



mph.sdcongress.ir

هشتمین همایش بین المللی مدیریت، روانشناسی و علوم انسانی با رویکرد توسعه پایدار

8th International Conference on
Management, Psychology & Humanities with sustainable development approach



روش جدید برای تشخیص استرس در محیط مجازی با استفاده از کلاس بندی SVM

تورج هاشمی^۱، مهدیه ملکی صدقی^۲

^۱ استاد گروه روانشناسی دانشگاه تبریز

^۲ دانشجوی کارشناسی ارشد روانشناسی بالینی دانشگاه تبریز

ایمیل: drconf@yahoo.com

چکیده

استرس حالتی است که ناشی از وارد شدن فشارهای روحی و جسمی به فرد است. استرس حالتی ناشی از فشار است با اندک اغمازی می توان استرس را همان فشار عصبی در نظر گرفت که انسان در مقابل محرک های ناسازگار بیرونی از خود بروز می دهد. تشخیص استرس یک موضوع در حال رشد در زمینه پردازش زبان طبیعی است. به منظور پیش بینی سلامت ذهنی برای بهره بردن از توسعه سیستم های توصیه و ارزیابی های خودکار سلامت ذهنی در تحقیقات قبلی، مطالعه تشخیص استرس اثبات شده است. علاوه بر این، کاربرد گسترده رسانه اجتماعی بعنوان یک منبع داده احتمالی برای توسعه چنین مدل هایی بکار گرفته شده است. در این مقاله تلاش شده است تا تشخیص دهد که آیا کاربران فضای مجازی تحت استرس هستند. لذا یک دیتاست مرتبط با Dreddit شامل پست های یک پلتفرم رسانه اجتماعی محبوب Reddit استفاده کردیم. ما یک مدل یادگیری ماشینی شامل ماشین بردار پشتیبان (SVM)، را پیشنهاد شد. ارزیابی نهایی مدل و درستی را نتیجه داد، و هر دو توسط SVM امتیاز داده شدند.

کلمات کلیدی: یادگیری ماشینی؛ تشخیص استرس؛ رسانه اجتماعی؛ پردازش زبان طبیعی

مقدمه

استرس به صورت یک احساس عاطفی یا تنش فیزیکی تعریف می شود که از هر رویداد ناامید کننده یا رویدادی نتیجه می شود که باعث عصبانیت یا عصبی شدن فرد می شود. استرس زیاد بر مغز و بدن تاثیر بدی می گذارد، اما استرس کم می تواند باعث بهبود عملکرد شود. استرس می تواند در تمام گروه های سنی وجود داشته باشد و تمام افراد می توانند در دوره های مختلف زندگی استرس را تجربه کنند. براساس داده های گالوپ در سال ۲۰۲۲، ده کشور بیشترین نرخ استرس را دارند که در شکل زیر مشاهده می کنید. سازمان بهداشت جهانی بیان کرده که سالانه هشتصد هزار نفر به دلیل استرس و افسردگی خودکشی می کنند. یکی از دلایل نرخ بالای خودکشی استفاده زیاد از رسانه های اجتماعی است زیرا این رسانه های سلامت ذهنی شخص را برهم می زنند. سازمان بهداشت جهانی در سال ۲۰۱۹ نرخ مرگ و میر در اثر خودکشی را در دنیا از هر ۱۰۰ هزار نفر، ۳٫۴ برآورد می کند. اعتیاد به رسانه های اجتماعی در دوره مدرن بسیار زیاد است و افراد زیادی تحت تاثیر رسانه های اجتماعی قرار دارند. افراد از رسانه های اجتماعی برای نشان دادن احساسات مختلف خود مانند ناراحتی، عصبانیت یا استرس استفاده می کنند. از طریق الگوریتم پردازش زبان طبیعی (NLP) می توان شرایط ذهنی یک

فرد را براساس متنی که در شبکه اجتماعی پست می کند، درک کنیم.

۱۰ کشور برتر با بیشترین نرخ استرس



شکل (۱) کشور های دارای بیشترین استرس

NLP داده های متنی مثل اسناد یا نوشته ها را با استفاده از روش های محاسباتی تحلیل می کند. هدف این تحقیق این است که برای کاهش نرخ خودکشی در افراد، یک سیستم تشخیص استرس اولیه ایجاد کند. در این تحقیق سعی شده است که استرس افراد را در هنگام نوشتن پست های آنها تشخیص داده شد. در این تحقیق از مجموعه داده Dreddit استفاده شد. از NLP برای پردازش متن و از شش الگوریتم یادگیری ماشینی ناوی بیز، ماشین بردار پشتیبان، درخت تصمیم، فراوانی اصطلاح-فراوانی سند معکوس، جنگل تصادفی و بسته کلمات برای طبقه بندی متن استفاده شد. برخی تحقیقات تلاش کردند تا با استفاده از یادگیری ماشینی و روش های مختلف دیگر استرس را تشخیص دهند. تحقیق موسسه ملی فناوری سوراتکل هند با استفاده از مجموعه داده

WESAD افراد با استرس و بدون استرس را مشخص می کند. در این تحقیق از KNN، درخت تصمیم، AB و SVM استفاده شد. تحقیق دیگر در دانشگاه چین بر روی ده نفر در ویب و توییتر برای یافتن بهترین روش تشخیص استرس آزمایش انجام داد. روش های استفاده شده در این تحقیق، LRC با امتیاز ۰٫۸۳، SVM با امتیاز ۰٫۸۲، RF با امتیاز ۰٫۸۴، و FGM با بیشترین رتبه ۰٫۹۳، هستند. انفورماتیک های پزشکی BMC و تصمیم گیری با استفاده از روش های یادگیری ماشینی برای تشخیص استرس تلاش کردند. موسسه علوم کامپیوتر و مهندسی صنعتی و مدیریتی در استکهلم سوئد برای سنجش استرس کارکنان شرکت های مختلف سعی کردند تا از نظریه های یادگیری ماشینی بر روی نقاط داده برای یک شرکت با نام لینکورا AB استفاده کنند. الگوریتم K-means و خوشه های دینامیکی تجمعی سلسله مراتبی آنها با استفاده از پلتفرم ابری گوگل نتایج امیدوار کننده ای را نشان می دهد. همچنین دانشگاه آزاد هلنیک نیز تحقیق درباره تشخیص استرس انجام داد و از روش های AB و یادگیری عمیق استفاده کرد. جدول زیر بخشی از تحقیقات انجام شده در زمینه تشخیص استرس را نشان می دهد.



mph.sdcongress.ir

هشتمین همایش بین المللی مدیریت، روانشناسی و علوم انسانی با رویکرد توسعه پایدار

8th International Conference on Management, Psychology & Humanities with sustainable development approach



مدل	داده	معیار	عملکرد
تحلیل تشخیص خطی	WESAD	درستی	۹۳
ناوری بیزی	۲۰۰ نفر	درستی	۸۶
LRC	ویبو	درستی	۸۷
KNN	WESAD	درستی	۸۳/۵۰

آماده سازی مجموعه داده و توسعه مدل ها در بخش های زیر بحث می شوند. در دنیای آنلاین ما، کاربران رسانه های اجتماعی روزانه تعداد باورنکردنی توییت، پست و پیام می دهند و ماهیت به هم پیوسته و پر اطلاعات زندگی ما استرس را برجسته تر می کند و به راحتی از همیشه قابل مشاهده است.

با بسیاری از پلتفرمها مانند توییتر، ردیت و فیس بوک، جامعه علمی به حجم عظیمی از دادهها برای مطالعه نگرانیها و استرسهای روزانه مردم در سراسر جهان دسترسی دارد. استرس یک پدیده تقریباً جهانی است و ما شواهدی مبنی بر شیوع و افزایش اخیر آن داریم. به عنوان مثال، انجمن روانشناسی آمریکا (APA) مطالعات سالانه ای را برای ارزیابی استرس در ایالات متحده انجام داده است از سال ۲۰۰۷۲ که تجربیات گسترده استرس مزمن را نشان می دهد. استرس یک تجربه ذهنی است که اثرات و حتی تعریف آن می تواند از فردی به فرد دیگر متفاوت است. به عنوان یک خط پایه، APA استرس را به عنوان واکنشی به تقاضاها و فشارهای موجود و آینده تعریف می کند که می تواند در حد اعتدال مثبت باشد. محققان سلامت و روانشناسی به طور گسترده ارتباط بین استرس بیش از حد و سلامت جسمی و روانی را مورد مطالعه قرار داده اند.

مواد و روش ها

طرح هایی که در اینجا پیشنهاد می شوند، مدل های مختلف پیش بینی یادگیری عمیق برای زبان انگلیسی هستند. مجموعه داده مورد استفاده از Dreddit بدست می آید و از Reddit شامل 190K پست از پنج دسته مختلف انجمن ردیت جمع آوری می شود. دلیل استفاده ما از مجموعه داده، فقدان مجموعه داده مربوط به اندونزی است که برای استفاده در ردیت در دسترس است. ما برای کاهش داده های غیرضروری از تعدادی رشته کد به نام «cleaning» استفاده می کنیم که حذف کلمه را متوقف می کند، علائم نگارشی را حذف می کند، URL را حذف می کند، حروف بزرگ را به کوچک تغییر می دهد، بازنشر پست و ایموجی را حذف می کند. پست های کاربری که در این تحقیق استفاده شده از داده های ردیت بر روی پست های ردیت انتخاب می شود. ستون متن و برچسب را از سایر ستون ها جدا می کنیم، زیرا می خواهیم پست ها را در مجموعه داده و برچسب ها را با توجه به پست ها آموزش دهیم.

ID	برچسب	پیام
0	بدون استرس	مثبت
1	بدون استرس	مثبت
2	با استرس	منفی
3	با استرس	منفی
4	بدون استرس	مثبت

این تحقیق پس از داده های متنی و داده های برچسب، متن ساده را در پیش پردازش داده ها اجرا می کند تا پست های موجود در مجموعه داده را پاکسازی کند تا اطمینان حاصل شود که ابر کامپیوتری می تواند متون را بدون هیچ مشکل یا زمان بندی یا بدون زمان بندی کند. زمانی که مراحل فوق انجام شوند، مجموعه داده مورد استفاده برای آموزش یا پردازش آماده است. این مجموعه داده را با متن ساده، کلمات توقف و روش های توکن سازی پردازش کرد.



mph.sdcongress.ir

هشتمین همایش بین المللی مدیریت، روانشناسی و علوم انسانی با رویکرد توسعه پایدار

8th International Conference on Management, Psychology & Humanities with sustainable development approach



مدل پیشنهادی

پست های قبلا حذف شده در مدل یادگیری ماشینی در این طبقه بندی در نظر گرفته می شوند تا عملکرد یادگیری ماشینی بهبود یابد. مجموعه داده آموزشی پست ها برای این مدل به انگلیسی از مدل های از قبل آموزش دیده یعنی ماشین بردار پشتیبان، درخت تصمیم، ناوی بیز، جنگل تصادفی، بسته کلمات و TF-IDF استفاده می کند. این روش ها توانایی طبقه بندی متن با تمام مدل های مورد استفاده را دارند تا داده های مربوط به اشخاص با استرس و بدون استرس را پیدا کنند. پست های پردازشی را به گروه A و گروه B تقسیم می کنیم. دلیلی که وجود دارد این است که بجای اینکه تمام پست ها در یک مدل کنار یکدیگر قرار گیرند، در دو گروه قرار می گیرند تا درک مدل ها آسانتر شود. در این گروه پس از اجرای متن ساده نسبت به پست های ردیت، به صورت تصادفی ۸۰٪ مجموعه داده انتخاب شد و آنها را در SVM، ناوی بیز و درخت تصمیم قرار گرفت.

ارزیابی مدل

هر شش مدل با تنظیمات و عدد نسبت ورودی مشابه یعنی ۸۰٪ آموزش و ۲۰٪ آزمایش به صورت جداگانه آموزش داده می شوند. برای گروه A، پس از عبور متون از طریق تابع متن ساده، توکن سازی و کلمات توقف انجام می شود و آموزش و آزمایش با SVM، ناوی بیز و درخت تصمیم انجام می شود. برای گروه B، پس از عبور متون از طریق تابع مشابه، آموزش و آزمایش با جنگل تصادفی، TF-IDF و بسته کلمات انجام می شود. SVM، داده های آموزشی را با اعداد برچسب دار «۱» و «۰» مشخص می کند، همانند آنهایی که استرس دارند یا آنهایی که استرس ندارند. ناوی بیزها با داده های آموزشی مشابه مشخص می شوند، داده ها از جمله داده های با برچسب را پردازش و ذخیره می کنند. درخت تصمیم با داده های آموزشی مشابه مشخص می شود، برچسب های داده را بررسی می کند. جنگل تصادفی با داده های آموزشی مشابه مشخص می شود و بر این اساس شاخه هایی را برای نتیجه «۱» یا «۰» ایجاد می کند. بسته کلمات با داده های آموزشی مشخص می شود و داده های آموزشی را به صورت کامل پردازش می کند، ساختارها و کلمات را از هر یک از پست ها در مجموعه داده آموزشی بررسی می کند. TF-IDF با داده های آموزشی مشخص می شود و فرایند از طریق TF با استفاده از فرمول TF انجام می شود. پس از اینکه تمام مدل ها آموزش دیدند، و ما تمام روش های مورد استفاده را با ایجاد یک مدل خط لوله ای وارد کردیم، در آنجا خروجی یک روش است که به صورت مستقیم با سایر روش ها استفاده می شود تا بتوان آن را به صورت مستقیم پردازش کند.

TP	FP
FN	TN

امتیاز F_1 و درستی برای ارزیابی در هر مدل محاسبه می شوند. پیش بینی درست باید مبتنی بر برچسب ها در ماتریس درهم ریختگی در شکل بالا باشد.

نتیجه و بحث

در آزمایش موردنظر از Dreddit استفاده شد. درستی نتایج تمام روش های مورد استفاده آزمایش شد. تعارض بین فردی: سوء استفاده و حوزه های اجتماعی پوسترهایی که در زیرمجموعه های مورد سوء استفاده قرار می گیرند، عمدتاً بازماندگان یک رابطه یا موقعیت آزاردهنده هستند که داستان ها و حمایت هایشان را به اشتراک می گذارند، در حالی که پوسترهایی در subreddit اجتماعی در مورد هر گونه مشکل در یک رابطه (اغلب اما نه منحصرأ عاشقانه) پست می کنند و به دنبال مشاوره برای نحوه مدیریت این موقعیت هستند. بیماری روانی: حوزه های اضطراب و اختلال استرس پس از سانحه پوسترهای موجود در این subreddits به دنبال مشاوره در مورد مقابله با بیماری روانی و علائم آن، به اشتراک گذاشتن حمایت و موفقیت، جستجوی تشخیص و غیره هستند. نیاز مالی، پوسترهای موجود در زیرمجموعه های مالی معمولاً از پوسترهای دیگر کمک مالی یا مادی می گیرند.

ارزیابی عملکرد

این مقاله عملکرد تمام روش های SVM، ناوی بیز و DT را همانند یادگیری ماشینی پیچیده آزمایش می کند تا پست های مجموعه داده مورد استفاده را طبقه بندی کند. ما پس از آزمایش عملکرد و بررسی امتیاز F_1 ، درستی، فراخوانی، و دقت تصمیم گرفتیم تا تلاش کنیم که گروه B با جایگذاری ورودی های دستی براساس پیش بینی ها عمل کند، به طوری که ما نتیجه واقعی مدل را بتوانیم مشاهده کنیم. نتیجه ورودی های دستی نشان داد که در TF-IDF، سه پاسخ درست و سه پاسخ نادرست وجود دارد، که پاسخ می دهد که چرا امتیاز TF-IDF کمتر از SVM، ناوی بیز و جنگل تصادفی است. نتایج نشان می دهند که به طور کلی SVM از نظر عملکرد با درستی ۷۰٪ و امتیاز F_1 ۸۵٪ بیشترین است، این موضوع نشان می دهد که SVM در پیش بینی استرس کاملاً شایسته است. جدول زیر را مشاهده کنید.

روش	فراخوانی	درستی	امتیاز	دقت
SVM	۹۰	۷۷	۸۵	۷۹
TF	۹۵	۶۵	۷۵	۶۴
RF	۸۵	۶۱	۶۳	۶۹
NB	۷۳	۶۹	۶۱	۶۲

نتیجه گیری

استرس یک تجربه تقریباً جهانی برای انسان است، به ویژه در دنیای آنلاین. در حالی که استرس می تواند یک محرک باشد، استرس بیش از حد با بسیاری از پیامدهای منفی سلامتی مرتبط است و شناسایی آن را در طیف وسیعی از حوزه ها مفید می کن با این حال، تحقیقات محاسباتی موجود معمولاً فقط استرس را در حوزه هایی مانند گفتار یا در ژانرهای کوتاه مانند توئیتر مطالعه می کند Dreddit را ارائه می کنیم، مجموعه متنی جدیدی از داده های رسانه های اجتماعی چند دامنه ای طولانی برای شناسایی استرس مجموعه داده ما شامل ۱۹۰ هزار پست از پنج دسته مختلف از جوامع Reddit است. ما علاوه بر این، ۳،۵ هزار بخش از پست های ۳K را با استفاده از Amazon Mechanical Turk برچسب گذاری می کنیم ما روش های یادگیری با نظارت اولیه را برای شناسایی استرس، هم عصبی و هم سنتی، ارائه می کنیم و پیچیدگی و تنوع داده ها و ویژگی های هر دسته را تحلیل می کنیم. تشخیص استرس یک زمینه تحقیقاتی که در حال تکامل است. در این تحقیق یک مدل یادگیری ماشینی برای پیش بینی استرس رسانه اجتماعی پیشنهاد می شود. روش های مورد استفاده در مدل های یادگیری ماشینی SVM، NB، DT، TF-IDF، جنگل تصادفی و بسته کلمات هستند. آزمایشات در این مقاله نشان می دهند که SVM عملکرد خوبی با امتیاز F_1 ۸۰٪، درستی ۷۵،۰۰٪، فراخوانی ۹۲،۰۰٪ و دقت ۷۱،۰۰٪ دارد تا یک ویژگی شخصی را پیش بینی کند. آزمایشات ما امتیازهای خوبی بدست آوردند اما مطابق با انتظار ما نیستند. بنابراین توسعه این پیش بینی های شخصی باید به صورت مداوم برای بهبود اثربخشی آنها توسعه یابد.

منابع

- [1] Bashir H, Bhat SA. Effects of Social Media on Mental Health: A Review.; 2017.
- [2] Turcan E, McKeown K. Dreddit: A Reddit dataset for stress analysis in social media. In arXiv preprint.; 2019.
- [3] Bobade P, Vani M. Stress Detection with Machine Learning and Deep Learning using Multimodal Physiological Data. Proceedings of the 2nd International Conference on Inventive Research in Computing Applications, ICIRCA 2020. Institute of Electrical and Electronics Engineers Inc. 2020; p. 51–57
- [4] Garg P, Santhosh J, Dengel A, Ishimaru S. Stress detection by machine learning and wearable sensors. International Conference on Intelligent User Interfaces, Proceedings IUI; 2021. p. 43–45
- [5] Ahuja R, Banga A. Mental stress detection in university students using machine learning algorithms. Procedia Computer Sci. Elsevier B.V.; 2019; p. 349–353



mph.sdcongress.ir

هشتمین همایش بین المللی
مدیریت، روانشناسی و علوم انسانی
با رویکرد توسعه پایدار

8th International Conference on
Management, Psychology & Humanities with sustainable development approach



- [6] Lin H, Jia J, Qiu J, Zhang Y, Shen G, Xie L, et al. Detecting stress based on social interactions in social networks. In IEEE Trans Knowl Data Eng. 2017. 29; p. 1820–1833
- [7] Li R, Liu Z Stress detection using deep neural networks. BMC Med Inform Decis Mak.; 2020. Available from: <https://doi.org/10.1186/s12911-020-01299-4>
- [8] Nijhawan T, Attigeri G, Ananthakrishna T Stress detection using natural language processing and machine learning over social interactions. In J Big Data.; 2022. Available from:
- [9] Walambe R, Nayak P, Bhardwaj A, Kotecha K Employing Multimodal Machine Learning for Stress Detection. 2021. J Healthc Eng. Available from:
- [10] Lal DK, Mathew C, Nazar J, Professor A, Jyoti K. Mental Stress Detection in Students using Machine Learning Algorithms. 2021.
- [18] Madjar N, Lindblom F. KTH SKOLAN FÖR ELEKTROTEKNIK OCH DATAVETENSKAP. Machine Learning implementation for Stress-Detection.; 2020.