پیش‌بینی افکار خودکشی بر روی داده‌های متنی شبکه‌های اجتماعی با استفاده از یادگیری عمیق

محمدجواد خدمتی یاوند[۱]، کاظم فولادی قلعه[۲]

[۱] دستیار پژوهشی آزمایشگاه پژوهشی فضای سایبر دانشگاه تهران؛ دانش‌آموخته مقطع کارشناسی مهندسی کامپیوتر، دانشکده مهندسی دانشکدگان فارابی دانشگاه تهران؛ دانشجوی کارشناسی ارشد مهندسی کامپیوتر گرایش هوش مصنوعی و رباتیک دانشگاه امام حسین (ع)، تهران، ایران
mjkhedmati@ihu.ac.ir

[۲] استادیار گروه مهندسی کامپیوتر، دانشکده مهندسی دانشکدگان فارابی دانشگاه تهران؛ سرپرست آزمایشگاه پژوهشی فضای سایبر دانشگاه تهران، ایران
kfouladi@ut.ac.ir

چکیده

افکار خودکشی یکی از مسائلی است که به شدت نیازمند توجه و پیشگیری است. در حال حاضر، شبکه‌های اجتماعی مانند توئیتر (ایکس) و ردیت به عنوان یک منبع غنی از داده‌های افراد برای مطالعه و تحلیل این مسئله مورد توجه قرار گرفته‌اند. در این مقاله، ما با استفاده از مدل Bidirectional LSTM راه‌حلی برای پیش‌بینی افکار خودکشی بر اساس داده‌های متنی این شبکه‌ها ارائه می‌دهیم. سیستم‌های هوش مصنوعی مختلف، از جمله سیستم‌های مبتنی بر متن، برای شناسایی افراد در معرض خطر خودکشی پیشنهاد شده‌اند. به منظور رسیدگی به این چالش، ما داده‌های متنی را جمع‌آوری، پردازش و تحلیل می‌کنیم. سپس با استفاده از سیستم‌های تشخیص خطر افکار خودکشی مبتنی بر پردازش زبان طبیعی (NLP)، مدل خود را آموزش می‌دهیم تا بتواند افکار خودکشی را پیش‌بینی کند. پس از آموزش مدل، با ارزیابی دقیق بر روی داده‌های تست، به دقت ۹۷ درصد در تشخیص متون با مضمون خودکشی از غیر خودکشی دست یافتیم. همچنین، از مدل‌های یادگیری عمیق و کلاسیک نیز بهره‌مند شده و نتایج به صورت جامع مقایسه و تحلیل شده‌اند. هدف اصلی تحقیق ما، دستیابی به عملکرد بهتر از کارهای تحقیقاتی قبلی است تا بتوانیم نشانه‌های اولیه را با دقت بالا تشخیص داده و از اقدام به خودکشی جلوگیری کنیم.

کلمات کلیدی: خودکشی، داده متنی، یادگیری ماشینی، یادگیری عمیق، Bidirectional LSTM

۱ مقدمه

در سطح جهانی، خودکشی یکی از علل اصلی مرگ‌ومیر به‌ویژه در میان جوانان است. از این رو شناسایی افراد در معرض خطر افکار خودکشی ضروری است [۶]، [۷]. ابزارهای سنتی برای ارزیابی خطر افکار خودکشی بر شناسایی عوامل خطر خودکشی مانند تشخیص‌های روانپزشکی، بی‌قراری و رفتار خودکشی متمرکز شده‌اند، اما توانایی این ابزارها برای پیش‌بینی افکار و رفتارهای خودکشی با استفاده از عوامل خطر خودکشی دچار شکاف شده است. بنابراین، استفاده از هوش مصنوعی برای توسعه ابزارهای ارزیابی دقیق افکار خودکشی پیشنهاد شده است. تاکنون چندین سیستم هوش‌مصنوعی برای تشخیص افکار خودکشی خاص اختلال در افراد مبتلا به افسردگی یا اسکیزوفرنی پیاده سازی شده است، در حالی که سایر سیستم‌های هوش‌مصنوعی شناسایی افکار خودکشی در میان کاربران رسانه‌های اجتماعی را هدف قرار داده‌اند. در واقع، بسیاری از این سیستم‌های هوش‌مصنوعی با استفاده از الگوریتم‌های یادگیری ماشینی یا یادگیری عمیق که بر روی انواع ویژگی‌های زبانی آموزش داده شده‌اند توسعه یافته‌اند. اگرچه سیستم‌های هوش‌مصنوعی، به‌ویژه سیستم‌های تشخیص افکار خودکشی مبتنی بر متن، عملکرد دلگرم‌کننده‌ای از خود نشان داده‌اند، ادغام معمول آنها در تنظیمات مراقبت‌های بهداشتی نیاز به ارزیابی و اعتبارسنجی بیشتری دارد. استفاده از این سیستم‌های مبتنی بر متن به دلیل آموزش ناهمگونی داده‌ها، ارزیابی کیفیت متناقض، عدم مقایسه با روش‌های استاندارد بالینی و عدم وجود ارزیابی‌های قابل اعتماد با مشکل مواجه شده است. یکی از گام‌های اولیه برای غلبه بر این مشکلات، ایجاد یک مجموعه‌داده معیار از نمونه‌های متنی است که می‌تواند بر اساس روشی استاندارد و سیستماتیک جمع‌آوری شود. هدف ما پیشنهاد روشی است که بتواند با استفاده از داده‌های شبکه‌های اجتماعی ایکس (توئیتر) و ردیت، افکار منجر به خودکشی را پیش‌بینی کند. با ترکیب روش‌های داده‌کاوی و یادگیری ماشینی، ما مدل‌هایی را آموزش می‌دهیم که قادرند بر اساس ویژگی‌های استخراج شده از پست‌ها/توئیت‌ها، افکار خودکشی را تشخیص دهند و آن را پیش‌بینی کنند. در این مقاله، ابتدا به بررسی پژوهش‌های قبلی در حوزه پیش‌بینی افکار خودکشی با استفاده از داده‌های ایکس (توئیتر) و ردیت می‌پردازیم. سپس، روش پیشنهادی خود را شرح می‌دهیم که شامل مراحل جمع‌آوری، پردازش و استخراج ویژگی‌ها، ساختار مدل و الگوریتم‌های استفاده شده است. سپس، نتایج آزمایش‌ها را بررسی و تجزیه و تحلیل می‌کنیم و مقایسه‌ای بین الگوریتم‌ها ارائه می‌دهیم.

۲ مروری بر کارهای مشابه

در سال‌های اخیر، رسانه‌های اجتماعی به‌طور فزاینده‌ای محبوب شده‌اند و تعداد کاربران فعال همچنان در حال افزایش است. اکنون چندین پدیده مانند خودکشی در شبکه‌های اجتماعی قابل مشاهده است. برای رسیدگی به خودکشی و کاهش میزان مرگ و میر مرتبط، آزمایش‌ها و پژوهش‌های قابل توجهی در تأکید بر قدرت تأثیرگذاری رسانه‌های اجتماعی بر افکار خودکشی ایجاد شده است. با پیشرفت‌های اخیر مدل‌های شبکه‌های عصبی در پردازش زبان طبیعی، سهم جدیدی در تشخیص افکار خودکشی از پیاده‌سازی معماری‌های یادگیری عمیق پیچیده‌تر برای پیشی گرفتن از سیستم‌های یادگیری ماشین سنتی‌تر پدید آمده

است. شبکه عصبی بازگشتی (RNN) به خوبی برای مدل سازی توالی طراحی شده است [۱]. به طور خاص، حافظه کوتاه مدت طولانی (LSTM) یکی از مدل های مؤثری است که می‌تواند اطلاعات مفید را از وابستگی طولانی مدت حفظ کند.

رامیت ساونی، پراچی مانچاندا، پونیت ماتور، راج سینگ و راجیو راتن شاه در تحقیقات خود نشاند داند به کمک مدل C-LSTM در مقایسه با سایر طبقه‌بندی‌کننده‌های یادگیری عمیق از جمله LSTM و RNN و یادگیری ماشین مانند رگرسیون لجستیک و ماشین بردار پشتیبان به دقت ۸۱٫۲ درصد برای تشخیص افکار خودکشی برسند [۲].

شاکسینگ جی، سلینا پینگ یو، سای فو فنگ، شیروی پان و گودونگ لونگ شش طبقه‌بندی‌کننده شامل چهار طبقه‌بندی‌کننده نظارت‌شده سنتی و دو مدل شبکه عصبی را با هم مقایسه می‌کنند. آنها به‌جای استفاده از مدل‌های پایه با ویژگی‌های ساده برای تشخیص افکار خودکشی، ویژگی‌های اطلاعاتی را از چندین دیدگاه، از جمله ویژگی‌های آماری، نحوی، زبانی، ویژگی‌های جاسازی کلمه، و ویژگی‌های موضوعی را شناسایی می‌کنند. آنها طبقه‌بندی‌کننده‌های مختلف از هر دو دیدگاه یادگیری عمیق و سنتی، مانند ماشین بردار پشتیبان، جنگل تصادفی، درخت طبقه‌بندی تقویت‌شده با گرادیان (GBDT)، XGBoost و شبکه عصبی پیش‌خور چندلایه (MLFFNN) را با مدل شبکه عصبی حافظه کوتاه‌مدت طولانی (LSTM) مقایسه می‌کنند و معیارهایی را برای تشخیص افکار خودکشی در SuicideWatch Reddit برای ارتباط در مورد خودکشی با دقت ۹۱٫۰۸ درصد برای مدل LSTM ارائه می‌دهند [۳].

در سال ۲۰۱۹، آتیکا مبارک، عبدالمجید بن حمادو و سلما جاموسی، تحقیقاتی در زمینه شناسایی پروفایل‌های کاربرانی که در معرض خطر خودکشی هستند، انجام دادند. این تحقیق بر روی جامعه‌ای که در شبکه اجتماعی توئیتر فعالیت می‌کردند، تمرکز داشت. آنها یک مدل تشخیصی با استفاده از مجموعه‌ای از ویژگی‌های گوناگون مانند زبانی، احساسی، چهره، خط زمانی و ویژگی‌های عمومی تعریف کردند تا بتوانند پروفایل‌های توئیتر را شناسایی کنند. آنها چندین روش یادگیری ماشین را برای این امر آزمایش کردند و نتیجه حاصل از استفاده از مدل طبقه‌بندی‌کننده جنگل تصادفی، دقتی به میزان ۸۳ درصد داشت [۴].

تهازن ح. الدیانی، صالح نگی السوباری، علی صالح الشبامی، حسن الکهطانی و زیاد ع.ت احمد، در تحقیقات خود یک روش مبتنی بر تحقیقات تجربی برای ساختن یک سیستم تشخیص افکار خودکشی با استفاده از مجموعه داده‌های عمومی Reddit، رویکردهای جاسازی کلمه، مانند TF-IDF و Word2Vec، برای نمایش متن، و الگوریتم‌های یادگیری عمیق ترکیبی و یادگیری ماشین پیشنهاد کردند. برای طبقه‌بندی یک شبکه عصبی کانوولوشنال و مدل حافظه کوتاه‌مدت طولانی دوجهته (CNN-BiLSTM) و مدل یادگیری ماشینی XGBoost برای طبقه‌بندی پست‌های اجتماعی به‌عنوان خودکشی یا غیر خودکشی با استفاده از ویژگی‌های متنی و مبتنی بر LIWC-22 با انجام دو آزمایش استفاده شد. مقایسه نتایج آزمایش نشان داد که هنگام استفاده از ویژگی‌های متنی، مدل CNN-BiLSTM از مدل XGBoost بهتر عمل می‌کند و به دقت تشخیص افکار خودکشی ۹۵ درصد در مقایسه با دقت ۹۱٫۵ درصد دست می‌یابد. برعکس، هنگام استفاده از ویژگی‌های LIWC، XGBoost عملکرد بهتری نسبت به CNN–BiLSTM نشان داد [۱۰].

در مقاله امیر حسین یزدآور، محمد سعید مهدوی نژاد، گونمیت باجاج، ویلیام رومین، آمیت شث، امیر

حسن منجمی، کریشناپراساد تیروناراین، جان ام مددار، آنی مایرز، جیوتیشمن پاتاک، و پاسکال هیتزلر، با وجود تناقضات مطالعات مشاهده‌ای سنتی که از طریق پرسشنامه و نظرسنجی‌های خودگزارش‌دهی انجام می‌شود، به تشخیص علائم افسردگی از طریق توئیت‌ها پرداخته شده است. آنان از داده‌های بزرگ چندوجهی برای تشخیص رفتارهای افسردگی با استفاده از تحلیل و بهره‌برداری از ویژگی‌های گوناگون از جمله جمعیت‌شناسی در سطح فردی استفاده کرده‌اند. با پیشرفت چارچوب چندوجهی و بهره‌گیری از تکنیک‌های آماری برای ترکیب مجموعه‌های ناهمگون از ویژگی‌های به‌دست‌آمده از پردازش داده‌های بصری، متنی و تعامل کاربر، آنها رویکردهای پیشرفته فعلی را برای شناسایی افراد افسرده در توئیتر را به ۹۰ درصد بهبود دادند [۸].

در سال ۲۰۲۱، نینگ وانگ، فن لو، یووراج شیوتاره، وارشا د بدا، سابالکشمی، ر.چندرامولی و الن لی یک ساختار یادگیری عمیق را برای تشخیص افکار خودکشی معرفی نمودند. آن‌ها سه مدل یادگیری ماشین را برای تشخیص خودکشی در افرادی که در فواصل زمانی ۳۰ روز و ۶ ماه قبل اقدام به خودکشی کرده‌اند، با استفاده از داده‌های پست‌های رسانه‌های اجتماعی بررسی کردند. با استناد به تئوری سه مرحله‌ای خودکشی و مبتنی بر کارهای پیشین در زمینه احساسات و استفاده از ضمایر در افراد مبتلا به افکار خودکشی، آن‌ها سه مجموعه از ویژگی‌های دست‌ساز برای تشخیص خطر خودکشی تدوین و استخراج کردند. نتایج آزمایش‌های گسترده نشان داد که برخی از روش‌های سنتی یادگیری ماشینی با امتیاز F-1 برابر با ۰٫۷۴۱ و امتیاز F-2 برابر با ۰٫۸۳۳ در زیرکار ۱ (پیش‌بینی اقدام به خودکشی ۳۰ روز قبل)، از روش پایه بهتر عمل کردند. هرچند، روش یادگیری عمیق با امتیاز F-1 برابر با ۰٫۷۳۷ و امتیاز F-2 برابر با ۰٫۸۴۳ در زیرکار ۲ (پیش‌بینی خودکشی ۶ ماه قبل)، نسبت به روش پایه کارایی بهتری از خود نشان داد [۹].

مایکل مسفین تادس، هونگ فای لین، بو شو و لیانگ یانگ، از یک مدل ترکیبی LSTM-CNN برای ارزیابی و مقایسه با سایر مدل‌های طبقه‌بندی جهت پیش‌بینی افکار خودکشی کاربران Reddit استفاده نمودند. تحقیقات نشان می‌دهد که معماری شبکه عصبی ترکیبی با تکنیک‌های جاسازی کلمه می‌تواند بهترین نتایج طبقه‌بندی ارتباط را به دست آورد. علاوه بر این، نتایج تحقیقات آنها از قدرت و توانایی معماری‌های یادگیری عمیق برای ساختن یک مدل مؤثر برای ارزیابی خطر خودکشی در وظایف مختلف طبقه‌بندی متن پشتیبانی می‌کند. در نهایت آنها توانستند مدل خود را با دقت ۹۳٫۴ درصد به سرانجام رسانند [۵].

در سال ۲۰۲۲، رضا الحق، نعیم الاسلام، مایدالاسلام و مدد منجور الاحسن، با ارائه یک تحلیل مقایسه‌ای، به معرفی چندین مدل یادگیری ماشینی و یادگیری عمیق برای شناسایی افکار خودکشی از پلتفرم رسانه اجتماعی توئیتر پرداختند. هدف اصلی این تحقیق، به دست آوردن یک مدل با عملکرد بهتر از کارهای پژوهشی قبلی بود تا بتواند نشانه‌های اولیه را با دقت بالا تشخیص داده و از اقدام به خودکشی جلوگیری نماید. در این تحقیق، از پیش پردازش متن و رویکردهای استخراج ویژگی مانند CountVectorizer و جاسازی کلمه استفاده شد و چندین مدل یادگیری ماشینی و یادگیری عمیق برای آموزش به‌کار گرفته شد. آزمایش‌های این تحقیق بر روی مجموعه داده‌ای شامل ۴۹۱۷۸ نمونه انجام گرفت که این نمونه‌ها از توئیت‌های زنده توسط ۱۸ کلمه کلیدی مربوط به خودکشی و غیرخودکشی با استفاده از Python Tweepy API بازیابی شده‌اند. یافته‌های این تحقیق نشان داد که مدل جنگل تصادفی می‌تواند بالاترین امتیاز طبقه‌بندی را در بین

الگوریتم‌های یادگیری ماشین با دقت ۹۳ درصد و امتیاز F-1 برابر با ۰/۹۲ به‌دست آورد. به‌علاوه، آموزش طبقه‌بندی‌کننده‌های یادگیری عمیق با جاسازی کلمه، عملکرد مدل‌های یادگیری ماشینی را افزایش می‌دهد، به‌گونه‌ای که مدل BiLSTM به یک دقت ۹۳/۶ درصد و امتیاز F-1 برابر با ۰/۹۳ رسید [۱۱].

در سال ۲۰۱۹، یک تیم پژوهشی شامل سواتی جین، سورج پراکاش نارایان، روپش کومار دوانگ، اوتکارش بهارتیا، نالینی مینا و وارون کومار، سیستمی را برای تجزیه و تحلیل افسردگی و تشخیص افکار خودکشی پیشنهاد کردند. این سیستم با استفاده از داده‌های بلادرنگ دانش‌آموزان و والدین، با هدف پیش‌بینی اقدامات خودکشی بر اساس سطح افسردگی، طراحی شده است. با پر کردن پرسش‌نامه‌هایی مشابه PHQ-9 (پرسشنامه سلامت والدین)، داده‌های معنی‌دار با ویژگی‌های مرتبط مانند سن، جنس، نظم در مدرسه و غیره جمع‌آوری شده و با استفاده از الگوریتم‌های ماشین طبقه‌بندی، آموزش داده شده است. سپس با توجه به‌شدت افسردگی، در پنج مرحله طبقه‌بندی شده‌اند: حداقل یا هیچ، خفیف، متوسط، متوسط شدید و شدید. حداکثر دقت یعنی ۸۳/۸۷ درصد با استفاده از طبقه‌بندی‌کننده XGBoost در این مجموعه داده به دست آمده است. همچنین، داده‌ها در قالب توئیت نیز جمع‌آوری شده و با استفاده از الگوریتم‌های طبقه‌بندی، دسته‌بندی شد که آیا فردی که توئیت کرده، در افسردگی است یا خیر. طبقه‌بندی‌کننده رگرسیون لجستیک حداکثر دقت را به‌عنوان ۸۶/۴۵ درصد برای همان نشان می‌دهد [۶].

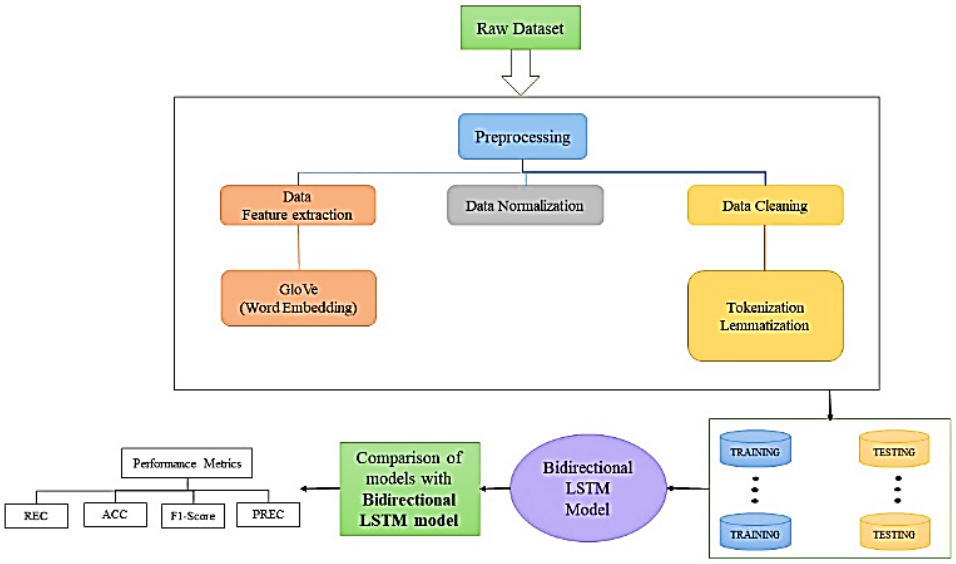
سماح فوده، تایهوا لی، کوین منچینسکی، تد بورگت، اندرو هریس، جورجتا ایلیتا، ساتیان رائو، جاناتان جمل و دانیلا رایکو در سال ۲۰۱۹، ۱۲۰۶۶ توئیت عمومی را از ۳۸۷۳ کاربر از طریق رابط برنامه‌نویسی برنامه توئیتر (API) جمع‌آوری کردند. آنها برچسب‌های «خطر بالا» یا «در معرض خطر» را برای کاربران بر اساس استفاده از اصطلاحات افکار خودکشی ایجاد نمودند و از سه الگوریتم کشف موضوع برای یافتن عوامل خطر خودکشی در میان کاربران استفاده کردند، که متعاقباً برای طبقه‌بندی کاربران به «خطر بالا» یا «در خطر» استفاده شد. الگوریتم‌های اعمال شده شامل تحلیل معنایی پنهان، تخصیص دیریکله پنهان، فاکتورگیری ماتریس غیرمنفی، درخت تصمیم و خوشه‌بندی K-means بود. آنها با استفاده از یک مدل طبقه‌بندی درخت تصمیم توانستند به ۸۴/۴ درصد در دقت، نمره ۰/۹۱۲ در حساسیت و ۰/۸۲۹ در طبقه‌بندی کاربران به گروه‌های "HighRisk" و "AtRisk" دست یابند [۷].

در سال ۲۰۲۲، عثمان نسیم، مطلوب خوشی، جینمن کیم و آدام جی، دست به ارائه یک روش نوین در شناسایی خطر خودکشی در رسانه‌های اجتماعی زده‌اند. این روش، یک روش نمایش متن ترکیبی است که نمایش‌های متنی در سطح کلمه و سند را برای توضیح شناسایی خطر خودکشی در رسانه‌های اجتماعی ادغام می‌کند. آنها پس از جمع‌آوری داده‌های مربوط به خودکشی در رسانه‌های اجتماعی، با استفاده از الگوریتم‌های ماشین طبقه‌بندی، روش پیشنهادی خود را به یک رمزگذار مبتنی بر ترانسفورمر با طبقه‌بندی ترتیبی برای تعیین خطر خودکشی داده می‌شود. نتایج نشان می‌دهد که روش پیاده‌سازی شده با نمره F-Score برابر با ۰/۷۹ در مجموعه داده‌های خودکشی عمومی، بهتر از مدل پایه است [۱۲].

جدول ۱ خلاصه‌ای از مطالب ذکر شده در این بخش را نمایش می‌دهد.

جدول ۱: خلاصه‌ای از کارهای مشابه بررسی شده

ردیف	عنوان مقاله	مدل استفاده شده	دقت/ نمره F-1
۱	[2] Exploring and Learning Suicidal Ideation Connotations on Social Media with Deep Learning	C-LSTM	۸۱٫۲
۲	[3] Supervised Learning for Suicidal Ideation Detection in Online User Content	LSTM	۹۱٫۰۸
۳	[4] Suicidal Profiles Detection in Twitter	Random Forest	۸۳
۴	[5] Detection of Suicide Ideation in Social Media Forums Using Deep Learning	LSTM-CNN	۹۳٫۴
۵	[6] A Machine Learning based Depression Analysis and Suicidal Ideation Detection System using Questionnaires and Twitter	Logistic Regression	۸۶٫۴۵
۶	[9] Learning Models for Suicide Prediction from Social Media Posts	C-Att, KNN	۰٫۷۳۷, ۰٫۷۴۱
۷	[10] Detecting and Analyzing Suicidal Ideation on Social Media Using Deep Learning and Machine Learning Models	CNN-BiLSTM	۹۵
۸	[11] A Comparative Analysis on Suicidal Ideation Detection Using NLP, Machine, and Deep Learning	Random Forest	۹۳
۹	[8] Multimodal mental health analysis in social media	N/A	۹۰
۱۰	[7] Using Machine Learning Algorithms to Detect Suicide Risk Factors on Twitter	Random Forest	۸۴٫۴
۱۱	[12] Hybrid Text Representation for Explainable Suicide Risk Identification on Social Media	Transformer	۰٫۷۹

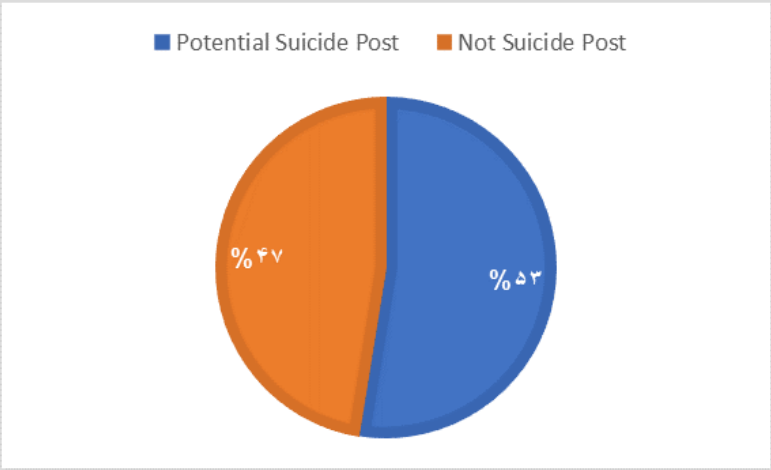


شکل ۱: فلوچارت روش‌شناسی

۳ روش پیشنهادی

در ابتدا، به عنوان بخشی از مرحله پیش‌پردازش داده‌ها، متن‌ها را به حروف کوچک تبدیل کرده و سپس با حذف نشانی‌های وب (URLها) و کاراکترهای خاص، داده‌ها به فرمی پاک و فیلتر شده تبدیل می‌شود. در مرحله بعد، متن را نشانه‌گذاری کرده و برای کاهش کلمات به شکل پایه خود، فرآیند ریشه‌یابی و واژه‌سازی را انجام می‌دهیم. این مرحله به کاهش ابعاد داده‌ها و بهبود کارایی مدل‌ها کمک می‌کند.

برای استخراج ویژگی‌ها و بردارسازی کلمات، از GloVe-100d بهره می‌گیریم که به‌جای یادگیری ویژگی‌های واژه‌ها از طریق شبکه‌های عصبی، از فراوانی کلمات در یک مجموعه‌داده بزرگ استفاده می‌کند. همچنین، همان‌گونه که در شکل ۱ قابل مشاهده است، به کمک مدل Bidirectional LSTM اقدام به مدل‌سازی و پیش‌بینی افکار خودکشی از روی داده‌های متنی می‌کنیم. در روند آموزش مدل Bidirectional LSTM، چندین مدل یادگیری ماشین، از جمله رگرسیون لجستیک، ماشین بردار پشتیبان، جنگل تصادفی، نایو بیز چندجمله‌ای، یادگیری گروهی و مدل تقویت سازگار، به‌همراه چندین مدل یادگیری عمیق مانند شبکه عصبی حافظه‌ی کوتاه‌مدت طولانی ساده (LSTM) و شبکه عصبی واحد بازگشتی دروازه‌گذاری‌شده (GRU) نیز تحت آموزش قرار می‌گیرند تا بتوانیم مدل‌های آموزش‌دیده را ارزیابی و مقایسه کنیم. در نهایت، مدل‌هایی که مورد استفاده قرار گرفته‌اند، با توجه به معیارهای ارزیابی چون Accuracy، Precision، Recall و F1-score ارزیابی می‌شوند تا عملکرد آنها به‌دقت بررسی شود.



شکل ۲: میزان سهم هر گروه از متن‌ها با مضمون خودکشی و غیرخودکشی

۱.۳ جمع‌آوری

هدف از آموزش مدل‌ها، تشخیص احتمال اقدام به خودکشی بر اساس متن منتشر شده توسط کاربران شبکه‌های اجتماعی ایکس (توئیتر) و ردیت است. بدین منظور دو گروه متن استخراج شده است که این مجموعه‌داده را گروهی از متن‌ها با مضمون خودکشی و گروهی دیگر خنثی و مثبت تشکیل داده است. در مجموع ۱۱۳۳۸ متن انگلیسی که کاربران انتشار داده‌بودند جمع‌آوری شده است که سهم گروه اول ٪۵۳ شامل ۵۹۵۸ متن و گروه دوم ٪۴۷ شامل ۵۳۸۰ متن می‌باشد (شکل ۲).

۲.۳ مدل یادگیری کلاسیک

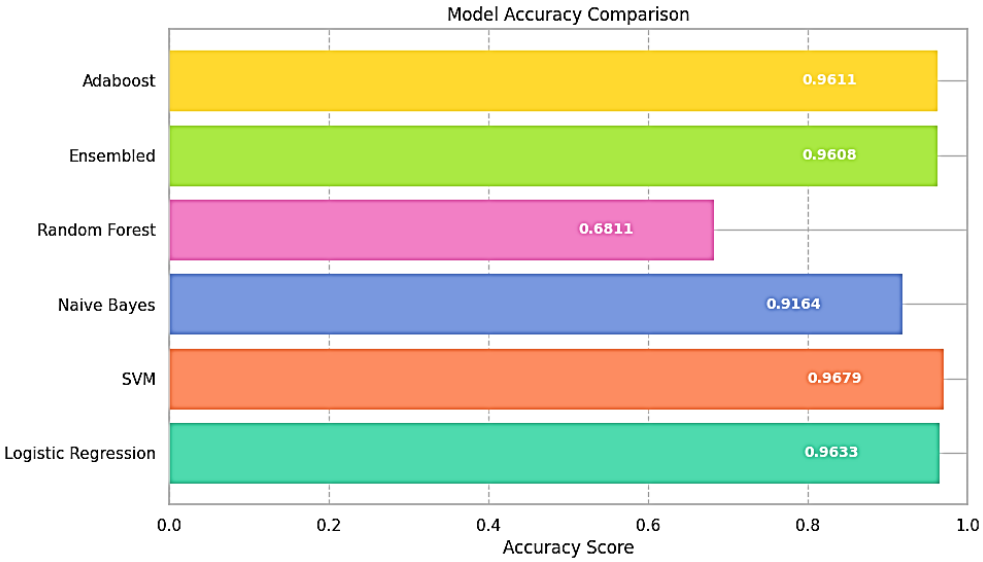
با استفاده از مدل‌های یادگیری ماشینی از جمله رگرسیون لجستیک، ماشین بردار پشتیبان، جنگل تصادفی، نایو-بیز چندجمله‌ای، یادگیری گروهی و مدل تقویت سازگار، به آموزش و تست بر روی داده‌های متنی خود پرداختیم. نتایج این آزمایش‌ها به صورت شکل ۳ قابل مشاهده است.

۳.۳ مدل یادگیری عمیق

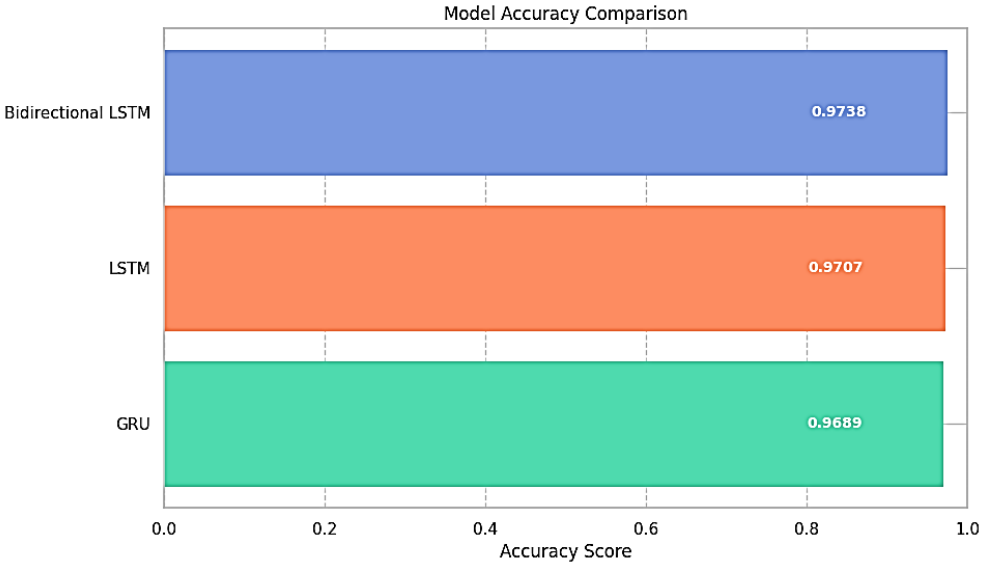
با استفاده از مدل‌های یادگیری عمیق از جمله شبکه عصبی حافظه‌ی کوتاه‌مدت طولانی ساده (LSTM) و شبکه عصبی واحد بازگشتی دروازه‌گذاری شده (GRU)، به آموزش و تست بر روی داده‌های متنی خود پرداختیم. نتایج این آزمایش‌ها به صورت شکل ۴ قابل مشاهده است.

۴.۳ مدل Bidirectional LSTM

در این مقاله، مدل شبکه عصبی Bidirectional LSTM (شبکه حافظه کوتاه‌مدت طولانی دوجهته) برای دسته‌بندی متن بررسی خواهد شد. این مدل به‌منظور ارتقای دقت و عملکرد دسته‌بندی در متون طراحی



شکل ۳: میزان دقت هر یک از مدل‌های یادگیری کلاسیک



شکل ۴: میزان دقت هر یک از مدل‌های LSTM و GRU

شده است. در ادامه به بررسی هر بخش از این مدل خواهیم پرداخت.

مدل شبکه عصبی مورد بررسی، از چندین لایه تشکیل شده است (شکل ۵) که به‌طور خاص برای پردازش متون ساخته شده‌اند:

- لایه Embedding

این لایه در ابتدای شبکه قرار دارد و برای تبدیل کلمات به بردارهای عددی با ابعاد مشخص استفاده می‌شود. در اینجا، Embedding با ابعاد emb_dim=۱۰۰ از واژگان موجود در دیتاست استفاده می‌کند

- لایه Spatial Dropout

به منظور کاهش اتصالات موقت در ورودی‌های فضایی، از SpatialDropout۱D با نرخ dropout برابر با ۰.۸ استفاده می‌شود. این کار کمک می‌کند تا مدل از بیش‌برازش جلوگیری کند.

- لایه Bidirectional LSTM

این لایه‌ها دو لایه LSTM با همکاری از جلو و عقب، به منظور درک بهتر وابستگی‌های زمانی در داده‌ها استفاده می‌شوند.

- لایه Dropout

پس از هر لایه LSTM، یک لایه Dropout با نرخ ۰٫۵ اعمال می‌شود تا از بیش‌برازش جلوگیری شود.

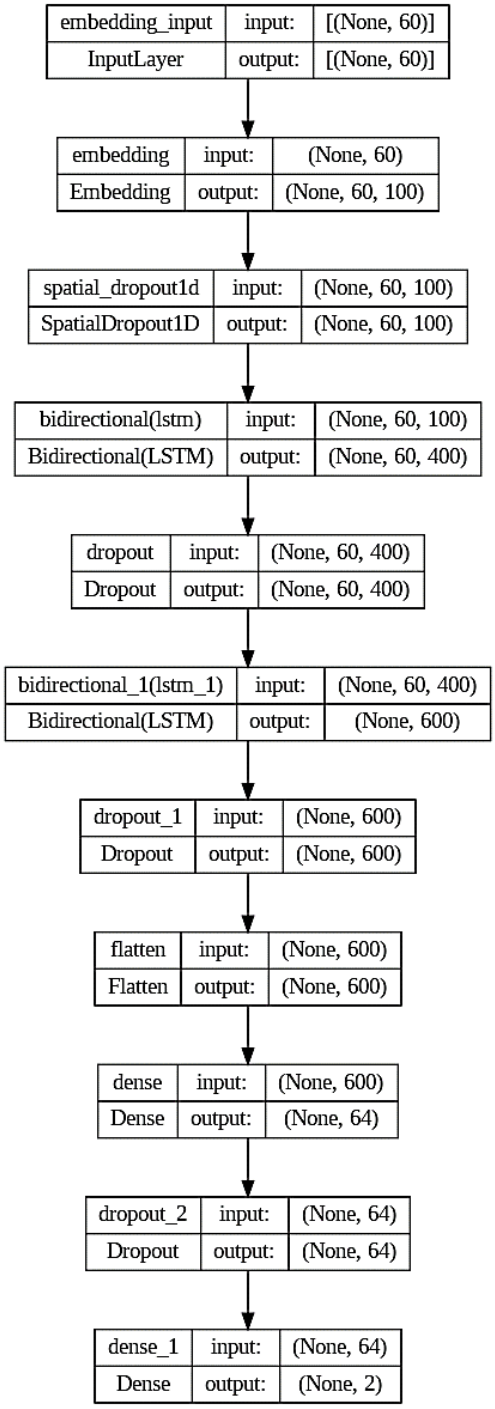
- لایه Flatten و Dense

از لایه Flatten برای تبدیل خروجی‌های چند بعدی از LSTM به یک بردار یک بعدی استفاده می‌شود. از لایه Dense با ۶۴ واحد و تابع فعال‌سازی relu برای استخراج ویژگی‌ها و کاهش ابعاد استفاده می‌شود.

- لایه خروجی

لایه آخر، یک لایه Dense با دو واحد و تابع فعال‌سازی softmax است که برای دسته‌بندی دو کلاسه استفاده می‌شود.

جهت بهینه‌سازی مدل، از بهینه‌ساز Adam و تابع هزینه binary_crossentropy استفاده شده است، که برای دسته‌بندی دو کلاسه مناسب می‌باشد. بهینه‌ساز Adam به‌عنوان یکی از پرکاربردترین بهینه‌سازها در شبکه‌های عصبی شناخته می‌شود، زیرا نرخ یادگیری آن به طور وفقی تنظیم می‌شود تا بهترین عملکرد را برای مدل فراهم آورد. همچنین، معیارهای ارزیابی مانند درستی (accuracy) به‌منظور پیگیری عملکرد مدل در طول فرآیند آموزش به کار گرفته شده‌اند.



شکل ۵: لایه‌های مدل شبکه عصبی Bidirectional LSTM

۴ معیارهای ارزیابی

معیارهای طبقه‌بندی متون با مضمون خودکشی از غیر خودکشی مختلفی برای ارزیابی مدل بهینه‌سازی شده Bidirectional LSTM استفاده می‌شود. این معیارها شامل درستی (ACC)، بازخوانی (Recall)، دقت (PREC) و امتیاز F1 (F1-score) برای ارزیابی عملکرد مدل پیشنهادی می‌باشند. این معیارها به‌طور گسترده در چندین مطالعه مورد استفاده قرار گرفته‌اند و به‌عنوان معیارهای رایج برای دسته‌بندی و طبقه‌بندی تصاویر در نظر گرفته می‌شوند. در ادامه، جزئیات بیشتر در مورد این معیارها بیان خواهد شد.

معیار درستی (ACC) نشان‌دهنده نسبت نتایج صحیح (هم مثبت و هم منفی) به تعداد کلی نمونه‌های مورد مطالعه است. این معیار با تعریف زیر استفاده می‌شود:

$$ACC = \frac{TP + TN}{TP + TN + FP + FN}. \quad (۱)$$

معیار فراخوان (SN)، که با نام نرخ مثبت واقعی (TPR) نیز شناخته می‌شود، تعداد پیش‌بینی‌های مثبت واقعی را نسبت به تعداد کل موارد مثبت کمی می‌کند. این معیار به عنوان نسبت پیش‌بینی‌های مثبت واقعی به تعداد موارد مثبت محاسبه می‌شود:

$$SN = \frac{TP}{TP + FN}. \quad (۲)$$

معیار دقت پیش‌بینی منفی (SP) با تقسیم تعداد منفی‌های واقعی بر تعداد کل منفی‌ها محاسبه می‌شود. این معیار به عنوان نرخ منفی واقعی (TNR) نیز شناخته می‌شود:

$$SP = \frac{TN}{TN + FP}. \quad (۳)$$

معیار دقت (PREC)، همچنین به عنوان ارزش پیش‌بینی مثبت (PPV) شناخته می‌شود، نسبت تعداد پیش‌بینی‌های مثبت واقعی به تعداد کل پیش‌بینی‌های مثبت انجام شده را محاسبه می‌کند:

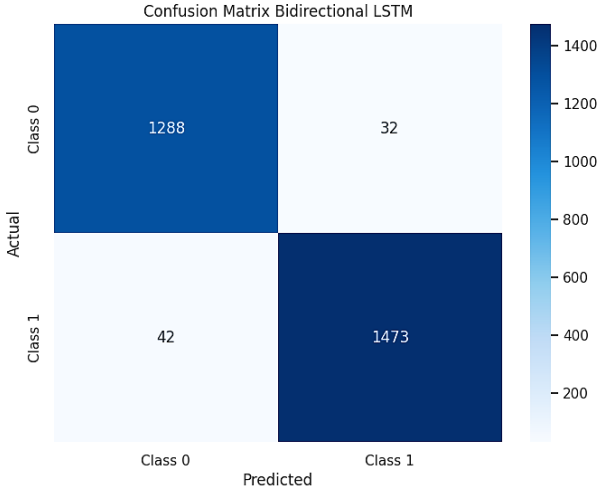
$$PREC = \frac{TP}{TP + FP}. \quad (۴)$$

امتیاز F1 (F1-Score) برای ارزیابی مدل استفاده می‌شود، زیرا میانگین هارمونیک دقت و یادآوری است و می‌تواند بینش بهتری نسبت به عملکرد مدل ارائه دهد:

$$F۱\text{-score} = ۲ \times \frac{PREC \times SN}{PREC + SN}. \quad (۵)$$

۵ ارزیابی

تکنیک ماتریس درهم‌ریختگی (Confusion Matrix) یک روش دیگر برای خلاصه کردن عملکرد الگوریتم‌های طبقه‌بندی است. این روش امکان محاسبه تعداد دقیق طبقه‌بندی‌هایی که توسط الگوریتم



شکل ۶: ماتریس درهم‌ریختگی برای داده‌های تست‌شده

جدول ۲: ارزیابی مدل Bidirectional LSTM

Metric	Accuracy	Precision	Recall	F1-score
Value	0.9738	0.9739	0.9738	0.9739

اشتباه گرفته شده‌اند را فراهم می‌کند. در این روش، پیش‌بینی‌های صحیح و نادرست به کلاس‌های مختلف تقسیم می‌شوند و نتایج به صورت خلاصه از طریق یک ماتریس، همانند شکل ۵، برای هر مجموعه داده نمایش داده می‌شود. این ماتریس می‌تواند به ما کمک کند تا به‌طور دقیق‌تر عملکرد الگوریتم طبقه‌بندی را ارزیابی کرده و خطاهای موجود در پیش‌بینی‌ها را شناسایی کنیم.

با توجه به ماتریس درهم‌ریختگی (شکل ۶) برای تست مدل Bidirectional LSTM نتایج ازیابی همانند جدول ۲ حاصل می‌شود.

نمودارهای درستی و اتلاف شبکه بر روی داده‌های آموزشی و آزمایشی شبکه را نیز می‌توان در شکل ۷ مشاهده کرد.

۶ نتیجه‌گیری

تحقیقات اخیر نشان داده است که استفاده از مدل‌های یادگیری ماشینی و یادگیری عمیق در زمینه تشخیص افکار خودکشی در پلتفرم شبکه‌های اجتماعی ایکس و ردیت بسیار مؤثر است. با توجه به بررسی مدل‌های کلاسیک و یادگیری عمیق با مدل پیشنهادی می‌توان مشاهده کرد که مدل پیشنهادی دقت مطلوب و نزدیک با مدل‌های تست شده دارد، می‌توان در جدول ۳ مشاهده کرد.

نشان دادیم که استفاده از مدل‌های پردازش زبان طبیعی (NLP) و یادگیری ماشینی و یادگیری عمیق در

جدول ۳: مقایسه مدل‌های آموزش داده شده

	F1-score	Recall	Precision	Accuracy
Logistic Regression	0.9633	0.9631	0.9633	0.9633
SVM	0.9677	0.9679	0.9679	0.9679
Naïve Bayes	0.9144	0.9164	0.9164	0.9164
Random Forest	0.6666	0.6811	0.6811	0.6811
Ensembled	0.9605	0.9608	0.9608	0.9608
Adaboost	0.961	0.9611	0.9611	0.9611
LSTM	0.9706	0.9707	0.971	0.9707
GRU	0.9688	0.9689	0.9689	0.9689
Bidirectional LSTM	**0.9739**	**0.9738**	**0.9739**	**0.9738**

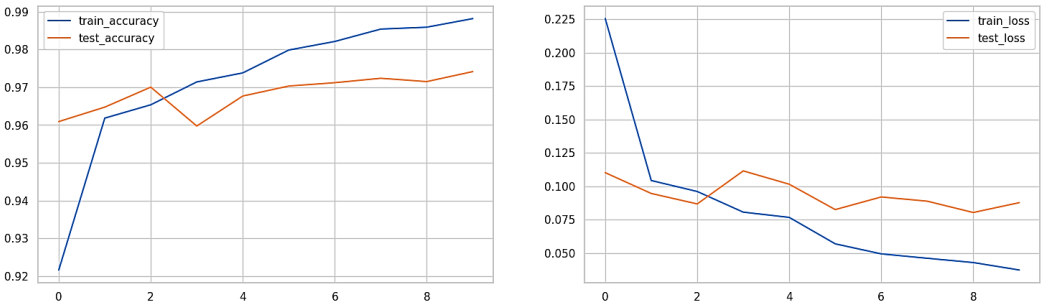
تشخیص افکار خودکشی و پیشگیری از خودکشی بسیار مؤثر است و دارای پتانسیل بالقوه‌ای برای آینده است. همچنین، می‌توان با بهره‌گیری از داده‌های صوتی و تصویری، قدرت سیستم‌های تشخیص خطر خودکشی را افزایش داد و بهبود و بهینه‌سازی بیشتری را به این صنعت بخشید.

در این مقاله بیشترین تمرکز ما بر روی استفاده از داده‌های متنی بوده است، اما به دنبال آن هستیم که در آینده به پیش‌بینی استفاده از داده‌های صوتی و تصویری نیز جهت بهبود سیستم‌های تشخیص خطر افکار خودکشی بپردازیم. این توسعه قادر است در ساخت پایه‌ی استواری برای سیستم‌های جلوگیری از اقدام به خودکشی نقش مؤثری ایفا کند و بتواند به‌طور قابل توجهی در کاهش خطرات خودکشی تأثیرگذار باشد.

مراجع

[1] C. Wang, F. Jiang, and H. Yang, “A Hybrid Framework for Text Modeling with Convolutional RNN”, Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, 13-Aug-2017.

[2] R. Sawhney, P. Manchanda, P. Mathur, R. Shah, and R. Singh, “Exploring and Learning Suicidal Ideation Connotations on Social Media with Deep Learning”, Proceedings of the



شکل ۷: نمودارهای درستی و اتلاف شبکه بر روی داده‌های آموزشی و آزمایشی

9th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis. Association for Computational Linguistics, 2018.

[3] S. Ji, C. P. Yu, S. Fung, S. Pan, and G. Long, "Supervised Learning for Suicidal Ideation Detection in Online User Content", Complexity, vol. 2018. Hindawi Limited, pp. 1–10, 09-Sep-2018.

[4] A. Mbarek, S. Jamoussi, A. Charfi, and A. Ben Hamadou, "Suicidal Profiles Detection in Twitter", Proceedings of the 15th International Conference on Web Information Systems and Technologies. SCITEPRESS - Science and Technology Publications, 2019.

[5] M. M. Tadesse, H. Lin, B. Xu, and L. Yang, "Detection of Suicide Ideation in Social Media Forums Using Deep Learning", Algorithms, vol. 13, no. 1. MDPI AG, p. 7, 24-Dec-2019.

[6] S. Jain, S. P. Narayan, R. K. Dewang, U. Bhartiya, N. Meena, and V. Kumar, "A Machine Learning based Depression Analysis and Suicidal Ideation Detection System using Questionnaires and Twitter", 2019 IEEE Students Conference on Engineering and Systems (SCES). IEEE, May-2019.

[7] S. Fodeh, T. Li, K. Menczynski, T. Burgette, A. Harris, G. Ilita, S. Rao, J. Gemmell, and D. Raicu, "Using Machine Learning Algorithms to Detect Suicide Risk Factors on Twitter", 2019 International Conference on Data Mining Workshops (ICDMW). IEEE, Nov-2019.

[8] A. H. Yazdavar, M. S. Mahdavinejad, G. Bajaj, W. Romine, A. Sheth, A. H. Monadjemi, K. Thirunarayan, J. M. Meddar, A. Myers, J. Pathak, and P. Hitzler, "Multimodal mental health analysis in social media", PLOS ONE, vol. 15, no. 4. Public Library of Science (PLoS), p. e0226248, 10-Apr-2020.

[9] N. Wang, L. Fan, Y. Shivtare, V. Badal, K. Subbalakshmi, R. Chandramouli, and E. Lee, "Learning Models for Suicide Prediction from Social Media Posts", Proceedings of the Seventh Workshop on Computational Linguistics and Clinical Psychology: Improving Access. Association for Computational Linguistics, 2021.

[10] T. H. H. Aldhyani, S. N. Alsubari, A. S. Alshebami, H. Alkahtani, and Z. A. T. Ahmed, "Detecting and Analyzing Suicidal Ideation on Social Media Using Deep Learning and Machine Learning Models", International Journal of Environmental Research and Public Health, vol. 19, no. 19. MDPI AG, p. 12635, 03-Oct-2022.

[11] R. Haque, N. Islam, M. Islam, and M. M. Ahsan, "A Comparative Analysis on Suicidal Ideation Detection Using NLP, Machine, and Deep Learning", Technologies, vol. 10, no. 3. MDPI AG, p. 57, 29-Apr-2022.

[12] U. Naseem, M. Khushi, J. Kim, and A. G. Dunn, "Hybrid Text Representation for Explainable Suicide Risk Identification on Social Media", IEEE Transactions on Computational Social Systems. Institute of Electrical and Electronics Engineers (IEEE), pp. 1–10, 2022.