نوع مقاله: پژوهشی

DOR:

مجله علمی
پژوهش‌های میان‌رشته‌ای زنان
دوره پنجم، شماره ۱، بهار ۱۴۰۲
(شماره ۱۴ پیاپی)
صفحات ۴۴-۳۵

Woman Interdisciplinary
Researches Journal
Vol. 5, No. 1, Spring 2023
Pages 35-44

پیش بینی اختلال اضطراب فراگیر در بین زنان دانشجو با استفاده از رویکرد جنگل تصادفی

زهرا غلامی[۱]، حبیبه زارع[۲*]

۱- دانشجوی کارشناسی ارشد روانشناسی، دانشگاه آزاد اسلامی، فیروزآباد، ایران.

Zroyaee@yahoo.com

۲- استادیارگروه زیست شناسی، دانشگاه پیام نور، تهران، ایران. (نویسنده مسئول)

drhzare@pnu.ac.ir

تاریخ دریافت: [۱۴۰۲/۹/۲] تاریخ پذیرش: [۱۴۰۲/۱۰/۱۵]

چکیده

سلامت روان یکی از بزرگترین چالش ها برای نسل کنونی است. اختلال اضطراب فراگیر (GAD) یکی از بسیاری از مشکلات سلامت روان است. افراد مبتلا به این اختلال نگرانی ها و تنش های اغراق آمیزی را در مورد رویدادهای روزمره تجربه می کنند. گزارش شده است که حدود ۵ درصد از جمعیت کشورهای توسعه یافته به GAD مبتلا هستند و زنان دو برابر بیشتر از مردان به این بیماری مبتلا می شوند و یک اتفاق رو به رشد در بین زنان بالاخص زنان دانشجو است. این پژوهش با هدف پیش بینی اختلال اضطراب فراگیر در بین زنان دانشجو با رویکرد جنگل تصادفی، انجام شده است. از روش داده کاوی جهت پیش بینی استفاده شد.جامعه پژوهشی را زنان دانشجوی دانشگاه آزاد شیرازتشکیل دادند. تعداد ۱۵۰ نفر از دانشجویان زن به روش تصادفی ساده انتخاب و با پرسشنامهDSM-IV , مورد ارزیابی قرار گرفتند. در این فرآیند، الگوریتم جنگل تصادفی برای تولید مدل پیش‌بینی پیشنهاد شده است. NetBeans IDE ابزاری بود که برای ساخت این پیاده سازی استفاده شد. جاوا زبان برنامه نویسی انتخاب شده برای کدگذاری این نمونه اولیه بود و از کتابخانه WEKA در این پیاده سازی استفاده شد. نتایج نشان داد که دقت پیش‌بینی با روش جنگل تصادفی بالای ۰.۹ است که نشان می‌دهد رویکرد جنگل تصادفی قادر به پیش‌بینی دقیق اختلال اضطراب فراگیر GAD است. برای ارزیابی ویژگی، رویکرد جنگل تصادفی در پیش‌بینی دقیق فردی که از GAD رنج نمی‌برد سازگاری نشان می‌دهد. نتایج به‌دست‌آمده از نمونه اولیه در مقایسه با خط پایه که در ابزار R پیاده‌سازی شده است، نسبتاً سازگار است. به طور خلاصه، رویکرد جنگل تصادفی عملکرد پیش‌بینی بالایی تولید می‌کند و می‌تواند روابط مهم بین پارامتر پیشنهادی و پارامتر وابسته را استخراج کند.

واژگان کلیدی: داده کاوی،اختلال اضطراب فراگیر؛ جنگل تصادفی.

۱- مقدمه

در عصر پرشتاب امروز، همه از جمله دانش آموزان، والدین، کارمندان و کارفرمایان در تلاش هستند تا رقابتی باقی بمانند. سبک زندگی رقابتی باعث می شود افراد مشتاق به موفقیت دست یابند و باعث می شود چالش ها، ناامیدی ها و خواسته های متعدد را مدیریت کنند. در این محیط تحت فشار، اختلالات اضطرابی به آرامی به سلامت روان آنها هجوم آورده است.

اختلالات اضطرابی می تواند اشکال مختلفی داشته باشد از جمله اختلال اضطراب فراگیر، فوبیا، اختلال هراس و اختلال اضطراب اجتماعی. اختلالات اضطرابی با اضطراب مکرر و بیش از حد و سایر علائم ناتوان کننده مشخص می شوند. اختلال اضطراب فراگیر (GAD)، یکی از شایع ترین اختلالات اضطرابی، به عنوان نگرانی و استرس بیش از حد در مورد رویدادها و مشکلات روزمره از جمله مسائل بی اهمیت در زندگی روزمره تعریف می شود. طبق آمار انجمن اضطراب و افسردگی آمریکا (Torpy, 2011) ۶.۸ میلیون بزرگسال یا ۳.۱ درصد از جمعیت ایالات متحده، تحت تأثیر اختلال اضطراب فراگیر هستند. افراد مبتلا به این اختلال فاجعه را پیش بینی می کنند و نسبت به پول، سلامتی، خانواده، کار و مسائل دیگر بدبین هستند. آنها تحت تأثیر چرخه ای از نگرانی هستند که می تواند عملکرد روزانه آنها را مختل کند.

جالب توجه است که این اختلال در زنان دو برابر بیشتر از مردان شایع است. از این رو، این پژوهش قصد دارد به طور خاص بر اختلال اضطراب فراگیر در بین زنان دانشجو تمرکز کند. اهداف این تحقیق عبارتند از:

- بررسی، اتخاذ و تقویت الگوریتم جنگل تصادفی برای پیش‌بینی اختلال اضطراب فراگیر در بین زنان
- توسعه، اجرا و ارزیابی رویکرد جنگل تصادفی برای تعمیم اختلال اضطراب

۲- مرور مبانی نظری و پیشینه

۲-۱- داده کاوی

امروزه داده کاوی در جریان اصلی است. این عملی است که در اواسط دهه ۱۹۹۰ به عنوان یک رویکرد جدید برای تجزیه و تحلیل داده ها و کشف دانش ظهور کرد (Yoo, Alafaireet, Marinov, Pena-Hernandez, Gopidi, Chang et al., 2012). پیشرفت فناوری محاسباتی و ذخیره سازی عظیم داده ها، علاقه محققان را برای کاوش در حوزه داده کاوی افزایش داده است. روش های مختلف داده کاوی به عنوان اقدامات متقابل برای بسیاری از مشکلات مانند ابداع استراتژی بازاریابی پیشرفته، تدوین برنامه های تجاری صرفه جویی در هزینه، تشخیص تقلب و شناسایی ایمیل های هرزنامه معرفی شده اند.

کاربرد داده کاوی را می توان به دو حوزه عمده طبقه بندی کرد که عبارتند از: پیش بینی خودکار روندها و رفتارها و کشف خودکار الگوهای ناشناخته قبلی (Yoo et al., 2012). فیاد، پیاتتسکی-شیپارو و اسمیت* (۱۹۹۶) نشان داد که اهداف پیش‌بینی و توصیف را می‌توان با استفاده از انواع روش‌های خاص داده‌کاوی به دست آورد. نباید مرزی وجود داشته باشد که مشخص کند هر تکنیک صرفاً در خدمت یک هدف از هدف پیشگویی یا توصیفی است. بنابراین، انتخاب روش های مناسب داده کاوی باید بر اساس ماهیت هدف (خروجی)، ویژگی ورودی، نیازهای محاسباتی روش ها، تحمل مقادیر از دست رفته، نقاط پرت و تعداد کمی از نقاط داده و قابلیت توضیح مدل باشد (Van, Gay, Kennedy, Barin & Leijdekkers, 2010).

*. Fayyad, Piatetsky-Shapiro & Smyth

روش های داده کاوی به دسته های مختلفی تقسیم می شوند. این مطالعه بر روی روش های داده کاوی خوشه بندی، طبقه بندی و تداعی تمرکز دارد. تجزیه و تحلیل خوشه‌ای یا خوشه‌بندی، یک تکنیک داده کاوی است که هدف آن گروه‌بندی اشیاء داده به کلاس‌ها یا خوشه‌ها است، به طوری که اشیاء درون یک خوشه شباهت زیادی به یکدیگر دارند، اما با اشیاء در همه خوشه‌های دیگر بسیار متفاوت هستند. هدف طبقه بندی پیش بینی کلاس اشیاء در مواردی است که برچسب کلاس ناشناخته است (Aminudin, Fadiawati & Tania, 2015). قواعد ارتباط دانشی را نشان می دهد که در مجموعه داده ها به عنوان پیامدهای احتمالی و مربوط به محاسبه مجموعه های مکرر است (Fayyad et al., 1996).

۲-۲- داده کاوی در حوزه بهداشت و درمان

فرآیند داده کاوی به طور گسترده در زمینه های متعددی به کار گرفته شده است و در حوزه مراقبت های بهداشتی رواج بیشتری یافته است زیرا نیاز به یک روش تحلیلی کارآمد برای شناسایی اطلاعات ناشناخته و با ارزش در داده های مراقبت های بهداشتی وجود دارد (Tomar & Agarwal, 2013). در مراقبت های بهداشتی، در میان سایر عملکردها، فرآیند داده کاوی برای شناسایی تقلب در بیمه سلامت و پیش بینی خطرات بیماری، هزینه های پزشکی بیماران و مدت اقامت بیماران در بیمارستان به کار گرفته شده است. نتیجه عمل داده کاوی در این زمینه برای ارائه مزایایی برای کل اکوسیستم مراقبت های بهداشتی در نظر گرفته شده است.

برای هدف پیش‌بینی در رابطه با مراقبت‌های بهداشتی، روش طبقه‌بندی در مقایسه با روش‌های داده‌کاوی خوشه‌بندی و ارتباط مناسب‌تر تلقی می‌شود. طبقه بندی دارای ویژگی های سادگی، سرعت طبقه بندی نمونه های بدون برچسب و نمایش گرافیکی بصری است. جدای از آن، مدل پیش بینی آن می تواند به راحتی توسط متخصصان حوزه تایید و درک شود (Matthiesen, 2010).

۲-۳- جنگل تصادفی

جنگل تصادفی یکی از روش های طبقه بندی است. این روشی برای تجمیع توانایی پیش‌بینی طبقه‌بندی‌کننده‌های متعدد است که به عنوان طبقه‌بندی گروهی شناخته می‌شود. طبقه‌بندی گروهی یک روش یادگیری ماشینی است که هدف آن دستیابی به دقت پیش‌بینی بهتر در مقایسه با مدل‌های منفرد با استفاده از مزایای مدل‌های متعدد است (Oza, 2009). تمایل دارد مدل های پایه ای را تولید کند که در عین حال مکمل باشند و در نتیجه یک طبقه بندی جامع ایجاد شود. بسیاری از الگوریتم های یادگیری ماشین سنتی یک مدل واحد مانند درخت تصمیم و شبکه عصبی تولید می کنند. با این حال، روش یادگیری گروهی چندین مدل تولید می کند. لوکیس و ماراگوداکیس* (۲۰۱۰) آزمایشی را نشان داد تا نشان دهد چگونه یک روش مجموعه ای می تواند عملکرد طبقه بندی کننده را بهبود بخشد. آزمایش میزان خطای محاسبه‌شده ۰/۰۶ را در طبقه‌بندی گروهی نشان داد که در مقایسه با طبقه‌بندی باینری ۰/۳۵ در آزمایش کمتر بود.

جنگل تصادفی بسیاری از طبقه‌بندی‌کننده‌های پایه درختی را ایجاد می‌کند، که در آنها هر درخت به مقادیر یک بردار ورودی تصادفی بستگی دارد که به طور مستقل از کل مجموعه داده با جایگزینی و با توزیع یکسان همه درختان در جنگل نمونه‌برداری شده است (Loukis & Maragoudakis, 2010). تصادفی بودن جنگل تصادفی از نمونه ها و متغیرهای تصادفی منتخب نشات می گیرد.

روش‌های خوشه‌بندی معمولاً زمانی انجام می‌شوند که اطلاعات بسیار کمی در مورد داده‌ها شناخته شده باشد یا هیچ اطلاعاتی وجود نداشته باشد (Fayyad et al., 1996). اغلب برای پیش‌بینی و طبقه‌بندی مسائل استفاده نمی‌شود، اما روش خوبی برای تشخیص یک الگوی پنهان از مجموعه داده است. موضوع اصلی مربوط به قوانین انجمن معدن در مجموعه داده های مراقبت های بهداشتی، تعداد

*. Loukis & Maragoudakis

زیادی از قوانین کشف شده است که اکثر آنها بی ربط هستند. برای ایجاد مشکل بیشتر، مرتبط ترین قوانین با معیارهای با کیفیت بالا فقط در مقادیر پشتیبانی پایین ظاهر می شوند (Yoo et al., 2012).

رویکرد جنگل تصادفی دقت خوبی در عملکرد کلی خود نشان داده است. خلیلیا، چاکرابورتی و پاپسیو* (۲۰۱۱) نشان داد که جنگل تصادفی† از ماشین بردار پشتیبان، کیسه‌بندی و تقویت از نظر ناحیه زیر منحنی‡ مشخصه عملیاتی گیرنده§ برای پیش‌بینی خطرات بیماری ناشی از داده‌های بسیار نامتعادل عملکرد بهتری داشت. با این حال، محدودیتی در داده‌های به‌دست‌آمده وجود دارد که ممکن است منجر به استفاده چندین بار از داده‌های یکسان برای یک بیمار شود که ممکن است باعث سوگیری جزئی در نتیجه پیش‌بینی شود.

به طور کلی، الگوریتم جنگل تصادفی دقت پیش‌بینی بالایی ایجاد می‌کند. از این رو، در این پژوهش، این الگوریتم در حوزه دیگری از سلامت که همان سلامت روان است، برای تعیین تناسب و عملکرد آن اعمال می شود.

۲-۴- اختلال اضطراب فراگیر** (GAD)

اضطراب یک واکنش طبیعی است مانند احساس اضطراب قبل از نشستن برای معاینه، عصبی بودن هنگام شرکت در مصاحبه و ناراحتی به دلیل مواجهه با مشکل مالی. اینها همه موقعیت های اضطراب آور هستند. با این حال، افرادی که دارای شرایط اضطرابی مزمن مانند داشتن نگرانی مداوم، بیش از حد و غیر واقعی در مورد چیزهای روزمره مانند مسئولیت های شغلی، سلامت خانواده یا مسائل بی اهمیت مانند کارها و قرار ملاقات ها و همراه با علائم جسمی هستند، می توانند به عنوان مبتلا به اختلال اضطراب فراگیر در نظر گرفته شوند. مبتلایان نگرانی و اضطراب بیش از حد را تجربه می کنند و اغلب انتظار بدترین اتفاقات را دارند حتی زمانی که هیچ دلیل واضحی برای نگرانی وجود ندارد. به طور خلاصه، آنها قادر به مهار نگرانی های خود نیستند.

GAD بر اساس کتابچه راهنمای تشخیصی و آماری اختلالات روانی، تعریف شده است که یک طبقه بندی استاندارد از اختلالات روانی است که توسط متخصصان بهداشت روان در جهان استفاده می شود. بر اساس راهنمای تشخیصی و آماری اختلالات روانی (American Psychiatric Association, 2013) این اختلال با ویژگی های انتظارات دلهره آمیز تشخیص داده می شود که بیش از روزها برای یک دوره حداقل شش ماهه، در مورد تعدادی از رویدادها یا موضوعات رخ می دهد. نگرانی باعث پریشانی یا اختلال عملکردی می شود و با حداقل سه مورد از علائم زیر مانند بی قراری، خستگی آسان، مشکل در تمرکز، تحریک پذیری، تنش عضلانی و اختلال خواب همراه است (Behar, DiMarco, Hekler, Mohlman & Staples, 2009). با این حال، تشخیص GAD تنها در صورتی امکان‌پذیر است که فرد معیارهای تشخیصی سایر اختلالات اضطرابی را در آن دوره نداشته باشد (American Psychiatric Association, 2013).

۳- روش‌شناسی

از آنجایی که پارامتر نشانه‌های افسردگی را هدف قرار داده بود، پارامترها از BDI، یکی از پرکاربردترین تست‌های روان‌سنجی برای اندازه‌گیری شدت افسردگی گرفته شد. آنها توسط یک روانپزشک ثبت شده تأیید شدند تا از ارتباط آن در ارزیابی اختلال اضطراب

*. Khalilia, Chakraborty & Popescu
†. Random Forest
‡. AUC
§. ROC
**. Generalized Anxiety Disorder

فراگیر اطمینان حاصل شود. پارامترها در جدول ۱ نشان داده شده است که شامل ۲۴ پارامتر شامل جزئیات شخصی و علائم افسردگی است.

جدول ۱- پارامترهای مورد استفاده

اطلاعات شخصی	۱. سن (عددی)
	۲. شغل (اسمی)
	۳. دوران کودکی ناخوشایند
	تجربه (اسمی)
علائم افسردگی	۴. اندوه
	۵. بدبینی
	۶. شکست گذشته
	۷. از دست دادن لذت
	۸. احساس گناه
	۹. احساس تنبیه
	۱۰. بیزاری از خود
	۱۱. انتقاد از خود
	۱۲. افکار یا آرزوهای خودکشی
	۱۳. گریه کردن
	۱۴. آشفتگی
	۱۵. از دست دادن علاقه
	۱۶. بلاتکلیف
	۱۷. بی ارزشی
	۱۸. از دست دادن انرژی
	۱۹. تغییر در الگوی خواب
	۲۰. تحریک پذیری
	۲۱. تغییر در اشتها
	۲۲. مشکل تمرکز
	۲۳. خستگی ناشی از خستگی
	۲۴. از دست دادن علاقه به رابطه جنسی

داده های این پژوهش از طریق پیمایش جمع آوری شده است. در این نظرسنجی، هر شرکت کننده از دانشجویان زن دانشگاه آزاد شیراز موظف بود یک عدد را پر کند. پرسشنامه ای که مشتمل بر گروه هایی از عبارات برای توصیف احساسات وی بر اساس پارامترهایی است که در بالا ذکر شد. این پرسشنامه بر اساس پرسشنامه DSM V (American Psychiatric Association, 2013) طراحی شده است. این نظرسنجی با ۱۵۰ پاسخ دهنده زن دانشجو برای ایجاد پایگاه داده GAD ایجاد کرد. بر اساس داده‌های جمع‌آوری‌شده، تنها داده‌های پاسخ‌دهندگان زن برای تمرکز بر زنان در این پژوهش استخراج شد.

پس از جمع آوری داده ها، کیفیت داده ها باید تایید شود. بنابراین، روش پیش پردازش داده ها انجام شد. در جمع آوری داده ها، مشکلی که با آن مواجه شد، ورود نامناسب داده ها مانند ارائه پاسخ نامربوط در نظرسنجی بود. تاپل با ورود نامناسب داده حذف شد تا داده‌های پر سر و صدا را هموار کند یا با محتمل‌ترین مقدار آن را پر کند که یکی از محبوب‌ترین استراتژی‌ها برای مقابله با این موضوع بود. علاوه بر این، تابع یافتن و جایگزینی برای رسیدگی به ناسازگاری در قالب داده‌ای که از نظرسنجی به‌دست آمده بود، استفاده شد. در مکانیزم تبدیل داده ها، داده های جمع آوری شده به فرم های مناسب برای استفاده در فرآیند پیاده سازی تبدیل شد.

داده‌های جمع‌آوری‌شده که به صورت بیانیه بود، سپس با فرمول‌بندی فرمول‌هایی برای ایجاد امتیاز برای هر تاپل با استفاده از Microsoft Excel به یک امتیاز تبدیل شد. مقیاس نمره برای هر پاسخ با ارزیابی وضعیت افسردگی پاسخگو از ۰ تا ۳ امتیاز اختصاص یافت. افرادی که بیش از ۲۸ امتیاز کسب کردند، مبتلا به افسردگی شدید در نظر گرفته شدند که احتمالاً می تواند نشان دهنده GAD باشد. تعمیم تکنیکی بود که برای سازماندهی داده های اولیه در دسته های سطح بالاتر آن استفاده می شد. قالب مورد نیاز برای ورود به فرآیند پیاده سازی به صورت arff. از این رو، مجموعه داده پردازش شده به نوع فایل arff. تبدیل شد.

تجزیه و تحلیل داده ها یک مرحله مقدماتی قبل از اینکه مجموعه داده از طریق فرآیند داده کاوی منتقل شود برای بازرسی و شناسایی هر گونه مجموعه داده سوگیری احتمالی که ممکن است منجر به عملکرد ضعیف فرآیند داده کاوی شود، بود. پس از تجزیه و تحلیل مجموعه داده تمیز شده، عدم تعادل کلاس در جایی که موارد مثبت کلاس اقلیت بودند، شناسایی شد.

نسبت نمونه های مثبت به موارد منفی به دست آمده از داده های جمع آوری شده ۱:۱۰ بود. بر اساس این نسبت، روش کم‌نمونه‌گیری حذف شد، زیرا موارد مثبت در مقایسه با موارد منفی به طور قابل‌توجهی کم بود. از این رو، SMOTE به جای آن برای افزایش نمونه های اقلیت با ایجاد نمونه های "مصنوعی" در مجموعه داده به عنوان ابزاری برای کاهش مجموعه داده های سوگیری استفاده شد. در این فرآیند، از ابزار فیلتر Weka برای تولید توزیع داده های بی طرفانه استفاده شد و مجموعه داده به دست آمده برای استفاده در فرآیند بعدی ذخیره شد. پس از فرآیند SMOTE، داده های به دست آمده به ۱: ۲.۳ تقسیم بندی شدند. سپس، این مجموعه داده برای جلوگیری از خوشه‌بندی داده‌های سنتز شده تصادفی شد. این مرحله از طریق تابع تصادفی در ابزار Weka به دست آمد و مجموعه داده پردازش شده نهایی برای فرآیند بعدی ذخیره شد.

در فرآیند طراحی و اجرا ، الگوریتم جنگل تصادفی برای تولید مدل پیش‌بینی پیشنهاد شده است.NetBeans IDE ابزاری بود که برای ساخت این پیاده سازی استفاده شد. جاوا زبان برنامه نویسی انتخاب شده برای کدگذاری این نمونه اولیه بود و از کتابخانه WEKA در این پیاده سازی استفاده شد. پارامترهای اجرای این جنگل تصادفی عبارتند از:

الف) نمونه گیری بوت استرپ: به اندازه مجموعه آموزشی.

ب) تعدادی درخت: ۱۰ و ۱۰۰ درخت تصادفی برای مقایسه همبستگی تعداد درختان و نتیجه تولید شده رشد می کنند.

پ) تعداد ویژگی های انتخاب شده به طور تصادفی برای انتخاب در یک گره خاص:

floor (log2(N))+1 = floor(log2(25))+1= 5

N تعداد کل ویژگی ها در داده ها، از جمله ویژگی کلاس است.

ت) اطلاعات برای انتخاب تقسیم به منظور تقسیم فضای داده های طبقه بندی شده به دست می آید. به دست آوردن اطلاعات و شاخص جینی هر دو معیارهای تقسیم ناخالصی هستند. تفاوت در تابع ناخالصی است. هر درخت به بیشترین میزان ممکن رشد می

کند، هیچ هرس اعمال نمی شود، و پیش بینی کلی بر اساس اکثریت آرای طبقه طبقه بندی شده است. علاوه بر این، تخمین احتمال GAD تنظیم شده به عنوان احتمال پیش‌بینی ۰/۸>= به عنوان یک وضعیت جدی در نظر گرفته می‌شود، در حالی که بین ۰.۵ تا ۰.۸ یک وضعیت متوسط و کمتر از ۰/۵ یک وضعیت خفیف در نظر گرفته می‌شود. شاخص های عملکرد Random Forest نیز در نمونه اولیه گنجانده شده است.

پس از اجرای فرآیند، نتایج تجربی به دست آمد. به منظور ایجاد درک بهتر از روش پیشنهادی، تعداد درختان در جنگل‌های تصادفی، داده‌های پردازش‌شده و داده‌های پردازش نشده در نظر گرفته شد تا به نمونه اولیه جنگل تصادفی وارد شود و نتایج تولید شده در آن مورد بحث قرار گیرد.

۴- یافته‌ها

نتایج حاصل از ابزار R بر اساس پیاده سازی تصادفی جنگل در زیر نشان داده شده است. واریانس داده های مجموعه داده متعادل و نامتعادل در جدول ۲ نشان داده شده است.تحلیل عملکرد برای این مطالعه در جدول ۳ نشان داده شده است.

جدول ۲- مجموعه داده متعادل و نامتعادل مورد استفاده

پیاده سازی تصادفی جنگل R	تعداد درخت	تعداد ویژگی	مجموعه داده
مورد۱	۱۰۰	۵	مجموعه داده متعادل
مورد۲	۱۰	۵	مجموعه داده متعادل
مورد۳	۱۰۰	۵	مجموعه داده نامتعادل
مورد۴	۱۰	۵	مجموعه داده نامتعادل

جدول ۳- تجزیه و تحلیل عملکرد از پیاده سازی R

تجزیه و تحلیل عملکرد	دقت	حساسیت	ویژگی
مورد۱	۰.۹۹۳۱	۰.۹۷۷۳	۱
مورد۲	۰.۹۳۷۹	۰.۹۳۰۲	۰.۹۷۵۹
مورد۳	۰.۹۴۶۴	۰.۴۵۴۵	۱
مورد۴	۰.۹۲۸۵	۰.۳۶۳۶	۰.۹۹۰۱

حساسیت (true positive rate) به معنی نسبتی از موارد مثبت است که آزمایش آن‌ها را به درستی به عنوان مثبت علامت‌گذاری می‌کند .ویژگی(true negative rate) به معنی نسبتی از موارد منفی است که آزمایش آن‌ها را به درستی به عنوان منفی علامت‌گذاری می‌کند.

۵- بحث و نتیجه‌گیری

نتایج نشان داد که مجموعه داده متعادل از دقت، حساسیت و ویژگی بالاتری در مقایسه با مجموعه داده نامتعادل در هر دو پیاده سازی برخوردار است. با اعمال SMOTE به مجموعه داده، اثر مثبت نمونه‌گیری داده‌ها مشاهده شد. این نتیجه نشان می‌دهد که گنجاندن نمونه‌گیری داده‌ها عملکرد طبقه‌بندی را بهبود می‌بخشد.

نتایج اجرای نمونه اولیه نشان داد که افزایش تعداد درختان بر دقت، حساسیت و ویژگی تأثیری ندارد. با این حال، برای اجرای R، در مورد ۱۰۰ درخت در مقایسه با ۱۰ درخت دقت بهتری را نشان داد. به طور کلی، درختان بیشتر معمولاً دقت فزاینده ای ایجاد می کنند. با این حال، این اثر زمانی که به یک نقطه خاص برسد، صاف می شود. برای توضیح نتیجه به‌دست‌آمده از اجرای نمونه اولیه، این ممکن است به این دلیل باشد که مجموعه داده برای این تحقیق به طور قابل‌توجهی کوچک بود، جایی که تعداد درخت‌ها عملکرد طبقه‌بندی را به طور قابل‌توجهی بهبود نمی‌بخشد.

از آنجایی که این تحقیق برای دستیابی به هدف پیش‌بینی اختلال اضطراب فراگیر انجام شده است، نشان داده شده است که تمام دقت پیش‌بینی بالای ۰/۹ است که نشان می‌دهد رویکرد جنگل تصادفی قادر به پیش‌بینی دقیق GAD است. از نظر حساسیت، نتایج نوسان بین مجموعه داده متعادل و مجموعه داده نامتعادل را نشان می دهد. آنها نشان می دهند که مجموعه داده نامتعادل می تواند بر عملکرد طبقه بندی کننده تأثیر بگذارد. برای ارزیابی ویژگی، رویکرد جنگل تصادفی در پیش‌بینی دقیق فردی که از GAD رنج نمی‌برد سازگاری نشان می‌دهد. نتایج به‌دست‌آمده از نمونه اولیه در مقایسه با خط پایه که در ابزار R پیاده‌سازی شده است، نسبتاً سازگار است. به طور کلی، رویکرد جنگل تصادفی عملکرد پیش‌بینی خوبی ایجاد کرده است که با نتایج به‌دست‌آمده ثابت شده است، همانطور که در جدول ۲ نشان داده شده است. به منظور آشکار کردن رابطه بین متغیرهای وابسته و مستقل، نتایج به‌دست‌آمده خستگی یا خستگی، افکار یا آرزوهای خودکشی، از دست دادن را توصیف می‌کنند. به عنوان پیشرو در تابلوی امتیاز متغیر اهمیت دارد، که توجیه می کند چرا این پارامترها نقش مهمی در پیش بینی اینکه آیا یک فرد دچار GAD شده است یا خیر. می توان این گونه تفسیر کرد که اگر فردی بیش از حد معمول خسته باشد و تمایل بیشتری به خودکشی داشته باشد و نسبت به افراد و چیزها از دست داده باشد، این علائم نشان می دهد که او دچار اختلال اضطراب فراگیر شده است. بنابراین، این می تواند بینش هایی را برای بخش سلامت روان فراهم کند تا بتواند GAD را در مراحل اولیه تشخیص دهد و آنها را قادر سازد تا اقداماتی را برای درمان بیمار انجام دهند.

به طور خلاصه، رویکرد جنگل تصادفی عملکرد پیش‌بینی بالایی تولید می‌کند و می‌تواند روابط مهم بین پارامتر پیشنهادی و پارامتر وابسته را استخراج کند.

۶- تقدیر و تشکر

بدینوسیله از دکتر زهره رویایی استاد گروه کامپیوتر که ما را در انجام این تحقیق یاری کردند، صمیمانه تشکر می‌کنیم.

۷- منابع

1- American Psychiatric Association, D. S. M. T. F., & American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders: DSM-5* (Vol. 5, No. 5). Washington, DC: American psychiatric association.
2- Aminudin, M. A., Fadiawati, N., & Tania, L. (2015). Pengembangan LKS berbasis multipel representasi pada materi klasifikasi materi. *Jurnal Pendidikan dan Pembelajaran Kimia*, *4*(2), 720-731.
3- Behar, E., DiMarco, I. D., Hekler, E. B., Mohlman, J., & Staples, A. M. (2009). Current theoretical models of generalized anxiety disorder (GAD): Conceptual review and treatment implications. *Journal of anxiety disorders*, *23*(8), 1011-1023.

4- Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). From data mining to knowledge discovery in databases. *AI magazine*, *17*(3), 37-37.
5- Khalilia, M., Chakraborty, S., & Popescu, M. (2011). Predicting disease risks from highly imbalanced data using random forest. *BMC medical informatics and decision making*, *11*, 1-13.
6- Loukis, E., & Maragoudakis, M. (2010, June). Heart murmurs identification using random forests in assistive environments. In *Proceedings of the 3rd International Conference on Pervasive Technologies Related to Assistive Environments* (pp. 1-6).
7- M. M. Antony, "Recommended Readings and DVDs Anxiety Disorders , Depression , and Related Problems," vol. 2631, no. August, pp. 1–17, 2013.
8- Matthiesen, R. (Ed.). (2010). *Bioinformatics methods in clinical research*. Totowa, NJ, USA: Humana Press.
9- Oza, N. C. (2009). Ensemble data mining methods. In *Encyclopedia of Data Warehousing and Mining, Second Edition* (pp. 770-776). IGI Global.
10- Tomar, D., & Agarwal, S. (2013). A survey on Data Mining approaches for Healthcare. *International Journal of Bio-Science and Bio-Technology*, *5*(5), 241-266.
11- Torpy, J. M. (2011). Generalized anxiety disorder. *JAMA*, *305*(5), 522.
12- Van, A., Gay, V. C., Kennedy, P. J., Barin, E., & Leijdekkers, P. (2010, December). Understanding risk factors in cardiac rehabilitation patients with random forests and decision trees. In *Conferences in Research and Practice in Information Technology Series*.
13- Yoo, I., Alafaireet, P., Marinov, M., Pena-Hernandez, K., Gopidi, R., Chang, J. F., & Hua, L. (2012). Data mining in healthcare and biomedicine: a survey of the literature. *Journal of medical systems*, *36*, 2431-2448.

Predicting Generalized Anxiety Disorder Among Female Students Using Random Forest Approach

Zahra Gholami[1], Habibeh Zare[2]*

1. Master's student in psychology, Islamic Azad University, Firozabad, Iran.

Zroyaee@yahoo.com

2. Assistant Professor, Department of Biology, Payam Noor University, Tehran, Iran. (Corresponding Author)

drhzare@pnu.ac.ir

Abstract

Mental health is considered one of the major challenges for the generations. Generalized anxiety disorder (GAD) is one of many mental health complications. However, individuals with the disorder experience hyperbolic concerns and tensions regarding daily events. Furthermore, it is reported that approximately 5% of the population of developed countries suffer from GAD. Additionally, women are affected by this disease twice as often as men, and it is an increasing disorder among women, particularly female students. This paper aims to predict generalized anxiety disorder among female students using the random decision forest algorithm. The data mining method was utilized for prediction. Female students of Shiraz Azad University developed the research community. Therefore, 150 female students were selected by simple random method and tested with a DSM-IV questionnaire. Accordingly, a random forest algorithm is proposed to generate a prediction model. Moreover, NetBeans IDE was applied for operationalization. Java was the programming language to code the prototype, and the WEKA library was involved in the operation. However, the results showed that the prediction accuracy with the random forest algorithm exceeds 0.9, which indicates that the algorithm is likely to predict GAD accurately. The random decision forest algorithm consistently predicts an individual not suffering from GAD. The results are relatively consistent compared to the baseline employed in the R. However, the random decision forest algorithm produces high predictive performance and may display significant relationships between the proposed and dependent parameters.

Keywords: Data mining, Generalized Anxiety Disorder (GAD), Random Decision Forest.