

پیش بینی سلامت روان نوجوانان ایرانی بر اساس تغذیه، فعالیت فیزیکی و اوقات فراغت با استفاده از تکنیک‌های داده کاوی

عسل آقاداتودیان جلفائی^۱، مریم جهانبخش^۲، محمد ستاری^۲، روبیا کلیشادی^۳

۱- کارشناس ارشد، گروه مدیریت و فن آوری اطلاعات سلامت، دانشگاه علوم پزشکی اصفهان، اصفهان، ایران

۲- استادیار، مرکز تحقیقات فن آوری اطلاعات در امور سلامت، دانشگاه علوم پزشکی اصفهان، اصفهان، ایران

۳- استاد، مرکز تحقیقات رشد و نمو کودکان، پژوهشکده پیشگیری اولیه از بیماری‌های غیرواگیر، دانشگاه علوم پزشکی اصفهان، اصفهان، ایران

چکیده

مقدمه و اهداف: پژوهش حاضر با هدف پیش‌بینی سلامت روان بر اساس سه عامل تغذیه، فعالیت فیزیکی و اوقات فراغت در دانش‌آموزان گروه سنی نوجوانان با بهره‌گیری از تکنیک‌های داده کاوی انجام شده است. **روش کار:** مطالعه‌ی تحلیلی حاضر بر روی ۱۴۲۷۴ داده موجود در پایگاه داده کاسپین ۵ انجام شد. مطابق با روش CRISP-DM داده کاوی در ۶ مرحله با استفاده از تکنیک‌های درخت تصمیم، k نزدیک‌ترین همسایه، بی‌زین ساده و جنگل تصادفی در نرم افزار رپیدماینر صورت گرفت. **یافته‌ها:** از میان ۴ تکنیک داده کاوی مورد استفاده جهت پیش‌بینی سلامت روان نوجوانان بر اساس تغذیه، فعالیت فیزیکی و اوقات فراغت، تکنیک جنگل تصادفی بیشترین صحت (۹۱/۷۲) و تشخیص پذیری (۸۲/۷۳) و تکنیک k نزدیک‌ترین همسایه بیشترین حساسیت (۹۶/۳۰) را داشت. علاوه بر این بر اساس تکنیک جنگل تصادفی، قانون با بالاترین میزان پشتیبان نشان داد نوجوانی که در مقطع تحصیلی دبیرستان است و هر روز هفته صبحانه، ناهار و شام میل می‌کند و به صورت هفتگی چای و قهوه میل می‌کند و در هفته ۲ ساعت در مدرسه ورزش می‌کند همچنین در هفته اخیر ۴ روز به مدت ۳۰ دقیقه فعالیت بدنی داشته و با سرویس به مدرسه می‌رود با اطمینان ۱۰۰ درصد از سلامت روان مطلوب برخوردار است. **نتیجه‌گیری:** براساس تکنیک جنگل تصادفی که بهترین عملکرد را داشته، تغذیه بیشترین تأثیر را بر سلامت روان نوجوانان ایرانی دارد، لذا باید جهت تأمین و بسترسازی مناسب آموزش به والدین و نوجوانان در رابطه با تغذیه مناسب و افزایش آگاهی در زمینه سلامت روان نوجوانان اندیشیده شود.

اطلاعات مقاله

تاریخ دریافت

۱۴۰۲/۰۶/۲۰

تاریخ پذیرش

۱۴۰۲/۰۹/۲۷

نویسنده رابط

مریم جهانبخش

ایمیل نویسنده رابط

jahanbakhsh.him@gmail.com

نشانی نویسنده رابط

مرکز تحقیقات فن آوری اطلاعات در

امور سلامت، دانشگاه علوم پزشکی

اصفهان، اصفهان

واژگان کلیدی: سلامت روان،

نوجوانان، داده کاوی، تغذیه، فعالیت

فیزیکی، اوقات فراغت

مقدمه

بر اساس تعریف سلامتی که روزی به عنوان نبود بیماری و اکنون برخورداری از آسایش جسمی، روانی و اجتماعی تلقی می‌شود؛ سلامت روان یکی از مهم‌ترین ابعاد سلامتی است که نقش برجسته ای در تضمین پویایی و کارآمدی هر جامعه ایفا می‌نماید (۱-۳). از نظر سازمان جهانی بهداشت نیز سلامت روان یکی از محورهای مهم ارزیابی سلامتی در جوامع مختلف است (۲).

این در حالی است که اختلالات روانی یکی از سه علت اولیه سال‌های از دست رفته‌ی عمر به دلیل ناتوانی است که در سال ۲۰۲۰ سهم آن از بار کلی بیماری‌ها ۵۰ درصد افزایش یافته و از ۱۰/۵ درصد به ۱۵ درصد کل بار بیماری‌ها رسیده است (۱، ۷-۴). طبق آمار سازمان جهانی بهداشت نیز بیش از پانصد میلیون نفر در دنیا مبتلا به یکی از

انواع اختلالات روانی هستند که در این میان نزدیک به پنجاه میلیون نفر از اختلالات شدید روانی رنج می‌برند (۸). در واقع می‌توان گفت تقریباً از هر ۱۰ نفر یک نفر مبتلا به اختلالات روانی است یا از هر ۴ خانواده یک خانواده عضوی با اختلالات روانی دارد (۹، ۱۰). در ایران نیز این آمار از سایر کشورها کمتر نبوده و در سال‌های اخیر روند فزاینده ای داشته است؛ چنانچه ۲۸/۲ درصد در مردان و ۳۱/۳ درصد در زنان گزارش شده است (۱۱).

همچنین بر خلاف تصور عمومی رشد صعودی ابتلا به ناخوشی‌های روانی محدود به جمعیت بالغ و بزرگسال کشورها نیست بلکه کودکان و نوجوانان بیش از سایر جمعیت‌ها نسبت به این نوع اختلالات آسیب پذیر هستند (۱۲)؛ به طوری که شایع‌ترین بیماری‌های دوران نوجوانی اختلالاتی چون اضطراب، افسردگی و پریشانی است (۱۳). بر

این تکنیک‌ها می‌تواند صنعت سلامت را از تحلیل عمیق داده‌ها بهره مند سازد، به توسعه تحقیقات پزشکی و تصمیم‌گیری‌های علمی در زمینه‌های پیشگیری و درمان منتج شود و پیش‌آگهی و درک عمیق‌تری از طبقه‌بندی بیماری‌ها به ویژه در رابطه با مشکلات سلامت روان ارائه دهد (۲۹، ۳۰).

لیکن در مطالعه حاضر با امید به اینکه بررسی عوامل مؤثر بر سلامت روان در نوجوانان با استفاده از فرآیند داده کاوی بتواند شاخص خوبی برای پیش‌بینی و حتی پیشگیری این بیماری‌ها در بزرگسالی باشد به پیش‌بینی سلامت روان بر اساس سه عامل تغذیه، فعالیت فیزیکی و اوقات فراغت در دانش‌آموزان گروه سنی نوجوانان با بهره‌گیری از تکنیک‌های داده کاوی پرداخته شده است.

روش کار

مطالعه تحلیلی- کاربردی حاضر در سال ۱۴۰۱ بر روی داده‌های پیمایش کاسپین صورت گرفته است. پیمایش کاسپین برگرفته از پیمایش جهانی سلامت دانش‌آموزان^۱ است که تا کنون در ۵ دوره با هدف بررسی وضعیت و روند تغییر رفتارهای سلامت و عوامل حفاظت‌کننده در دانش‌آموزان در ایران صورت گرفته است. آخرین دوره این پیمایش تحت عنوان کاسپین ۵ در سال‌های ۹۴-۱۳۹۳ در مناطق شهری و روستایی ۳۱ استان به روش نمونه‌گیری خوشه‌ای (خوشه‌ای چند مرحله‌ای) با حجم نمونه ۴۸۰ نفر (۴۸ خوشه‌ی ۱۰ نفری) در هر استان اجرا شده و داده‌های مربوط به ۱۴۲۷۴ دانش‌آموز در رده سنی ۱۸-۷ سال به دست آمده است. در پژوهش حاضر نیز جامعه آماری شامل اطلاعات دانش‌آموزان ۷ تا ۱۸ سال کشوری، ثبت شده از طریق پرسشنامه کاسپین ۵، موجود در پایگاه داده پژوهشکده پیشگیری اولیه از بیماری‌های غیرواگیر دانشگاه علوم پزشکی اصفهان است که به روش مشاهده و مطالعه استخراج گردیده و سپس بر مبنای روش استاندارد CRISP-DM در ۶ مرحله به شرح ذیل تحلیل گشت:

۱. **شناسایی سیستم:** در این مرحله ابتدا به شناخت پژوهشکده پیشگیری اولیه از بیماری‌های غیرواگیر دانشگاه

اساس آمارهای منتشر شده نیز شیوع بالای مشکلات روانی در نوجوانان ۱۰ تا ۲۰ درصد برآورد شده است (۱۴)؛ در ایران نیز ۳۹ درصد دانش‌آموزان در رده سنی نوجوانان مبتلا به مشکلات رفتاری - عاطفی می‌باشند (۱۵). علاوه بر این، بسیاری از آسیب‌های روانی در دوران بزرگسالی در واقع ناشی از مشکلات در دوران نوجوانی است (۱۶). به گونه‌ای که بر اساس آمارها نیمی از اختلالات روانی از سن ۱۴ سالگی آغاز می‌شوند (۱۷). از این رو توجه و تمرکز بر روی سلامت روان نوجوانان از اهمیت بالایی برخوردار است (۱۸).

غالباً نوجوانی را دوره سلامتی می‌پندارند و این قشر از جامعه با همین پیش فرض از خدمات مراقبتی پیشگیری بی بهره مانده‌اند (۱۶). در حالی که در این دوره سنی سه عامل مؤثر بر سلامت روان وجود دارد که می‌تواند به عنوان رفتارهای پیشگیری‌کننده نیز مدنظر قرار گیرد. این سه عامل عبارتند از: تغذیه مناسب، سبک زندگی پرتحرک و اوقات فراغت (۱۹). بر اساس شواهد یک رژیم غذایی متنوع و متراکم از مواد مغذی به عبارتی الگوی غذایی سالم با سلامت روان بهتر در نوجوانان همراه است در حالی که رژیم‌های غذایی کم ارزش و وعده‌های غذایی نامنظم با اختلال در سلامت روان همراه می‌باشد (۲۰، ۲۱). علاوه بر این بر اساس مطالعات، علائم افسردگی و اضطراب در نوجوانان می‌تواند با افزایش سطح فعالیت فیزیکی آن‌ها کاهش یابد (۲۲، ۲۳). همچنین میان فعالیت‌های بی تحرک در اوقات فراغت همچون مشاهده تلویزیون و کار با رایانه با سلامت روان رابطه معکوس گزارش شده است (۲۴). Marateb و همکارانش در مطالعه بر روی داده‌های کاسپین بر اهمیت عوامل مذکور تاکید داشته‌اند (۲۵).

لذا با توجه به شیوع بالای اختلالات روانی، حساسیت بالای دوره نوجوانی، تأثیر عادات رفتاری بر سلامت روان و نقش برجسته سلامت روان در رشد و تکامل اجتماعی - روانی نوجوانان بررسی عوامل تأثیرگذار بر سلامت روان و پیش‌بینی به‌هنگام اختلالات روانی در نوجوانان با استفاده از تکنیک‌های هوشمندی همچون تکنیک‌های داده کاوی نسبت به روش‌های سنتی تشخیصی از اطمینان بالاتری برخوردار است (۱۵، ۱۹، ۲۶).

داده کاوی که به عنوان فرآیندی برای یافتن الگوها و ارتباطات در پایگاه‌های داده به همراه استفاده از اطلاعات برای ساختن مدل‌های پیش‌بینی‌کننده تعریف شده است بیش از یک دهه است با هدف شناسایی سوء استفاده‌های مالی وارد عرصه سلامت شده و به تدریج در حوزه بالینی نیز مورد توجه قرار گرفته است (۲۷، ۲۸). استفاده از

¹ GSHS :Global School-Based Student Health Survey.

۲. اطلاعات دانش‌آموزانی که در شهر زندگی می‌کنند مورد بررسی قرار گرفت. زیرا طی مشورت‌های صوت گرفته با متخصصان سلامت تفاوت در سبک زندگی شهرنشینی و روستائین می‌تواند بر عادات و سلامت روان افراد مؤثر باشد.

■ کاهش ابعاد: جهت کاهش ابعاد متغیرهای مورد بررسی، بر اساس نظرات متخصصین بالینی و داده کاوی برخی از سوالات که ماهیت کمی پیوسته داشته‌اند مانند ساعت خوردن صبحانه، نهار و شام از داده‌ها حذف گردید. سپس بر اساس پرسشنامه استاندارد پاسخ سوالات در بخش‌های تغذیه، اوقات فراغت و فعالیت فیزیکی کدگذاری شد و عملیات جمع کردن، میانگین گرفتن و رند کردن برخی از متغیرها صورت گرفت. این امر منجر به کاهش تعداد متغیرها از ۴۳ به ۲۶ متغیر گشت. در نهایت، انتخاب ویژگی، به عنوان یکی از مراحل کاهش ابعاد، بر اساس سه درخت ID3، CHAID و C4.5 در نرم افزار رپیدماینر اجرا گردید که با توجه به اینکه در هر سه درخت همه‌ی ۲۶ ویژگی استفاده گردید، در مطالعه حاضر نیز تمام ۲۶ متغیر به عنوان ورودی مورد استفاده قرار گرفت.

■ پاک سازی داده: درصد فراوانی داده‌های از دست رفته برای هر یک از متغیرها محاسبه شد و با در نظر گرفتن یک قاعده کلی که اگر داده‌های از دست رفته در هر متغیر ۵ درصد و یا کمتر باشد، بدین شرط که حذف رکورد بر کاهش چشمگیر حجم نمونه منتهی نشود و حذف کمتر از ۱۵ درصد نمونه کل باشد، می‌توان از آن‌ها چشم پوشی کرد (۳۲)؛ داده‌های از دست رفته در تمامی متغیرها، به استثنای احساس غمگینی (درصد داده‌های از دست رفته بیش از ۵ درصد)، به روش حذف کردن از پایگاه داده پاکسازی شد. داده‌های از دست رفته در متغیر احساس غمگینی نیز با میانگین داده‌های موجود جایگذاری شد.

■ نرمال سازی: به منظور جلوگیری از تأثیر بازه مقادیر متفاوت بر تکنیک‌های داده کاوی، تمامی متغیرها بر اساس تابع نرمالسازی حداقل- حداکثر در بازه ۰ تا ۱ نگاشت شد.

■ برچسب گذاری: در نهایت، در مراحل کشف و استخراج دانش، خوشه بندی می‌تواند برای پیش پردازش داده‌ها و یا

علوم پزشکی اصفهان و پیمایش کاسپین پرداخته شد و سپس تعریف مسئله، تعیین اهداف، تعیین گستره تحقیق و شناخت ابزارهای مورد نیاز بر اساس مطالعات و با مشورت صاحب نظران در حوزه‌ی سلامت روان و داده کاوی صورت گرفت. در نهایت رپیدماینر، که نرم افزاری توانمند در مباحث داده کاوی، یادگیری ماشین، تحلیل و پیش‌بینی می‌باشد به عنوان ابزار داده کاوی انتخاب گردید.

۲. شناخت داده: تمام مجموعه داده کاسپین برگرفته از پرسشنامه مبتنی بر برنامه نظام مراقبت سلامت دانش‌آموزان (GSHS) کاسپین شامل: ارتباط با همسالان و دوستان، وضعیت جثه، وضعیت مدرسه، عادات‌های تغذیه‌ای، عادات‌های بهداشتی، فعالیت فیزیکی، فعالیت‌های اوقات فراغت، آسیب‌ها و جراحات، پرسش‌هایی در مورد خشونت، پرسش‌هایی در مورد زورگویی و قلدری، مواجهه با دود دخانیات، سوالاتی در مورد خانواده، سوالات سلامت و ناخوشی، مصرف دخانیات، رضایت از زندگی، عملکرد تحصیلی، مواجهه با نور خورشید، تعامل با محیط و آلودگی‌های محیطی و یافته‌های معاینه‌های بالینی مورد ارزیابی قرار گرفت و پس از برگزاری جلسات بارش افکار با صاحب نظران حوزه‌های تغذیه، فعالیت فیزیکی، فعالیت‌های اوقات فراغت و سلامت و ناخوشی به عنوان متغیرهای غالب و قابل بکارگیری در پژوهش تعیین گردید.

۳. آماده سازی و پیش پردازش داده‌ها: در این مرحله، با هدف افزایش کیفیت داده‌های مورد استفاده، مجموعه‌ای از فعالیت‌ها صورت گرفته است که شامل:

■ استخراج داده: داده‌های مربوط به بخش‌های ذکر شده در قسمت قبل در قالب فایل Excel از پایگاه داده پژوهشکده پیشگیری اولیه از بیماری‌های غیرواگیر دانشگاه علوم پزشکی اصفهان استخراج گردید. سپس نمونه گیری هدفمند بر اساس دو معیار زیر صورت گرفت:

۱. تنها اطلاعات دانش‌آموزان در رده سنی ۱۳ تا ۱۸ سال در نظر گرفته شده است. زیرا بر اساس متون علمی و نظر سازمان جهانی بهداشت، بیشتر علائم اختلالات روانی در دوران نوجوانی آغاز خواهد شد (۱۷) و دوران نوجوانی نیز به گروه سنی ۱۳ تا ۱۸ سال اطلاق می‌شود (۳۱).

نزدیکترین همسایه از "تشابه ویژگی" برای پیش بینی مقادیر نقاط داده جدید استفاده می‌کند؛ که به این معنی است که به نقطه داده جدید بر اساس میزان مطابقت آن با نقاط مجموعه آموزشی، یک مقدار تخصیص می‌دهد. یک الگوریتم جنگل تصادفی از چندین درخت به نام درخت تصمیم (Decision Tree)، تشکیل شده است. جنگل تشکیل شده توسط الگوریتم جنگل تصادفی از طریق Bagging Classifier، آموزش داده می‌شود. Bagging یک متا الگوریتم مجموعه‌ای است که دقت الگوریتم‌های یادگیری ماشین را افزایش می‌دهد.

۵. **ارزیابی:** در عملگر Cross Validation داده‌های بخش آموزش مدل را می‌سازد و داده‌های بخش آزمایش مدل ایجاد شده را بر اساس معیارهای مختلف مورد ارزیابی قرار می‌دهد. در این مطالعه از معیارهای صحت، حساسیت و تشخیص پذیری جهت ارزیابی عملکرد هر تکنیک در نرم افزار رپیدماینر استفاده گردید.

معیار صحت با فرمول ذیل بدست می‌آید:

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN}$$

معیار حساسیت و تشخیص پذیری هم به ترتیب با فرمول‌های ذیل بدست می‌آید.

$$\text{Sensitivity} = \frac{TP}{TP+FN}$$

$$\text{Specificity} = \frac{TN}{TN+FP}$$

۶. **توسعه:** در این مرحله ابتدا نتایج حاصل از درخت تصمیم و جنگل تصادفی که به صورت قوانین میان متغیرهای تغذیه، فعالیت فیزیکی و اوقات فراغت با سلامت روان قابل دستیابی بود بر اساس در نظر گرفتن حد آستانه کمینه ۷۵ برای دو معیار پشتیبان و اطمینان جداسازی گردید. سپس قوانین حاصل به صورت جدول بندی در اختیار ۸ نفر از متخصصان حوزه‌های اطفال، تغذیه، روانپزشکی کودک و نوجوان، گوارش و کبد کودک و نوجوان و آموزش بهداشت و ارتقا سلامت که توسط پژوهشکده پیشگیری اولیه از بیماری‌های غیرواگیر به عنوان خبره معرفی شدند قرار گرفت. هر یک از این متخصصین در حوزه‌هایی همچون تغذیه و تأثیرات مواد غذایی

آماده سازی آن‌ها بکار برده شود (۳۳). در این پژوهش نیز تمامی داده‌های اولیه بدون برچسب سلامت روان بوده است؛ از این رو از ۹ متغیر مرتبط با بخش سلامت و ناخوشی (شامل: احساس بی‌ارزش بودن، زود عصبانی شدن و یا بدخلقی، احساس دلشوره، بدخواهی، احساس گیجی و منگی، احساس غمگینی، احساس نگرانی، وضعیت سلامتی فرد از دیدگاه خود و وضعیت زندگی فرد از دیدگاه خود) جهت تعیین برچسب سلامت روان دانش‌آموزان، در تکنیک پرکاربرد خوشه بندی K-Means در نرم افزار رپیدماینر استفاده گردید. علاوه بر این تعداد دسته‌ها در این تکنیک با استفاده از شاخص ارزیابی دیویس- بولدین و نظرات متخصصان بالینی تعیین گردید.

۴. **مدل سازی و اجرای تکنیک:** در این مرحله به اجرای تکنیک‌های طبقه بندی پرداخته شد. ابتدا در نرم افزار رپیدماینر با استفاده از عملگر Cross Validation داده‌ها به صورت تصادفی به ۱۰ قسمت مساوی تقسیم شده و در هر تکرار یک قسمت به عنوان آزمایش و نه قسمت دیگر به عنوان آموزش انتخاب شدند. سپس تکنیک‌های درخت تصمیم، تکنیک بی‌زین ساده، تکنیک k نزدیک‌ترین همسایه و تکنیک جنگل تصادفی جهت کشف الگوهای رابطه‌ای میان داده‌ها و پیش‌بینی سلامت روان دانش‌آموزان، مورد استفاده قرار گرفت.

درخت تصمیم از تعدادی گره و شاخه تشکیل شده است که در آن نمونه‌ها را به نحوی طبقه بندی می‌کند که از ریشه به سمت پایین رشد می‌کند و در نهایت به گره‌های برگ می‌رسد. هر گره داخلی یا غیر برگ با یک ویژگی مشخص می‌شود. این ویژگی سوالی را در رابطه با مثال ورودی مطرح می‌کند. در هر گره داخلی به تعداد جواب‌های ممکن با این سوال شاخه وجود دارد که هر یک با مقدار آن جواب مشخص می‌شوند. برگ‌های این درخت با یک کلاس و یا یک طبقه از جواب‌ها مشخص می‌شوند.

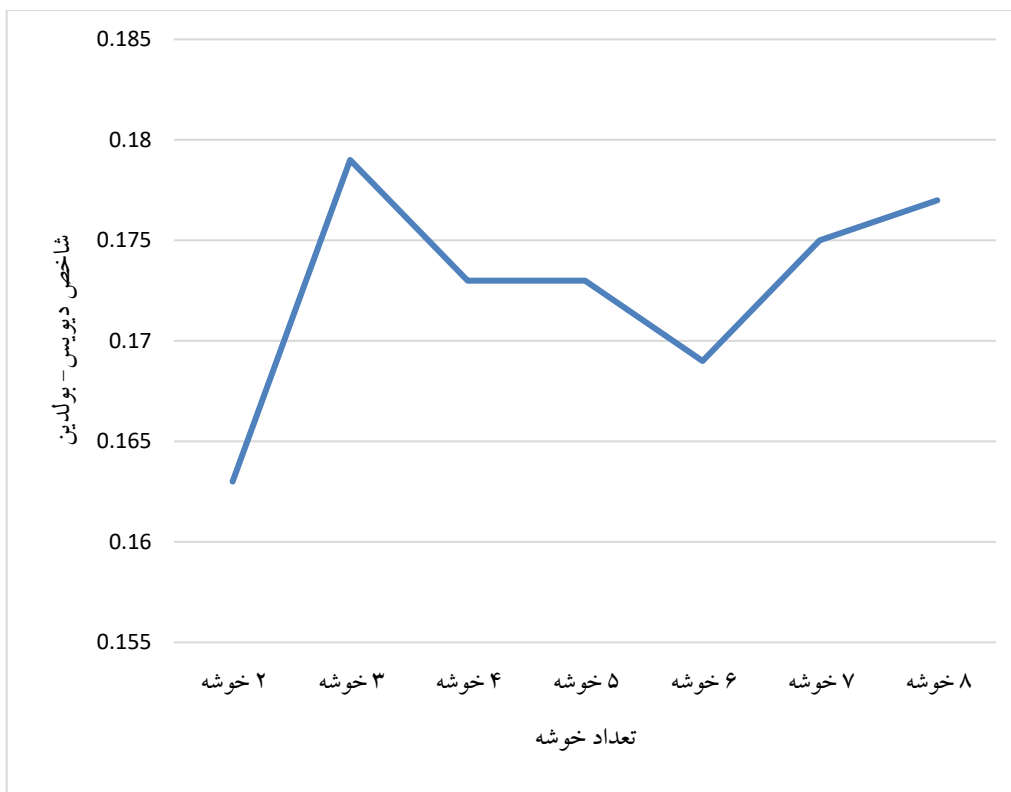
بی‌زین ساده، یک روش طبقه بندی بر اساس قضیه بی‌زین با فرض استقلال در بین پیش بینی کننده‌ها است. به زبان ساده، یک طبقه بند بی‌زین ساده فرض می‌کند که وجود یک ویژگی خاص در یک کلاس با وجود ویژگی دیگر ارتباطی ندارد. الگوریتم K

لازم به ذکر است که در جهت حفظ ملاحظات اخلاقی در این پژوهش اصل رازداری و محرمانگی داده‌ها رعایت گردیده (داده‌های مورد مطالعه فاقد اطلاعات هویتی بوده و دانش‌آموزان بر اساس کدهای شناسه یکتا قابل تمایز بوده‌اند.) و نتایج بدون هیچ گونه دخل و تصرف منتشر گردید.

یافته‌ها

در ابتدا جهت خوشه بندی داده‌ها از تکنیک K-Means استفاده شد که نتایج حاصل از تعیین تعداد خوشه بهینه در این تکنیک بر اساس شاخص ارزیابی دیویس-بولدین به شرح ذیل می باشد:

مختلف بر نوجوانان، اختلالات اضطرابی نوجوانان، تأثیر فعالیت فیزیکی بر سلامت روان، تأثیر انواع ویتامین‌ها بر رشد و روان نوجوانان، اثرات فضای مجازی بر سلامت نوجوانان مطالعات فراوانی داشته‌اند از این رو خبرگان با توجه به پیشنهاد اولیه توسط پژوهشگرده پیشگیری اولیه از بیماری‌های غیر واگیر و بررسی حوزه فعالیت و مطالعات هر یک به مطالعه حاضر وارد گردیدند. در نهایت قوانینی که با موافقت بیش از ۷۵ درصد از خبرگان همراه بود، گزارش گردید و بر اساس معیارهای صحت، حساسیت و تشخیص پذیری که در مراحل قبل به دست آمده بود بهترین تکنیک در جهت پیش‌بینی سلامت روان بر اساس سه عامل تغذیه، اوقات فراغت و فعالیت فیزیکی انتخاب و ارائه گردید.



نمودار شماره ۱- شاخص ارزیابی دیویس-بولدین در تعداد خوشه متفاوت

گردیده، شاخص ارزیابی دیویس-بولدین به کمترین مقدار دست یافته است. از این رو با توجه به شاخص ارزیابی و نظرات خبرگان در حوزه سلامت، تعداد خوشه ۲ جهت خوشه بندی سلامت روان دانش‌آموزان انتخاب گردید.

شاخص دیویس-بولدین میانگین شباهت بین هر خوشه با شبیه‌ترین خوشه به آن را محاسبه می‌کند، لذا هرچه مقدار شاخص کمتر باشد خوشه‌های مناسب‌تری تولید می‌شود. براساس نمودار ۱ زمانی که تعداد خوشه برابر با ۲ انتخاب

جدول شماره ۱- نقاط مرکزی متغیرهای سلامت و ناخوشی در هر یک از خوشه‌ها

نام متغیر	میانگین در خوشه اول	میانگین در خوشه دوم
احساس بی ارزش بودن	۰/۵۴۵	۰/۰۱۷
زودعصبانی شدن یا بدخلقی	۰/۳۶۶	۰/۰۲۸
احساس دلشوره	۰/۷۲۷	۰/۱۸۹
بد خوابی	۰/۶۲۱	۰/۱۸۹
احساس گیجی و منگی	۰/۴۶۳	۰/۰۷۱
احساس غمگینی	۰/۱۴۶	۰/۰۴۴
احساس نگرانی	۰/۳۸۶	۰/۱۵۲
وضعیت سلامتی فرد از دیدگاه خود	۰/۳۷۶	۰/۲۰۹

با اجرای تکنیک درخت تصمیم بر روی داده‌های پیش پردازش شده، بر اساس معیار انتخاب صفت بهره اطلاعاتی^۱، از میان ۲۸۱ قانون سلامت روان دانش‌آموزان و عوامل تغذیه، فعالیت فیزیکی و اوقات فراغت به دست آمده ۱۶ قانون با در نظر گرفتن حد آستانه کمینه برای معیار پشتیبان معادل با ۷۵ و حد آستانه کمینه برای معیار اطمینان برابر با ۷۵ درصد استخراج گردید و به نظرخواهی از خبرگان گذاشته شد. در نهایت از میان ۱۶ قانون به دست آمده ۷ قانون ذیل مورد تأیید قرار گرفت.

بر اساس نتایج حاصل از جدول ۱ میان نقاط مرکزی در دو خوشه‌ی اول و دوم تفاوت چشمگیری مشاهده شده که این امر خود نمایانگر متمایز بودن دو خوشه می‌باشد. همچنین با توجه به بالاتر بودن تمامی نقاط مرکزی در خوشه اول نسبت به خوشه‌ی دوم به ترتیب این دو خوشه به عنوان کلاس سلامت روان نامطلوب و مطلوب نام گذاری شد. کلاس سلامت روان مطلوب با حجمی معادل ۳۱۶۸ نمونه، تقریباً دو برابر کلاس سلامت روان نامطلوب با حجم ۱۵۸۱ بود.

¹ Information Gain.

جدول شماره ۲ - قوانین درخت تصمیم

شماره قانون	قانون درخت تصمیم	وضعیت سلامت روان	معیار پشتیبان	معیار اطمینان (درصد)
۱	نوجوانی که به ندرت نوشیدنی گازدار مصرف می‌کند، روزانه لبنیات میل می‌کند، در هفته اخیر ۲ روز به مدت ۳۰ دقیقه فعالیت بدنی داشته و روزانه حدود یک ساعت با رایانه کار می‌کند.	سلامت روان مطلوب	۱۸۲	۹۸/۴ درصد
۲	نوجوانی که به ندرت سوسیس، کالباس، پیتزا و همبرگر میل می‌کند و در هفته اخیر ۳ روز به مدت ۳۰ دقیقه فعالیت بدنی داشته و روزانه یک ساعت تلویزیون تماشا می‌کند.	سلامت روان مطلوب	۸۷	۹۴/۳ درصد
۳	نوجوانی که در هفته اخیر ۴ روز به مدت ۳۰ دقیقه فعالیت بدنی داشته و روزانه یک ساعت صرف انجام تکالیف می‌کند همچنین در هفته ۲ ساعت در مدرسه ورزش می‌کند.	سلامت روان مطلوب	۷۵	۹۴/۷ درصد
۴	نوجوانی که به ندرت سوسیس، کالباس، پیتزا و همبرگر میل می‌کند و در هفته اخیر ۴ روز به مدت ۳۰ دقیقه فعالیت بدنی داشته و روزانه ۲ ساعت صرف انجام تکالیف می‌کند.	سلامت روان مطلوب	۸۷	۹۴/۳ درصد
۵	نوجوانی که روزانه چای و قهوه میل می‌کند و در هفته اخیر ۶ روز به مدت ۳۰ دقیقه فعالیت بدنی داشته است.	سلامت روان مطلوب	۸۳	۹۵/۲ درصد
۶	نوجوانی که در هفته اخیر ۷ روز به مدت ۳۰ دقیقه فعالیت بدنی داشته و پیاده به مدرسه می‌رود، در هفته یک ساعت در مدرسه ورزش می‌کند و روزانه یک ساعت صرف انجام تکالیف می‌کند.	سلامت روان مطلوب	۸۷	۹۸/۹ درصد
۷	نوجوانی که هرگز چیپس، پفک و چوبشور میل نمی‌کند و در هفته اخیر ۷ روز به مدت ۳۰ دقیقه فعالیت بدنی داشته و پیاده به مدرسه می‌رود و در هفته ۲ ساعت در مدرسه ورزش می‌کند.	سلامت روان مطلوب	۹۹	۹۸ درصد

۹۱/۷۲ درصد را نشان داد. لذا جهت بررسی قوانین موجود میان سلامت روان نوجوانان با تغذیه، فعالیت فیزیکی و اوقات فراغت تعداد درخت معادل با ۸ تعیین گردید. سپس از میان قوانین به دست آمده قوانینی که حد آستانه کمینه برای معیار پشتیبان معادل با ۷۵ و حد آستانه کمینه برای معیار اطمینان برابر با ۷۵ درصد را دارا بودند استخراج گردید. در نهایت ۵۸ قانون به دست آمده جهت بررسی نظرات خبرگان در اختیار آن‌ها قرار گرفت و مطابق با جدول زیر ۸ قانون استخراج شد.

بر اساس قوانین ذکر شده در جدول ۲ تعداد روز فعالیت بدنی حداقل به مدت ۳۰ دقیقه در هفته بیشترین تکرار را در قوانین داشته که می‌تواند نشان دهنده تأثیر بالای میزان فعالیت بدنی بر سلامت روان نوجوانان باشد.

علاوه بر این جهت کاوش قوانین حاصل از تکنیک جنگل تصادفی ابتدا صحت ۲ تا ۱۵ درخت مورد بررسی قرار گرفت که تعداد ۸ درخت در جنگل تصادفی بالاترین صحت معادل با

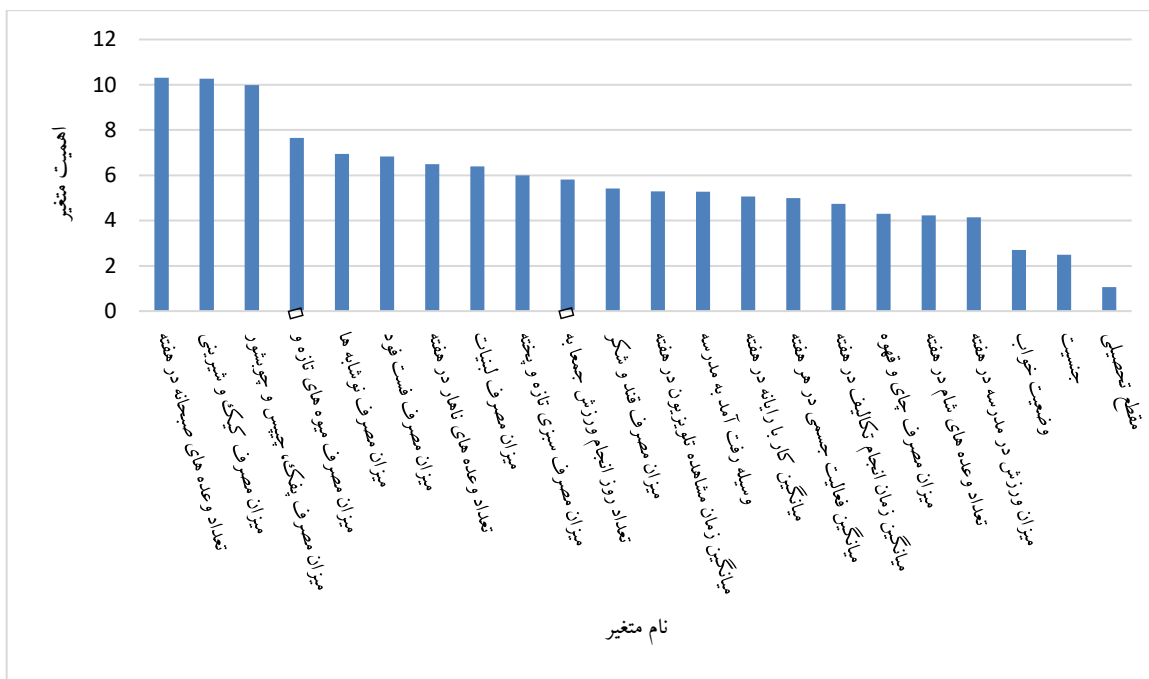
جدول شماره ۳- قوانین جنگل تصادفی

شماره قانون	قانون جنگل تصادفی	وضعیت سلامت روان	میزان معیار پشتیبان	درصد معیار اطمینان
۱	نوجوانی که در مقطع تحصیلی دبیرستان است و هر روز هفته صبحانه، نهار و شام میل می‌کند و به صورت هفتگی چای و قهوه میل می‌کند و در هفته ۲ ساعت در مدرسه ورزش می‌کند همچنین در هفته اخیر ۴ روز به مدت ۳۰ دقیقه فعالیت بدنی داشته و با سرویس به مدرسه می‌رود.	سلامت روان مطلوب	۱۵۸	۱۰۰ درصد
۲	نوجوانی که در هفته ۶ روز نهار و ۵ روز شام میل می‌کند و به ندرت چیپس، پفک و چوبشور میل می‌کند ولی به صورت روزانه میوه تازه، خشک و آبمیوه طبیعی میل می‌کند.	سلامت روان مطلوب	۱۵۴	۱۰۰ درصد
۳	نوجوانی که هر روز هفته صبحانه، نهار و شام میل می‌کند و به صورت هفتگی نوشابه، ماءالشعیر، چای و قهوه میل می‌کند و در هفته ۲ ساعت در مدرسه ورزش می‌کند همچنین با سرویس به مدرسه می‌رود و روزانه یک ساعت تلویزیون تماشا می‌کند.	سلامت روان مطلوب	۱۲۱	۱۰۰ درصد
۴	نوجوانی که به صورت هفتگی چای و قهوه میل می‌کند ولی به ندرت نوشابه و ماءالشعیر مصرف می‌کند و در هفته اخیر ۴ روز به مدت ۳۰ دقیقه فعالیت بدنی داشته و روزانه حدود یک ساعت با رایانه کار می‌کند همچنین با سرویس به مدرسه می‌رود و کمتر از ۹ ساعت در روز می‌خوابد.	سلامت روان مطلوب	۱۰۱	۱۰۰ درصد

شماره قانون	قانون جنگل تصادفی	وضعیت سلامت روان	میزان معیار پشתיبان	درصد معیار اطمینان
۵	نوجوانی که ۷ روز هفته صبحانه و شام میل می‌کند و به ندرت کیک، شیرینی، شکلات، چای و قهوه میل می‌کند و در یک هفته اخیر روزانه فعالیت جسمی داشته است.	سلامت روان مطلوب	۹۴	۱۰۰ درصد
۶	نوجوانی که در مقطع تحصیلی دبیرستان است و هر روز هفته صبحانه، ناهار و شام میل می‌کند و به صورت هفتگی چای و قهوه میل می‌کند و در هفته ۲ ساعت در مدرسه ورزش می‌کند همچنین در هفته اخیر ۲ روز به مدت ۳۰ دقیقه فعالیت بدنی داشته و هرگز با رایانه کار نمی‌کند علاوه بر این کمتر از ۹ ساعت در روز می‌خوابد.	سلامت روان مطلوب	۹۲	۱۰۰ درصد
۷	نوجوانی که در مقطع تحصیلی دبیرستان است و هر روز هفته صبحانه، ناهار و شام میل می‌کند و به ندرت چای و قهوه میل می‌کند همچنین روزانه حدود یک ساعت با رایانه کار می‌کند و ۹ ساعت در روز می‌خوابد.	سلامت روان مطلوب	۸۹	۱۰۰ درصد
۸	نوجوانی که ۶ روز در هفته صبحانه و شام میل می‌کند و به ندرت چای، قهوه، پیتزا، همبرگر، سوسیس و کالباس مصرف می‌کند و در یک هفته اخیر هرگز فعالیت جسمی نداشته است همچنین کمتر از ۹ ساعت در روز می‌خوابد.	سلامت روان نامطلوب	۷۵	۱۰۰ درصد

علاوه بر این یکی از ویژگی‌های مهم جنگل تصادفی، عملکرد بالای آن در اندازه‌گیری اهمیت متغیرها در پیش‌بینی متغیر هدف است. در این پژوهش نیز اهمیت هریک از متغیرهای تغذیه، فعالیت فیزیکی و فعالیت اوقات فراغت در پیش‌بینی سلامت روان نوجوانان در نمودار زیر نشان داده شده است.

باتوجه به جدول فوق، قانون شماره یک با بیشترین میزان پشתיبان نشان می‌دهد نوجوانی که در مقطع تحصیلی دبیرستان است و هر روز هفته صبحانه، ناهار و شام میل می‌کند، به صورت هفتگی چای و قهوه می‌خورد، در هفته ۲ ساعت در مدرسه ورزش می‌کند، همچنین در هفته اخیر ۴ روز به مدت ۳۰ دقیقه فعالیت بدنی داشته و با سرویس به مدرسه می‌رود با اطمینان ۱۰۰ درصد از سلامت روان مطلوب برخوردار است.



نمودار شماره ۲- اهمیت هریک از متغیرها بر اساس شاخص نسبت بهره در جنگل تصادفی

جنسیت و وضعیت خواب نوجوانان کمترین تأثیر را در پیش‌بینی سلامت روان داشته‌اند.

در نهایت با بررسی ماتریس درهم ریختگی، عملکرد چهار تکنیک پرکاربرد داده کاوی شامل درخت تصمیم، بی‌زین، k نزدیک‌ترین همسایه و جنگل تصادفی مورد مقایسه قرار گرفت.

با توجه به نمودار ۲، در پیش‌بینی سلامت روان نوجوانان ایرانی بر اساس سه عامل تغذیه، فعالیت فیزیکی و فعالیت اوقات فراغت با استفاده از جنگل تصادفی به ترتیب سه متغیر تعداد وعده‌های صبحانه، میزان مصرف کیک و شیرینی و میزان مصرف پفک، چیپس و چوبشور بیشترین تأثیر را داشته است. این درحالی است که متغیرهایی همچون مقطع تحصیلی،

جدول شماره ۴- حساسیت، تشخیص پذیری و صحت هر یک از تکنیک‌ها

تکنیک‌ها	حساسیت(درصد)	تشخیص پذیری (درصد)	صحت(درصد)
درخت تصمیم	۹۴/۰۳	۷۹/۸۸	۳۲/۸۹±۱/۰۸
بی‌زین	۸۷/۱۵	۵۱/۹۹	۷۵/۴۵±۱/۳۸
K نزدیک‌ترین همسایه	۹۶/۳۰	۸۲/۰۹	۹۱/۵۸±۱/۱۵
جنگل تصادفی	۹۶/۲۱	۸۲/۷۳	۹۱/۷۲±۱/۳۳

صحت و تشخیص پذیری و تکنیک k نزدیک‌ترین همسایه بیشترین حساسیت را دارد. به عبارت دیگر تکنیک k نزدیک‌ترین همسایه در پیش‌بینی صحیح کلاس سلامت روان

با توجه به جدول فوق، از میان ۴ تکنیک مورد استفاده جهت پیش‌بینی سلامت روان نوجوانان بر اساس تغذیه، فعالیت فیزیکی و اوقات فراغت، تکنیک جنگل تصادفی بیشترین

مطلوب و تکنیک جنگل تصادفی در پیش‌بینی صحیح کلاس سلامت روان نامطلوب بهتر عمل کرده است.

بحث

در میان قوانین حاصل از درخت تصمیم اصلی‌ترین عامل مؤثر بر پیش‌بینی سلامت روان نوجوانان مربوط به متغیر فعالیت بدنی است. در واقع اولین گویه‌ی تأثیرگذار که توانسته است شاخه نخست درخت را تشکیل دهد تعداد روزی است که نوجوانان در هر هفته حداقل به مدت ۳۰ دقیقه فعالیت فیزیکی داشته‌اند.

Kleszczewska و همکارانش نیز در مطالعه‌ای همسو با پژوهش حاضر نشان دادند که فعالیت بدنی با سلامت روان نوجوانان در ارتباط است. همچنین فعالیت فیزیکی جهت تعدیل علائم افسردگی بویژه در شهرهای بزرگ و خانواده‌های تک‌والد و یا فقیر اثر محافظتی دارد (۳۴).

بر خلاف آن در مطالعه Louise Bell و همکارانش، با استفاده از آنالیزهای چند متغیره، شواهد محکمی مبنی بر ارتباط میان میزان و یا شدت فعالیت بدنی با سلامت روان و یا علائم اختلالات روانی در نوجوانان نشان داده نشد. با این حال در یافته‌ها گزارش کرده‌اند که افزایش میزان فعالیت بدنی ممکن است با کاهش خطر مشکلات عاطفی در نوجوانان همراه باشد (۲۳).

بر اساس قوانین حاصل از جنگل تصادفی نیز مهم‌ترین عوامل مؤثر در پیش‌بینی سلامت روان نوجوانان ایرانی بر اساس تغذیه، فعالیت فیزیکی و فعالیت اوقات فراغت به ترتیب متعلق به تعداد وعده مصرف صبحانه، میزان مصرف کیک، شیرینی و شکلات و میزان مصرف پفک، چیپس و چوبشور است که هر سه در گروه تغذیه قرار دارند لذا می‌توان به اهمیت تغذیه بر سلامت روان نوجوانان تأکید کرد. Marateb و همکارانش طی داده کاوی بر روی کاسپین ۴ با استفاده از الگوریتم‌های داده کاوی نیز بر اهمیت تغذیه بر سلامت روان نوجوانان نسبت به سایر عوامل دست یافتند (۲۵).

Ingram و همکارانش نشان دادند که در دوران همه‌گیری بیماری کرونا تغییر در رژیم غذایی، خواب و فعالیت بدنی بیشترین تأثیر را بر سلامت روان افراد داشته‌اند (۳۵).

علاوه بر این Kulkarni و همکارانش در مطالعه خود تأثیر تغذیه بر سلامت روان نوجوانان را نشان داده‌اند به گونه‌ای که تأکید داشته‌اند هم افزایش مصرف غذای‌های سالم و هم کاهش مصرف غذاهای ناسالم بر سلامت روان نوجوانان تأثیرات مثبت می‌گذارد (۳۶).

در نهایت مقایسه شاخص‌های صحت، حساسیت و تشخیص پذیری تکنیک‌های مختلف داده کاوی در پایگاه داده سلامت روان نوجوانان ایرانی گویای آن بود که تکنیک جنگل تصادفی با صحت ۹۱/۷۲ درصد، تشخیص پذیری ۸۲/۷۳ درصد بهترین عملکرد را جهت پیش‌بینی سلامت روان نوجوانان ایرانی بر اساس تغذیه، فعالیت فیزیکی و اوقات فراغت داشته‌است. اما لازم به ذکر است که تکنیک k نزدیک‌ترین همسایه با اختلاف ۰/۰۷ درصد از تکنیک جنگل تصادفی در پیش‌بینی سلامت روان مطلوب بهتر بوده است به گونه‌ای که ۹۶/۳۰ درصد از افراد با سلامت روان مطلوب را به درستی طبقه بندی کرده است. علاوه بر این، به دلیل بالاتر بودن میزان معیار حساسیت نسبت به تشخیص پذیری در تمامی تکنیک‌ها می‌توان اذعان داشت هر ۴ تکنیک در پیش‌بینی نوجوانان با سلامت روان مطلوب نسبت به افراد با سلامت روان نامطلوب توانا تر عمل کرده‌اند.

در پژوهشی همسو با این مطالعه Dooshima و همکارانش جهت پیش‌بینی خطر اختلالات روانی در نیجریه نشان دادند تکنیک درخت تصمیم با صحت ۸۳/۳ درصد نسبت به تکنیک بیزین با صحت ۷۶/۷ درصد بهتر عمل می‌کند (۳۷).

همچنین مطالعه Srividya و همکارانش گویای آن است که جهت پیش‌بینی سلامت روان تکنیک‌های جنگل تصادفی و مدل ترکیبی^۱ (شامل ماشین بردار پشتیبان، k نزدیک‌ترین همسایه، درخت تصمیم و رگرسیون لجستیک) بالاترین صحت را داشته‌اند اما بیزین و درخت تصمیم عملکرد خوبی را نداشته‌اند (۳۸).

Vaishnavi و همکارانش نیز در مطالعه‌ای غیر همسو با مطالعه حاضر از میان پنج تکنیک رگرسیون لجستیک، k نزدیک‌ترین همسایه، درخت تصمیم، جنگل تصادفی و پشته^۲

¹ Ensemble (Bagging).

² Stacking Technique.

بر اساس این تکنیک تغذیه بیشترین تأثیر را بر سلامت روان نوجوانان ایرانی دارد. لذا با توجه به نقش مؤثر سلامت روان نوجوانان در پیشرفت آن‌ها الزام است در جهت تأمین و بسترسازی مناسب آموزش به والدین و نوجوانان در رابطه با تغذیه مناسب و افزایش آگاهی در زمینه سلامت روان نوجوانان اندیشیده شود. علاوه بر این با استفاده بیشتر از تکنیک‌های داده کاوی در تحلیل پایگاه داده سلامت در داخل کشور می‌توان به نتایج ارزشمندی دست یافت که در جهت تشخیص و پیش‌بینی به‌هنگام بیماری‌ها به ویژه اختلالات روانی به عنوان سیستم تصمیم یار یاریگر کادر سلامت باشد.

تشکر و قدردانی

این مطالعه با کد اخلاق (IR.MUI.RESEARCH.REC.1399.277) در معاونت تحقیقات و فن آوری دانشگاه علوم پزشکی اصفهان تصویب شده است. نویسندگان بر خود لازم می‌دانند از مدیریت و کارشناسان پژوهشکده پیشگیری اولیه از بیماری‌های غیرواگیر دانشگاه علوم پزشکی اصفهان که در ارائه داده‌های اولیه همکاری نموده‌اند، تشکر و قدردانی نمایند.

جهت پیش‌بینی سلامت روان، تکنیک پشته با بالاترین صحت معادل ۸۱/۷۵ درصد را پیشنهاد دادند (۳۹). علاوه بر این Gonçalvesa و همکارانش در مطالعه‌ای همسو با این پژوهش بدین نتیجه دست یافتند که از میان چهار تکنیک درخت تصمیم، جنگل تصادفی، k نزدیک‌ترین همسایه و درختان تقویت شده گرادیان جهت پیش‌بینی سلامت روان دو تکنیک جنگل تصادفی و درخت تصمیم بیشترین صحت و k نزدیک‌ترین همسایه کمترین صحت را داشته است (۴۰).

نتیجه‌گیری

رشد و توسعه حجم داده‌های ثبت شده در پایگاه‌های داده به ویژه پایگاه‌های داده سلامت استخراج اطلاعات معنی دار و دانش نوین را به یک الزام تبدیل نموده که جهت رفع این نیاز مستلزم به استفاده از عناصر هوشمند مانند داده کاوی هستیم. لذا در این پژوهش بر آن شدیم تا برخی از تکنیک‌های داده کاوی را در حوزه سلامت روان مورد بررسی قرار دهیم. در نهایت یافته‌های حاصل مبین آن بود که از میان ۴ تکنیک پایه و پرکاربرد درخت تصمیم، بی‌زین، k نزدیک‌ترین همسایه و جنگل تصادفی بر اساس برآیند کلی صحت، حساسیت و تشخیص پذیری بهترین عملکرد را در پیش‌بینی سلامت روان بر اساس سه عامل تغذیه، فعالیت فیزیکی و اوقات فراغت می‌توان به تکنیک جنگل تصادفی اختصاص داد.

References

1. Yaychi NM, Hasanzade R, Farmarzi M, Homayoni A. Mediating role of health promoting behaviors on the relationship between self-efficacy and mental health in adolescent. *Journal of Education and Community Health* 2019;6(2):63-9.
2. Mostafavirad F, Dehghan Ardekani Z. Predicting Mental Health Based on the Personality Characteristics and Body image of High School Female Adults Students. *Quarterly Journal of Woman and Society* 2019; 10(37): 331-364.
3. Lajordi S, Nurbala Tafti A. Mental Health Indicators and Modalities to Improve Them. *Iranian Journal of Culture and Health Promotion* 2018; 2(2): 160-169.
4. Ghasemi F, Ebrahimi A, Samouei R. A Review of Mental Health Indicators in National Studies. *Journal of Isfahan Medical School* 2018; 36(470): 209-215.
5. Akbari M, Najafi S, Nadrian H. The Relationship Between Social Capital Perceived Social Support and The Women's Mental Health in Sanandaj: A Community-Based Study. *Iran J Health Educ Health Promote* 2017; 5(3):155-163.
6. Alizade Z, Rejali M, Feizi A, Afshar H, Hassanzade Kashtali A, Adibi P. Investigation of psychological disorders profile (anxiety, depression and psychological distress) in adult population of Isfahan province. *Journal of Torbat Heydariyeh University of Medical Sciences* 2016; 3(4):42-48

7. Merikangas KR, Nakamura EF, Kessler RC. Epidemiology of mental disorders in children and adolescents. *Dialogues in Clinical Neuroscience*. 2009; 11(1): 20-7.
8. Mousavi SI, Alizadeh Chaharborj T, Sheikh Mohammad R, Yavari Kermani M. Frequency of psychiatric disorder symptoms in diabetic Patients of Yasuj city in 2014. *Journal Of Neyshabur University Of Medical Sciences* 2016; 4(1):65-71.
9. Rahbari S, Riahi L, Tabibi J. Comparison of Mental Health Services in Iran and Selected Countries. *Journal of Health Promotion Management* 2018; 7 (3) :10-19.
10. Mehregan N, Rashid K, Ghasemifar S, Sohrabi Vafa H, Ghasemifar N. Estimating the mental health index in Iran provinces using fuzzy logic. *Rooyesh* 2019; 8(5) :187-192.
11. Tavousi M, Haeri Mehrizi AA, Hashemi A, Naghizadeh F, Montazeri A. Mental health in Iran: a nationwide cross-sectional study. *Payesh* 2016; 15 (3) :233-239.
12. Kaveh M, Shojaezadeh D, Shahmohammadi D, Eftekhari Ardebili H, Rahimi A, Bolhari J. Organizing mental health services in schools: assessing teachers' roles. *Journal of School of Public Health and Institute of Public Health Research* 2003; 1 (4) :1-14.
13. Karimi G, Vard B, Riyahi R, Motlagh ME, Heshmat R, Kelishadi R. Association between family dinner frequency and mental health in children and adolescents; the CASPIAN-V

- study. *International Journal of Food Sciences and Nutrition*. 2020; 71(5): 628-34.
- 14 .Baksheev GN, Robinson J, Cosgrave EM, Baker K, Yung AR. Validity of the 12-item General Health Questionnaire (GHQ-12) in detecting depressive and anxiety disorders among high school students. *Psychiatry Research*. 2011; 187(1): 291-6.
- 15 .Shakib Z, tahmasebi R, Noroozi A. Determinant factors of mental health based on social cognitive theory among high school girl students of Bushehr. *Iranian Journal of Health Education and Health Promotion* 2014; 2 (2) :131-142.
- 16 .Sadeghian E, Moghadari Kosha M, Gorji S. The study of mental health status in high school female students in Hamadan city. *Avicenna Journal of Clinical Medicine* 2010; 17 (3) :39-45.
- 17 .Child and adolescent mental health [Internet]. World Health Organization; Available from: https://www.who.int/mental_health/maternal-child/child_adolescent/en/ Accessed August 10, 2022.
- 18 .Biddle SJH, Asare M. Physical activity and mental health in children and adolescents: a review of reviews. *BJSM*. 2011; 45(11): 886-95.
- 19 .Sedigh Arfaee F, Hadianpour Z, & Mohseni Habibabadi Z. The Investigation of the Relationship between Sleep Quality, Physical Activity, and Dietary Pattern with the Mental Health of University of Kashan Students. *SOREN Student Sports & Health Open Researches E-Journal* 2021; 2(3):14–21.
- 20 .Oellingrath IM, Svendsen MV, Hestetun I. Eating patterns and mental health problems in early adolescence – a cross-sectional study of 12–13-year-old Norwegian schoolchildren. *PHN*. 2014; 17(11): 2554-62.
- 21 .O'neil A, Quirk SE, Housden S, Brennan SL, Williams LJ, Pasco JA, et al. Relationship between diet and mental health in children and adolescents: a systematic review. *American journal of public health*. 2014; 104(10): 31-42.
- 22 .ten Have M, de Graaf R, Monshouwer K. Physical exercise in adults and mental health status: Findings from the Netherlands Mental Health Survey and Incidence Study (NEMESIS). *J PSYCHOSOM RES*. 2011; 71(5): 342-8.
- 23 .Bell SL, Audrey S, Gunnell D, Cooper A, Campbell R. The relationship between physical activity, mental wellbeing and symptoms of mental health disorder in adolescents: a cohort study. *International Journal of Behavioral Nutrition and Physical Activity*. 2019; 16(1): 1-12.
- 24 .Hamer M, Stamatakis E, Mishra GD. Television- and Screen-Based Activity and Mental Well-Being in Adults. *AJPM*. 2010; 38(4): 375-80.
- 25 .Marateb, H.R., Tasdighi, Z., Mohebian, M.R. et al. Classification of psychiatric symptoms using deep interaction networks: the CASPIAN-IV study. *Scientific reports* 2021; 11: 15706-15721.
- 26 .Parsapour F, Peymani J. Using Data Mining Techniques for Intelligent Diagnosis of Severity of Depressive Disorder. *Journal of Health and Biomedical Informatics* 2020; 7 (3) :252-262.
- 27 .Samad Soltani T, Langarizadeh M, Zolnoori M. Data mining and analysis: reporting results for patients with asthma. *Payavard* 2015; 9 (3) :224-234.
- 28 .Gholamhosseini L, Damroodi M. Evaluation of data mining applications in the health system. *Paramedical Sciences and military health* 2015; 10 (1) :39-48.
- 29 .Qaderpour N. Data mining with data algorithms in health trends. *Journal of Science and engineering elites* 2016; 4 (2): 103-9.
- 30 .Alonso SG, de la Torre-Díez I, Hamrioui S, López-Coronado M, Barreno DC, Nozaleda LM, et al. Data Mining Algorithms and Techniques in Mental Health: A Systematic Review. *J MED SYST*. 2018; 42(9): 1-15.
- 31 .Viksvveen P, Bjønness SE, Berg SH, Cardenas NE, Game JR, Aase K, et al. User involvement in adolescents' mental healthcare: protocol for a systematic review. *BMJ*. 2017; 7(12): 1-7.
- 32 .Karimi R. Easy guide to statistical analysis with SPSS. Tehran: 2014 ,318 p (in persian).
- 33 .Ismaili M. Concepts and techniques of data mining. Niازه Danesh. 2017. 600p (in persian)
- 34 .Kleszczewska D, Mazur J, Siedlecka J. Family, school and neighborhood factors moderating the relationship between physical activity and some aspects of mental health in adolescents. *International Journal of Occupational Medicine and Environmental Health*. 2019; 32(4): 423-39.
- 35 .Ingram J, Maciejewski G, Hand CJ. Changes in Diet, Sleep, and Physical Activity Are Associated with Differences in Negative Mood During COVID-19 Lockdown. *Front Psychol*. 2020; 11: 588604 p
- 36 .Kulkarni AA, Swinburn BA, Utter J. Associations between diet quality and mental health in socially disadvantaged New Zealand adolescents. *EJCN*. 2015; 69(1): 79-83.
- 37 .Dooshima MP, Chidozie EN, Ademola BJ, Sekoni OO, Adebayo IP. A predictive model for the risk of mental illness in Nigeria using data mining. *IJI*. 2018; 6(1): 5-16.
- 38 .Srividya M, Mohanavalli S, Bhalaji N. Behavioral Modeling for Mental Health using Machine Learning Algorithms. *Journal of Medical Systems*. 2018; 42(5): 1-12.
- 39 .Vaishnavi K, Nikhitha Kamath U, Ashwath Rao B, Subba Reddy NV. Predicting Mental Health Illness using Machine Learning Algorithms. *AICECS; 2022/01/01; India: IOP Publishing; 2022. p. 012021.*
- 40 .Gonçalves C, Ferreira D, Neto C, Abelha A, Machado J. Prediction of Mental Illness Associated with Unemployment Using Data Mining. *Procedia Computer Sci*. 2020; 177: 556-61.

Tehran University of
Medical Sciences

Iranian Epidemiological Association

Original Article

Prediction of Mental Health Among Iranian Adolescents Based on Nutrition, Physical Activity and Leisure Time Using Data Mining Techniques

Asal Aghadavodian Jolfaee¹, Maryam Jahanbakhsh², Mohamad Sattari², Roya Kelishadi³

1- M.Sc., Dept. of Management and Health Information Technology, Isfahan University of Medical Sciences, Isfahan, Iran

2- Assistant Professor, Health Information Technology Research Center, Isfahan University of Medical Sciences, Isfahan, Iran

3- Professor, Child Growth and Development Research Center, Research Institute for Primordial Prevention of Non-Communicable Disease, Isfahan University of Medical Sciences, Isfahan, Iran

Article Information

Received

11 September 2023

Accepted

18 December 2023

Corresponding author

Maryam Jahanbakhsh

Corresponding author E-mail

jahanbakhsh.him@gmail.com

Keywords:

Mental health, Adolescent, Data mining, Nutrition, Physical activity, Leisure time

Abstract

Background and Objectives: The present research was conducted to predict mental health based on three factors: nutrition, activity, and leisure time, among students in the adolescent age group, using data mining techniques.

Methods: The present analytical study was conducted on 14274 data available in the Caspian 5 database. According to the CRISP-DM method, data mining was done in 6 steps using decision trees, k nearest neighbors, simple Bayesian and random forest techniques in Rapidminer software.

Results: Among the four data mining techniques used to predict the mental health of adolescents based on nutrition, physical activity and leisure time, the random forest technique has the highest accuracy (91.72) and specificity (82.73) and the k-nearest neighbors technique has the highest sensitivity (96.30). In addition, based on random forest techniques, the rule with the highest level of support showed that an adolescent who is in high school, eats breakfast, lunch, and dinner every day, drinks tea and coffee weekly, exercises 2 hours a week at school, also, he has 4 days of physical activity for 30 minutes in the last week, and he goes to school with the service, with 100% confidence has good mental health.

Conclusion: Based on the random forest technique, which has shown the best performance, nutrition has the greatest impact on the mental health of Iranian adolescents. So, it is necessary to think about providing a suitable platform for training parents and adolescents regarding proper nutrition and increasing awareness in the field of adolescent mental health.

Copyright © 2023 The Authors. Published by Tehran University of Medical Sciences.



This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International license (<https://creativecommons.org/licenses/by-nc/4.0/>). Non-commercial uses of the work are permitted, provided the original work is properly cited.