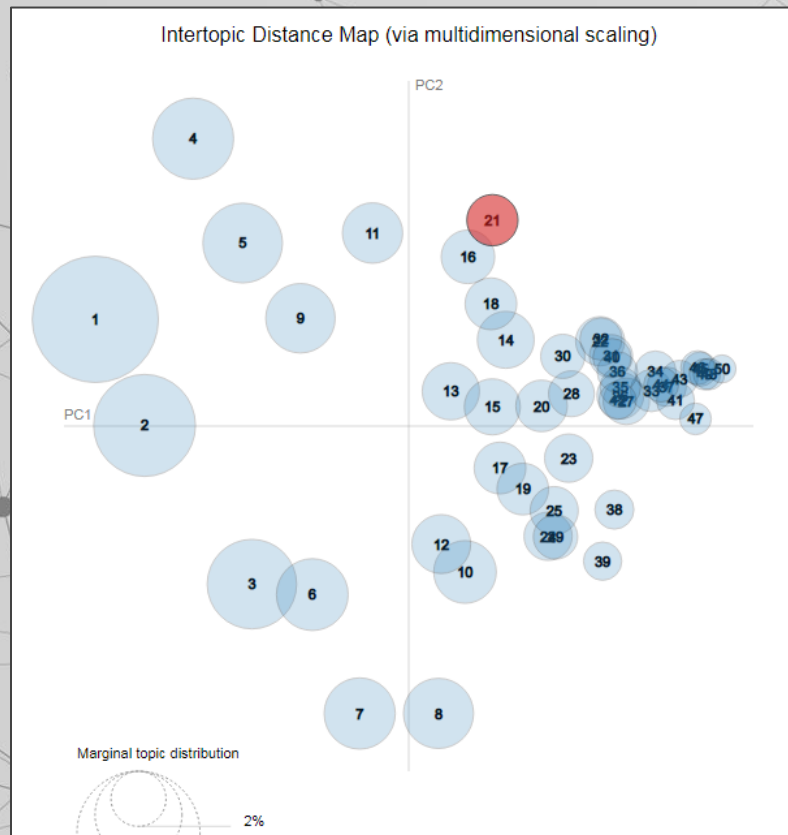
احساس خبر



موضوع خبر (دسته‌بندی)

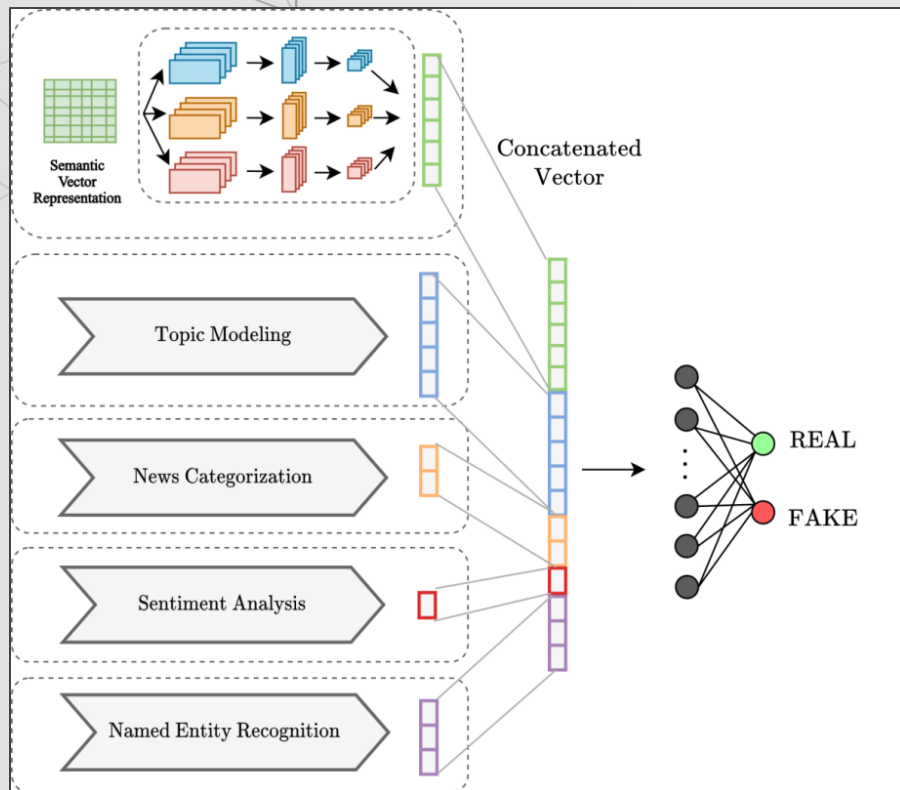


اسامی نامدار



موضوع خبر (تخصیص نهان دریکله)

دسته‌بندی اخبار با استفاده از فراداده‌ها



- موضوع خبر (تخصیص نهان دریگله): یک بردار ۵۰ بعدی شامل زیرموضوعات هر خبر
- موضوع خبر (دسته‌بندی): یک بردار ۶ بعدی
- تک‌روشن نماینده موضوع خبر
- احساس خبر: یک بردار ۲ بعدی تک‌روشن
- اسامی نامدار: یک بردار ۱۵ بعدی شامل تعداد اسامی خاص موجود در متن خبر

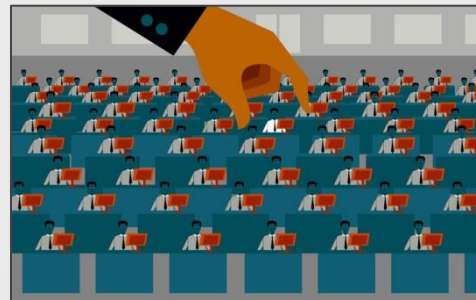
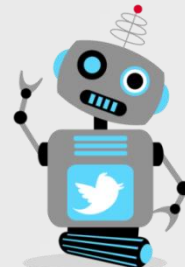
نتایج تشخیص اخبار جعلی فارسی با استفاده از بازنمایی پارس برت و دسته بند پیچشی

فرداده	فراخوانی	صحت	معیار اف ۱	دقت
-	۸۹,۱۳	۹۳,۷۱	۹۱,۳۶ +۱,۷۴	۹۱,۶۴ +۱,۵۹
موجودیت‌های نامدار	۹۱,۱۶	۹۴,۲۸	۹۲,۶۹	۹۲,۹۹
احساس خبر	۹۱,۵۲	۹۲,۵۷	۹۲,۰۴	۹۲,۴۵
موضوع (دسته‌بندی)	۹۶,۸۷	۸۸,۵۷	۹۲,۵۳	۹۳,۲۶
موضوع (تخصیص نهان دیریکله)	۹۱,۴۲	۹۱,۴۲	۹۱,۴۲	۹۱,۹۱
تمام ویژگی‌ها	۹۳,۶۴	۹۲,۵۷	۹۳,۱۰	۹۳,۵۳

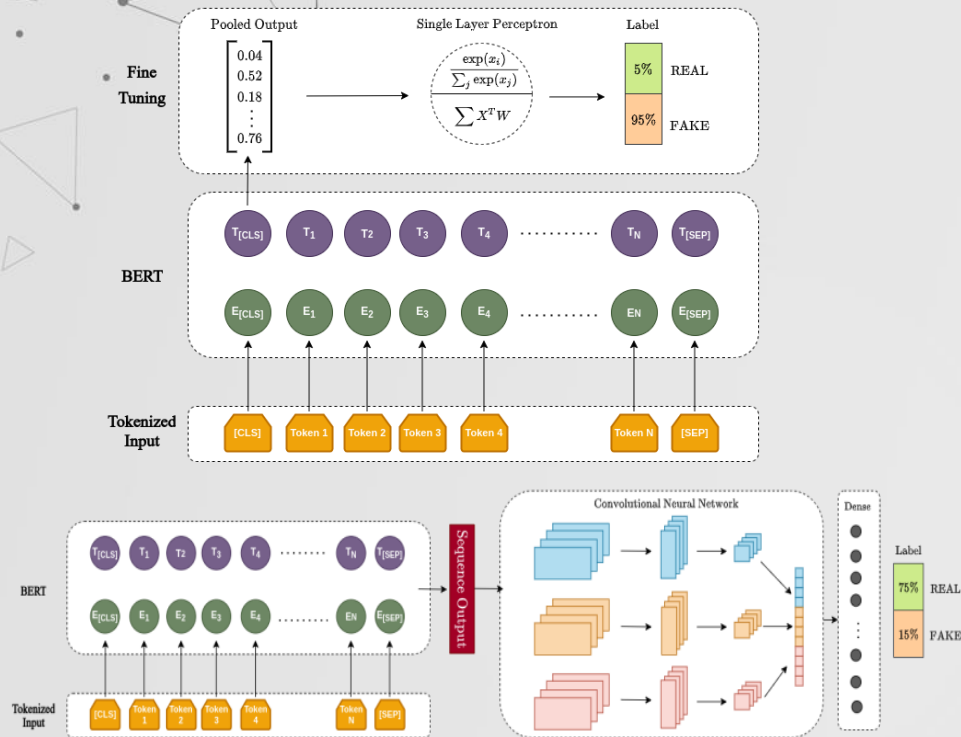
۶. جمع بندی



خبر جعلی، اطلاعات ساختگی است که از ساختار محتوای اخبار رسانه‌ها تقلید می‌کند اما از لحاظ فرایند انتشار و اهداف با آن‌ها متفاوت است (لیزر و همکاران، ۲۰۱۸).



Misleading Information	Label	Removal
Disputed Claim	Label	Warning
Unverified Claim	No action	No action*
	Moderate	Severe
Propensity for Harm		



۵۹۲ وبسایت متمایز

میانگین طول اخبار: ۳۰۳ توکن

کوتاه‌ترین خبر: ۹ توکن

طولانی‌ترین خبر: ۷,۱۷۲ توکن

استخراج اخبار اصیل از ۵ خبرگزاری معتبر: ایرنا،

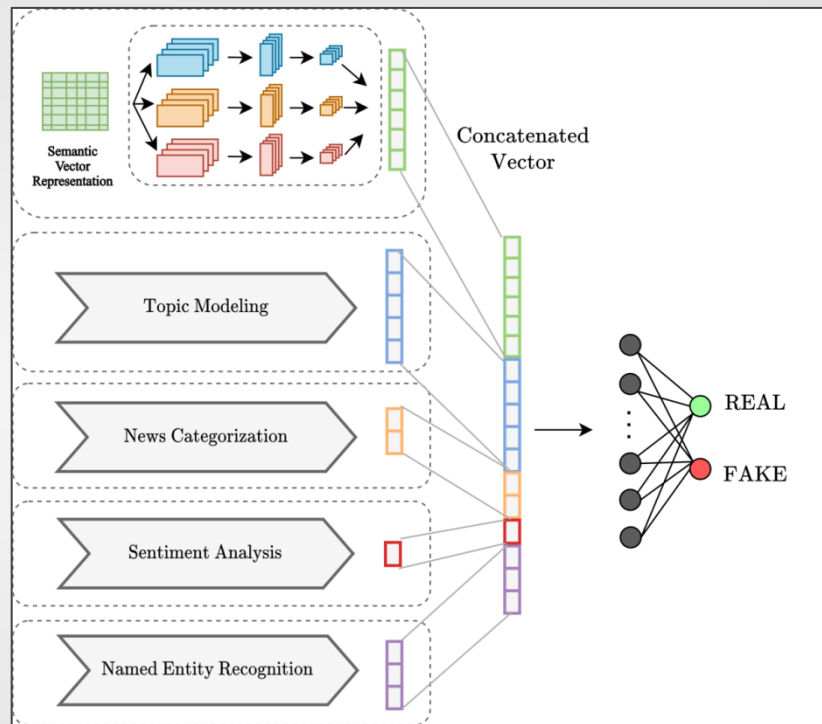
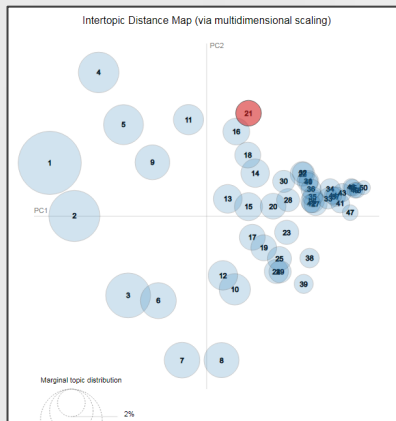
ایسنا، همشهری آنلاین، فارس، مهر.

دی ۱۳۸۸



مهر ۱۳۹۹





با تشکر از **توجه** شما

