



[❖](#_heading=h.ipnfr4lfg4r7) مقدمه 3

[∙](#_heading=h.4vdnfg9g85gq) تعریف استرس و اثرات آن بر سلامت روان و رفتار انسان 3

[∙](#_heading=h.rztdphuxveci) اهمیت تشخیص استرس در شبکه‌های اجتماعی 4

[∙](#_heading=h.lnjcpskweh6g) نقش هوش مصنوعی در تحلیل زبان و تشخیص استرس 5

[❖](#_heading=h.8cycthb8q6zx) جمع‌آوری و آماده‌سازی داده‌ها 8

[∙](#_heading=h.l6x0mix8udh4) روش‌های جمع‌آوری داده‌های آموزشی برای تشخیص استرس 8

[1. جمع‌آوری داده‌های رسانه‌های اجتماعی 8](#_heading=h.4k8kufuyapvg)

[2. پرسشنامه‌ها و نظرسنجی‌ها 8](#_heading=h.c7gaqex67hcu)

[3. جمع‌آوری داده از چت‌ها و پیام‌ها 9](#_heading=h.myorfoq7xr0h)

[4. آزمایش‌های کنترل‌شده 9](#_heading=h.5cemzmp7lovi)

[5. تکنیک‌های پردازش زبان طبیعی 9](#_heading=h.jfunqf7xegzf)

[6. استفاده از تطبیق عبارات منظم 9](#_heading=h.1tioufq7lzb)

[7. برچسب‌گذاری دستی 10](#_heading=h.84d5kyzgsq0t)

[8. جمع‌آوری داده‌های چندوجهی 10](#_heading=h.ntdjyd70t93v)

[∙](#_heading=h.fqwkwn92pa89) معرفی مجموعه داده‌های موجود برای تشخیص استرس 10

[مجموع داده ردیت 10](#_heading=h.iqbt3k4hqf87)

[مجموعه داده TensiStrength 14](#_heading=h.vgy4p12609kp)

[مجموعه‌داده DAIC-WOZ 16](#_heading=h.eed2cs1edv6)

[∙](#_heading=h.uubp7ijoqmbf) شناسایی عبارات مرتبط با استرس از پرسشنامه‌ها 18

[∙](#_heading=h.nh694zza9lu) گسترش به زبان محاوره و عبارات متداول توییتر 24

[∙](#_heading=h.z2u38gd4von7) قطبیت زبانی: تعریف معیارهای عبارات استرس‌زا و آرامش‌بخش 25

[تعریف معیارهای عبارات استرس‌زا 25](#_heading=h.a6htb9d8hxp3)

[تعریف معیارهای عبارات آرامش‌بخش 25](#_heading=h.cbqhg6f6iprr)

[اهمیت قطبیت زبانی در تشخیص استرس 26](#_heading=h.pr7e7vh4e3a7)

[∙](#_heading=h.6vlen1c01fp) فرایند جمع‌آوری و دسته‌بندی 6 میلیون داده 26

[❖](#_heading=h.mnbixk98q3ew) تحلیل داده‌ها 27

[∙](#_heading=h.spzif266x5rg) ساختار جملات دارای استرس 27

[∙](#_heading=h.yhbc0qx88dz6) تحلیل زبانی کلمات پرتکرار در توییت‌های استرس‌زا 29

[روش تحلیل 29](#_heading=h.3phstryi4kn5)

[یافته‌ها 29](#_heading=h.17mpse12wdwm)

[❖](#_heading=h.nh9c29k3dsj6) مدل‌سازی و تشخیص استرس 32

[∙](#_heading=h.m42zpco5l2wg) معرفی مدل‌های کلاسیک و یادگیری عمیق 32

[∙](#_heading=h.v0e2342g7qlc) استخراج ویژگی‌های متنی و مدل‌سازی تشخیص استرس 34

[فیلتر و پالایش واژگان 34](#_heading=h.9396gvyh8cxd)

[∙](#_heading=h.gf050b2tq9eu) مدل‌سازی با استفاده از رگرسیون لجستیک 34

[نتایج ارزیابی 35](#_heading=h.e0ojlldpsptn)

[∙](#_heading=h.uoq09k5twk65) طرز کار مدل فست‌تکست برای تشخیص استرس در متن‌ها 35

[∙](#_heading=h.fxzq617kjdwx) استفاده از مدل BERT برای تشخیص استرس در متون فارسی 36

[∙](#_heading=h.vvzuk13w9yb5) استفاده از مدل SVM برای تشخیص استرس در متون فارسی 38

[∙](#_heading=h.hnjtbarr1mko) استفاده از روش Ensemble و ترکیب مدل‌ها برای تشخیص استرس در متون فارسی 38

[❖](#_heading=h.ofpjqvb5vy0c) نتیجه‌گیری و مسیرهای آینده 39

[∙](#_heading=h.vwz3b4l3jy27) جمع‌بندی اهمیت موضوع و دستاوردها 39

[∙](#_heading=h.rh5bg7y1hrd) پیشنهادهایی برای تحقیقات آینده و بهبود روش‌ها 40

[❖](#_heading=h.e8ss6aazbpfj) منابع 42

# مقدمه

## تعریف استرس و اثرات آن بر سلامت روان و رفتار انسان

استرس یک واکنش طبیعی بدن و ذهن به فشارها و چالش‌های زندگی است که زمانی رخ می‌دهد که افراد احساس می‌کنند منابع یا توانایی‌های آن‌ها برای مقابله با موقعیت‌های پیش رو کافی نیست. این واکنش شامل جنبه‌های روانی، جسمانی و رفتاری است که به‌منظور حفظ بقا و بازگرداندن تعادل در شرایط سخت فعال می‌شود. اگرچه استرس می‌تواند در کوتاه‌مدت مفید باشد و فرد را برای مواجهه با چالش‌ها آماده کند، اما در صورتی که به شکل مزمن ادامه یابد یا به‌درستی مدیریت نشود، اثرات منفی جدی بر سلامت روان و جسم خواهد داشت. [1]

استرس از طریق فعال‌سازی سیستم عصبی خودمختار و محور هیپوتالاموس-هیپوفیز-آدرنال (HPA)، تغییرات فیزیولوژیکی متعددی را در بدن ایجاد می‌کند. در کوتاه‌مدت، این تغییرات می‌توانند فرد را هوشیارتر و آماده‌تر کنند. بااین‌حال، در صورت تداوم استرس، بدن در معرض ترشح مداوم هورمون‌هایی مانند کورتیزول قرار می‌گیرد که این امر می‌تواند به افزایش خطر بیماری‌های جسمانی مانند فشار خون بالا، بیماری‌های قلبی و ضعف سیستم ایمنی منجر شود. همچنین، استرس مداوم ممکن است باعث اختلالات خواب، مشکلات گوارشی و دردهای عضلانی شود. [2]

در زمینه سلامت روان، استرس نقش قابل‌توجهی در بروز اختلالاتی مانند اضطراب و افسردگی دارد. افرادی که استرس را تجربه می‌کنند، اغلب احساس خستگی روانی و کاهش انگیزه را گزارش می‌دهند. سبک‌های مقابله‌ای افراد می‌تواند اثرات استرس را تعدیل کند. افرادی که از شیوه‌های سالم مقابله مانند حل مسئله، استفاده از حمایت اجتماعی و تمرین‌های آرامش‌بخش بهره می‌برند، معمولاً کمتر تحت تأثیر منفی استرس قرار می‌گیرند. در مقابل، سبک‌های ناسالم مانند اجتناب، انکار یا سوءمصرف مواد، اثرات زیانباری بر سلامت روان و رفتار افراد دارند.

از نظر رفتاری، استرس می‌تواند باعث کاهش تمرکز، اختلال در تصمیم‌گیری و افزایش رفتارهای پرخطر شود. در محیط‌های شغلی، استرس مزمن ممکن است به کاهش بهره‌وری، افزایش فرسودگی شغلی و ایجاد تعارضات اجتماعی منجر شود. در زندگی روزمره نیز افرادی که قادر به مدیریت استرس نیستند، معمولاً تعاملات اجتماعی کمتری دارند و بیشتر دچار انزوا می‌شوند. [3]

به‌طورکلی، استرس بخشی جدایی‌ناپذیر از زندگی انسان‌هاست، اما نحوه برخورد با آن تأثیر بسزایی در پیامدهای آن دارد. آموزش و تمرین مهارت‌های مقابله‌ای می‌تواند به افراد کمک کند تا تأثیرات منفی استرس را به حداقل برسانند و سلامت روان و جسم خود را حفظ کنند.

## اهمیت تشخیص استرس در شبکه‌های اجتماعی

شبکه‌های اجتماعی به بخش مهمی از زندگی روزمره افراد تبدیل شده‌اند و بستری فراهم کرده‌اند که افراد در آن احساسات، تجربیات و افکار خود را به اشتراک می‌گذارند. با توجه به این که حجم زیادی از اطلاعات شخصی و اجتماعی در این فضا تولید می‌شود، شبکه‌های اجتماعی به یکی از منابع ارزشمند برای تحلیل وضعیت روانی افراد، از جمله استرس، تبدیل شده‌اند. تشخیص استرس در این فضا اهمیت زیادی دارد، چرا که می‌تواند در پیشگیری، درمان و حتی سیاست‌گذاری‌های کلان تأثیر بسزایی داشته باشد.

یکی از دلایل اصلی اهمیت تشخیص استرس در شبکه‌های اجتماعی، توانایی آن‌ها **در شناسایی علائم اولیه استرس** است [4]. بسیاری از افراد ممکن است استرس خود را در تعاملات روزمره یا جلسات مشاوره آشکار نکنند، اما در فضای مجازی به دلیل حس ناشناس بودن یا راحتی در بیان، احساسات خود را به‌صورت صریح‌تری بیان می‌کنند. تحلیل این داده‌ها می‌تواند به متخصصان سلامت روان کمک کند تا افراد در معرض خطر را شناسایی و به آن‌ها کمک کنند.

علاوه بر این، تحلیل استرس در شبکه‌های اجتماعی در مقیاس وسیع می‌تواند بینش‌های اجتماعی و جمعیت‌شناختی ارزشمندی ارائه دهد. این تحلیل‌ها می‌توانند درک بهتری از عوامل استرس‌زا در جوامع مختلف، تأثیر رویدادهای اجتماعی یا سیاسی بر روان جامعه، و روندهای کلی مرتبط با سلامت روان ارائه کنند. به عنوان مثال، در دوران همه‌گیری کووید-19، تحلیل شبکه‌های اجتماعی نشان داد که اضطراب و استرس در سطح جهانی به‌شدت افزایش یافته است [5].

در زمینه جنگ شناختی نیز تشخیص استرس در شبکه‌های اجتماعی نقش کلیدی دارد. دشمنان در جنگ‌های شناختی می‌توانند از اطلاعات روان‌شناختی کاربران برای گسترش استرس و اضطراب در جوامع هدف استفاده کنند. شناسایی زودهنگام این فعالیت‌ها از طریق تحلیل استرس در شبکه‌های اجتماعی می‌تواند به مقابله با این تهدیدات کمک کند.

تشخیص استرس در شبکه‌های اجتماعی همچنین می‌تواند برای طراحی ابزارهای هوش مصنوعی و سیستم‌های پیش‌بینی سلامت روان مورد استفاده قرار گیرد. مدل‌های پیشرفته‌ای که بر پایه یادگیری ماشین و تحلیل زبان طبیعی توسعه یافته‌اند، می‌توانند به تحلیل خودکار داده‌ها و شناسایی الگوهای مرتبط با استرس کمک کنند. این سیستم‌ها در نهایت می‌توانند به عنوان ابزارهای پشتیبانی تصمیم‌گیری برای متخصصان سلامت روان عمل کنند.

در مجموع، تشخیص استرس در شبکه‌های اجتماعی نه تنها به درک بهتر سلامت روان افراد و جوامع کمک می‌کند، بلکه ابزار ارزشمندی برای مداخله‌های پیشگیرانه، ارتقای کیفیت زندگی و حتی مقابله با تهدیدات شناختی است. با توجه به پیچیدگی‌های موجود در تحلیل داده‌های شبکه‌های اجتماعی، بهره‌گیری از تکنیک‌های هوش مصنوعی می‌تواند به تشخیص دقیق‌تر و سریع‌تر استرس کمک کند.

## نقش هوش مصنوعی در تحلیل زبان و تشخیص استرس

هوش مصنوعی ابزار قدرتمندی برای تحلیل زبان و استخراج اطلاعات روان‌شناختی از متن است. در زمینه تشخیص استرس، هوش مصنوعی با تحلیل نوشته‌های کاربران در شبکه‌های اجتماعی، امکان شناسایی خودکار نشانه‌های استرس به منظور ارائه راهکارهای مداخله‌ای را فراهم می‌کند. این سیستم‌ها با استفاده از مدل‌های پردازش زبان طبیعی قادرند تا الگوهای زبانی مرتبط با استرس را شناسایی کرده و به تشخیص اضطراب و تنش‌های روانی کمک کنند.

یکی از ویژگی‌های رایج در افراد تحت استرس، استفاده از کلمات و عبارات منفی و ضمایر اول شخص مفرد است. برای مثال، عباراتی مانند «استرس دارم»، «مضطربم»، «نمی‌دانم چه کنم» و «نمی‌توانم تحمل کنم» به‌طور مکرر در نوشته‌های افراد مضطرب دیده می‌شوند. این نشانه‌ها می‌توانند به‌عنوان شاخص‌های اولیه استرس شناسایی شوند. مدل‌های پردازش زبان طبیعی با تجزیه و تحلیل دقیق این ویژگی‌ها می‌توانند به‌سرعت و دقت بالا تغییرات احساسی در متن را شناسایی کنند و آن‌ها را به عنوان نشانه‌هایی از استرس در نظر بگیرند.

در مراحل پیشرفته‌تر، مدل‌های یادگیری ماشین و یادگیری عمیق وارد عمل می‌شوند. این مدل‌ها با آموزش بر روی مجموعه‌های بزرگی از داده‌های متنی، قادر به شناسایی الگوهای پیچیده‌تری هستند که ممکن است برای انسان آشکار نباشد. به‌ویژه مدل‌های مبتنی بر شبکه‌های عصبی عمیق به دلیل توانایی خود در مدل‌سازی روابط زمانی در متن، می‌توانند تغییرات ظریف در نحوه بیان افراد در طول زمان را شناسایی کرده و به تحلیل دقیق‌تری از وضعیت روانی افراد برسند.

مدل‌های مبتنی بر ترنسفورمر، به‌ویژه مدل‌هایی مانند BERT و GPT، به‌دلیل توانایی درک زمینه کلمات در متن، انقلاب بزرگی در تحلیل زبان ایجاد کرده‌اند. این مدل‌ها می‌توانند متن‌ها را در سطح جمله و پاراگراف تحلیل کرده و معنای دقیق‌تری از نوشته‌ها استخراج کنند. برای تشخیص استرس، این مدل‌ها قادرند ترکیب‌های خاص کلمات، تغییرات احساسی در متن، و حتی لحن کلی نوشته‌ها را شناسایی کنند.

هوش مصنوعی همچنین می‌تواند در شخصی‌سازی تحلیل‌ها نقش داشته باشد. به‌عنوان مثال، الگوریتم‌ها می‌توانند رفتار زبانی خاص هر کاربر را یاد بگیرند و تغییرات غیرعادی در الگوهای زبانی فردی را به‌عنوان نشانه‌ای از استرس شناسایی کنند. این قابلیت می‌تواند در توسعه سیستم‌های نظارت بر سلامت روان شخصی مفید واقع شود و به متخصصان روان‌شناسی کمک کند تا به مداخله‌های به‌موقع و مؤثر دست یابند.

علاوه بر تشخیص استرس، هوش مصنوعی می‌تواند ابزارهای کاربردی برای مداخله‌های روان‌شناختی فراهم کند. به‌عنوان مثال، چت‌بات‌های مبتنی بر هوش مصنوعی می‌توانند به افرادی که علائم استرس نشان می‌دهند، پیشنهادات و راهکارهایی ارائه دهند. این چت‌بات‌ها می‌توانند به‌صورت خودکار به گفتگوهای حمایتی بپردازند، تکنیک‌های آرام‌سازی را پیشنهاد دهند، یا کاربران را به منابع معتبر ارجاع دهند. علاوه بر این، با استفاده از هوش مصنوعی می‌توان به ارائه آگاهی‌های آموزشی در مورد سبک‌های مقابله با استرس و روش‌های بهبود سلامت روان نیز پرداخته شود.

در سطح اجتماعی، هوش مصنوعی به تحلیل و پایش روندهای جمعیتی استرس کمک می‌کند. تحلیل داده‌های شبکه‌های اجتماعی می‌تواند الگوهای گسترده‌تری از استرس در جمعیت‌ها را آشکار کند. به‌عنوان مثال، در دوران بحران‌های اقتصادی یا رویدادهای اجتماعی، ممکن است استرس در افراد افزایش یابد. تحلیل این داده‌ها می‌تواند به سیاست‌گذاران و متخصصان سلامت عمومی در طراحی برنامه‌های پیشگیری و کاهش استرس کمک کند و از این طریق، مداخلات مؤثری برای بهبود وضعیت اجتماعی و روانی جوامع ایجاد شود.

در نهایت، هوش مصنوعی نه تنها ابزاری برای شناسایی و تحلیل استرس است، بلکه پتانسیل ارائه راهکارهایی برای بهبود سلامت روان در مقیاس فردی و اجتماعی را نیز دارد. با ادامه پیشرفت در فناوری‌های هوش مصنوعی، این ابزارها نقشی کلیدی در درک و مدیریت استرس ایفا خواهند کرد و به کمک آن‌ها می‌توان اقدامات مؤثری در راستای بهبود کیفیت زندگی افراد و جوامع انجام داد.

## تاثیر تشخیص استرس بر مدیریت بحران و جنگ شناختی

تشخیص استرس در مواقع بحران و جنگ، نقش بسیار مهمی در مدیریت بحران و دفاع شناختی دارد. استرس ناشی از بحران‌های اجتماعی، اقتصادی یا سیاسی می‌تواند بر سلامت روانی افراد تأثیر بگذارد و حتی باعث کاهش کارایی و تصمیم‌گیری در سطح فردی و جمعی شود. در چنین شرایطی، تشخیص سریع و دقیق استرس می‌تواند به تصمیم‌گیرندگان کمک کند تا اقداماتی مؤثر برای کاهش تنش‌ها و بحران‌ها انجام دهند و از گسترش مشکلات روانی جلوگیری کنند.

در مواقع جنگ یا بحران‌های اجتماعی، افراد و گروه‌ها ممکن است به‌طور ویژه‌ای در معرض استرس‌های شدید قرار گیرند. استرس ناشی از تهدیدات فیزیکی، روانی یا اجتماعی می‌تواند باعث رفتارهای غیرقابل پیش‌بینی، تصمیم‌گیری‌های اشتباه، و کاهش توانایی‌های فردی و گروهی شود. به‌ویژه در زمینه جنگ شناختی، که هدف آن تضعیف اراده و روحیه افراد است، تشخیص استرس می‌تواند به‌عنوان ابزاری برای شناسایی آسیب‌پذیری‌های روانی افراد و گروه‌ها استفاده شود.

هوش مصنوعی می‌تواند با شناسایی نشانه‌های استرس از طریق تحلیل زبان و رفتار، به‌سرعت وضعیت روانی افراد را ارزیابی کرده و به‌طور خودکار راهکارهایی برای کاهش تنش‌ها ارائه دهد. این تحلیل‌ها می‌توانند در سطح فردی، گروهی و اجتماعی انجام شوند و به رهبران و مسئولین کمک کنند تا به‌طور مؤثرتر بر بحران‌ها و مشکلات روانی مدیریت کنند. به‌طور خاص، در شرایط جنگ شناختی، که هدف دشمن تحریک استرس و اضطراب در جامعه است، تشخیص استرس به‌عنوان یک ابزار پیشگیرانه می‌تواند از گسترش نفوذ روانی دشمن جلوگیری کند.

از دیگر فواید تشخیص استرس در این زمینه، توانایی شناسایی الگوهای رفتاری و زبانی است که نشان‌دهنده استرس یا اضطراب در افراد و جوامع است. این اطلاعات می‌توانند به سیاست‌گذاران و سازمان‌های مسئول کمک کنند تا مداخلات روان‌شناختی به‌موقع انجام دهند و از تأثیرات منفی جنگ و بحران بر روان جامعه جلوگیری کنند. در این راستا، فناوری‌های هوش مصنوعی می‌توانند به‌عنوان یک ابزار قدرتمند در تحلیل رفتارهای فردی و اجتماعی، شناسایی مشکلات روانی و ارائه راهکارهای مؤثر در مدیریت بحران و جنگ‌های شناختی ایفای نقش کنند.

# جمع‌آوری و آماده‌سازی داده‌ها

این بخش به تشریح روش‌های مختلف جمع‌آوری داده‌های متنی مرتبط با استرس و چگونگی آماده‌سازی آن‌ها برای تحلیل می‌پردازد.

## روش‌های جمع‌آوری داده‌های آموزشی برای تشخیص استرس

برای جمع‌آوری داده‌های آموزشی در راستای تشخیص استرس از متون تولید شده توسط کاربران، چندین روش مختلف وجود دارد که می‌توان از آن‌ها به صورت جداگانه یا ترکیبی استفاده کرد. این روش‌ها به‌طور کلی به ساخت مجموعه داده‌های قوی برای آموزش مدل‌های یادگیری ماشین کمک می‌کنند که به تشخیص نشانه‌های استرس می‌پردازند. در ادامه برخی از رایج‌ترین روش‌ها آورده شده است:

### 1. جمع‌آوری داده‌های رسانه‌های اجتماعی

در این روش، داده‌ها از پست‌ها و نظرات کاربران در پلتفرم‌هایی نظیر توییتر، فیس‌بوک و رددیت جمع‌آوری می‌شوند. این داده‌ها از طریق API‌ های موجود برای هر پلتفرم به دست می‌آیند و به‌طور معمول برای شناسایی نشانه‌های استرس از طریق کلمات یا هشتگ‌های مرتبط با استرس فیلتر می‌شوند. تحقیقات نشان داده‌اند که شبکه‌های اجتماعی به‌عنوان یک منبع غنی برای شناسایی حالات روانی از جمله استرس عمل می‌کنند، زیرا کاربران اغلب افکار و احساسات خود را در این فضاها به اشتراک می‌گذارند. این داده‌ها به‌طور مستقیم از فعالیت‌های آنلاین افراد به دست می‌آیند و به محققان اجازه می‌دهند تا نشانه‌های استرس را در متن‌های کوتاه شناسایی کنند. [7]

### 2. پرسشنامه‌ها و نظرسنجی‌ها

در این روش، پرسشنامه‌ها و نظرسنجی‌ها به افراد ارسال می‌شود که در آن‌ها سوالاتی در مورد تجربه‌های استرس‌زا و سطح استرس فرد مطرح می‌شود. این روش به محققان این امکان را می‌دهد که داده‌های متنی و ارزیابی‌های سطح استرس را به‌طور همزمان جمع‌آوری کنند. از آنجایی که این روش شامل ارزیابی مستقیم از خود فرد است، داده‌ها به‌طور دقیق‌تر برچسب‌گذاری می‌شوند و می‌توانند برای آموزش مدل‌های هوش مصنوعی در شناسایی استرس مفید باشند. همچنین، این روش اجازه می‌دهد تا داده‌ها به‌طور دقیق‌تر بررسی شوند و سطح استرس افراد از طریق مقیاس‌های استاندارد ارزیابی شود. [8]

### 3. جمع‌آوری داده از چت‌ها و پیام‌ها

این روش به جمع‌آوری داده‌ها از چت‌ها و پیام‌های متنی بین افراد، معمولاً با استفاده از سیستم‌های پیام‌رسان مانند WhatsApp یا Facebook Messenger، پرداخته می‌شود. این داده‌ها معمولاً به صورت ناشناس جمع‌آوری می‌شوند و ممکن است شامل ارتباطات با سطح استرس مختلف باشند. برای این که داده‌ها به‌طور مؤثر تحلیل شوند، معمولاً سطح استرس خودکار از طریق ابزارهای تحلیل متن مانند الگوریتم‌های یادگیری ماشین برچسب‌گذاری می‌شود. این روش به دلیل اینکه معمولاً شامل ارتباطات خصوصی افراد است، می‌تواند اطلاعات دقیقی درباره استرس فردی و تغییرات آن در طول زمان فراهم کند [9]

### 4. آزمایش‌های کنترل‌شده

در این روش، آزمایش‌هایی طراحی می‌شود که در آن از شرکت‌کنندگان خواسته می‌شود در مورد تجربیات استرس‌زای خود بنویسند یا گزارشی از مشکلات استرس‌زا در زندگی‌شان ارائه دهند. این تحقیقات به محققان این امکان را می‌دهد که داده‌های برچسب‌خورده با دقت بالا به دست آورند، زیرا استرس در این آزمایش‌ها به‌طور عمدی شبیه‌سازی می‌شود. این داده‌ها می‌توانند برای آموزش مدل‌های هوش مصنوعی در شناسایی استرس به‌طور دقیق‌تر استفاده شوند، چرا که سطح استرس در این نوع آزمایش‌ها به‌طور مستقیم توسط خود شرکت‌کنندگان اعلام می‌شود و بنابراین برچسب‌گذاری داده‌ها دقیق‌تر است [10]

### 5. تکنیک‌های پردازش زبان طبیعی

استفاده از ابزارهایی مانند LIWC (Linguistic Inquiry and Word Count) به‌منظور تحلیل متون و استخراج ویژگی‌های مرتبط با استرس یکی دیگر از روش‌های معمول است. این ابزارها به‌طور خاص بر روی ویژگی‌های خاص زبان مانند کلمات منفی، افعال مرتبط با احساسات و ضمایر اول شخص تمرکز دارند. با استفاده از این ابزارها، می‌توان الگوهای زبانی خاصی که نشان‌دهنده استرس هستند را شناسایی و استخراج کرد. این روش به‌ویژه برای بررسی داده‌های متنی از رسانه‌های اجتماعی، پیام‌ها یا سایر متون کوتاه مناسب است و می‌تواند به‌طور مؤثر نشانه‌های استرس را شناسایی کند. [10]

### 6. استفاده از تطبیق عبارات منظم

در این روش، از الگوهای عباراتی برای شناسایی محتوای مرتبط با استرس در مجموعه‌های بزرگ داده‌ها استفاده می‌شود. این الگوها می‌توانند به‌ویژه برای شناسایی عباراتی که حاوی کلمات کلیدی استرس‌زا هستند، مفید باشند. این روش معمولاً برای جمع‌آوری داده‌ها از فروم‌های علمی یا سایت‌های مربوط به استرس استفاده می‌شود و می‌تواند به‌طور خودکار عباراتی که به استرس اشاره دارند را شناسایی کند. [11]

### 7. برچسب‌گذاری دستی

در این روش، از کارشناسان خبره استفاده می‌شود تا متون جمع‌آوری‌شده را مرور کرده و آن‌ها را بر اساس نشانه‌های استرس برچسب‌گذاری کنند. این روش به‌ویژه برای ایجاد مجموعه‌های داده با کیفیت بالا مفید است، چرا که نشانه‌های استرس به‌طور دقیق توسط افراد متخصص شناسایی می‌شوند. البته این روش زمان‌بر است و نیاز به منابع انسانی دارد، اما معمولاً داده‌های جمع‌آوری‌شده دقیق‌تر و قابل اعتمادتر خواهند بود [11].

### 8. جمع‌آوری داده‌های چندوجهی

این روش به جمع‌آوری داده‌ها از چندین منبع مختلف مانند داده‌های متنی همراه با نشانه‌های فیزیولوژیکی (مانند ضربان قلب یا تغییرات در عرق پوست) می‌پردازد. این ترکیب داده‌ها می‌تواند به تحلیل دقیق‌تر استرس کمک کند، چرا که مدل‌ها می‌توانند از اطلاعات چندگانه برای تشخیص استرس استفاده کنند. به‌ویژه در حوزه‌های بالینی، این داده‌ها می‌توانند به‌طور مؤثری نشان‌دهنده تغییرات فیزیکی و روانی مرتبط با استرس باشند [12].

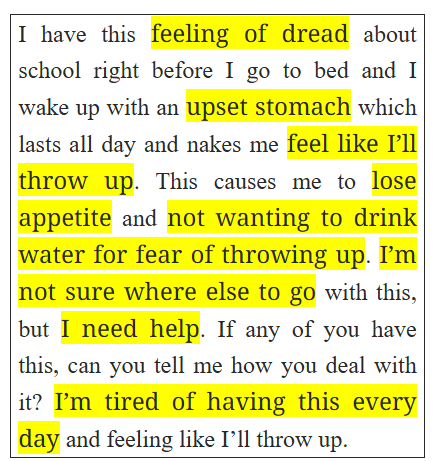
## معرفی مجموعه داده‌های موجود برای تشخیص استرس

### مجموع داده ردیت

Reddit یک وب‌سایت/رسانه اجتماعی است که کاربران در آن در جوامع موضوع‌محور به نام subreddit پست ارسال می‌کنند و دیگر کاربران می‌توانند روی این پست‌ها نظر بدهند و رأی دهند. طبیعت طولانی این پست‌ها، آنها را به منبع مناسبی برای مطالعه جزئیات پدیده‌هایی مانند استرس تبدیل کرده است. برای جمع‌آوری داده‌های مربوط به استرس، دسته‌بندی‌های زیرمجموعه‌هایی انتخاب شده‌اند که اعضای آن‌ها احتمالاً درباره موضوعات پراسترس بحث می‌کنند:

1. تعاملات میان‌فردی: حوزه‌های سوءاستفاده و اجتماعی. در زیرمجموعه‌های سوءاستفاده، افراد غالباً تجربیات خود از روابط یا موقعیت‌های آزاردهنده را به اشتراک می‌گذارند، در حالی که در زیرمجموعه اجتماعی، کاربران مشکلات روابط (غالباً اما نه فقط عاشقانه) را بیان کرده و درخواست راهنمایی می‌کنند.
2. بیماری‌های روانی: حوزه‌های اضطراب و PTSD. کاربران در این زیرمجموعه‌ها درباره مقابله با علائم بیماری‌های روانی، درخواست حمایت و تشخیص صحبت می‌کنند.
3. نیاز مالی: حوزه مالی. کاربران این زیرمجموعه‌ها اغلب از دیگران درخواست کمک مالی یا مادی می‌کنند.

این مجموعه شامل ده زیرمجموعه در پنج حوزه است که جزئیات آن در جدول آمده است. با استفاده از API PRAW، تمام پست‌های موجود از این زیرمجموعه‌ها بین 1 ژانویه 2017 تا 19 نوامبر 2018 جمع‌آوری شده‌اند؛ در مجموع 187444پست. سپس به 3553 بخش از این پست‌ها برچسب‌های استرس دودویی اختصاص داده شده است تا یک مجموعه آموزشی نظارتی و نیمه‌نظارتی تشکیل شود. یک مثال از بخش برچسب‌گذاری شده در شکل 1 ارائه شده است.



در این مثال، عبارات برجسته به عنوان نشانه‌هایی از استرس شناسایی شده‌اند: نویسنده به علائم فیزیکی شایع (حالت تهوع)، ترس و اضطراب آشکار و زبان نشان‌دهنده بی‌کمکی و رفتارهای کمک‌خواهانه اشاره می‌کند.

طول متوسط پست‌ها در این مجموعه 420 کلمه است که بسیار طولانی‌تر از داده‌های میکروبلاگ‌هایی مانند توییتر است.

| **دامنه** | **نام زیرمجموعه** | **تعداد پست‌ها** | **میانگین تعداد کلمات در هر پست** | **بخش‌های برچسب‌گذاری شده** |
| --- | --- | --- | --- | --- |
| **سوءاستفاده** | r/domesticviolence | 1529 | 365 | 388 |
|  | r/survivorsofabuse | 1372 | 444 | 315 |
| **جمع کل** |  | 2901 | 402 | 703 |
| **اضطراب** | r/anxiety | 58130 | 193 | 650 |
|  | r/stress | 1078 | 107 | 78 |
| **جمع کل** |  | 59208 | 191 | 728 |
| **مالی** | r/almosthomeless | 547 | 261 | 99 |
|  | r/assistance | 9243 | 209 | 355 |
|  | r/food\_pantry | 343 | 187 | 43 |
|  | r/homeless | 2384 | 143 | 220 |
| **جمع کل** |  | 12517 | 198 | 717 |
| **PTSD** | r/ptsd | 4910 | 265 | 711 |
| **اجتماعی** | r/relationships | 107908 | 578 | 694 |
| **مجموع کل** |  | 187444 | 420 | 3553 |

فرایند برچسب‌گذاری داده‌ها در این مجموعه داده‌ها به شکل زیر انجام شده است[[1]](#footnote-0):

1. **تقسیم پست‌ها به بخش‌های پنج‌جمله ای:** پست‌های طولانی به بخش‌های پنج‌جمله‌ای تقسیم شدند تا برچسب‌گذاران بتوانند آن‌ها را به راحتی بررسی کنند. این تقسیم‌بندی کمک کرد تا تمرکز بیشتری بر روی بخش‌های خاص و جزئیات استرس در متن صورت گیرد.
2. **آزمون صلاحیت برچسب‌گذاران**: قبل از شروع برچسب‌گذاری، برچسب‌گذاران باید آزمون صلاحیت را گذرانده و توانایی خود در تشخیص استرس در متن‌ها را نشان می‌دادند. این اطمینان می‌دهد که فقط افراد با توانایی مناسب به برچسب‌گذاری می‌پردازند.
3. **تعریف دقیق استرس**: دستورالعمل‌ها شامل یک تعریف دقیق از استرس بر اساس «دیکشنری آکسفورد» بود که بیان می‌کرد استرس به عنوان «فشار ذهنی یا احساسی ناشی از شرایط سخت» است. برچسب‌گذاران به این امر توجه داشتند که باید احساس منفی نویسنده نسبت به استرس را شناسایی کنند، نه فقط شناسایی خود وضعیت استرس‌زا.
4. **برچسب‌گذاری توسط چندین برچسب‌گذار**: هر بخش از پست‌ها حداقل توسط پنج برچسب‌گذار برچسب‌گذاری شد تا دقت برچسب‌گذاری افزایش یابد و نتیجه نهایی قابل اعتمادتر باشد.
5. **برچسب‌گذاری با انتخاب «استرس»، «بدون استرس» یا «نمی‌توان تشخیص داد»:** برچسب‌گذاران برای هر بخش از پست‌ها باید یکی از سه گزینه «استرس»، «بدون استرس» یا «نمی‌توان تشخیص داد» را انتخاب می‌کردند. این روش به کمک متنی که در آن بحث استرس یا بی‌استرسی مطرح شده، کمک کرد تا برچسب‌گذاری دقیق‌تری انجام گیرد.
6. **حذف داده‌های نامعتبر**: پس از دریافت برچسب‌ها از برچسب‌گذاران، بخش‌هایی که توسط برچسب‌گذاران نادرست برچسب‌گذاری شده بودند یا در آن‌ها حداقل نیمی از برچسب‌گذاران گزینه «نمی‌توان تشخیص داد» را انتخاب کرده بودند، حذف شدند تا داده‌های دقیق‌تری باقی بماند.
7. **انتخاب برچسب نهایی بر اساس اکثریت آرا**: در نهایت، برای هر بخش از پست‌ها، برچسب نهایی بر اساس اکثریت آرا از برچسب‌گذاران انتخاب شد. اگر اکثریت نظر مشابهی می‌دادند، آن برچسب به عنوان برچسب نهایی در نظر گرفته می‌شد.
8. **ارزیابی توافق برچسب‌گذاران**: برای ارزیابی کیفیت برچسب‌گذاری‌ها، توافق بین برچسب‌گذاران با استفاده از معیار کاپا[[2]](#footnote-1) سنجیده شد که نشان‌دهنده میزان توافق بین برچسب‌گذاران است. این میزان توافق در این مجموعه داده‌ها برابر با 0.47 بود که به عنوان توافق «متوسط» ارزیابی شد.

معیار کاپا (κ) فلایس یکی از متداول‌ترین روش‌های ارزیابی توافق بین چندین برچسب‌گذار است. این معیار به ما کمک می‌کند تا میزان هم‌رای بودن ارزیابی‌های مختلف برچسب‌گذاران را اندازه‌گیری کنیم.

مقدار کاپا بین 0 تا 1 متغیر است. 0 نشان‌دهنده عدم توافق کامل و 1 نشان‌دهنده توافق کامل است. مقدار بین 0.2 تا 0.4 معمولاً به عنوان «توافق ضعیف»، 0.4 تا 0.6 به عنوان «توافق متوسط»، و 0.6 تا 0.8 به عنوان «توافق خوب» در نظر گرفته می‌شود. در مثال این مجموعه داده‌ها مقدار کاپا برابر با 0.47 بود که نشان‌دهنده «توافق متوسط» است. این میزان نشان می‌دهد که برچسب‌گذاران معمولاً در بسیاری از موارد هم‌نظر بودند، اما هنوز برخی تفاوت‌ها وجود داشت.

### مجموعه داده TensiStrength

#### جمع‌آوری داده‌ها:

الف. انتخاب کلمات کلیدی: هدف شناسایی توییت‌های مرتبط با استرس یا آرامش.

کلمات کلیدی استفاده‌شده:

مرتبط با استرس: "استرس"، "اضطراب"، "نگران"، "غرق"، "وحشت"، "عصبی"، "تنش"، "ناامید"، "افسرده"، "غمگین".

مرتبط با آرامش: "آرامش"، "سکون"، "راحت"، "صلح‌آمیز"، "خوشحال"، "شاد"، "راضی"، "مطمئن".

ب. جمع‌آوری توییت‌ها: استفاده از API توییتر

#### پیش‌پردازش:

الف. پاک‌سازی متن:

* حذف URLها، هشتگ‌ها، و نام‌ها برای تمرکز بر محتوای اصلی.
* تبدیل متن به حروف کوچک برای یکسان‌سازی.
* حذف کاراکترهای خاص و علائم نگارشی به‌جز ایموتیکون‌ها.

ب. اصلاحات زبانی:

اصلاح اشتباهات املایی: مثلاً تبدیل "woooorrried" به "worried").

تبدیل ایموجی به حالت‌های احساسی معادل (مثلاً :) = مثبت، :( = منفی).

#### راهنمایی‌های برچسب‌گذاری

الف. مقیاس ارزیابی: ایجاد مقیاس پنج‌نقطه‌ای برای برچسب‌گذاری

1- : بدون استرس (محتوای خنثی)

2- : استرس خفیف

3- : استرس متوسط

4- : استرس زیاد

5- : استرس بسیار زیاد

#### ب. فرآیند برچسب‌گذاری

انتخاب سه داور مستقل با تجربه در زمینه بهداشت روان یا روانشناسی،

ارائه دستورالعمل‌های دقیق برای ارزیابی توییت‌ها بر اساس محتوا و احساسات.

#### ج. نمونه برچسب‌گذاری‌ها

مثال استرس‌زا: "دیگه نمی‌تونم این فشار کاری رو تحمل کنم!"

برچسب: 4- (استرس بالا)

مثال غیر استرس‌زا: "یه روز عالی در پارک داشتم."

برچسب: 1- (بدون استرس)

#### اصلاح خودکار و توسعه فرهنگ‌نامه

الف. ایجاد فرهنگ‌نامه: ایجاد فرهنگ‌نامه‌ای از اصطلاحات مرتبط با استرس و آرامش و اختصاص نمره عددی به هر اصطلاح بر اساس بار معنایی آن.

ب. تنظیمات مبتنی بر قواعد: اجرای قواعدی برای تنظیم نمرات بر اساس زمینه

نقیض‌ها (مثلاً "آرام نیستم" باید نمره استرس را افزایش دهد).

اصطلاحات اغراق‌آمیز (مثلاً "خیلی استرس دارم!" ممکن است نشان‌دهنده استرس بالاتر باشد).

#### ترکیب نهایی مجموعه داده

الف. تبدیل به برچسب‌های دودویی: تبدیل مقیاس پنج‌نقطه‌ای به برچسب‌های دودویی

توییت‌های با نمره 1- به برچسب 0 (غیر استرس) تبدیل می‌شوند.

توییت‌های با نمره 2- تا 5- به برچسب 1 (استرس) تبدیل می‌شوند.

ب. تعداد نهایی و توزیع

اطمینان از توزیع متوازن برچسب‌ها بین استرس و غیر استرس.

#### ارزیابی و کنترل کیفیت

الف. توافق بین داوران: محاسبه توافق بین داوران با استفاده از معیار Fleiss' Kappa برای ارزیابی دقت و ثبات در برچسب‌گذاری.

### مجموعه‌داده DAIC-WOZ

این مجموعه‌داده‌ها[[3]](#footnote-2) شامل داده‌های چندوجهی است که شامل متن، صوت و ویدئو می‌باشد. داده‌های متنی شامل رونویسی دقیق مکالمات میان کاربران و مصاحبه‌کنندگان است که به صورت فایل‌های متنی ذخیره شده و شامل جزئیات مربوط به هر مکالمه مانند نوع سؤال، پاسخ‌دهنده و زمان‌بندی مکالمات هستند. داده‌های صوتی شامل فایل‌های صوتی با فرمت WAV هستند که ویژگی‌هایی همچون زیر و بم صدا، شدت صدا، نرخ گفتار و مکث‌ها را می‌توان از آن‌ها استخراج کرد. این داده‌ها برای تحلیل‌های مربوط به علائم استرس و افسردگی مورد استفاده قرار می‌گیرند. داده‌های ویدئویی شامل فایل‌های ویدئویی هستند که حرکات چهره، وضعیت بدن و رفتارهای غیرکلامی کاربران را ضبط می‌کنند. ویژگی‌های قابل استخراج از این داده‌ها شامل حرکات چهره، حرکات چشم و وضعیت بدن می‌شود. برچسب‌گذاری داده‌ها با استفاده از مقیاس PHQ-8 انجام می‌شود که برای سنجش شدت افسردگی و استرس از ۰ تا ۲۴ متغیر است. علاوه بر این، ویژگی‌های اضافی از جمله ویژگی‌های زبانی، صوتی و فیزیولوژیکی از داده‌ها استخراج شده و برای تحلیل‌های پیشرفته‌تر مورد استفاده قرار می‌گیرند. ساختار چندوجهی این مجموعه‌داده امکان تحلیل دقیق و جامع‌تری از وضعیت روان‌شناختی افراد را فراهم می‌آورد و در پژوهش‌های مربوط به پردازش زبان طبیعی و تشخیص اختلالات روانی کاربرد دارد.

روش‌های موجود برای جمع‌آوری داده‌های آموزشی از شبکه‌های اجتماعی برای تشخیص استرس با مشکلاتی همراه هستند که دقت نتایج را تحت تأثیر قرار می‌دهند. یکی از مشکلات اصلی استفاده از هشتگ‌هاست. هشتگ‌ها ممکن است نشان‌دهنده استرس نباشند، زیرا بسیاری از کاربران ممکن است از هشتگ‌ها برای بیان استرس خود استفاده نکنند یا به‌طور غیرمستقیم و طنزآمیز به استرس اشاره کنند. علاوه بر این، پیدا کردن لیست کاملی از هشتگ‌ها که مرتبط با استرس باشند کار دشواری است و همیشه نمی‌توان تمام هشتگ‌های رایج را پیدا کرد. این موضوع باعث می‌شود که برخی از داده‌ها از دست بروند.

همچنین، جست‌وجو با کلمات کلیدی می‌تواند نتایج دقیقی نداشته باشد، زیرا بسیاری از کلمات ممکن است به‌طور غیرمستقیم با استرس ارتباط داشته باشند، اما در واقع هیچ ارتباطی با آن نداشته باشند. علاوه بر این، این کلمات باید به عبارات رایج در شبکه‌های اجتماعی تبدیل شوند که این فرآیند گاهی پیچیده و زمان‌بر است. استفاده از پرسشنامه‌ها و نظرسنجی‌ها هم با وجود اینکه می‌تواند داده‌های دقیقی فراهم کند، محدود به گروه خاصی از افراد است که این پرسشنامه‌ها را پر کرده‌اند، و این ممکن است موجب تعصب در داده‌ها شود. در این میان، افراد با استرس کمتر ممکن است تمایل کمتری به پر کردن این پرسشنامه‌ها داشته باشند.

داده‌های جمع‌آوری شده از شبکه‌های اجتماعی نیز می‌توانند ناقص باشند، زیرا بسیاری از کاربران ممکن است احساسات خود را بیان کنند اما از هشتگ‌ها یا کلمات کلیدی خاص استفاده نکنند، که باعث می‌شود داده‌های مربوط به استرس از دست بروند یا نادرست جمع‌آوری شوند. برای رفع این مشکلات، ما روشی جدید را پیشنهاد می‌دهیم. این روش مبتنی بر استفاده از عبارات کلیدی است که بر اساس سوالات پرسشنامه‌هایی که در زمینه استرس طراحی شده‌اند انتخاب شده‌اند. این پرسشنامه‌ها به شناسایی انواع مختلف استرس مانند استرس فیزیکی، روانی، اجتماعی، و استرس ناشی از کار کمک می‌کنند. پس از شناسایی این عبارات کلیدی، ما آن‌ها را به‌گونه‌ای گسترش می‌دهیم که با زبان و سبک نگارش رایج در شبکه‌های اجتماعی همخوانی داشته باشند. این روش کمک می‌کند تا استرس در متن‌هایی که ممکن است به‌طور مستقیم از هشتگ‌ها یا کلمات کلیدی استفاده نکنند، شناسایی شود. به این ترتیب، داده‌های جمع‌آوری شده از شبکه‌های اجتماعی دقیق‌تر و گسترده‌تر خواهند بود و می‌توانند به شناسایی بهتر استرس در انواع مختلف نوشته‌های کاربران کمک کنند.

## شناسایی عبارات مرتبط با استرس از پرسشنامه‌ها

در اینجا جدول مربوط به سوالات پرسشنامه استرس به تفکیک دسته‌های مختلف استرسً ارائه شده است:

| دسته‌بندی نشانه استرس | سوالات پرسشنامه |
| --- | --- |
| استرس فیزیکی | آیا در مواقع استرس احساس درد در بدن یا سردرد دارید؟ |
| آیا در هنگام استرس تپش قلب یا مشکلات تنفسی را تجربه می‌کنید؟ |
| آیا در مواقع استرس احساس ضعف یا خستگی می‌کنید؟ |
| آیا از شدت استرس دچار دل درد یا مشکلات گوارشی می‌شوید؟ |
| آیا در مواقع استرس احساس سنگینی در بدن یا خستگی غیرعادی دارید؟ |
| استرس روانی و احساسی | آیا در شرایط استرس‌زا احساس ناامیدی یا غمگینی می‌کنید؟ |
| آیا در شرایط استرس‌زا دچار احساس اضطراب یا نگرانی بیش از حد می‌شوید؟ |
| آیا در مواقع استرس احساس می‌کنید که کنترل خود را از دست می‌دهید؟ |
| آیا احساس می‌کنید که نمی‌توانید با افکار منفی مقابله کنید؟ |
| آیا در هنگام استرس، ذهن شما دائماً مشغول نگرانی‌ها و مشکلات است؟ |
| استرس ناشی از روابط اجتماعی | آیا در مواقع استرس احساس تنهایی یا قطع ارتباط با دیگران دارید؟ |
| آیا هنگام استرس دچار مشکل در برقراری ارتباط با دیگران می‌شوید؟ |
| آیا در شرایط استرس‌زا احساس می‌کنید که دیگران شما را درک نمی‌کنند؟ |
| آیا در مواقع استرس احساس می‌کنید که دیگران از شما دور شده‌اند؟ |
| استرس ناشی از کار و مسئولیت‌ها | آیا در اثر حجم زیاد کار احساس فشار می‌کنید؟ |
| آیا در هنگام استرس احساس می‌کنید که کارها از کنترل شما خارج شده است؟ |
| آیا احساس می‌کنید که کارها زیاد است و شما نمی‌توانید به موقع انجام دهید؟ |
| آیا تحت فشار زمانی برای انجام کارها احساس استرس می‌کنید؟ |
| استرس ناشی از افکار منفی و خودآسیب‌رسان | آیا در مواقع استرس افکار منفی یا خودآسیب‌رسانی به ذهن شما می‌آید؟ |
| آیا در مواقع استرس به خودکشی یا آسیب به خود فکر می‌کنید؟ |
| آیا افکار منفی در مورد آینده شما را به شدت ناراحت می‌کند؟ |
| استرس ناشی از زمان و فشارهای زمانی | آیا در مواقع شلوغ یا فشار زمانی احساس استرس می‌کنید؟ |
| آیا در شرایطی که زمان کمی دارید، دچار استرس می‌شوید؟ |
| آیا همیشه احساس می‌کنید که زمان کافی برای انجام کارها ندارید؟ |
| استرس ناشی از تغییرات و بحران‌ها | آیا در هنگام تغییرات بزرگ زندگی (مثل تغییر شغل یا محل زندگی) احساس استرس می‌کنید؟ |
| آیا در شرایط بحرانی یا هنگام بحران‌های زندگی، احساس اضطراب می‌کنید؟ |
| استرس ناشی از نگرانی‌ها و ترس‌ها | آیا به طور مداوم نگران آینده یا مشکلات احتمالی هستید؟ |
| آیا در مواقع استرس احساس می‌کنید که نمی‌توانید از پس مشکلات برآیید؟ |
| آیا ترس از آینده یا از دست دادن کنترل شما را نگران می‌کند؟ |
| استرس ناشی از مشکلات جسمی و سلامت | آیا در مواقع استرس از لحاظ جسمی احساس ناتوانی می‌کنید؟ |
| آیا استرس باعث مشکلات خواب یا بی‌خوابی شما شده است؟ |
| استرس ناشی از مشکلات مالی و اقتصادی | آیا در شرایط استرس از مسائل مالی نگران هستید؟ |
| آیا استرس شما ناشی از مشکلات اقتصادی یا فشار مالی است؟ |

این جدول شامل سوالاتی است که برای شناسایی انواع مختلف استرس، از جمله استرس فیزیکی، روانی، اجتماعی، مالی و دیگر انواع استرس طراحی شده‌اند. این سوالات می‌توانند به عنوان معیارهایی برای شناسایی نشانه‌های استرس در تحلیل داده‌های متنی در شبکه‌های اجتماعی یا سایر منابع به کار روند.

سپس بر اساس این دسته‌ها و سوالات عباراتی برای جستجو در شبکه‌های اجتماعی ایجاد کردیم.

| **دسته‌بندی نشانه استرس** | **سوالات پرسشنامه** | **عبارات جستجو شده** |
| --- | --- | --- |
| **استرس فیزیکی** | آیا در مواقع استرس احساس درد در بدن یا سردرد دارید؟ | سرم شلوغ، سرم شلوغه، بدن درد، فشار زیادی رومه، تپش قلب |
| آیا در هنگام استرس تپش قلب یا مشکلات تنفسی را تجربه می‌کنید؟ | تپش قلب، نفس تنگی، نمی‌توانم نفس بکشم |
| آیا در مواقع استرس احساس ضعف یا خستگی می‌کنید؟ | زود خسته میشم، احساس ضعف |
| آیا از شدت استرس دچار دل درد یا مشکلات گوارشی می‌شوید؟ | دل درد، درد معده |
| آیا در مواقع استرس احساس سنگینی در بدن یا خستگی غیرعادی دارید؟ | خون میخورم، از هم پاشیده شدم، فشار زیاد |
| **استرس روانی و احساسی** | آیا در شرایط استرس‌زا احساس ناامیدی یا غمگینی می‌کنید؟ | از دستم ناراحت شد، دل کندم، نمی‌توانم ادامه بدم، احساس ناامنی می‌کنم |
| آیا در شرایط استرس‌زا دچار احساس اضطراب یا نگرانی بیش از حد می‌شوید؟ | اضطراب دارم، نگرانم، استرس دارم |
| آیا در مواقع استرس احساس می‌کنید که کنترل خود را از دست می‌دهید؟ | کنترل ندارم، از دستم در رفت، حالم بهم می‌خوره |
| آیا احساس می‌کنید که نمی‌توانید با افکار منفی مقابله کنید؟ | ذهنم آشفته، افکار منفی زیاد |
| آیا در هنگام استرس، ذهن شما دائماً مشغول نگرانی‌ها و مشکلات است؟ | فکر و خیال زیاد، نگرانی، ذهنم درگیر |
| **استرس ناشی از روابط اجتماعی** | آیا در مواقع استرس احساس تنهایی یا قطع ارتباط با دیگران دارید؟ | احساس تنهایی، درک نکردن از طرف دیگران |
| آیا هنگام استرس دچار مشکل در برقراری ارتباط با دیگران می‌شوید؟ | نمی‌توانم حرف بزنم، احساس بی‌اطمینانی |
| آیا در شرایط استرس‌زا احساس می‌کنید که دیگران شما را درک نمی‌کنند؟ | دیگه هیچی مثل قبل نمیشه، از هم پاشیدیم |
| آیا در مواقع استرس احساس می‌کنید که دیگران از شما دور شده‌اند؟ | از هم دور شدیم، همه کرونا بگیرن |
| **استرس ناشی از کار و مسئولیت‌ها** | آیا در اثر حجم زیاد کار احساس فشار می‌کنید؟ | کار زیاد دارم، کارها عقب افتاده، باید سریع تمام کنم |
| آیا در هنگام استرس احساس می‌کنید که کارها از کنترل شما خارج شده است؟ | از کنترل من خارج شده، هیچ کاری نمی‌رسم، تحت فشارم |
| آیا احساس می‌کنید که کارها زیاد است و شما نمی‌توانید به موقع انجام دهید؟ | نمی‌رسم، کار مونده، بار سنگینی هست |
|  | آیا تحت فشار زمانی برای انجام کارها احساس استرس می‌کنید؟ | زمان کم دارم، فورا باید، دیرم شد |
| **استرس ناشی از افکار منفی و خودآسیب‌رسان** | آیا در مواقع استرس افکار منفی یا خودآسیب‌رسانی به ذهن شما می‌آید؟ | خودکشی کنم، گند زدم، نمی‌توانم ادامه بدهم |
| آیا در مواقع استرس به خودکشی یا آسیب به خود فکر می‌کنید؟ | خودکشی کنم، از خودم خسته شدم |
| آیا افکار منفی در مورد آینده شما را به شدت ناراحت می‌کند؟ | آینده رو هوا، نمی‌دونم چه کار کنم |
| **استرس ناشی از زمان و فشارهای زمانی** | آیا در مواقع شلوغ یا فشار زمانی احساس استرس می‌کنید؟ | کارهای عقب افتاده، باید سریع انجام بدم، وقت ندارم |
| آیا در شرایطی که زمان کمی دارید، دچار استرس می‌شوید؟ | وقت ندارم، عجله دارم، باید سریع انجام بدم |
| آیا همیشه احساس می‌کنید که زمان کافی برای انجام کارها ندارید؟ | همیشه دیرم می‌شه، نمی‌رسم |
| **استرس ناشی از تغییرات و بحران‌ها** | آیا در هنگام تغییرات بزرگ زندگی (مثل تغییر شغل یا محل زندگی) احساس استرس می‌کنید؟ | تغییرات بزرگی میاد، شرایط سخت |
| آیا در شرایط بحرانی یا هنگام بحران‌های زندگی، احساس اضطراب می‌کنید؟ | بحران دارم، شرایط از دستم در رفته |
| **استرس ناشی از نگرانی‌ها و ترس‌ها** | آیا به طور مداوم نگران آینده یا مشکلات احتمالی هستید؟ | نگرانم، ترس از آینده، ترس دارم |
| آیا در مواقع استرس احساس می‌کنید که نمی‌توانید از پس مشکلات برآیید؟ | احساس ناتوانی، فشار زیادی رومه |
| آیا ترس از آینده یا از دست دادن کنترل شما را نگران می‌کند؟ | ترس بی دلیل، فشار زیاد، نگرانی از آینده |
| **استرس ناشی از مشکلات جسمی و سلامت** | آیا در مواقع استرس از لحاظ جسمی احساس ناتوانی می‌کنید؟ | خون میخورم، حالم بهم می‌خوره، احساس ضعف |
| آیا استرس باعث مشکلات خواب یا بی‌خوابی شما شده است؟ | بی‌خوابی، خوابم نمیبره، کم خوابی |
| **استرس ناشی از مشکلات مالی و اقتصادی** | آیا در شرایط استرس از مسائل مالی نگران هستید؟ | مشکلات اقتصادی، فشار مالی، وضعیت خراب مالی |
| آیا استرس شما ناشی از مشکلات اقتصادی یا فشار مالی است؟ | مسائل مالی، مشکل اقتصادی، فشار اقتصادی |

## گسترش به زبان محاوره و عبارات متداول توییتر

در روش پیشنهادی ما، یکی از مهم‌ترین نکات این است که عبارات کلیدی را به زبان محاوره‌ای و عباراتی که در شبکه‌های اجتماعی، به ویژه توییتر، متداول هستند، گسترش دهیم. زبان محاوره‌ای در شبکه‌های اجتماعی به‌طور معمول با زبان رسمی یا علمی تفاوت دارد و کاربران معمولاً از عبارات غیررسمی و کوتاه استفاده می‌کنند. بنابراین، برای شناسایی دقیق استرس در نوشته‌ها و پیام‌های کاربران، باید این عبارات به‌گونه‌ای انتخاب و گسترش یابند که با سبک گفتاری و نوشتاری رایج در توییتر و سایر شبکه‌های اجتماعی هماهنگ باشند.

برای مثال، عباراتی مانند "استرس دارم" یا "مضطربم" در زبان رسمی ممکن است به‌طور مستقیم و بدون تغییر یافت شوند. اما در زبان محاوره‌ای و در شبکه‌های اجتماعی، کاربران ممکن است از عبارات کوتاه‌تر یا ترکیبی استفاده کنند که معنای مشابهی دارند، مانند "الان اعصاب ندارم" یا "دلم نمی‌کشه". این نوع تغییرات در زبان ممکن است در صورت استفاده از کلمات کلیدی تنها، شناسایی نشود.

گسترش عبارات به زبان محاوره‌ای به این معناست که باید از معادل‌های رایج و عبارات متداول کاربران در شبکه‌های اجتماعی برای بیان استرس استفاده کنیم. برای این منظور، باید درک عمیقی از سبک نگارش توییتر و شبکه‌های اجتماعی داشته باشیم و این سبک نگارش را در جست‌وجوهای خود لحاظ کنیم. برای نمونه، در توییتر ممکن است کاربران از اختصارات، ایموجی‌ها، یا حتی عبارات با اشتباهات تایپی استفاده کنند که به‌طور معمول در جست‌وجوهای کلمات کلیدی پوشش داده نمی‌شوند، اما در تحلیل دقیق‌تر باید مد نظر قرار گیرند.در نهایت، این روش گسترش عبارت‌ها باعث می‌شود تا داده‌های جمع‌آوری شده از شبکه‌های اجتماعی، به‌ویژه توییتر، دقیق‌تر و به‌روزتر باشند و ما بتوانیم استرس را به‌طور مؤثرتری شناسایی کنیم.

## قطبیت زبانی: تعریف معیارهای عبارات استرس‌زا و آرامش‌بخش

قطبیت زبانی به معنای شناسایی و تحلیل احساسات و عواطف موجود در متن‌ها است. در زمینه شناسایی استرس و آرامش، این قطبیت می‌تواند به تعیین اینکه یک عبارت نشانه استرس‌زا است یا آرامش‌بخش، کمک کند. به‌طور کلی، قطبیت زبانی در تحلیل احساسات به بررسی کلمات و عبارات به منظور تعیین بار احساسی مثبت یا منفی آن‌ها می‌پردازد.

### تعریف معیارهای عبارات استرس‌زا

عبارات استرس‌زا به آن دسته از عبارات اطلاق می‌شوند که معمولاً احساساتی مانند اضطراب، نگرانی، ترس، ناامیدی، فشار، و خستگی را به همراه دارند. این عبارات ممکن است نشان‌دهنده‌ی اضطراب ذهنی یا فشارهای جسمی و روانی باشند. معیارهایی که می‌توانند عبارات استرس‌زا را شناسایی کنند عبارتند از:

استفاده از کلمات منفی: عباراتی که حاوی کلمات منفی مانند "نمی‌توانم"، "دیر"، "مشکل"، "بی‌حالی"، "ضعف"، و "ترس" هستند.

فعل‌های مرتبط با احساسات منفی: افعال مانند "اضطراب داشتن"، "ترسیدن"، "گریه کردن"، "احساس ضعف کردن" می‌توانند نشانه‌های استرس باشند.

تکرار و تاکید بر مشکلات: عبارات با تأکید زیاد بر مشکلات و منفی‌نگری مثل "هرگز نمی‌توانم"، "هیچ وقت"، "همیشه"، می‌توانند نشان‌دهنده‌ی استرس باشند.

توصف‌های بدنی ناخوشایند: عباراتی که به مشکلات جسمی مرتبط با استرس اشاره دارند، مانند "درد"، "تپش قلب"، "دست‌پاچگی"، "بی‌خوابی"، "دل‌درد"، معمولاً نشانه‌های استرس‌زا هستند.

### تعریف معیارهای عبارات آرامش‌بخش

عبارات آرامش‌بخش به عباراتی اطلاق می‌شوند که احساسات مثبت، راحتی، آسایش، و آرامش را به همراه دارند. معیارهایی که می‌توانند عبارات آرامش‌بخش را شناسایی کنند عبارتند از:

استفاده از کلمات مثبت: عباراتی که حاوی کلمات مثبت مانند "آرامش"، "شاد"، "خوشحال"، "راحت"، "آرام"، "مطمئن" هستند.

استفاده از ضمایر اول شخص مثبت: استفاده از ضمایر اول شخص مثبت مانند "من احساس راحتی می‌کنم" یا "من خوشحال هستم" نشان‌دهنده وضعیت روانی مثبت و آرامش‌بخش است.

فعل‌های مرتبط با احساسات مثبت: افعالی مانند "آرام گرفتن"، "استراحت کردن"، "خندیدن"، "لذت بردن"، "خواب خوب داشتن" معمولاً نشانه‌های آرامش و رضایت هستند.

توصیف‌های بدنی مثبت: عباراتی که به بهبود و راحتی بدنی اشاره دارند مانند "حس خوب داشتن"، "بدن سبک شدن"، "حالت خوب بودن" می‌توانند نشان‌دهنده احساس آرامش باشند.

نشانه‌های از خودمراقبتی: عباراتی که حاکی از توجه به خود و سلامت روان و جسم هستند، مثل "ورزش کردن"، "مدیتیشن"، "استراحت کردن" معمولاً به آرامش و سلامتی جسمی و روانی اشاره دارند.

### اهمیت قطبیت زبانی در تشخیص استرس

در تحلیل استرس از طریق متن‌ها، قطبیت زبانی به‌ویژه در شبکه‌های اجتماعی و چت‌ها می‌تواند به شناسایی عبارات استرس‌زا و آرامش‌بخش کمک کند. استفاده از این معیارها به‌ویژه در محیط‌های آنلاین که زبان محاوره‌ای غالب است، برای سیستم‌های تحلیل احساسات و تشخیص استرس ضروری است. این سیستم‌ها می‌توانند بر اساس قطبیت زبانی، متن‌ها را به دو دسته‌ی "استرس‌زا" و "آرامش‌بخش" تقسیم کرده و به‌طور خودکار سطح استرس کاربران را تشخیص دهند.

در نهایت، تحلیل قطبیت زبانی نه تنها برای شناسایی استرس‌زا بودن یک عبارت بلکه برای درک دقیق‌تر تغییرات در وضعیت روانی کاربران در طول زمان، به‌ویژه در واکنش به شرایط خاص، مفید است.

## فرایند جمع‌آوری و دسته‌بندی 6 میلیون داده

برای جمع‌آوری داده‌های مورد نیاز برای تحلیل استرس در توییت‌ها، ابتدا لیستی از عبارات نشان‌دهنده استرس و آرامش تهیه کردیم. این عبارات بر اساس سوالات پرسشنامه و نشانه‌های روان‌شناختی انتخاب شدند که معمولاً نشان‌دهنده وضعیت استرس‌زا یا آرامش‌بخش هستند. این عبارات شامل کلماتی مانند "استرس دارم"، "مضطربم"، "نگرانم"، "احساس اضطراب می‌کنم" و مشابه این‌ها برای استرس و کلماتی مانند "خوشحالم"، "آرامش دارم"، "راحت هستم" برای آرامش بودند.

در مرحله بعد، با استفاده از این لیست از عبارات، جستجوهایی را در توییتر انجام دادیم. برای هر عبارت، توییت‌هایی که حاوی این کلمات بودند را استخراج کردیم. همچنین، در این مرحله از تنظیم قطبیت نیز استفاده کردیم. به این معنی که:

برای توییت‌هایی که حاوی عبارات استرس‌زا بودند، قطبیت را به طور عمده روی "منفی" یا "خنثی" تنظیم کردیم.

برای توییت‌هایی که حاوی عبارات آرامش‌بخش بودند، قطبیت را روی "مثبت" یا "خنثی" قرار دادیم.

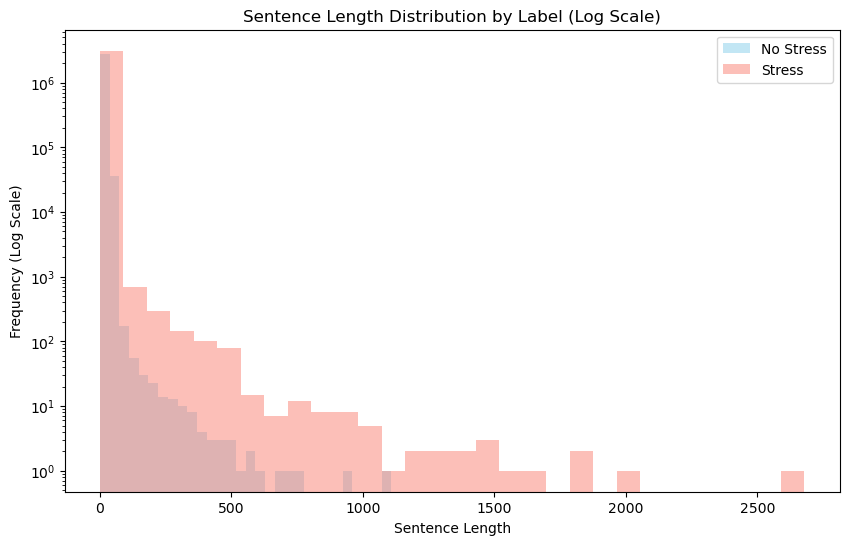
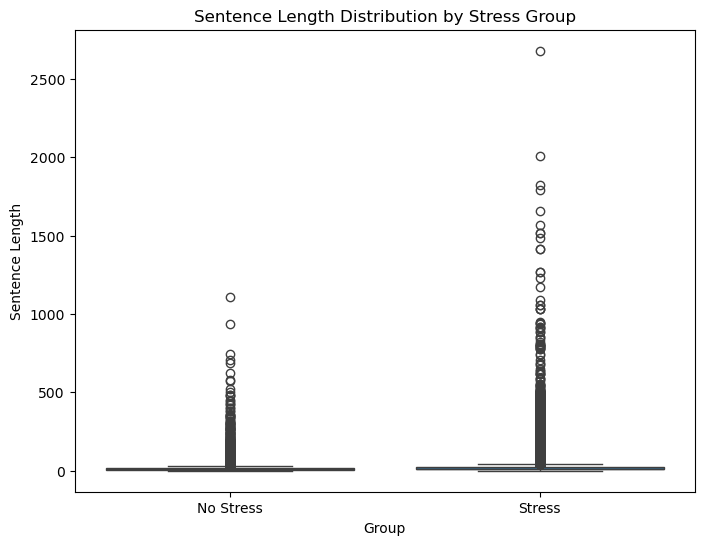
این کار به ما کمک کرد تا بتوانیم توییت‌هایی را که به وضوح نشان‌دهنده استرس یا آرامش بودند، شناسایی و جدا کنیم. بعد از این، داده‌های مورد نظر را از توییت‌های منتشر شده بین سال‌های 1397 تا 1403 استخراج کردیم که در مجموع حدود 6 میلیون داده از توییت‌ها به دست آمد. این داده‌ها به ما این امکان را داد که مدل‌هایی برای تشخیص استرس در متون ایجاد کنیم و از آن‌ها برای تحلیل وضعیت روان‌شناختی افراد در شبکه‌های اجتماعی استفاده کنیم.

در نهایت، این مجموعه داده‌ها، که شامل توییت‌های حاوی عبارات استرس‌زا و آرامش‌بخش است، به عنوان پایگاه داده برای آموزش مدل‌های تشخیص استرس استفاده شد و به ما این امکان را می‌دهد که رفتارهای زبانی مرتبط با استرس و آرامش را شبیه‌سازی کرده و مدل‌هایی برای شناسایی استرس در شبکه‌های اجتماعی ایجاد کنیم.

# تحلیل داده‌ها

## ساختار جملات دارای استرس

تحلیل طول جملات در داده‌های دو گروه استرس‌زا و غیر استرس‌زا نشان می‌دهد که تفاوت‌های زبانی قابل‌توجهی میان این دو گروه وجود دارد. در گروه جملات استرس‌زا، میانگین طول جمله برابر با ۱۵.۳۵ کلمه است، در حالی که در گروه جملات غیر استرس‌زا این مقدار به ۱۲.۴۴ کلمه کاهش می‌یابد. میانه‌ی طول جمله برای گروه استرس‌زا برابر با ۱۳ کلمه و برای گروه غیر استرس‌زا برابر با ۱۰ کلمه است. همچنین، پرتکرارترین طول جمله (مد) در گروه استرس‌زا ۸ کلمه و در گروه غیر استرس‌زا ۴ کلمه است.



نمودارهای هیستوگرام و جعبه‌ای توزیع طول جملات را در هر دو گروه نشان می‌دهند. در گروه استرس‌زا، توزیع طول جملات به سمت مقادیر بیشتر جابه‌جا شده است و پراکندگی بیشتری در داده‌ها مشاهده می‌شود. این در حالی است که در گروه غیر استرس‌زا، بیشتر جملات در طول‌های کوتاه‌تر متمرکز شده‌اند.

این نتایج نشان می‌دهد که پیچیدگی زبانی در جملات استرس‌زا بیشتر از جملات غیر استرس‌زا است. افراد در شرایط استرس ممکن است تلاش کنند احساسات یا وضعیت خود را با جزئیات بیشتری بیان کنند و از عبارات توصیفی یا ساختارهای زبانی پیچیده‌تر استفاده کنند. در مقابل، جملات کوتاه‌تر در گروه غیر استرس‌زا می‌تواند نشان‌دهنده‌ی تمایل به بیان ساده‌تر و مستقیم‌تر باشد، که با آرامش ذهنی هماهنگ است.

تفاوت میان میانه و مد نیز گویای این است که توزیع طول جملات در هر دو گروه کاملاً متقارن نیست و در گروه استرس‌زا برخی جملات بسیار بلند وجود دارند که میانه را به سمت مقادیر بالاتر جابه‌جا کرده‌اند.

## تحلیل زبانی کلمات پرتکرار در توییت‌های استرس‌زا

در این بخش، به تحلیل کلمات پرتکرار در توییت‌های استرس‌زا پرداخته‌ایم. هدف این تحلیل شناسایی ویژگی‌های زبانی است که بیشترین ارتباط را با احساسات استرس‌زا دارند. برای این منظور، از مدل TF-IDF برای استخراج کلمات کلیدی و محاسبه اهمیت آن‌ها در دو دسته توییت‌های استرس‌زا و بدون استرس استفاده کردیم. سپس این کلمات بر اساس اهمیت مرتب شدند تا کلمات با بیشترین تأثیر در تجربه استرس شناسایی شوند.

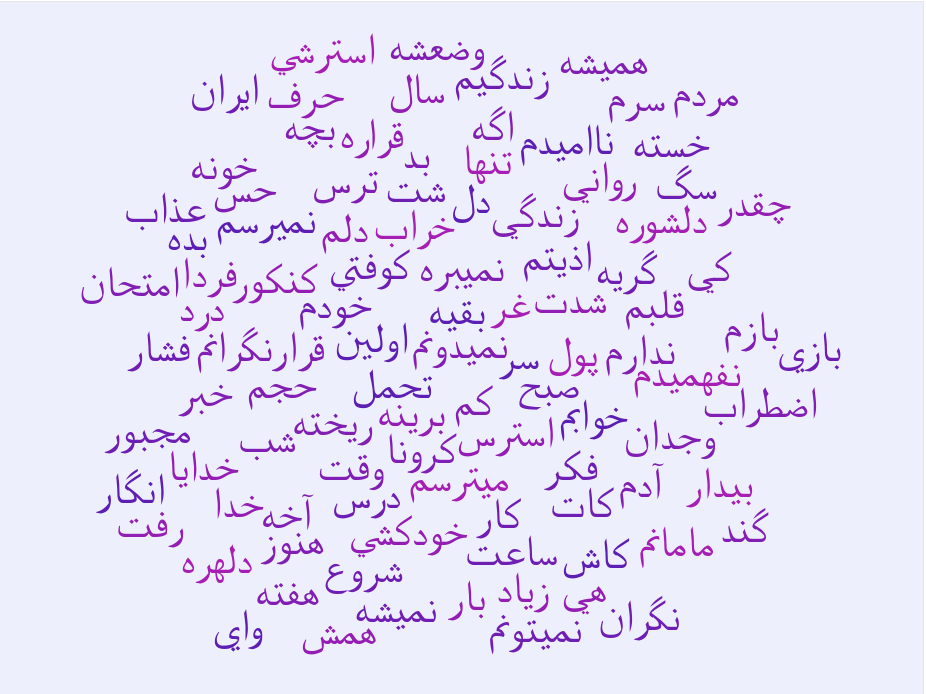
### روش تحلیل

ابتدا با استفاده از TF-IDF Vectorizer، کلمات پرتکرار در هر دو دسته توییت‌های استرس‌زا و بدون استرس استخراج شدند. سپس، امتیاز اهمیت هر کلمه در هر دسته محاسبه و نرمال‌سازی شد. این امتیازات نرمال‌شده نشان‌دهنده میزان تأثیر هر کلمه در ایجاد استرس یا عدم استرس در متون هستند. در نهایت، امتیازهای اهمیت کلمات استرس‌زا و بدون استرس با یکدیگر مقایسه شد تا امتیاز مطلق هر کلمه در توییت‌های استرس‌زا محاسبه گردد.

### یافته‌ها

در توییت‌های استرس‌زا، کلماتی مانند استرس، نگرانم، خوابم نمی‌بره، دلهره و ترس به‌طور مکرر ظاهر می‌شوند که همگی به‌طور مستقیم با احساسات منفی و اضطراب‌های روانی مرتبط هستند. این کلمات به وضوح اشاره دارند به فشارهای ذهنی، نگرانی‌ها و ترس‌هایی که افراد تجربه می‌کنند.

برای تحلیل بیشتر و نمایش بصری این کلمات، از ابر کلمات استفاده خواهیم کرد. ابر کلمات به‌عنوان ابزاری برای نمایش کلمات پرتکرار و مهم در یک فضای بصری طراحی می‌شود که می‌تواند توجه مخاطب را به عوامل اصلی و برجسته در داده‌ها جلب کند. در این ابر کلمات، اندازه هر کلمه نمایانگر میزان اهمیت آن در ایجاد استرس در توییت‌ها است.



برای تحلیل دقیق‌تر و تفسیر این کلمات کلیدی استرس‌زا، به چند نکته توجه می‌کنیم:

#### کلمات مرتبط با اضطراب و نگرانی

کلمات مانند **استرس**، **نگرانم**، **فشار**، **دلهره**، **اضطراب**، **ترس**، **نگران** و **درد** به وضوح به احساسات منفی و استرس‌زا اشاره دارند. این کلمات در اکثر مواقع با احساسات اضطراب، نگرانی و فشار ذهنی همراه هستند و می‌توانند به‌طور مستقیم بر روی کیفیت زندگی و رفاه روانی فرد تأثیر بگذارند. این کلمات نشان‌دهنده وضعیتی هستند که فرد تحت تأثیر استرس و فشار زیاد قرار دارد.

#### کلمات مرتبط با مشکلات جسمی و ذهنی

کلمات **خوابم نمی‌بره، کوفتی، درد، خسته، گرفتاری و چه وضعشه** اشاره به مشکلات جسمی و روانی ناشی از استرس دارند. این کلمات نشان می‌دهند که استرس نه تنها بر وضعیت روحی فرد تأثیر می‌گذارد، بلکه می‌تواند مشکلات جسمی مانند بی‌خوابی، خستگی و درد نیز به همراه داشته باشد. این اثرات جسمی معمولاً در افراد دچار استرس‌های مزمن دیده می‌شوند.

#### کلمات مرتبط با بی‌اعتمادی و ناکامی

کلمات مانند **ندارم، نمیدونم، نمی‌توانم، نمی‌شه، بی‌خیال و نمی‌تونم** نشان‌دهنده احساس بی‌اعتمادی به خود یا وضعیت زندگی فرد هستند. این کلمات معمولاً در مواقعی به کار می‌روند که فرد احساس می‌کند نمی‌تواند به اهداف خود دست یابد یا شرایط دشوار است و نمی‌تواند آن‌ها را مدیریت کند. این نوع احساسات می‌توانند بر روحیه فرد تأثیر منفی بگذارند و به افسردگی و ناامیدی منجر شوند.

#### کلمات مرتبط با زمان و آینده

کلمات **فردا، ساعت، صبح، شب و امتحان** نشان‌دهنده نگرانی‌های زمانی هستند. فرد ممکن است به طور مداوم درباره آینده و چالش‌هایی که پیش رو دارد، فکر کند. این نگرانی‌ها می‌توانند به استرس‌های روزانه و یا فشارهای ناشی از مسئولیت‌های پیش رو اشاره داشته باشند.

#### کلمات مرتبط با ترس از آینده و وضعیت‌های ناپایدار

کلمات **ترس، ناامیدم، خودکشی، کرونا، خراب، حجم** اشاره به نگرانی‌های بزرگ و اغلب در مورد مسائل شدید و بحرانی دارند. این کلمات نشان‌دهنده احساس ترس و اضطراب در مواجهه با بحران‌ها یا مشکلات بزرگ هستند.

#### کلمات مرتبط با تعاملات اجتماعی و حمایت

کلمات **خودم، خدا، خدایا، آدم، مردم و مامانم** نشان‌دهنده تعاملات اجتماعی و حمایتی فرد هستند. این کلمات می‌توانند به روابط فرد با دیگران یا حمایت‌های اجتماعی اشاره کنند که در مواقع استرس‌زا بسیار مهم هستند. فرد ممکن است به دنبال حمایت از خانواده یا دیگران باشد تا بتواند با استرس خود مقابله کند.

#### کلمات مرتبط با احساسات منفی شدید

کلمات **گند، واي، بازی، عذاب و گريه** نشان‌دهنده احساسات منفی شدید و ناراحتی هستند. این کلمات معمولاً در موقعیت‌هایی به کار می‌روند که فرد تحت فشار شدید روانی قرار دارد و از شرایط موجود راضی نیست. این احساسات می‌توانند به افزایش شدت استرس منجر شوند.

#### کلمات مرتبط با مسائل اجتماعی و فرهنگی

کلمات **ایران، کنکور، تحمل و پول** نشان‌دهنده فشارهای اجتماعی و فرهنگی هستند. این کلمات ممکن است به چالش‌ها و توقعات اجتماعی، فرهنگی و اقتصادی اشاره داشته باشند که فرد برای مواجهه با آن‌ها تحت استرس قرار دارد. برای مثال، فشارهای اجتماعی مانند امتحانات (مانند کنکور) یا مشکلات اقتصادی می‌توانند عامل استرس‌زا باشند.

این کلمات پرتکرار در توییت‌های استرس‌زا به وضوح نشان‌دهنده چالش‌ها و فشارهایی هستند که فرد در موقعیت‌های استرس‌زا تجربه می‌کند. تحلیل این کلمات می‌تواند کمک کند تا درک بهتری از عوامل ایجاد استرس و نگرانی در زندگی افراد به دست آورده و از آن در طراحی ابزارهای پیش‌بینی استرس یا ارائه راهکارهای درمانی و حمایتی استفاده شود. این تحلیل همچنین نشان می‌دهد که استرس در اکثر مواقع با مشکلات جسمی، ذهنی، اجتماعی و فرهنگی پیوند خورده است و می‌تواند بر تمامی جنبه‌های زندگی فرد تأثیر بگذارد. البته تحلیل‌های بیشتری در مورد کلمات پرتکرار توییت‌های استرسی می‌توان ارائه کرد که به خواننده واگذار می‌کنیم.

# مدل‌سازی و تشخیص استرس

## معرفی مدل‌های کلاسیک و یادگیری عمیق

| **مدل** | **داده** | **زبان** | **دقت** | **توضیحات** |
| --- | --- | --- | --- | --- |
| BERT | TensiStrength (Twitter) | انگلیسی | 83.6% | مدل BERT برای تشخیص استرس مزمن در توییت‌ها با کلمات کلیدی استرس تنظیم شده و از SVM بهتر عمل کرده است. |
| Logistic Regression + Bag of Words | Dreaddit (Reddit) | انگلیسی | دقت: 77.78%،  F1: 0.79 | با استفاده از ویژگی‌های Bag of Words در پست‌های آکادمیک Reddit برای تشخیص استرس به کار گرفته شده است. |
| Bidirectional LSTM (BLSTM) با Attention | Dreaddit (Reddit) | انگلیسی | مشخص نشده؛ به عنوان مؤثر گزارش شده | تمرکز بر جاسازی‌های پیوسته لغوی برای تشخیص استرس. |
| KC-Net | Dreaddit (Reddit) | انگلیسی | مشخص نشده؛ مؤثر برای تشخیص زودهنگام | شبکه یادگیری تضادآگاه از دانش که دانش حالت ذهنی را با یادگیری تضادآمیز ترکیب می‌کند. |
| Naive Bayes Classifier | Dreaddit (Reddit) | انگلیسی | دقت: 65% | به عنوان مدل پایه برای طبقه‌بندی استرس در پست‌های Reddit استفاده شده است. |
| Random Forest Classifier | Dreaddit (Reddit) | انگلیسی | دقت: 90% | روش جمعی (Ensemble) برای تشخیص استرس در دامنه‌های مختلف Reddit. |
| CNN-LSTM | TensiStrength (Twitter) و Reddit | انگلیسی | مشخص نشده؛ عملکرد قوی گزارش شده | ترکیب CNN برای استخراج ویژگی و LSTM برای وابستگی‌های متوالی. |

## استخراج ویژگی‌های متنی و مدل‌سازی تشخیص استرس

برای تشخیص استرس از متن‌های کاربران شبکه‌های اجتماعی، یکی از گام‌های اساسی، تبدیل متن‌های خام به داده‌های عددی قابل تحلیل است. این کار با استفاده از روش‌های آماری مانند TF-IDF انجام می‌شود. TF-IDF، ، تکنیکی برای سنجش اهمیت کلمات در متن‌ها نسبت به کل مجموعه داده است. این روش به مدل‌های یادگیری ماشین کمک می‌کند تا الگوهای نهفته در متن‌ها را شناسایی کنند و واژگان کلیدی مرتبط با استرس را تشخیص دهند.

### فیلتر و پالایش واژگان

یکی از چالش‌های تحلیل متن‌های فارسی، وجود نویزهای زبانی است. برای مقابله با این چالش، فرآیندی برای حذف کلمات غیرمرتبط، اعداد و کلمات بسیار کوتاه طراحی شد. همچنین، از یک لیست از پیش تعریف‌شده برای حذف کلمات توقف مانند «و»، «که»، و «در» استفاده شد. این فیلتر به مدل کمک می‌کند تا فقط بر واژگان معنادار تمرکز کند.

برای بهبود کیفیت واژگان، لیست ویژگی‌ها استخراج شده و به صورت دستی پالایش شد. در این فرآیند، واژگان کم‌کاربرد یا نامرتبط حذف و واژگان کلیدی برای تشخیص استرس شناسایی شدند.

## مدل‌سازی با استفاده از رگرسیون لجستیک

پس از آماده‌سازی داده‌ها، از یک مدل یادگیری ماشین ساده به نام رگرسیون لجستیک برای پیش‌بینی استرس استفاده شد. این مدل به دلیل سرعت بالا و شفافیت عملکرد، انتخاب شد. پس از آموزش مدل با داده‌های پردازش‌شده، مدل قادر بود میزان استرس موجود در متن‌های کاربران را تشخیص دهد.

### نتایج ارزیابی

برای ارزیابی عملکرد مدل، از معیارهایی مانند دقت (Accuracy)، فراخوان (Recall)، و دقت پیش‌بینی (Precision) استفاده شد. نتایج مدل به صورت زیر خلاصه شده است:

| **معیار** | **مقدار** |
| --- | --- |
| **دقت (Accuracy)** | 82 |
| **دقت پیش‌بینی (Precision)** | 84 |
| **فراخوان (Recall)** | 84 |
| **امتیاز F1** | 83.5 |

## طرز کار مدل فست‌تکست برای تشخیص استرس در متن‌ها

مدل فست‌تکست برای تشخیص استرس در متن‌ها با استفاده از مجموعه‌ای از داده‌های آموزشی و پارامترهای خاصی آموزش داده شد. ابتدا داده‌های آموزشی که شامل متون و برچسب‌های مربوط به آن‌ها بودند، به فرمت خاصی تبدیل شدند که مدل فست‌تکست قادر به پردازش آن‌ها باشد. این داده‌ها شامل برچسب‌های استرس‌زا و غیر استرس‌زا بودند که به متون اضافه شدند. مدل با استفاده از این داده‌ها ویژگی‌های معنایی و ساختاری متن‌ها را شناسایی کرده و الگوهایی را که می‌توانند نشان‌دهنده استرس باشند، یاد می‌گرفت.

برای ایجاد این مدل، از پارامترهای زیر استفاده شد:

تعداد دوره‌های آموزش (epoch=100): این پارامتر مشخص می‌کند که مدل چند بار باید روی داده‌ها آموزش ببیند. با تنظیم این مقدار به 100، مدل برای 100 دوره روی داده‌ها تمرین کرد تا الگوهای معنایی و ساختاری مناسب را یاد بگیرد.

نرخ یادگیری (lr=0.1): نرخ یادگیری مدل تعیین‌کننده سرعت به‌روزرسانی وزن‌های مدل در طول فرایند آموزش است. با تنظیم نرخ یادگیری به 0.1، مدل با سرعت معقولی به بهبود دقت خود پرداخت.

استفاده از ۴-گرام (wordNgrams=4): این پارامتر به مدل این امکان را می‌دهد که ترکیبات ۴ کلمه متوالی را به‌عنوان ویژگی‌ها در نظر بگیرد. استفاده از ۴-گرام به مدل کمک کرد تا روابط پیچیده‌تری میان کلمات و معانی آن‌ها شناسایی کند.

نمایش جزئیات آموزش (verbose=2): این تنظیم باعث شد که در طول فرایند آموزش، مدل گزارش‌های مفصلی از وضعیت آموزش و پیشرفت آن به‌طور پیوسته نمایش دهد.

پس از آموزش، مدل برای پیش‌بینی استرس در متون جدید استفاده شد. مدل فست‌تکست برای هر متن وارد شده، احتمال تعلق آن به دسته‌های مختلف (استرس‌زا یا غیر استرس‌زا) را پیش‌بینی کرد. این پیش‌بینی‌ها شامل برچسب‌ها و احتمال‌های مربوط به آن‌ها بودند که نشان می‌دهند چقدر مدل به این نتیجه رسیده است که متن مربوطه استرس‌زا است. در نهایت، برای ارزیابی عملکرد مدل از معیارهایی چون دقت، بازخوانی و F1-score استفاده شد تا میزان توانایی مدل در شناسایی متون استرس‌زا سنجیده شود.

نتایج حاصل از این مدل به شرح زیر است:

| **معیار** | **مقدار** |
| --- | --- |
| **دقت** | 0.89 |
| **بازخوانی** | 0.85 |
| **F1-Score** | 0.87 |

این نتایج نشان‌دهنده عملکرد خوب مدل در شناسایی متون استرس‌زا است. مدل توانسته است با دقت مناسبی، متون استرس‌زا را شناسایی کند و با استفاده از مقیاس‌های بازخوانی و F1-score، عملکرد قابل قبولی را در تشخیص استرس در متن‌ها از خود نشان دهد.

## استفاده از مدل BERT برای تشخیص استرس در متون فارسی

در این بخش، از مدل BERT برای تشخیص استرس در متون فارسی استفاده شده است. BERT (Bidirectional Encoder Representations from Transformers) یک مدل پیشرفته در پردازش زبان طبیعی است که به‌ویژه در مسائل طبقه‌بندی متن عملکرد بسیار خوبی از خود نشان داده است. مدل BERT فارسی که در این تحقیق مورد استفاده قرار گرفته، نسخه‌ی HooshvareLab/bert-fa-base-uncased است که به‌طور خاص برای پردازش متون فارسی آموزش داده شده است.

برای استفاده از این مدل، ابتدا داده‌ها باید به فرمت مناسب برای BERT تبدیل شوند. این تبدیل با استفاده از توکنایزر BERT انجام می‌شود. توکنایزر متون را به توکن‌ها (که می‌توانند کلمات یا بخش‌های کوچک‌تری از کلمات باشند) تقسیم کرده و سپس هر توکن به شناسه‌ای عددی تبدیل می‌شود. این شناسه‌ها به مدل این امکان را می‌دهند که متون را پردازش کرده و ویژگی‌های معنایی آن‌ها را استخراج کند. علاوه بر این، ماسک توجه (Attention Mask) برای مشخص کردن بخش‌های مهم متن و جلوگیری از پردازش بخش‌های غیرضروری به‌کار گرفته می‌شود.

فرآیند آموزش مدل شامل دو مرحله اصلی است. در مرحله اول، مدل BERT با لایه‌های ثابت (frozen) بارگذاری می‌شود و تنها لایه‌های نهایی آن که مسئول طبقه‌بندی هستند، آموزش می‌بینند. این مرحله به مدل کمک می‌کند که از دانش عمومی خود برای پردازش متن استفاده کرده و تنها برای تشخیص استرس در متن‌ها تنظیم شود. در مرحله دوم، لایه‌های BERT آزاد می‌شوند تا مدل بتواند برای شناسایی دقیق‌تر استرس، ویژگی‌های پیچیده‌تر و خاص‌تری را از متون استخراج کند.

برای جلوگیری از آموزش بیش از حد و اطمینان از بهبود عملکرد مدل، از تکنیک‌هایی مانند EarlyStopping و ModelCheckpoint استفاده شده است. EarlyStopping باعث می‌شود که در صورت عدم بهبود عملکرد مدل در مجموعه داده‌های ارزیابی، فرآیند آموزش متوقف شود. همچنین، با استفاده از ModelCheckpoint، بهترین مدل در طول فرآیند آموزش ذخیره می‌شود تا در صورت نیاز از آن استفاده شود.

نتایج حاصل از آموزش مدل BERT برای تشخیص استرس در متون فارسی به شرح زیر است:

| **معیار** | **مقدار** |
| --- | --- |
| **دقت** | 87 |
| **بازخوانی** | 85 |
| **F1-Score** | 86 |

## استفاده از مدل SVM برای تشخیص استرس در متون فارسی

مدل SVM (Support Vector Machine) یکی از الگوریتم‌های معروف در مسائل طبقه‌بندی است که برای شناسایی مرزهای تصمیم‌گیری میان کلاس‌های مختلف کاربرد فراوان دارد. این مدل به‌ویژه در داده‌های خطی و غیرخطی مؤثر است و می‌تواند به خوبی در شناسایی الگوهای پیچیده در متون به کار رود.

برای استفاده از مدل SVM در تشخیص استرس در متون فارسی، ابتدا ویژگی‌های متنی از طریق روش‌هایی مانند TF-IDF استخراج می‌شوند. این ویژگی‌ها نشان‌دهنده اهمیت کلمات در متن و میزان حضور آن‌ها در اسناد مختلف هستند. سپس داده‌ها به دو بخش آموزش و آزمون تقسیم می‌شوند و مدل SVM بر اساس داده‌های آموزشی مرزهای تصمیم‌گیری میان دو کلاس استرس و غیر استرس را یاد می‌گیرد.

مدل SVM با استفاده از داده‌های آموزش به‌طور خودکار بهترین مرز تصمیم‌گیری را تعیین کرده و سپس بر روی داده‌های آزمون ارزیابی می‌شود. این مدل به دلیل سادگی و کارایی بالا، در تشخیص استرس در متون فارسی نتایج قابل قبولی ارائه می‌دهد.

نتایج حاصل از استفاده از مدل SVM در تشخیص استرس به شرح زیر است:

| **معیار** | **مقدار** |
| --- | --- |
| **دقت** | 0.85 |
| **بازخوانی** | 0.83 |
| **F1-Score** | 0.84 |

## استفاده از روش Ensemble و ترکیب مدل‌ها برای تشخیص استرس در متون فارسی

در این تحقیق، علاوه بر مدل‌های جداگانه مانند BERT، SVM، رگرسیون لوجستیک و FastText، از یک روش انزِمبل برای ترکیب نتایج مدل‌ها استفاده شده است. در این روش، پیش‌بینی‌های چند مدل مختلف ترکیب می‌شوند تا عملکرد کلی بهبود یابد. برای این منظور، مدل BERT که بهترین دقت را داشت، دو برابر وزن بیشتری در فرآیند پیش‌بینی نسبت به مدل‌های دیگر دریافت کرد. این کار به این دلیل انجام شد که BERT قادر است ویژگی‌های پیچیده‌تر متنی را در مقایسه با مدل‌های دیگر استخراج کرده و دقت بالاتری را در شناسایی استرس ارائه دهد.

با این ترکیب، دقت نهایی مدل از دقت مدل BERT که از همه مدل‌ها بالاتر بود، نیز فراتر رفت. این نشان می‌دهد که ترکیب مدل‌ها با وزن‌دهی متفاوت به هرکدام، می‌تواند دقت پیش‌بینی را بهبود بخشد و شناسایی دقیق‌تر استرس در متون را ممکن سازد.

مدل‌های غیر BERT مانند SVM، رگرسیون لوجستیک و FastText سرعت آموزش و پیش‌بینی بالاتری دارند. بنابراین، استفاده از این مدل‌ها در ترکیب با BERT هیچ مشکلی از نظر زمان پردازش ایجاد نکرد و باعث شد که نه تنها دقت مدل افزایش یابد، بلکه مدل به‌طور کلی سریع‌تر و کارآمدتر شود.

نتایج حاصل از ترکیب مدل‌ها به شرح زیر است:

| **معیار** | **مقدار** |
| --- | --- |
| **دقت** | 90 |
| **بازخوانی** | 88 |
| **F1-Score** | 89 |

این نتایج نشان می‌دهند که استفاده از روش انزِمبل و وزن‌دهی متفاوت به مدل‌ها، توانسته است عملکرد مدل را به طور قابل توجهی بهبود بخشد و دقت پیش‌بینی را افزایش دهد.

# نتیجه‌گیری و مسیرهای آینده

## جمع‌بندی اهمیت موضوع و دستاوردها

تشخیص استرس از طریق تحلیل متون و نوشته‌های کاربران، راهکاری نوین در حوزه سلامت روان و تحلیل رفتار اجتماعی است. این روش، با بهره‌گیری از فناوری‌های هوش مصنوعی، به‌ویژه مدل‌های پردازش زبان طبیعی (NLP)، توانایی شناسایی دقیق‌تر وضعیت روانی افراد را فراهم می‌کند. اهمیت این موضوع به دلایل زیر برجسته است:

* پیشگیری و مدیریت استرس: شناسایی زودهنگام استرس می‌تواند به ارائه مداخلات به‌موقع و کاهش پیامدهای منفی آن منجر شود.
* تحلیل در مقیاس بزرگ: فناوری هوش مصنوعی امکان تحلیل گسترده‌تری را فراهم می‌کند که می‌تواند برای سیاست‌گذاران و متخصصان سلامت روان ابزار ارزشمندی باشد.
* کاربرد در جنگ شناختی و مدیریت بحران: تشخیص استرس در جوامع و افراد هدف می‌تواند به مقابله با تهدیدات شناختی و مدیریت بهتر بحران‌های روانی و اجتماعی کمک کند.

دستاوردهای این تحقیق شامل:

* ایجاد مجموعه‌داده‌ای جامع از توییت‌های فارسی مرتبط با استرس و آرامش.
* توسعه مدل‌هایی با دقت بالا برای تشخیص استرس از متن.
* تحلیل ویژگی‌های زبانی و معنایی مرتبط با استرس و آرامش در نوشته‌های کاربران.

## پیشنهادهایی برای تحقیقات آینده و بهبود روش‌ها

کنترل استرس و مقابله با تهدیدات جنگ شناختی به‌عنوان دو مسئله حیاتی در سلامت روان فردی و اجتماعی نیازمند رویکردهایی جامع و متنوع است. تحقیقات آینده در این حوزه می‌تواند بر توسعه ابزارهای پیشرفته، طراحی مداخلات هدفمند و ایجاد راهبردهای مؤثر برای تقویت تاب‌آوری شناختی متمرکز شود. یکی از اولویت‌های کلیدی در این زمینه، طراحی سامانه‌های پایش روان‌شناختی در مقیاس جمعی است. این سامانه‌ها، با استفاده از تحلیل زبانی و رفتاری کاربران در شبکه‌های اجتماعی، می‌توانند افزایش غیرطبیعی استرس یا گسترش پیام‌های اضطراب‌آفرین را شناسایی و به مسئولان هشدار دهند. تحلیل روندهای روانی در زمان واقعی، به‌ویژه در دوره‌های بحران اجتماعی یا سیاسی، می‌تواند به اتخاذ مداخلات سریع‌تر و هدفمندتر کمک کند.

علاوه بر شناسایی وضعیت روانی جمعیت‌ها، طراحی مداخلات فردی برای کاهش استرس نیز باید مورد توجه قرار گیرد. توسعه ابزارهای دیجیتال مانند چت‌بات‌های هوشمند که بتوانند با کاربران تعامل کرده و تکنیک‌های روان‌شناختی مؤثر نظیر مدیتیشن، تمرین‌های تنفس عمیق یا آموزش مهارت‌های مقابله‌ای را ارائه دهند، از جمله راهبردهای کاربردی است. همچنین، ایجاد پلتفرم‌های آموزشی که محتوای مرتبط با مدیریت استرس و ارتقای تاب‌آوری را در قالب‌های جذاب و قابل دسترس عرضه کنند، می‌تواند نقشی کلیدی در توانمندسازی افراد داشته باشد.در جنگ شناختی، آگاهی و تاب‌آوری افراد در برابر پیام‌های اضطراب‌آور دشمن بسیار اهمیت دارد. در این راستا، تولید محتواهای مقاومتی و آموزشی که افراد را با اثرات پیام‌های منفی و روش‌های مقابله با آن‌ها آشنا کند، ضروری به نظر می‌رسد. این محتواها می‌توانند از طریق رسانه‌های جمعی و شبکه‌های اجتماعی به‌صورت گسترده توزیع شوند و حس همبستگی و حمایت اجتماعی را تقویت کنند. علاوه بر این، توسعه مدل‌های تحلیل رفتارشناختی که بتوانند الگوهای حملات روانی دشمن را پیش‌بینی کرده و منابع این پیام‌ها را شناسایی کنند، می‌تواند دفاع روانی جوامع را بهبود بخشد.

یکی دیگر از زمینه‌های تحقیقاتی مهم، بررسی روندهای طولانی‌مدت تغییرات روانی در مواجهه با بحران‌ها است. این تحقیقات می‌توانند با تحلیل دینامیک استرس در بازه‌های زمانی طولانی، الگوهای رفتاری جمعیت‌ها را در واکنش به رویدادهای مختلف اجتماعی یا اقتصادی شناسایی کنند. نتایج این تحلیل‌ها می‌تواند مبنای طراحی راهکارهای پیشگیرانه و سیاست‌گذاری‌های کلان برای کاهش استرس در سطح جامعه قرار گیرد.

در زمینه فناوری، استفاده از مدل‌های پیشرفته زبانی مانند GPT-4 و LLama، امکان تحلیل معنایی عمیق‌تر و شناسایی الگوهای پیچیده زبانی را فراهم می‌کند. با این حال، این مدل‌ها باید از طریق تنظیم دقیق و آموزش با داده‌های محلی و بومی، متناسب با ویژگی‌های زبانی و فرهنگی جوامع هدف بهینه‌سازی شوند. همچنین، شفافیت در عملکرد و تفسیرپذیری نتایج این مدل‌ها برای افزایش اعتماد کاربران و متخصصان بسیار حیاتی است.

در نهایت، ترکیب علوم روانشناسی، زبان‌شناسی و فناوری اطلاعات در طراحی راهکارهای جامع‌تر برای مقابله با استرس و جنگ شناختی، می‌تواند نتایجی نوآورانه و عملی فراهم کند. همکاری میان‌رشته‌ای و مطالعات تطبیقی در جوامع مختلف به توسعه روش‌هایی متناسب با نیازهای فرهنگی و اجتماعی هر جامعه کمک خواهد کرد. این تلاش‌ها، نه‌تنها به بهبود سلامت روان افراد بلکه به تقویت مقاومت جمعی در برابر تهدیدات شناختی منجر خواهد شد.

# منابع

[1] S. Cohen, D. Janicki-Deverts, and G. E. Miller, “Psychological stress and disease,” *JAMA*, vol. 298, no. 14, pp. 1685-1687, Oct. 2007.

[2] B. S. McEwen, “Stress, adaptation, and disease. Allostasis and allostatic load,” *Ann N Y Acad Sci*, vol. 840, pp. 33-44, May 1998.

[3] S. Folkman, “Positive psychological states and coping with severe stress,” *Social Science & Medicine*, vol. 45, no. 8, pp. 1207-1221, Oct. 1997.

[4] I. C.-H. Fung, Z. T. H. Tse, and K.-W. Fu, “The use of social media in public health surveillance,” *Western Pac Surveill Response J*, vol. 6, no. 2, pp. 3-6, Jun. 2015.

[5] S. C. Guntuku *et al.*, “Tracking Mental Health and Symptom Mentions on Twitter During COVID-19,” *J Gen Intern Med*, vol. 35, no. 9, pp. 2798-2800, Sep. 2020.

[6] B. Wang, Y. Zhao, X. Lu, and B. Qin, “Cognitive distortion based explainable depression detection and analysis technologies for the adolescent internet users on social media,” *Front Public Health*, vol. 10, p. 1045777, 2022.

[7] “Stress detection using natural language processing and machine learning over social interactions,” *SpringerOpen*, vol. 9, no. 1, pp. 1-24.

[8] V. Arya and A. K. Mishra, “Machine learning approaches to mental stress detection: a review,” *Firouzabad Institute of Higher Education-Annals of Optimization Theory and Practice*, vol. 4, no. 2, pp. 55-67, 2021.

[9] “SMACS: Stress Management AI Chat System,” *SCITEPRESS*, vol. 2, pp. 167-174.

[10] S. Muñoz and C. A. Iglesias, “A text classification approach to detect psychological stress combining a lexicon-based feature framework with distributional representations,” *Information Processing & Management*, vol. 59, no. 5, p. 103011, Sep. 2022.

[11] M. Zhuang, D. Cheng, X. Lu, and X. Tan, “Postgraduate psychological stress detection from social media using BERT-Fused model,” *PLoS One*, vol. 19, no. 10, Oct. 2024.

[12] J. Zhang, H. Yin, J. Zhang, G. Yang, J. Qin, and L. He, “Real-time mental stress detection using multimodality expressions with a deep learning framework,” *Front Neurosci*, vol. 16, p. 947168, 2022.

1. https://ar5iv.labs.arxiv.org/html/1911.00133 [↑](#footnote-ref-0)
2. Fleiss’s Kappa [↑](#footnote-ref-1)
3. DAIC-WOZ (Distress Analysis Interview Corpus - Wizard of Oz) [↑](#footnote-ref-2)