

Detecting Depression Using Artificial Intelligence

Sam Hamzehloo

Assistant Professor, Faculty of Engineering,
Computer Group, Pasargad Higher Education
Institute, Shiraz, Iran.

**Mohammad Hashem
Imanpour ***

Ph.D. Student, Faculty of Engineering, Artificial
Intelligence Group, Islamic Azad University,
Shiraz Branch, Shiraz, Iran.

Abstract

Depression is one of the common disorders in today's societies. This disorder occurs in various intensities, and if diagnosed in the early stages, it can be easily treated with less cost and time. In this research, using data related to various personality tests and labeled by a reputable specialist obtained from the statistical population of visitors to Shiraz city, we aim to select a model for predicting the depression of visitors using classification algorithms. In this study, the information obtained from visitors is preprocessed first, and then, using data mining software and the WEKA tool, several classification algorithms are applied, and the results are compared based on the cross-validation method.

Keywords: artificial intelligence, depression, classification, cross-validation, WEKA

Received: 28/February/2024

Accepted: 04/May/2024

ISSN: 2980-8936

تشخیص افسردگی با استفاده از هوش مصنوعی

سام حمزه لو

استادیار دانشکده فنی و مهندسی، گروه کامپیوتر، مؤسسه آموزش عالی پاسارگاد، شیراز، ایران.

محمد هاشم ایمانپور*

دانشجوی دکتری، دانشکده فنی و مهندسی، گروه هوش مصنوعی، دانشگاه آزاد اسلامی، شیراز، ایران.

چکیده

بیماری افسردگی از جمله اختلالات مرسوم در جوامع امروزی به شمار می‌رود. این اختلال در شدت‌های مختلف بروز می‌نماید و اگر در مراحل اولیه تشخیص داده شود، به راحتی و با هزینه و زمانی کمتر درمان می‌شود. در این پژوهش با استفاده از داده‌های مربوط به تست‌های مختلف شخصیتی و برجسب‌گذاری شده توسط متخصص معتبر که از جامعه آماری مراجعه‌کنندگان شهر شیراز به دست آمده، قصد داریم با استفاده از الگوریتم‌های کلاسه‌بندی، مدلی برای پیش‌بینی افسردگی مراجعه‌کنندگان انتخاب کنیم. بدین ترتیب، ابتدا در این پژوهش اطلاعات به دست آمده از مراجعه‌کنندگان را پیش‌پردازش نموده و سپس، نرم‌افزار داده‌کاوی وکا را بر روی تعدادی از الگوریتم‌های کلاسه‌بندی اعمال نموده و نتایج به دست آمده را بر اساس روش اعتبارسنجی متقابل مقایسه می‌کنیم.

کلیدواژه‌ها: هوش مصنوعی، افسردگی، طبقه‌بندی، اعتبارسنجی متقابل، وکا

مقدمه

انسان در ایام مختلف عمر خود با پدیده‌ها، مسائل و مشکلات زیادی روبه‌رو شده و با آن‌ها مبارزه کرده است. عامل موفقیت در مواجهه با موارد مختلف، قدرت انرژی درونی ناشی از نشاط و تفکر درست و تصمیم‌های صحیح است. با افزایش مشکلات ناشی از زندگی ماشینی و رخوت و کرختی فعالیت‌های غیر فیزیکی و همچنین، افزایش استرس و فشارهای های عصبی، نسل حاضر بشر دست‌به‌گریبان عارضه افسردگی شده است. این عارضه باعث کاهش کیفیت زندگی افراد می‌شود، به‌نحوی که در جوانی احوالات پیری و ازپافتادگی پیدا می‌کنند. در صورت عدم تشخیص و درمان نشدن، این عارضه جای خود را به امراض روحی و روانی مزمن ازجمله اسکیزوفرنی می‌دهد به‌طوری که حتی ممکن است در مراحل پیشرفته هیچ راه درمانی وجود نداشته باشد و شخص مریض می‌بایست از اجتماع جدا و با هزینه‌های سرسام‌آور در مراکز نگهداری بستری شود. در زمینه آنالیز داده‌های بالینی، روش‌های سنتی وجود دارند که آزمون MMPI^۱، آزمون افسردگی بک و یا آزمون میلون ازجمله آن‌ها هستند. هزینه زمانی و ریالی انجام این آزمون‌ها بسیار زیاد بوده و برای اجرا در سطح جامعه اصلاً امکان‌پذیر نیست. بدین جهت، سیستم‌های مکانیزه هوشمند، نقش مشاوران و ارزیابان را بر عهده می‌گیرند. تولید یک راه‌حل مکانیزه هوشمند به مغزافزار به‌ویژه همراه با الگوریتم مناسب نیاز دارد.

پیشینه پژوهش

بختیاری و همکاران (۱۳۹۶)، از یک مدل برای سیستم خبره هوش مصنوعی به‌منظور تشخیص وجود اختلال افسردگی اساسی استفاده کرده‌اند که از نوع تحقیق و توسعه است. در این راستا، آن‌ها جهت کسب اطلاعات لازم برای تشخیص، پرسشنامه ویژه‌ای طراحی کردند که بر اساس ملاک‌های لازم در نسخه پنجم معیارهای راهنمای تشخیصی و آماری بیماری‌های روانی و پرسش‌هایی از تاریخچه شخصی و بررسی وضعیت روانی، اطلاعات موردنیاز را فراهم می‌کند. از هر کلاس، تعداد ۱۰۰۰ رکورد به‌صورت تصادفی انتخاب و الگوریتم‌ها و تکنیک‌های داده‌کاوی بر روی آن‌ها اجرا شد. درنهایت، خروجی سیستم طراحی شده، وجود افسردگی اساسی یا عدم آن است. با توجه به پژوهش انجام شده و نتایج به‌دست آمده می‌توان به این نتیجه رسید که داده‌های فراهم شده برای انجام پژوهش به نحو درستی جمع‌آوری شده و مدل ساخته شده از آن نیز از دقت قابل توجهی برخوردار است.

البته، بنا به گفته نویسندگان این پژوهش، مبنای برچسب‌گذاری داده‌ها بر اساس صرف ارزیابی پرسشنامه‌ای بوده و مصاحبه بالینی با هیچ‌یک از اعضای جامعه آماری انجام نشده است که یک نقطه ضعف اساسی برای این پژوهش محسوب می‌شود.

داوودی و مرادی (۱۳۹۷) به دنبال تمایز میان سطوح افسردگی با استفاده از تحلیل سیگنال مغزی هستند. مدل پیشنهادی، سیستم عمیق مبتنی بر قاعده با استفاده از قابلیت پشته بوده و تمرکز روی تفسیرپذیری قواعد در کنار دقت بالا است. سیستم‌های فازی، قابلیت مناسبی در طبقه‌بندی دادگان پزشکی با عدم قطعیت نشان داده‌اند. سیستم پیشنهادی از یک رویکرد خوشه‌بندی مقاوم بهره می‌برد که قادر است تعداد خوشه‌های بهینه برای هر لایه را به‌صورت بدون سرپرست تعیین نماید. در کنار آن، مدل پیشنهادی از یک ساختار سلسله‌مراتبی پشته‌ای بهره می‌برد؛ به این صورت که قواعد آموزش یافته تفسیرپذیر در لایه اول را با برچسب‌های زبانی یکسان برای تمام ورودی‌ها، به‌صورت خروجی لایه اول در کنار ورودی، به لایه بعد منتقل می‌نماید. وجود خروجی قواعد لایه‌های قبل در فضای ورودی لایه‌های بعد، معادل با قابلیت اطمینان در سیستم فازی با تالی خطی یا یک سیستم فازی با تالی غیرخطی است. دادگان مورد استفاده

پس از پیش‌پردازش، استخراج ویژگی‌های زمانی، فرکانسی و غیرخطی نظیر بعد نگاشت بازگشتی و کاهش بعد، به سیستم پیشنهادی ارائه شد. سیستم پیشنهادی با طبقه‌بندهای متداول نظیر شبکه عصبی، ماشین بردار پشتیبان، بیز ساده، درخت تصمیم و آنالیز افتراقی خطی مقایسه شد. نتایج صحت دادگان تست به‌دست آمده در ۳۰ تکرار (۴۹.۰۱ درصد در مقابل به ترتیب ۴۱/۳۲ درصد، ۴۰/۴۷ درصد، ۴۰/۰۱ درصد، ۳۸/۳۸ درصد و ۴۰/۲۸ درصد)، بیان‌گر قابلیت قابل توجه این مدل در تفکیک چهار سطح افسردگی است.

البته، طبقه‌بندی سیگنال انسفالومتری که با استفاده از گیرنده‌های مستقر بر روی پوست سر استحصال می‌شود، روش بسیار دقیقی برای شناسایی افسردگی محسوب نمی‌شود زیرا بیانگر حالات مراجعه‌کننده در کلیه ساعات شبانه‌روز نبوده و هنگام ثبت سیگنال ممکن است مراجعه‌کننده دچار استرس شدیدی شود که به نوبه خود روایی این معلومات را زیر سؤال می‌برد.

پارساپور و پیمانی (۱۳۹۷) با مراجعه به تعدادی کلینیک روان‌پزشکی در شهر تهران و بررسی پرونده ۴۲۰ نفر که به پرسشنامه مینه‌سوتای ۷۱ سؤالی پاسخ داده بودند، با استفاده از روش نمونه‌گیری در دسترس، این تعداد به‌عنوان نمونه انتخاب شدند (۳۰۰ نفر مبتلا به مراتب متفاوتی از افسردگی و ۱۲۰ نفر فاقد آن). پاسخنامه آزمون مینه‌سوتای ۷۱ سؤالی و تشخیص روان‌پزشک به‌عنوان داده برای ایجاد مدل توسط الگوریتم‌های K نزدیک‌ترین همسایه، درخت تصمیم و ماشین بردار پشتیبان مورد استفاده قرار گرفتند. ۷۰ درصد داده‌ها برای آموزش و ۳۰ درصد داده‌ها برای اعتبارسنجی مدل به کار گرفته شد که جهت تحلیل داده‌ها از نرم‌افزار Matlab استفاده کردند. نتایج ارزیابی‌ها نشان داد که الگوریتم درخت تصمیم با دقت ۱۶/۹۹ درصد، در مقایسه با دیگر الگوریتم‌ها به دقت بالاتری دست یافت. همچنین، با اجرای مدل‌های ایجادشده بر روی هر سؤال آزمون مینه‌سوتای ۷۱ سؤالی، تأثیر هر سؤال در ارزیابی مشخص شد. همان‌طور که در بخش نتیجه‌گیری این پژوهش مشخص شده، نتیجه نزدیک به ۱۰۰ درصد است. این موضوع، خطر وقوع بیش‌برازش در ساختار مدل برگزیده را گوشزد می‌کند.

وانالامپویا و لاله‌مینگیلانا^۱ (۲۰۲۲) معتقدند که سایت‌های شبکه‌های اجتماعی در ده سال گذشته توجه نسل جوان را به خود جلب کرده است. از طریق فناوری اینترنت، دسترسی و تعامل با سایت‌های شبکه‌های اجتماعی راحت و آسان می‌شود؛ بنابراین، سایت‌های شبکه‌های اجتماعی اکنون منبعی برای استخراج داده‌ها هستند زیرا افراد، سازمان‌ها و شرکت‌های بزرگ در حال تعامل و معاشرت با یکدیگر هستند. همچنین می‌توان اطلاعات بسیار زیادی در رابطه با وضعیت ذهنی، احساس و منفی‌گرایی کاربر استخراج کرد. در همین حال، سایت‌های شبکه‌های اجتماعی بدون اطلاع از مشکلات روانی، برای رفع نگرانی‌های خود از آن‌ها استفاده می‌کنند. از این‌رو، داده‌کاوی روش‌های مختلفی از تکنیک‌ها را برای بازیابی اطلاعات و دانش مرتبط ارائه می‌دهد و می‌تواند برای گزارش‌های آماری، یادگیری ماشینی و یادگیری عمیق استفاده شود. این تکنیک‌ها، نحوه مدیریت داده‌های بازیابی‌شده را در دوره تجزیه و تحلیل داده‌ها معرفی می‌کنند. این مطالعه با تجزیه و تحلیل داده‌های هر کاربر، وضعیت سلامت روانی کاربر را برجسته و آشکار کرده است. تجزیه و تحلیل داده‌ها شامل آموزش داده‌های هر کاربر برای به دست آوردن یک خروجی و داشتن یک مجموعه تست برای به دست آوردن کارایی و دقت است. این مطالعه، کلمات نشانگر افسردگی بسیاری را بیان کرده است که نقش مهمی در موفقیت آن دارند. در ضمن، چند روش را پیشنهاد کرده‌اند که در این مطالعه انجام شده است. تحلیل کارایی با استفاده از این روش پیشنهادی انجام شده است که حاوی بسته‌هایی از کلمات احساسی بوده و می‌تواند دقت، کارایی و مقیاس خطای تحلیل را افزایش دهد. هم این داده‌کاوی فناوری و هم یادگیری ماشینی می‌توانند وضعیت مشکل سلامت روان کاربر را به ما ارائه دهند.

البته در این پژوهش که در نشریه معتبر IEEE xplor منتشر شده است، نویسندگان بیشتر بر نوآوری در روش گردآوری اطلاعات فاقد اعتبار حقیقی و با فراوانی بالای فضای مجازی تأکید دارند و بر روش‌های آماری لغت‌شماری به‌عنوان ویژگی‌های مجموعه داده تأکید دارند که نتایج فاقد فرضیه مخالف و مستقل از نتیجه تحقیق مشابه بوده و تشکیک در صحت داده‌ها از موارد نقص آن ایده و پژوهش است.

کاستا و پدریرا^۱ (۲۰۲۳) معتقدند که افسردگی، اختلالی است که مشخصه آن بدبختی و عبوس بودن است. دوره زمانی برخی از علائم افسردگی با سایر علائم جسمی همپوشانی دارند. در نوشته آنان، تشخیص بیماری‌های با دشواری قابل توجه در تشخیص با استفاده از داده‌های هموارتر می‌شود. برای پیش‌بینی بیمارانی که به احتمال زیاد دچار افسردگی شده و در حال حاضر از افسردگی رنج می‌برند، از داده‌های مصنوعی استفاده شد. برای به دست آوردن نتایج، از مجموعه محبوب نرم‌افزار یادگیری ماشین WEKA استفاده شد.

همان‌طور که در نوشتار آمده است، جامعه آماری این نویسندگان، غیرواقعی و ساختگی بوده و به روایی بودن و اعتبار تحقیق بسیار صدمه زده است. به‌علاوه، جای خالی یک تحقیق علمی با داده‌های معتبر و تأییدشده توسط خبره روانشناسی به‌شدت حس می‌گردد.

دینال و همکاران^۲ (۲۰۱۶) معتقدند که تکنیک‌های داده‌کاوی در مقیاس بزرگ نظری با استفاده از الگوریتم‌های یادگیری ماشینی در تجزیه و تحلیل مجموعه داده‌های اپیدمیولوژیک بزرگ نویدبخش هستند. در پژوهش ایشان، استفاده از یک روش ترکیبی برای انتخاب متغیر نشان می‌دهد که داده‌های از دست‌رفته و طراحی نظرسنجی پیچیده برای شناسایی نشانگرهای زیستی کلیدی مرتبط با افسردگی از یک مطالعه اپیدمیولوژیک بزرگ در نظر گرفته شده است. آن‌ها در مطالعه خود از یک روش سه مرحله‌ای استفاده کردند که تلفیق چندگانه، یک الگوریتم رگرسیون تقویت‌شده یادگیری ماشین و رگرسیون لجستیک را برای شناسایی نشانگرهای زیستی کلیدی مرتبط با افسردگی در مطالعه ملی سلامت و تغذیه (۲۰۰۹-۲۰۱۰) مورد استفاده قرار داد. افسردگی با استفاده از پرسشنامه سلامت بیمار-۹ اندازه‌گیری شد و ۶۷ نشانگر زیستی مورد تجزیه و تحلیل قرار گرفت. متغیرهای کمکی در این مطالعه شامل جنسیت، سن، نژاد، مصرف سیگار، امنیت غذایی، نسبت درآمد فقر، شاخص توده بدنی، فعالیت بدنی، مصرف الکل، شرایط پزشکی و داروها بودند. مدل رگرسیون لجستیک چندگانه وزنی متناسب شده نهایی شامل عوامل مخدوش‌کننده و تعدیل‌کننده‌های احتمالی بود. پس از ایجاد ۲۰ مجموعه داده انتساب از توالی‌های رگرسیون زنجیره‌ای چندگانه، یادگیری ماشینی رگرسیون را تقویت و در ابتدا، ۲۱ نشانگر زیستی مرتبط با افسردگی را شناسایی کرد. با استفاده از روش‌های رگرسیون لجستیک سنتی، از جمله کنترل عوامل مخدوش‌کننده و تعدیل‌کننده‌های احتمالی، مجموعه نهایی سه نشانگر زیستی انتخاب شد. سه بیومارکر نهایی از روش انتخاب متغیر هیبریدی جدید، عرض توزیع گلوبول‌های قرمز (OR: ۱/۱۵؛ ۹۵ درصد فاصله اطمینان: (CI) ۱/۰۱، ۱/۳۰)، گلوکز سرم (OR: ۱/۰۱؛ ۹۵ درصد فاصله اطمینان: (CI) ۱/۰۰، ۱/۰۱) و بیلی روبین کل (OR: ۰/۱۲؛ ۹۵ درصد فاصله اطمینان: (CI) ۰/۲۸، ۰/۰۵). تداخلات قابل توجهی بین بیلی روبین تام با گروه مکزیکایی/آمریکایی/اسپانیایی $p = ۰/۰۱۶$ و سیگاری‌های فعلی $p < ۰/۱$ مشاهده شد. بدین ترتیب، این پژوهش از ویژگی‌های بسیاری در مجموعه داده خود استفاده نموده و در نهایت، معادلات زیست‌شیمی و پارامترهای نژادی را در پیش‌بینی افسردگی و شدت آن لحاظ کرده که در نوع خود قابل توجه بوده اما به‌شدت نامعتبر است!

وان لو و همکاران^۳ (۲۰۲۰) در نشریه معتبر الزویر بیان می‌کنند که دوره بیماری در افسردگی اساسی بسیار متنوع است و اگر پزشکان از رویکرد "یک اندازه مناسب برای همه" پیروی کنند، ممکن است منجر به درمان کم و

بیش از حد شود. فرصت‌های جدید در داده‌کاوی می‌تواند به مدل‌های پیش‌بینی منجر شود که می‌تواند به پزشکان در تصمیم‌گیری‌های درمانی متناسب با هر بیمار کمک کند. این مطالعه عملکرد یک الگوریتم داده‌کاوی قبلاً توسعه یافته را برای پیش‌بینی قسمت‌های آینده بر اساس اطلاعات بالینی در داده‌های جدید ارزیابی می‌کند. آن‌ها یک مدل پیش‌بینی را با استفاده از ویژگی‌های بالینی پایه در افرادی که مادام‌العمر به دو نمونه آزمایش مستقل گزارش کردند (تعداد = ۴۲۲۶)، اعمال کرده‌اند. آن‌ها عملکرد مدل را برای پیش‌بینی اپیزودهای آینده، اختلالات اضطرابی و ناتوانی در طول پیگیری (۱ تا ۹ سال پس از شروع مطالعه) ارزیابی کرده‌اند. علاوه بر این، عملکرد پیش‌بینی آن را با عوامل خطر شناخته‌شده برای یک دوره شدید بیماری مقایسه کرده‌اند. در نهایت، این مدل به‌طور مداوم اپیزودهای آینده را در هر دو نمونه آزمایش پیش‌بینی کرد (۰/۶۸-۰/۷۳، پیش‌بینی متوسط) و به همان اندازه دقیق، قسمت‌های اختلال اضطراب فراگیر، اختلال هراس و ناتوانی را نیز پیش‌بینی کرد (۰/۶۵-۰/۷۸). این مدل این پیامدها را با دقت بیشتری نسبت به عوامل خطر برای یک دوره شدید بیماری مانند سابقه خانوادگی و ترومای مادام‌العمر پیش‌بینی کرد.

البته به دلایل بالینی در اطلاعات مجموعه داده این گروه می‌توان ایراد وارد کرد؛ بدین جهت که دقت پیش‌بینی ممکن است برای زیرگروه‌های خاصی همچون بیماران بستری در بیمارستان یا بیماران با پیشینه فرهنگی مختلف، متفاوت باشد. در این بخش نیز جای یک مجموعه داده معتبر خالی است.

آلونسو و همکاران^۱ (۲۰۱۸) در پژوهش مروری منتشرشده توسط انتشارات اشپرینگل معتقدند که داده‌کاوی در پزشکی، یک زمینه نوظهور با اهمیت زیاد برای ارائه پیش‌آگهی و درک عمیق‌تر از طبقه‌بندی بیماری‌ها، به‌ویژه در زمینه‌های بهداشت روان است. در این مقاله، مروری بر آثار پژوهشی موجود در ادبیات، با اشاره به تکنیک‌ها و الگوریتم‌های داده‌کاوی در سلامت روان، به‌ویژه در شایع‌ترین بیماری‌ها مانند زوال عقل، آلزایمر، اسکیزوفرنی و افسردگی ارائه شده است. با در نظر گرفتن تاریخ انتشار ۱۰ سال گذشته (از ۲۰۰۸ تاکنون)، چندین معیار جستجو مانند «تکنیک‌ها» و «داده‌کاوی» و «سلامت روان»، «الگوریتم‌ها» و «داده‌کاوی» و «زوال عقل» و «اسکیزوفرنی» و «افسردگی» و غیره ایجاد شد و در مجموع، ۲۱۱ مقاله مرتبط با تکنیک‌ها و الگوریتم‌های داده‌کاوی به کاررفته در بیماری‌های اصلی سلامت روان یافت شد. ۷۲ مقاله به‌عنوان آثار مرتبط شناسایی شده که ۳۲ درصد آلزایمر، ۲۲ درصد زوال عقل، ۲۴ درصد افسردگی، ۱۴ درصد اسکیزوفرنی و ۸ درصد اختلالات دوقطبی هستند. بسیاری از مقالات، پیش‌بینی عوامل خطر در این بیماری‌ها را نشان می‌دهند. از بررسی مقالات پژوهشی تحلیل شده می‌توان گفت که استفاده از تکنیک‌های داده‌کاوی به کاررفته در بیماری‌هایی مانند زوال عقل، اسکیزوفرنی، افسردگی و غیره می‌تواند به تصمیم‌گیری بالینی، پیش‌بینی تشخیص و بهبود کیفیت زندگی بیمار کمک شایانی کند.

ایلگن و همکاران^۲ (۲۰۰۹) در پژوهشی به این مطلب اشاره کرده‌اند که اگرچه تحقیقات قبلی تعدادی از عوامل خطر جداگانه برای خودکشی را در میان بیماران مبتلا به افسردگی شناسایی کرده است اما اطلاعات کمی در مورد چگونگی تأثیر این عوامل برای اصلاح خطر خودکشی وجود دارد. با استفاده از یک تجزیه و تحلیل درخت تصمیم مبتنی بر تجربی برای نمونه ملی بزرگی از بیماران سیستم سلامت امور کهنه‌سربازان (VA) که برای افسردگی درمان شده‌اند، زیرگروه‌هایی را با نرخ‌های بالا یا پایین خودکشی شناسایی کرده‌اند. در روش آن‌ها، ۸۸۷۸۵۹ بیمار VA را که برای افسردگی درمان شده بودند، در بین ۱ آوریل ۱۹۹۹ و ۳۰ سپتامبر ۲۰۰۴ شناسایی و به‌طور تصادفی داده‌ها را به ۲ نمونه (نمونه‌های اولیه و تکراری) تقسیم کردند. یک درخت تصمیم برای نمونه اولیه با استفاده از پاریشن‌بندی بازگشتی ایجاد کرده و سپس آزمایش کردند که آیا گروه‌های توسعه یافته در نمونه اولیه با افزایش خطر خودکشی در نمونه تکراری مرتبط هستند یا خیر.

تجزیه و تحلیل داده‌های اکتشافی یک درخت تصمیم با زیرگروه‌های بیماران در سطوح مختلف خطر خودکشی تولید کرد. این‌ها با ترکیبی از عوامل از جمله تشخیص اختلال مصرف مواد هم‌زمان، جنسیت مرد، نژاد آفریقایی آمریکایی و بستری شدن در بیمارستان روان‌پزشکی در سال گذشته شناسایی شدند. گروه‌هایی که به‌عنوان بخشی از درخت تصمیم توسعه یافته‌اند، به‌دقت بین افراد دارای و بدون خودکشی در نمونه تکرار تمایز قائل شدند. بیماران که در معرض بالاترین خطر خودکشی قرار داشتند، کسانی بودند که دارای اختلال مصرف مواد بودند که غیر آفریقایی آمریکایی بوده و در ۱۲ ماه گذشته بستری شده بودند. بر این اساس به این نتیجه رسیدند که شناسایی بیماران افسرده در معرض افزایش خطر خودکشی از طریق بررسی تعاملات مرتبه بالاتر بین عوامل خطر بالقوه بهبود می‌یابد.

مطابق با این پژوهش، فقط تأثیر عواملی همچون افسردگی بر خودکشی بررسی شده و ملاک ما دقیقاً برعکس یعنی بررسی تأثیر میل به خودکشی در بین بیماران است.

ظفر و چیتنیس^۱ (۲۰۲۰) معتقدند که تشخیص افسردگی از سایت‌های شبکه‌های اجتماعی به‌طور گسترده در سال‌های گذشته مورد مطالعه قرار گرفته است. این سایت‌ها بستری را برای کاربران خود فراهم می‌کنند تا رویدادهای زندگی، احساسات و روال روزمره خود را به اشتراک بگذارند. بسیاری از محققان نشان دادند که محتوای تولیدشده توسط کاربران، روشی کارآمد برای آگاهی از وضعیت روانی آنهاست. با استخراج محتوای تولیدشده توسط کاربر می‌توان افسردگی را پیش‌بینی کرد. با جمع‌آوری تمام اطلاعات لازم و مرتبط از سایت‌های شبکه‌های اجتماعی از روی پست‌ها می‌توان روحیه یا منفی بودن فرد را پیش‌بینی کرد. این مقاله نظرسنجی بر تحقیقات قبلی انجام‌شده در مورد تشخیص سطوح افسردگی بر اساس محتوای سایت‌های شبکه‌های اجتماعی متمرکز است. البته این مقاله بر روی روش‌های آماری کلاسیک برای تشخیص افسردگی و یا منفی‌نگری در سطح جامعه هدف متمرکز است.

تو و همکاران^۲ (۲۰۲۳) در پژوهشی معتبر عنوان داشته‌اند که طب سوزنی سابقه طولانی در درمان اختلال افسردگی اساسی دارد اما انتخاب طب سوزنی برای اختلال افسردگی اساسی بسیار متفاوت است. این مطالعه با هدف بررسی ویژگی‌ها و اصول طب سوزنی برای اختلال افسردگی اساسی با تجزیه و تحلیل کار آزمایشی‌های بالینی طب سوزنی برای اختلال افسردگی اساسی با استفاده از تکنیک‌های داده‌کاوی انجام شد. نتایج نشان داد که ۶SP، ۶PC، ۳LR، ۲۰GV و ۲۹GV بیشترین استفاده را داشتند. نقاط طب سوزنی در نصف‌النهار یانگ بیشتر از آن‌هایی که در نصف‌النهار بین هستند، استفاده شد (با بیشترین کاربرد در کشتی فرماندار). درصد نقاط طبی خاص اعمال‌شده ۶۹/۳۹ درصد بود که بیشترین کاربرد آن نقاط پنج شو بود. فراوانی نقاط طب سوزنی استفاده‌شده در اندام تحتانی، بالاترین بود در حالی که سر، صورت و گردن بیشترین نقاط طب سوزنی استفاده‌شده را داشتند. ۲۹GV همراه با ۲۰GV بیشترین استفاده از گروه‌های نقطه طبی بودند. نقاط acupoint هسته مورد استفاده برای ۲۰MDD GV، ۶PC و ۳۶SP بودند. بر اساس تجزیه و تحلیل خوشه‌ای، ۵ گروه نقطه طبی وجود داشت. پرکاربردترین روش طب سوزنی، طب سوزنی دستی بود. دفعات درمان اغلب ۷ بار در هفته و مدت درمان ۴۲ روز بود.

همان‌طور که مشخص است، این پژوهش بر روی روش درمان و استفاده از داده‌کاوی در آن متمرکز است و فرض را بر تشخیص قطعی بالینی گذاشته است.

واشا و همکاران (۲۰۲۳) در پژوهشی جدید اعلام داشتند که افسردگی، سطح بعدی احساسات منفی است. زمانی که فردی حالت غمگینی دارد یا شرایط سختی را پشت سر می‌گذارد و او را رها نمی‌کند و مدام به او درد می‌آورد و دیگر نمی‌تواند آن را تحمل کند، به آن حالت افسردگی می‌گویند. آخرین مرحله افسردگی در خودکشی رخ می‌دهد. بر اساس گزارش سازمان بهداشت جهانی، در حال حاضر ۴/۴ درصد از مردم جهان از افسردگی رنج می‌برند. در سال

1. Zafar & Chitnis

2. Tu et al.

۲۰۲۱، چهارده هزار نفر در سراسر جهان خودکشی کردند و روزبه‌روز بر میزان خودکشی افزوده می‌شود؛ بنابراین، مطالعه آن‌ها در این زمینه است که افراد افسرده را با نظرات، پست‌ها یا متن‌هایشان در رسانه‌های اجتماعی پیدا کنیم. آن‌ها تقریباً ۱۰۰۰۰ داده را از پست‌ها جمع‌آوری کردند. الگوریتم‌های داده‌کاوی و یادگیری ماشین کار آن‌ها را آسان‌تر کرده و نقش مهمی در تشخیص آسان احساسات افراد دارند. آن‌ها از شش طبقه‌بندی‌کننده برای پیش‌بینی افسردگی و غیر افسردگی استفاده کرده و بهترین دقت را در ماشین بردار پشتیبان یافتند.

همان‌طور که در متن مقاله اشاره شده، داده‌های مورد پژوهش از طریق استاندارد طبقه‌بندی و برچسب‌گذاری نشده است و مسلماً همین امر در نتایج به‌دست آمده خطا ایجاد می‌کند.

در این مطالعه، لیم (۲۰۲۳) شناسایی عوامل مرتبط با بیماری اسکیمیک قلبی و گروه‌های آسیب‌پذیر را در زنان میان‌سال و مسن‌تر کره‌ای با استفاده از داده‌های نظرسنجی ملی سلامت و تغذیه کره انجام داده‌اند. از میان ۲۴۲۲۹ نفری که در نظرسنجی ۲۰۱۷-۲۰۱۹ شرکت کردند، ۷۲۴۹ زن میان‌سال ۴۰ سال و بالاتر در تحلیل نهایی وارد شدند. داده‌ها با استفاده از نرم‌افزارهای تحلیل آماری با استفاده از تحلیل مجذور کای، تحلیل رگرسیون لجستیک و تجزیه و تحلیل درخت تصمیم مورد تجزیه و تحلیل قرار گرفتند. شیوع بیماری اسکیمیک قلب در نتایج مطالعه ۲/۷۷ درصد بود. از جمله مواردی که انفارکتوس میوکارد یا آنژین تشخیص داده شد، عوامل مرتبط با بیماری اسکیمیک قلب در زنان میان‌سال و مسن‌تر به‌عنوان سن، سابقه خانوادگی، فشارخون بالا، دیس لیپیدمی، سکتة مغزی، آرتریت و افسردگی شناسایی شد. گروهی که بیشتر در برابر بیماری اسکیمیک قلب آسیب‌پذیر بودند، زنانی بودند که فشارخون بالا، سابقه خانوادگی بیماری اسکیمیک قلب داشتند و یائسه بودند. بر اساس این نتایج، مدیریت مؤثر باید با به‌کارگیری خدمات پزشکی سفارشی و خدمات مدیریت سلامت برای هر یک از عوامل مرتبط با در نظر گرفتن ویژگی‌های گروه‌های با خطرات احتمالی حاصل شود. این مطالعه می‌تواند به‌عنوان داده‌های اساسی که می‌تواند در تصمیم‌گیری سیاست ملی برای مدیریت بیماری‌های مزمن مفید باشد، استفاده شود.

همان‌طور که ملاحظه شد، این مقاله بیشتر در مورد استفاده از علم داده‌کاوی برای تشخیص اختلالات قلبی به کار رفته است و در این بین داده‌های بالینی مربوط به افسردگی نیز به چشم می‌خورد.

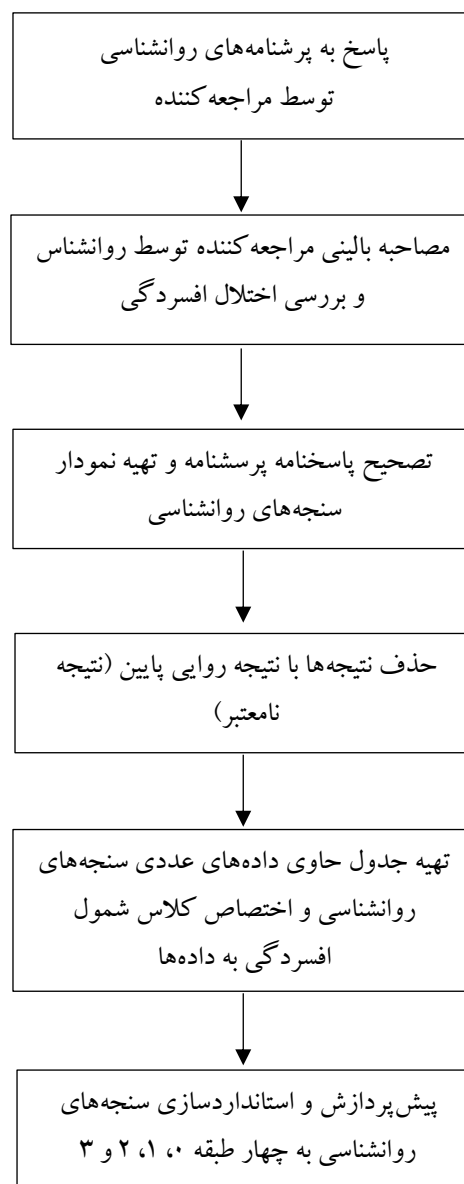
روش^۱

روش انتخابی میدانی - مدل سای تحلیلی است. در این تحقیق ما ابتدا به روش میدانی به مدت چند ماه به‌صورت میدانی نسبت به اخذ جواب پرسشنامه‌های آزمون‌های روانشناسی اقدام کردیم. دلیل انتخاب روش میدانی، جمع‌آوری این بود که داده‌های معتبر در مورد مراجعه‌کنندگان شیراز موجود نبود و می‌بایست این اقدام صورت پذیرد. در ادامه می‌بایست داده‌ها را به الگوریتم‌هایی می‌دادیم تا نتایج را به‌وسیله برخی مدل‌ها پیاده‌سازی کند. با توجه به اینکه داده‌ها دارای برچسب بودند، به الگوریتم‌هایی که معمولاً یادگیری با نظارت دارند، نیاز داشتیم. با توجه به اینکه داده‌ها دارای برچسب بلی یا خیر (باینری) بودند، به الگوریتم‌های با نظارت طبقه‌بندی‌کننده نیاز داشتیم تا مدل‌ها را پارامترگذاری مناسب کنند و بتوانند نتایج پیش‌بینی بهتری داشته باشند.

در این روش به مراجعه‌کنندگان پرسشنامه‌های MMPI II و آزمون میلون و همچنین، آزمون SCL ۹۰ داده شد و مراجعه‌کنندگان به شرط عدم درج اطلاعات هویتی‌شان پرسشنامه را پر کردند و سپس پرسشنامه‌ها تجمیع و توسط اصول تفسیر هر یک از آزمون‌ها نمودارهای مربوطه به انواع اختلال به‌صورت عددی رسم و در جداول مربوطه ثبت شد. در مرحله بعد، روانشناسان بعد از انجام مصاحبه بالینی با توجه به شرح حال و صحبت‌ها و علائم مراجعه‌کننده

اعلام داشتند که فرد مصاحبه‌شونده مبتلا به افسردگی است یا خیر. در این زمان، داده‌های هر فرد از این جامعه آماری برچسب‌گذاری شد. ضمناً لازم به ذکر است در جامعه آماری از کلیه آحاد اعم از زن و مرد و کوچک و نوجوان و همچنین در تمامی سطوح تحصیلی شرکت داشتند و گاه‌آ‌روانشناس با یک جلسه مصاحبه اعلام نظر نمی‌کرد. این پرسشنامه‌ها در پیوست این پژوهش موجود است.

بعد از تکمیل این داده‌ها به‌منظور پیش‌پردازش و یک‌دست شدن داده‌ها و با توجه به درصدهای هر اختلال که به‌عنوان ویژگی مجموعه داده در نظر گرفته می‌شد، هر ۲۵ درصد از یک سطح در نظر گرفته شد؛ بدین‌صورت که ۲۵-۰ درصد معادل عدد صفر، ۲۵-۵۰ درصد معادل عدد یک، ۵۰-۷۵ درصد معادل عدد دو و ۷۵-۱۰۰ معادل عدد سه در نظر گرفته شد. همچنین، برچسب‌ها نیز در صورت مثبت بودن ابتلا به افسردگی، عدد یک و عدم ابتلا به افسردگی طبق نظر روانشناس عدد صفر در نظر گرفته شد. این فرآیند در شکل شماره ۱ آمده است.



شکل ۱. مراحل جمع‌آوری و پیش‌پردازش داده‌ها

بعد از تکمیل مجموعه داده و انجام فرآیند حذف داده‌های با روایی پایین و همچنین عملیات پیش‌پردازش، هم‌اکنون نیاز داریم به کمک برخی الگوریتم‌های طبقه‌بندی‌کننده و دارای یادگیری با نظارت، مدل‌هایی برای پیش‌بینی داده‌های آتی تولید نموده و با مقایسه‌ای مدل‌ها بر حسب قواید اعتبار متقابل، مدل بهتر را شناسایی کنیم.

یافته‌ها

به‌طور کلی، خروجی الگوریتم رگرسیون خطی معمولاً به پیش‌بینی ویژگی‌ها می‌پردازد و ریسک افراد بر اساس سوابق را بر اساس تک‌تک ویژگی‌هایشان پیش‌بینی می‌کند. اصولاً این تکنیک با روش‌هایی که تاکنون انجام شده است، نتایجی متفاوت دارد. در جدول ۱، عوامل مؤثر در افسردگی این جامعه آماری به همراه میزان دخیل بودن اختلال موردنظر در افسردگی تشخیص داده‌شده به‌عنوان خروجی این الگوریتم آمده است.

جدول ۱. نتایج حاصل از رگرسیون خطی

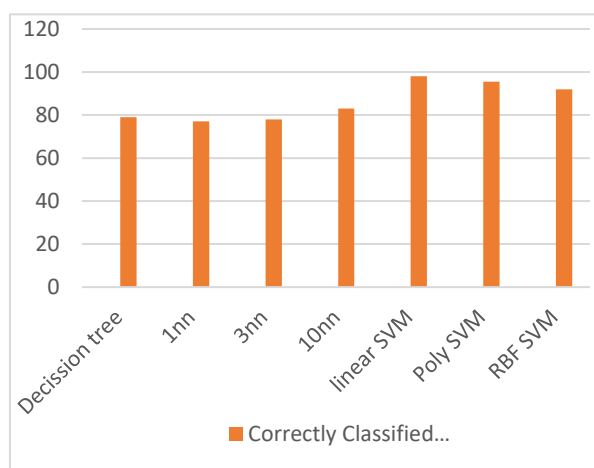
0.02080	غم
0.04660	دل‌سردی
0.03230	عزت نفس پایین
0.05210	حقارت
0.04010	گناه
0.09090	از دست دادن علاقه به زندگی
0.03400	از دست دادن انگیزه
0.03870	تصویر ضعیف از خود
0.06920	از دست دادن میل جنسی
0.05670	تنبل
0.03470	طلسم‌گریه
0.03850	درماندگی
0.03700	پرسیمیم
0.05340	هم‌زدن
0.03100	شکست گذشته
0.03860	کاهش تحمل درد
0.02980	تمایل به حمایت اجتماعی
0.03950	عقب‌ماندگی روانی حرکتی
0.02500	مغز پراکنده
0.06050	اختلال محرک
0.07820	سوء مصرف مواد
0.04750	هیپوکندریازیس
0.15010	انگیزه خودکشی
0.12890	ترومای دوران کودکی

یکی از پارامترهای مورد توجه در مقایسه مدل‌های یادگیری، مدت زمان انجام محاسبات هر مدل بوده که در جدول ۲، مدت زمان انجام هر پارامتر مشخص شده است.

جدول ۲. مدت زمان انجام

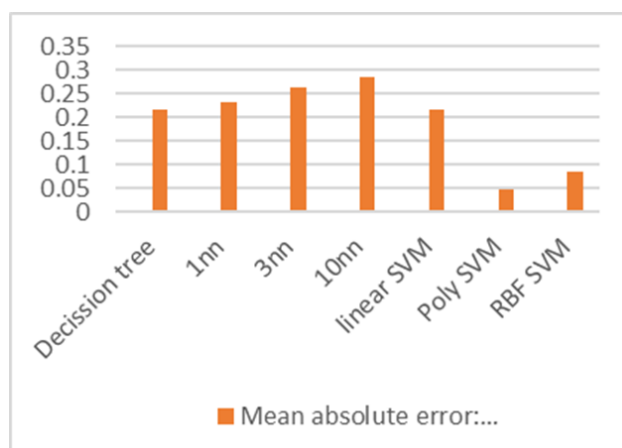
Time taken to build:	
Model name	value
Decision tree	0.12
1nn	0.01
3nn	0.01
10nn	0.01
linear SVM	0.06
Poly SVM	0.04
RBFSVM	0.06

پارامتر دیگر بسیار مهم در ارزیابی، مدل‌ها دقت بوده که در اینجا دقت مدل‌ها در تشخیص‌های درست و نادرست بیان شده است.

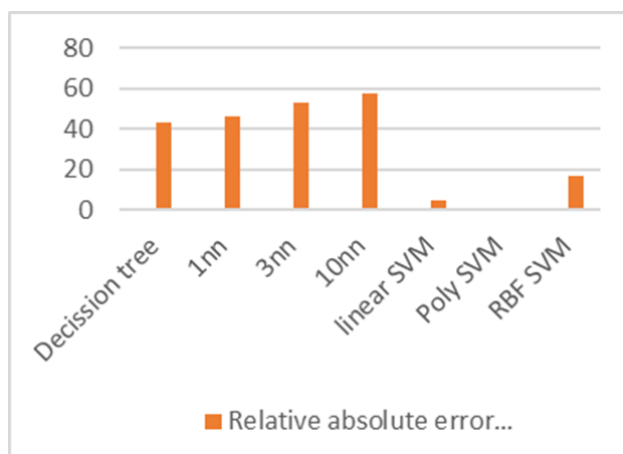


شکل ۲. دقت مدل

اندازه‌گیری خطا به صورت میانگین‌گیری از خطای مطلق محاسبه می‌شود. در مبحث خطای‌های مدل، MAE میانگین تفاوت بین مقدار واقعی و مقدار پیش‌بینی‌شده بر روی تمام نمونه‌های آموزش است.



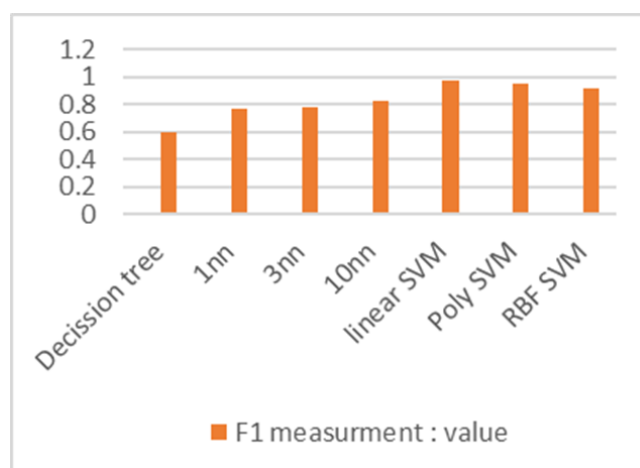
شکل ۳. نمودار میانگین خطای مطلق



شکل ۴. خطای مطلق نسبی

معیار F میانگین هارمونیک از دقت و یادآوری را به ما می‌دهد. به بیان ریاضی، امتیاز F میانگین وزن‌دار از دقت و یادآوری است. بهترین مقدار برای معیار F ، ۱ و بدترین مقدار، عدد صفر است. با کمک رابطه زیر می‌توان امتیاز F را محاسبه کرد:

$$F = 2 * (precision * recall) / (precision + recall) \quad (۱)$$

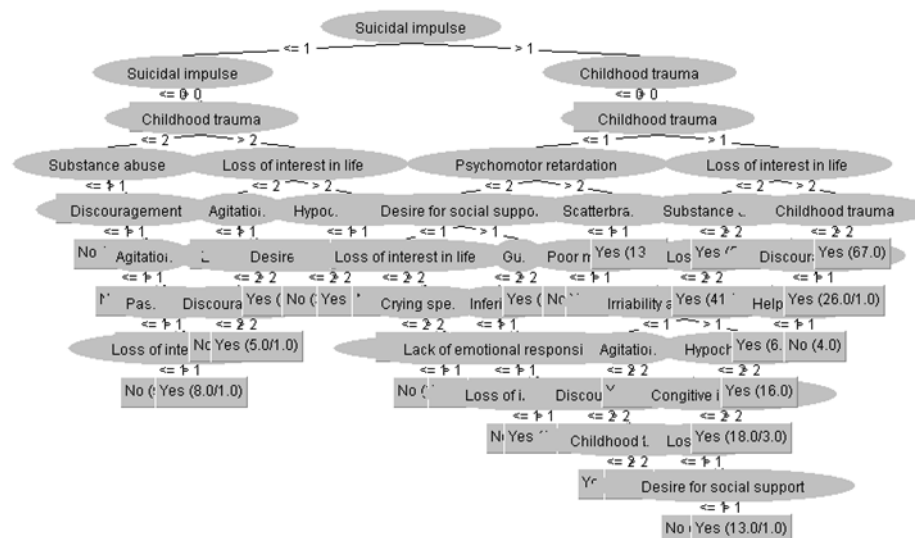


شکل ۵. معیار F

درخت تصمیم، یک تکنیک طبقه‌بندی بوده که از سه جزء گره ریشه، شاخه (لبه یا پیوند) و گره برگ تشکیل شده است. ریشه، شرایط آزمون را برای ویژگی‌های مختلف نشان می‌دهد، شاخه نشان‌دهنده تمام نتایج ممکن است که می‌تواند در آزمون وجود داشته باشد و گره‌های برگ حاوی برچسب کلاسی هستند که به آن تعلق دارد. گره ریشه در ابتدای درخت است که به آن بالای درخت نیز می‌گویند. طبقه‌بندی ۴۸۱ یک الگوریتم برای تولید یک درخت تصمیم است که توسط ۴/۵C تولید می‌شود. همچنین به عنوان یک طبقه‌بندی آماری شناخته می‌شود. در نهایت، مدل به دست آمده را در شکل ۶ نمایش داده‌ایم.

Weka Classifier Tree Visualizer: 02:48:55 - trees.J48 (600p)

Tree View



شکل ۶. خروجی درخت تصمیم

بحث و نتیجه‌گیری

در قسمت‌های قبل به معرفی متدهای داده‌کاوی مبتنی بر یادگیری ماشین پرداختیم و سپس منطبق بر روش‌های استاندارد جمع‌آوری داده، نسبت به جمع‌آوری داده‌های این پژوهش اقدام کردیم. داده‌های مصاحبه‌شوندگان نیز پس از مصاحبه بالینی توسط متخصص صلاحیت‌دار برچسب‌گذاری و سپس توسط مدل‌های ذکرشده به ماشین یادگیری سپرده شد و با استفاده از روش‌های اعتبارسنجی متقابل، خروجی مدل‌ها تعیین شد. پس از بررسی، مدل رگرسیون به علت خروجی متفاوت از گردونه رقابت کنار گذاشته شد و در پایان به بررسی و جمع‌بندی نتایج به‌دست‌آمده پرداختیم. با توجه به جمع‌بندی نتایج و مقایسه انواع پارامترهای کارایی متدهای عنوان‌شده و با توجه به شکل‌های ۲، ۳ و ۵، مشخصاً الگوریتم‌های مبتنی بر ماشین بردار پشتیبان بسیار کارآمدتر به نظر می‌رسند. همچنین از میان الگوریتم‌های دسته‌بندی ماشین بردار پشتیبان، مشخصاً ماشین بردار پشتیبان خطی موفق‌تر عمل کرده است.

پیشنهادها

در این پژوهش سعی بر انتخاب مدلی جهت ارائه یک سامانه پشتیبان تصمیم‌گیری با هدف شناسایی اختلال افسردگی در جامعه هدف بود که البته پیشنهاد نگارنده، تحقیق سایر پژوهشگران بر روی سایر اختلالات روانشناسی است. در بحث هوش ماشین و محاسبات نرم پیشنهاد می‌شود سایر پژوهشگران بر روی استفاده از الگوریتم‌های فراابتکاری جهت بهینه‌سازی (پارامترهای) مدل انتخابی (ماشین بردار پشتیبان خطی) تحقیق کنند تا کارایی مدل بیشتر شود.

منابع

- بختیار، معصومه، اسماعیل‌پور، منصور و ابراهیمی، محمداسماعیل. (۱۳۹۶). تشخیص افسردگی با استفاده از هوش مصنوعی. *کنگره انجمن روانشناسی ایران*.
- پارساپور فرشته، پیمانی جاوید. (۱۳۹۷) استفاده از تکنیک‌های داده‌کاوی جهت تشخیص هوشمند شدت اختلال افسردگی. *مجله انفورماتیک سلامت و زیست پزشکی*.

داودی، راحله و مرادی، محمدحسن. (۱۳۹۷). طراحی یک سیستم فازی عمیق مبتنی بر قاعده به منظور تعیین سطح افسردگی. نشریه علمی مهندسی پزشکی زیستی.

- Alonso, S. G., de La Torre-Díez, I., Hamrioui, S., López-Coronado, M., Barreno, D. C., Nozaleda, L. M., & Franco, M. (2018). Data mining algorithms and techniques in mental health: a systematic review. *Journal of medical systems*, 42, 1-15.
- Costa, V. G., & Pedreira, C. E. (2023). Recent advances in decision trees: An updated survey. *Artificial Intelligence Review*, 56(5), 4765-4800.
- Dipnall, J. F., Pasco, J. A., Berk, M., Williams, L. J., Dodd, S., Jacka, F. N., & Meyer, D. (2016). Fusing data mining, machine learning and traditional statistics to detect biomarkers associated with depression. *PloS one*, 11(2), e0148195.
- Ilgen, M. A., Downing, K., Zivin, K., Hoggatt, K. J., Kim, H. M., Ganoczy, D., ... & Valenstein, M. (2009). Exploratory data mining analysis identifying subgroups of patients with depression who are at high risk for suicide. *The Journal of clinical psychiatry*, 70(11), 6977.
- Lim, J. (2023). A Predictive Model of Ischemic Heart Disease in Middle-Aged and Older Women Using Data Mining Technique. *Journal of Personalized Medicine*, 13(4), 663.
- Tu, M., Xiong, S., Lv, S., Wu, X., Hu, H., Hu, R., ... & Shao, X. (2023). Acupuncture for major depressive disorder: a data mining-based literature study. *Neuropsychiatric Disease and Treatment*, 1069-1084.
- Van Loo, H. M., Bigdeli, T. B., Milaneschi, Y., Aggen, S. H., & Kendler, K. S. (2020). Data mining algorithm predicts a range of adverse outcomes in major depression. *Journal of Affective Disorders*, 276, 945-953.
- Vanlalawmpuia, R., & Lalhmingliana, M. (2020). Prediction of depression in social network sites using data mining. In *2020 4th International Conference on Intelligent Computing and Control Systems (ICICCS)* (pp. 489-495). IEEE.
- Vasha, Z. N., Sharma, B., Esha, I. J., Al Nahian, J., & Polin, J. A. (2023). Depression detection in social media comments data using machine learning algorithms. *Bulletin of Electrical Engineering and Informatics*, 12(2), 987-996.
- Zafar, A., & Chitnis, S. (2020). Survey of depression detection using social networking sites via data mining. In *2020 10th international conference on cloud computing, data science & engineering (confluence)* (pp. 88-93). IEEE.

