

دریافت مقاله: ۱۴۰۴/۰۴/۰۵

پذیرش مقاله: ۱۴۰۴/۰۶/۱۵

نوع مقاله: پژوهشی

شناسایی عوامل کلیدی در کیفیت خواب اشخاص با جنسیت، سنین و مشاغل مختلف با استفاده از یادگیری ماشین

علی نقاش اسدی*

استادیار، مهندسی کامپیوتر، دانشکده فنی فومن، دانشکدگان فنی دانشگاه تهران، فومن، ایران
پست الکترونیکی: naghahasadi@ut.ac.ir

رومینا عقیلی

کارشناسی، مهندسی کامپیوتر، دانشکده فنی فومن، دانشکدگان فنی دانشگاه تهران، فومن، ایران
پست الکترونیکی: romina.aghili@ut.ac.ir

امین مستوری نژاد

کارشناسی، مهندسی کامپیوتر، دانشکده فنی فومن، دانشکدگان فنی دانشگاه تهران، فومن، ایران
پست الکترونیکی: amin.mastouri@ut.ac.ir

چکیده

عامل در کیفیت خواب اشخاص در جنسیت، سنین، و مشاغل مختلف، متفاوت باشد. برای این منظور می‌توان از الگوریتم‌های یادگیری ماشین استفاده کرد. در این مقاله با استفاده از الگوریتم‌های Bagging و Boosting، مهم‌ترین عوامل موثر بر کیفیت خواب اشخاص با ویژگی‌های مختلف، پیش‌بینی می‌شود. نتایج نشان می‌دهند که مهم‌ترین عامل موثر بر کیفیت خواب اشخاص با جنسیت و سنین مختلف، مدت زمان خواب است؛ ولی مهم‌ترین عامل در تعیین کیفیت خواب اشخاص با مشاغل مختلف، سطح استرس است. مقادیر معیارهای ارزیابی به دست آمده از این الگوریتم‌ها، نشان‌دهنده دقت بالای نتایج به دست آمده است.

کلید واژه‌ها: یادگیری ماشین، کیفیت خواب، سبک زندگی، مدل‌های پیش‌بینی، عوامل کلیدی.

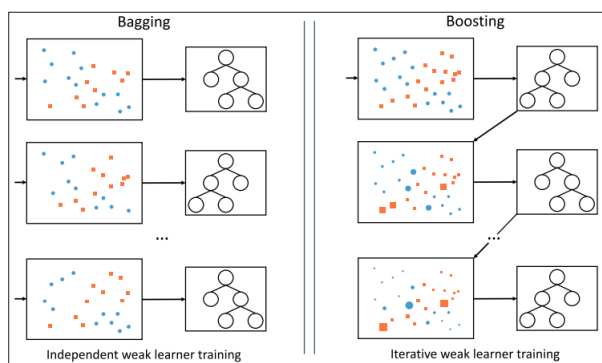
امروزه استفاده از هوش مصنوعی در حوزه‌های پزشکی و روان‌شناسی، بسیار گسترده شده است. در این حوزه، مقالات متعددی ارائه شده است، ولی اغلب آنها به موضوع کیفیت خواب و عوامل موثر بر آن نپرداخته‌اند. معدود مقالاتی که به این موضوع پرداخته‌اند، آن را به تفکیک اشخاص با ویژگی‌های مختلف بررسی نکرده‌اند. عوامل متعددی می‌توانند بر کیفیت خواب اشخاص تاثیرگذار باشند که برای مثال می‌توان به جنسیت، سن، شغل، سطح فعالیت بدنی، سطح استرس، نسبت وزن به قد، فشار خون، ضربان قلب و غیره اشاره کرد. بنابراین پیش‌بینی تاثیرگذارترین عامل بر کیفیت خواب اشخاص، موضوع بسیار مهمی است. همچنین ممکن است، مهم‌ترین

* نویسنده مسئول

۱. مقدمه

امروزه از داده‌ها به عنوان یک سرمایه یاد می‌شود. از داده‌ها می‌توان برای شناسایی روابط و الگوهای پنهان به منظور پیش‌بینی رویدادهای آینده استفاده کرد. یکی از مهم‌ترین حوزه‌هایی که می‌توان از داده‌های آن برای رفع چالش‌های آینده استفاده کرد، پزشکی و روان‌شناسی است. اشخاص مختلف، با مسائل پزشکی و روان‌شناسی متفاوتی مانند سرطان، افسردگی، آلزایمر، آسم، سنگ کلیه و غیره درگیر هستند. پژوهش‌های متعددی برای شناسایی عوامل موثر بر این مسائل و پیش‌بینی احتمال وقوع آنها در آینده برای هر شخص با ویژگی‌های مشخص ارائه شده است. برای این منظور معمولاً از فرآیند داده‌کاوی و الگوریتم‌های یادگیری ماشین استفاده می‌شود [۱]. در این فرآیند، ابتدا عوامل احتمالی موثر بر تشخیص یک بیماری شناسایی شده و سپس با استفاده از روش‌های مختلف آمارگیری، داده‌هایی جمع‌آوری می‌شود؛ سپس با استفاده از الگوریتم‌های یادگیری ماشین، مهم‌ترین عوامل موثر بر بیماری شناسایی شده و مدل‌های پیش‌بینی برای تشخیص آن ارائه می‌شود. یکی از موضوعات مهمی که در این پژوهش‌ها کمتر مورد بررسی قرار گرفته است، کیفیت خواب است. کیفیت خواب نشان‌دهنده میزان آرامش، پیوستگی، و کارایی فرآیند خواب است که بر سلامت فیزیکی و روانی افراد تأثیر می‌گذارد. معمولاً کیفیت خواب افراد بر اساس معیارهای مدت زمان و ساختار خواب، میزان بیداری‌های شبانه، عمق خواب، و توزیع مراحل مختلف خواب اندازه‌گیری می‌شود. این معیارها همواره تحت تأثیر عوامل فیزیولوژیکی، محیطی، روانی و سبک زندگی هستند. بنابراین می‌توان با شناسایی این عوامل، میزان تأثیر آنها را بر معیارهای کیفیت خواب ارزیابی کرده و به افراد به منظور بهبود کیفیت خوابشان کمک کرد. معدود مقالاتی که به این موضوع پرداخته‌اند، آن را برای اشخاص با ویژگی‌های مختلف بررسی نکرده‌اند. از نوآوری‌های این مقاله، بررسی کیفیت خواب اشخاص به تفکیک ویژگی‌های مختلف آنها است. به عبارت دیگر، در اکثر کارهای مرتبط، تعدادی عوامل موثر بر کیفیت خواب

شناسایی شده و آن عوامل برای همه اشخاص مورد بررسی قرار می‌گیرد. این در حالی است که در این مقاله، علاوه بر این کار، جامعه آماری بر اساس جنسیت، سن و شغل اشخاص تفکیک شده و برای هر گروه، مدل پیش‌بینی مجزا ارائه می‌شود تا مهم‌ترین عامل موثر بر کیفیت خواب در هر گروه شناسایی شود. عوامل متعددی مانند جنسیت، سن، شغل، سطح فعالیت بدنی، سطح استرس، نسبت وزن به قد، فشار خون، ضربان قلب و غیره می‌توانند بر کیفیت خواب اشخاص تأثیرگذار باشند؛ ولی کدام عامل بیشترین تأثیر را دارد؟ مهم‌ترین عامل موثر بر کیفیت خواب در اشخاص با جنسیت، سن، و مشاغل مختلف به چه صورت است؟ پاسخ به این پرسش‌ها نیازمند انجام فرایند داده‌کاوی و استفاده از الگوریتم‌های یادگیری ماشین است. برای این منظور در این مقاله از الگوریتم‌های Bagging و Boosting برای شناسایی مهم‌ترین عوامل موثر بر کیفیت خواب اشخاص با ویژگی‌های مختلف استفاده شده است. این الگوریتم‌ها از تجمیع نتایج چندین مدل یادگیری، برای ارائه یک مدل جامع و با دقت استفاده می‌کنند. این الگوریتم‌ها به این دلیل انتخاب شدند که در مقالات مرتبط، اثبات شده است که این الگوریتم‌ها نسبت به سایر الگوریتم‌های سنتی، دقت بالاتری دارند. در انتها با مشاهده مقادیر معیارهای ارزیابی، دقت نتایج به دست آمده از این الگوریتم‌ها اثبات می‌شود. در ادامه این مقاله و در بخش ۲، مفاهیم اصلی استفاده شده در تحقیق شامل الگوریتم‌های Bagging و Boosting و معیارهای ارزیابی، معرفی می‌شوند. سپس در بخش ۳، مقالات مرتبط با موضوع تحقیق معرفی شده و با کارهای انجام شده در این مقاله مقایسه می‌شوند. در بخش ۴، از الگوریتم‌های Bagging و Boosting به منظور شناسایی عوامل کلیدی در تعیین کیفیت خواب اشخاص با جنسیت، سن، و مشاغل مختلف استفاده می‌شود. همچنین در این بخش، نتایج معیارهای ارزیابی به دست آمده از الگوریتم‌ها، ارائه می‌شود. در انتها نیز، نتیجه‌گیری و کارهای آینده ارائه می‌شود.



شکل ۱: الگوریتم‌های Bagging و Boosting

۲-۲- معیارهای ارزیابی

برای ارزیابی نتایج به دست آمده از الگوریتم‌های یادگیری ماشین، معیارهای مختلفی وجود دارد. در این مقاله از معیارهای ضریب همبستگی^۵ (CC) با رابطه (۱)، میانگین خطای مطلق^۶ (MAE) با رابطه (۲)، خطای جذر میانگین مربعات^۷ (RMSE) با رابطه (۳)، درصد خطای مطلق نسبی^۸ (RAE) با رابطه (۴)، و درصد خطای نسبی جذر مربعات^۹ (RRSE) با رابطه (۵) برای مقایسه مقادیر به دست آمده از الگوریتم (y_i) و مقادیر واقعی (x_i) استفاده می‌شود [۷ و ۸]. هر چه مقادیر معیار ضریب همبستگی به مقدار یک نزدیک‌تر باشد، و هر چه مقادیر معیارهای میانگین خطای مطلق، خطای جذر میانگین مربعات، درصد خطای مطلق نسبی، و درصد خطای نسبی جذر مربعات به مقدار صفر نزدیک‌تر باشند، به معنی فاصله کم بین نتایج به دست آمده از الگوریتم و مقادیر واقعی است.

$$CC = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2}} \quad (1)$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |x_i - y_i| \quad (2)$$

- 5-Correlation Coefficient
- 6-Mean Absolute Error
- 7-Root Mean Squared Error
- 8-Relative Absolute Error
- 9-Root Relative Squared Error

۲. مفاهیم اصلی

۱-۲- الگوریتم‌های Bagging و Boosting

به منظور تحلیل و استخراج هوشمند اطلاعات از داده‌ها، الگوریتم‌های مختلفی وجود دارد که می‌توان آنها را به پیش‌بینی‌کننده^۱ و توصیف‌کننده^۲ طبقه‌بندی کرد [۲ و ۳]. در الگوریتم‌های پیش‌بینی‌کننده، هدف اصلی پیش‌بینی نتایج یا رویدادهای آینده با استفاده از داده‌های گذشته و بررسی الگوهای موجود است. از این نوع الگوریتم‌ها می‌توان به BaggingRegressor و XGBoost اشاره کرد [۴-۶]. در این الگوریتم‌ها، از تجمیع نتایج چندین مدل یادگیری، برای ارائه یک مدل جامع و دقیق استفاده می‌شود. به عبارت دیگر در این الگوریتم‌ها، از یک نوع الگوریتم بهینه‌سازی برای ایجاد یک مدل بهینه از تجمیع چندین زیرمدل ضعیف استفاده می‌شود. روش کار این الگوریتم‌ها به این صورت است که ابتدا چندین زیرمدل ساده که ممکن است خیلی دقیق نباشند ایجاد می‌کنند و سپس از روی آنها (Boosting) و یا با تجمیع آنها (Bagging) مدل دقیق ارائه می‌کنند. در این الگوریتم‌ها از هر نمونه در مجموعه داده به طور مؤثرتری استفاده می‌شود و بنابراین حتی اگر مجموعه داده خیلی بزرگ نباشد، مدل‌ها به طور قابل اعتمادی آموزش می‌بینند. مطابق با شکل ۱، در BaggingRegressor (مدل‌سازی با رأی‌گیری)، ابتدا زیرمجموعه‌های تصادفی از مجموعه داده اصلی ایجاد شده و سپس یک رگرسیون پایه برای هر زیرمجموعه ایجاد می‌شود. در انتها نتایج رگرسیون‌های پایه (با رأی‌گیری یا میانگین‌گیری) تجمیع می‌شوند تا یک پیش‌بینی نهایی و با دقت تشکیل شود. از سوی دیگر مطابق با شکل ۱، XGBoost، یک روش پیشرفته و بهینه‌شده از الگوریتم‌های تقویت گرادیان^۳ (GB) است. این روش به طور قابل‌توجهی دقت مدل را با استفاده از مجموعه‌ای از درخت‌های تصمیم^۴ کوچک که به‌طور متوالی برای کاهش خطاهای مدل‌های قبلی آموزش داده می‌شوند، بهبود می‌بخشد.

- 1-Predictive
- 2-Descriptive
- 3-Gradient Boosting
- 4-Decision Tree

تحقیق از الگوریتم‌های یادگیری ماشین استفاده نشده است. در مقاله [۱۰]، رابطه کیفیت خواب شامل کیفیت ذهنی خواب، دوره نهان خواب، طول دوره خواب، کارآمدی عادات خواب، مختل‌کننده‌های خواب، استفاده از داروهای خواب‌آور و سوء کارکرد روزانه فرد با سلامت روانی و فرسودگی شغلی مورد بررسی قرار گرفت. در این تحقیق اطلاعات کارکنان آتش‌نشانی شهر اهواز با استفاده از روش نمونه‌گیری تصادفی از طریق پرسش‌نامه دریافت شد. در این مقاله نیز از الگوریتم‌های یادگیری ماشین استفاده نشده است و صرفاً تحلیل داده آماری انجام شده است. در مقاله [۱۱ و ۱۲]، ارتباط استفاده بیش از حد از فضای مجازی و تلفن همراه با کیفیت خواب در بین دانشجویان مورد بررسی قرار گرفت. در این مقالات نیز از الگوریتم‌های یادگیری ماشین استفاده نشده است.

در مقاله [۱۳]، ارتباط عوامل مختلف بر کیفیت خواب مورد بررسی قرار گرفت و نتایج آن نشان داد که علاوه بر دمای پوست بالاتر و مدت خواب کافی، با میانگین ضربان قلب بالاتر در هنگام بیداری و فعالیت کمتر سیستم عصبی پاراسمپاتیک و سمپاتیک در هنگام خواب، احتمال کیفیت خواب بالاتر خواهد شد. در مقاله [۱۴]، از روش‌های یادگیری ماشین، به ویژه ماشین بردار پشتیبان^{۱۰} (SVM)، برای پیش‌بینی کیفیت خواب از طریق ضربان قلب و دمای پوست استفاده شده است. در این مقاله [۱۵]، با استفاده از روش‌های یادگیری ماشین مثل الگوریتم‌های تقویت گرادیان، بیز ساده^{۱۱}، ماشین بردار پشتیبان و درخت تصمیم دلایل اختلالات شدید خواب در نوجوانان چینی مورد بررسی قرار گرفت و نتایج نشان داد، از میان تمامی روش‌ها، مدل تقویت گرادیان بهترین عملکرد را دارد. در مقاله [۱۶]، از داده‌های الکتروانسفالوگرافی^{۱۲} جمع‌آوری شده از اشخاص مبتلا به اختلالات خواب استفاده می‌کند و یک تحلیل مقایسه‌ای از اثربخشی الگوریتم‌های یادگیری ماشین مثل ماشین‌های بردار پشتیبان، جنگل

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (x_i - y_i)^2}{n}} \quad (۳)$$

$$RAE = \frac{\sqrt{\sum_{i=1}^n (x_i - y_i)^2}}{\sqrt{\sum_{i=1}^n (x_i)^2}} \times 100 \quad (۴)$$

$$RRSE = \sqrt{\frac{\sum_{i=1}^n (x_i - y_i)^2}{\sum_{i=1}^n (x_i - \bar{x})^2}} \times 100 \quad (۵)$$

۳- کارهای مرتبط

در مقالات بسیاری از یادگیری ماشین در موضوعات پزشکی و روان‌شناسی استفاده شده است. در بیشتر این مقالات، ابتدا عواملی که به وسیله آنها احتمالاً می‌توان بیماری یک شخص را تشخیص داد، شناسایی شده و سپس با استفاده از روش‌های مختلف آمارگیری، داده‌هایی از اشخاص سالم و بیمار جمع‌آوری می‌شود؛ سپس با استفاده از الگوریتم‌های یادگیری ماشین، مهم‌ترین عوامل موثر بر بیماری شناسایی شده و مدل‌های پیش‌بینی برای تشخیص آن بیماری ارائه می‌شود. این فرایند در مقالات معدودی برای بررسی کیفیت خواب اشخاص انجام شده است. در جدول ۱، مقایسه‌ای بین کارهای پیشین و کار انجام شده در این مقاله ارائه شده است. در اکثر کارهای پیشین، یا عوامل بررسی شده برای پیش‌بینی کیفیت خواب متفاوت بوده است، و یا از الگوریتم‌های یادگیری ماشین متفاوتی استفاده شده است. با این حال در هیچ یک از کارهای پیشین، کیفیت خواب برای اشخاص با جنسیت، سنین، و مشاغل مختلف مورد بررسی قرار نگرفته است. برای مثال، در مقاله [۹]، وضعیت کیفیت خواب، خواب آلودگی و کیفیت زندگی دانشجویان و تعیین رابطه بین این عوامل با عوامل شخصیتی مورد بررسی قرار گرفت. در این تحقیق صرفاً اطلاعات دانشجویان دانشگاه علوم پزشکی مازندران جمع‌آوری شده و تحلیل داده آماری روی این داده‌ها ارائه شده است. به عبارت دیگر، در این

10-Support Vector Machines
11-Naïve Bayesian
12-Electroencephalogram

مقاله [۲۰]، از الگوریتم‌های XGBoost و AdaBoost برای پیش‌بینی بیماری‌های بی‌خوابی و آپنه خواب استفاده شد که مدل‌های پیش‌بینی ارائه شده توسط آنها توانستند، به ترتیب با دقت ۹۳/۴۹ و ۹۲/۷۳ درصد این بیماری‌ها را تشخیص دهند.

در مقاله حاضر، از الگوریتم‌های Bagging و Boosting برای شناسایی مهم‌ترین عوامل موثر بر کیفیت خواب اشخاص با ویژگی‌های مختلف استفاده می‌شود. استفاده از این الگوریتم‌ها به این دلیل است که در مقالات مرتبط، اثبات شده است که این الگوریتم‌ها نسبت به سایر الگوریتم‌های سنتی، دقت بالاتری دارند. دلیل این موضوع این است که در این الگوریتم‌ها، با تجمع چندین زیرمدل ضعیف، یک مدل بهینه ایجاد می‌شود. بنابراین یک نوع فرآیند بهینه‌سازی در این نوع الگوریتم‌ها وجود دارد.

۴- استفاده از یادگیری ماشین در شناسایی عوامل کلیدی در کیفیت خواب اشخاص

در این قسمت، عوامل کلیدی موثر بر کیفیت خواب اشخاص به تفکیک جنسیت، سنین، و مشاغل مختلف شناسایی می‌شود. برای این منظور با استفاده کتابخانه‌های موجود در زبان برنامه‌نویسی پایتون در ابزار colab، از الگوریتم‌های XGBoost و BaggingRegressor استفاده می‌شود. همان‌طور که در بخش کارهای مرتبط مشاهده شد، الگوریتم‌های مبتنی بر یادگیری گروهی^{۱۳} دقت بالاتری نسبت به سایر الگوریتم‌های یادگیری ماشین سنتی دارند و به همین دلیل، در این مقاله نیز از این الگوریتم‌ها برای شناسایی عوامل کلیدی موثر بر کیفیت خواب استفاده شده است. برای تنظیم پارامترهای کلیدی در این الگوریتم‌ها، از یک رویکرد جستجوی سیستماتیک بر اساس تغییر در مقادیر Test Size و Random State استفاده شده است. در این فرایند، داده‌ها با ترکیب‌های مختلف از این دو پارامتر، بارها به مجموعه‌های آموزش و آزمون تقسیم شده و عملکرد مدل‌ها با معیار ضریب همبستگی (CC) ارزیابی

تصادفی^{۱۴}، و K نزدیکترین همسایه^{۱۵} (KNN) و الگوریتم‌های یادگیری عمیق مثل شبکه‌های عصبی هم‌آمیختگی^{۱۶} (CNN) و شبکه‌های عصبی بازگشتی^{۱۷} (RNN) در تشخیص زودهنگام اختلالات خواب با استفاده از سیگنال‌های مغزی ارائه می‌دهد. نتایج نشان دادند که در شناسایی اختلالات خواب، مدل‌های یادگیری عمیق (CNN و RNN) نسبت به الگوریتم‌های یادگیری ماشین سنتی بهتر عمل می‌کنند. در مقاله [۱۷]، با استفاده از داده‌های جمع‌آوری شده از وضعیت سلامت خواب و سبک زندگی اشخاص، مدل‌های پیش‌بینی با استفاده از روش‌های مختلف یادگیری ماشین شامل ماشین تقویت گرادیان، AdaBoost، جنگل تصادفی، K نزدیکترین همسایه، رگرسیون لجستیک^{۱۸} (LR)، و بیز ساده گوسی^{۱۹} (GNB) برای پیش‌بینی کیفیت خواب اشخاص ارائه شد. نتایج نشان داد که مدل به دست آمده از GB به طور موثرتری نسبت به سایر روش‌ها، کیفیت خواب اشخاص را پیش‌بینی می‌کند. در مقاله [۱۸]، از الگوریتم بهینه‌سازی ژنتیک برای طبقه‌بندی اختلالات خواب با استفاده از الگوریتم‌های یادگیری عمیق مثل K نزدیکترین همسایه، ماشین بردار پشتیبان، درخت تصمیم، جنگل تصادفی، و شبکه‌های عصبی استفاده شده است. برای این منظور از مجموعه داده‌های سلامت خواب و سبک زندگی برای آموزش و آزمون مدل‌های پیشنهادی استفاده شد که نتایج نشان داد، شبکه‌های عصبی نسبت به سایر الگوریتم‌های آزمایش‌شده، دقت بالاتری دارد. در مقاله [۱۹]، اثربخشی مدل‌های مختلف یادگیری ماشین شامل رگرسیون لجستیک، درخت تصمیم، جنگل تصادفی، XGBoost، ماشین‌های بردار پشتیبان، و شبکه‌های عصبی، در شناسایی اختلالات خواب بر اساس داده‌های به دست آمده از شیوه زندگی و سلامت اشخاص، بررسی شده است. نتایج این تحقیق نشان داد که از نظر دقت، روش‌های جنگل تصادفی و XGBoost، بهتر از سایر مدل‌ها عمل کردند. در

13-Random Forest
14-K-Nearest Neighbors
15-Convolutional Neural Network
16-Recurrent neural network
17-Logistic regression
18-Gaussian Naive Bayes

جدول ۱. مقایسه بین کارهای مرتبط

مرجع	عوامل بررسی شده	تفکیک اشخاص	الگوریتم‌های استفاده شده
[۹]	کیفیت خواب، خواب آلودگی، کیفیت زندگی دانشجویان و عوامل شخصیتی	خیر	صرفاً تحلیل داده آماری
[۱۰]	کیفیت خواب، سلامت روانی و فرسودگی شغلی	خیر	صرفاً تحلیل داده آماری
[۱۱]	فضای مجازی و کیفیت خواب	خیر	صرفاً تحلیل داده آماری
[۱۲]	تلفن همراه و کیفیت خواب	خیر	صرفاً تحلیل داده آماری
[۱۳]	دمای پوست، مدت خواب، میانگین ضربان قلب، فعالیت سیستم عصبی	خیر	GB
[۱۴]	ضربان قلب و دمای پوست	خیر	SVM
[۱۵]	ویژگی‌های اساسی (سن، جنسیت، و وضعیت تاهل)، سبک زندگی (تعداد سیگار در روز، دفعات نوشیدن در هفته، هزینه ماهانه، و ترجیحات غذایی)، عادت ورزشی (زمان بی‌حرکی و دفعات ورزش در هفته)، بیماری مزمن، و سلامت روانی (اضطراب، افسردگی، اختلالات روانی، و غیره)	خیر	GB, NB, SVM, DT
[۱۶]	داده‌های الکتروانسفالوگرافی	خیر	SVM, RF, KNN, CNN, RNN
[۱۷]	وضعیت سلامت خواب و سبک زندگی اشخاص	خیر	GB, AdaBoost, RF, KNN, LR, GNB
[۱۸]	وضعیت سلامت خواب و سبک زندگی اشخاص	خیر	KNN, SVM, DT, RF, ANN
[۱۹]	وضعیت سلامت خواب و سبک زندگی اشخاص	خیر	LR, DT, RF, XGBoost, SVM, ANN
[۲۰]	وضعیت سلامت خواب و سبک زندگی اشخاص	خیر	XGBoost, AdaBoost
مقاله ما	وضعیت سلامت خواب و سبک زندگی اشخاص	بله	XGBoost, BaggingRegressor

شدند. بهترین مقادیر انتخابی، آنهایی هستند که بالاترین دقت را در پیش‌بینی کیفیت خواب نشان دهند. برای کاهش احتمال بیش‌برازش ۲۰، نتایج در چندین تکرار مستقل آزمایش شدند تا اطمینان حاصل شود، مدل‌ها بیش از حد به الگوهای خاص داده‌ها وابسته نشده‌اند. این اقدامات باعث شد تا مدل‌های پیش‌بینی علاوه بر دقت مناسب، از قابلیت تعمیم‌پذیری قابل قبولی نیز برخوردار باشند. در ادامه این بخش، ابتدا مجموعه داده مورد بررسی معرفی می‌شود و سپس با استفاده از الگوریتم‌های XGBoost و BaggingRegressor برای هر گروه از اشخاص، مدل‌سازی مجزایی انجام شده و نتایج به دست آمده از آنها برای هر گروه از اشخاص به صورت مجزا ارائه می‌شود.

۱-۴- معرفی مجموعه داده

مجموعه داده سلامت خواب و سبک زندگی شامل ۳۷۴ نمونه است که ۱۳ ویژگی مرتبط با کیفیت خواب و عادات روزانه را پوشش می‌دهد [۲۱]. مطابق با جدول ۲، این ویژگی‌ها شامل جنسیت، سن، شغل، مدت زمان

خواب، کیفیت خواب، سطح فعالیت بدنی، سطح استرس، شاخص توده بدنی^{۲۱} (BMI)، ضربان قلب، تعداد قدم‌های روزانه، مشکلات خواب و شاخص‌های مربوط به فشار خون (فشار خون سیستولیک و دیاستولیک) است. مقادیر بعضی از ویژگی‌ها مانند شاخص توده بدنی، ضربان قلب، فشار خون و غیره، با استفاده از ابزارهای اندازه‌گیری از اشخاص گرفته شده است و مقادیر سایر ویژگی‌ها، از طریق پرسش‌نامه جمع‌آوری شده است. از آنجایی که معمولاً الگوریتم‌های یادگیری ماشین با مقادیر عددی کار می‌کنند، مقادیر برخی از ویژگی‌ها (جنسیت، شاخص توده بدنی و اختلال خواب) که به صورت کیفی بودند، به مقادیر عددی نگاشت شدند.

پس از پاک‌سازی داده‌ها از مقادیر تهی و تکراری و کمی کردن مقادیر داده‌های کیفی، با استفاده از الگوریتم CFS^{۲۲} بررسی شد که هر زیرمجموعه از ویژگی‌ها به طور جداگانه به چه میزان می‌توانند ویژگی‌های دیگر را پیش‌بینی کنند. همچنین درجه افزونگی بین آنها ارزیابی شد. طبق بررسی انجام شده مشخص شد، با وجود آن

21-Body Mass Index

22-Correlation-based Feature Selection

20-Overfitting

جدول ۲. معرفی ویژگی‌های مجموعه داده سلامت خواب و سبک زندگی

نام ویژگی	معرفی ویژگی و مقادیر آن
شناسه شخص	یک شناسه برای هر شخص
جنسیت	● آقا (صفر) ● خانم (یک)
سن	سن شخص بر اساس سال در بازه ۲۷ تا ۵۹ سال
شغل	شغل یا حرفه شخص شامل حسابدار، دکتر، مهندس، وکیل، مدیر، پرستار، نماینده فروش، فروشنده، دانشمند، مهندس نرم‌افزار و معلم
مدت زمان خواب	تعداد ساعاتی که شخص در شبانه روز می‌خوابد
کیفیت خواب	رتبه‌بندی از کیفیت خواب از ۱ تا ۱۰
سطح فعالیت بدنی	تعداد دقیقی که شخص به صورت روزانه به فعالیت بدنی می‌پردازد
سطح استرس	رتبه‌بندی از سطح استرس تجربه‌شده توسط شخص از ۱ تا ۱۰
شاخص توده بدنی	یک معیار استاندارد برای ارزیابی وضعیت وزنی اشخاص است که بر اساس نسبت وزن به مجذور قد (به متر) محاسبه می‌شود. شاخص توده بدنی اشخاص در این مجموعه داده شامل مقادیر زیر است: ● طبیعی (یک): اشخاصی که وزن آنها نسبت به قدشان در محدوده سالم قرار دارد (BMI بین ۱۸/۵ تا ۲۴/۹). ● اضافه وزن (دو): اشخاصی که وزن آنها بیشتر از محدوده سالم است ولی به حد چاقی نرسیده (BMI بین ۲۴/۹ تا ۲۹/۹). ● چاق (سه): اشخاصی که وزن آنها به حدی از محدوده سالم فراتر رفته است که به عنوان چاق طبقه‌بندی می‌شوند (BMI بیشتر از ۳۰).
فشار خون سیستولیک	فشار خون در زمان انقباض عضله قلب و پمپاژ خون به شریان‌ها که این مقدار معمولاً به عنوان عدد بالایی در اندازه‌گیری فشار خون گزارش می‌شود.
فشار خون دیاستولیک	فشار خون در زمان استراحت عضله قلب، بین دو ضربان قلب که این مقدار به عنوان عدد پایینی در اندازه‌گیری فشار خون نشان داده می‌شود.
ضربان قلب	تعداد ضربان قلب شخص در دقیقه در حالت استراحت
تعداد قدم‌های روزانه	تعداد قدم‌هایی که شخص در روز بر می‌دارد
اختلال خواب	وجود/عدم وجود اختلال خواب در شخص که در این مجموعه داده شامل مقادیر زیر است: ● هیچکدام (صفر): شخص هیچ اختلالی در خواب ندارد. ● آپنه خواب (یک): شخص از وقفه تنفسی در طول خواب رنج می‌برد که منجر به اختلال در الگوهای خواب و خطرات بالقوه سلامتی می‌شود. ● بی‌خوابی (دو): شخص در به خواب رفتن یا ماندن در خواب با مشکل مواجه است که منجر به خواب ناکافی یا بی‌کیفیت می‌شود.

۲-۴- عوامل کلیدی در کیفیت خواب اشخاص با جنسیت مختلف

در مجموعه داده سلامت خواب و سبک زندگی، اطلاعات ۳۷۴ شخص جمع‌آوری شده است که ۱۸۵ نفر از آنها خانم، و ۱۸۹ نفر دیگر آقا هستند. در جدول ۳، حداقل، حداکثر و میانگین مقادیر ویژگی‌های اشخاص به تفکیک جنسیت آنها ارائه شده است. همان‌طور که مشاهده می‌شود، مقادیر حداقل و حداکثر ویژگی‌ها در آقایان و خانم‌ها تقریباً مشابه یکدیگر است ولی به طور میانگین:

- خانم‌ها: مدت زمان خواب، کیفیت خواب، شاخص توده بدنی، تعداد قدم‌های روزانه، اختلال خواب و فشار خون بیشتری داشتند.

که اهمیت بعضی از ویژگی‌ها از سایرین بیشتر است، ولی تمامی ویژگی‌ها برای پیش‌بینی ویژگی‌های دیگر دارای اهمیت هستند و بنابراین نمی‌توان هیچ یک از ویژگی‌ها را از مجموعه داده حذف کرد. علاوه بر این، مقادیر هر ویژگی نیز مورد بررسی قرار گرفت تا توازن بین مقادیر مختلف آن وجود داشته باشد. برای مثال، از ۳۷۴ نمونه جمع‌آوری شده، ۱۸۹ نفر آقا و ۱۸۵ نفر خانم هستند و ۱۲۶ نفر از آنها بین ۲۷ تا ۳۷ سال سن دارند، ۱۴۱ نفر از آنها بین ۳۸ تا ۴۸ سال سن دارند، و ۱۰۷ نفر از آنها بین ۴۹ تا ۵۹ سال سن دارند. این توازن برای سایر ویژگی‌ها نیز رعایت شده است.

جدول ۳. حداقل، حداکثر و میانگین مقادیر ویژگی‌های اشخاص با جنسیت مختلف

ویژگی‌ها	آقایان			خانم‌ها		
	حداقل	میانگین	حداکثر	حداقل	میانگین	حداکثر
مدت زمان خواب	۵/۹	۷/۰۴	۸/۱	۵/۸	۷/۲۳	۸/۵
کیفیت خواب	۴	۶/۹۷	۹	۴	۷/۶۶	۹
سطح فعالیت بدنی	۳۰	۵۹/۲	۹۰	۳۰	۵۹/۱۴	۹۰
سطح استرس	۳	۶/۰۸	۸	۳	۴/۶۷	۸
شاخص توده بدنی	۱ (طبیعی)	۱/۳۲ (نزدیک به طبیعی)	۳ (چاق)	۱ (طبیعی)	۱/۵۸ (بین طبیعی و اضافه وزن)	۳ (چاق)
ضربان قلب	۶۵	۷۱/۰۵	۸۶	۶۵	۶۹/۲۶	۸۴
تعداد قدم‌های روزانه	۳۰۰۰	۶۷۹۳/۶	۱۰۰۰۰	۳۳۰۰	۶۸۴۰/۵	۱۰۰۰۰
اختلال خواب	۰ (بدون اختلال)	۰/۴۹ (بین بدون اختلال و آپنه خواب)	۲ (بی‌خوابی)	۰ (بدون اختلال)	۰/۷۵ (نزدیک به آپنه خواب)	۲ (بی‌خوابی)
فشار خون سیستولیک	۱۱۷	۱۲۶/۹۴	۱۴۲	۱۱۵	۱۳۰/۲۰	۱۴۰
فشار خون دیاستولیک	۷۶	۸۳/۰۱	۹۲	۷۵	۸۶/۳۲	۹۵

جدول ۴. عامل کلیدی در کیفیت خواب اشخاص با جنسیت مختلف

شغل	میانگین کیفیت خواب (رتبه)	عامل کلیدی در تعیین کیفیت خواب بر اساس الگوریتم‌ها				
		XGBoost				
		RRSE	RAE	RMSE	MAE	CC
آقایان	۶/۹۷ (۳)	مدت زمان خواب				
		۱/۰	۸/۲۲۴۱	۰/۰۰۰۱	۰/۰۰۰۱	۰/۰۰۰۱
خانم‌ها	۷/۶۶ (۱)	مدت زمان خواب				
		۱/۰	۶/۶۴۸۱	۹/۷۷۱۶	۵/۸۵۳۵	۷/۶۹۵۶
همه جنسیت‌ها	۷/۳۱ (۲)	مدت زمان خواب				
		۰/۹۹۹۹	۰/۰۰۰۲	۰/۰۰۰۴	۰/۰۰۰۲	۰/۰۰۰۳

● آقایان: سطح فعالیت بدنی، سطح استرس، و ضربان قلب بیشتری داشتند.

۳-۴- عوامل کلیدی در کیفیت خواب اشخاص با سنین مختلف

در مجموعه داده سلامت خواب و سبک زندگی، اطلاعات ۳۷۴ شخص جمع‌آوری شده است که ۱۲۶ نفر از آنها بین ۲۷ تا ۳۷ سال سن دارند، ۱۴۱ نفر از آنها بین ۳۸ تا ۴۸ سال سن دارند، و ۱۰۷ نفر از آنها بین ۴۹ تا ۵۹ سال سن دارند. در جدول ۵، حداقل، حداکثر و میانگین مقادیر ویژگی‌های اشخاص به تفکیک سن آنها ارائه شده است. همان‌طور که مشاهده می‌شود، مقادیر حداقل و حداکثر ویژگی‌ها در سنین مختلف تقریباً مشابه یکدیگر است ولی به طور میانگین:

همان‌طور که در جدول ۴ مشاهده می‌شود، کیفیت خواب خانم‌ها نسبت به آقایان بیشتر است. بر اساس الگوریتم‌های XGBoost و BaggingRegressor، عامل کلیدی در تعیین کیفیت خواب اشخاص با جنسیت‌های مختلف، مدت زمان خواب است. مطابق با معیارهای ارزیابی، هر دو الگوریتم XGBoost و BaggingRegressor توانستند با دقت بسیار بالایی این نکته را تشخیص دهند. با این حال، الگوریتم BaggingRegressor به صورت جزئی، دقت بالاتری داشته است.

جدول ۵. حداقل، حداکثر و میانگین مقادیر ویژگی‌های اشخاص با سنین مختلف

ویژگی‌ها	۲۷ الی ۳۷ سال			۳۸ الی ۴۸ سال			۴۹ الی ۵۹ سال		
	حداقل	میانگین	حداکثر	حداقل	میانگین	حداکثر	حداقل	میانگین	حداکثر
مدت زمان خواب	۵/۸	۶/۹۴	۷/۹	۵/۹	۷/۰۲	۸	۶	۷/۵۱	۸/۵
کیفیت خواب	۴	۶/۸۲	۸	۵	۷/۲۳	۹	۶	۸	۹
سطح فعالیت بدنی	۳۰	۵۳/۲	۷۵	۳۰	۶۰/۴۷	۹۰	۳۰	۶۴/۴۸	۹۰
سطح استرس	۳	۶	۸	۳	۵/۳۳	۸	۳	۴/۷۲	۸
شاخص توده بدنی	۱ (طبیعی)	۱/۱۶ (نزدیک به طبیعی)	۳ (چاق)	۱ (طبیعی)	۱/۵۲ (بین طبیعی و اضافه وزن)	۳ (چاق)	۱ (طبیعی)	۱/۶۹ (بین طبیعی و اضافه وزن)	۳ (چاق)
ضربان قلب	۶۵	۷۱/۰۵	۸۵	۶۵	۶۹/۷	۸۴	۶۵	۶۹/۷۵	۸۶
تعداد قدم‌های روزانه	۳۰۰۰	۶۵۱۹	۱۰۰۰۰	۳۳۰۰	۶۸۲۰/۵	۱۰۰۰۰	۳۷۰۰	۷۱۶۲/۶	۱۰۰۰۰
اختلال خواب	۰ (بدون اختلال)	۰/۲ (نزدیک به بدون اختلال)	۲ (بی‌خوابی)	۰ (بدون اختلال)	۰/۹۳ (نزدیک به آپنه خواب)	۲ (بی‌خوابی)	۰ (بدون اختلال)	۰/۷۱ (نزدیک به آپنه خواب)	۲ (بی‌خوابی)
فشار خون سیستولیک	۱۱۵	۱۲۲/۳۲	۱۴۰	۱۱۵	۱۲۹/۷	۱۴۲	۱۱۸	۱۳۴/۳۸	۱۴۰
فشار خون دیاستولیک	۷۵	۸۰/۰۴	۹۰	۷۵	۸۵/۱۸	۹۵	۷۵	۸۹/۳۸	۹۵

جدول ۶. عامل کلیدی در کیفیت خواب اشخاص با سنین مختلف

شغل	میانگین کیفیت خواب (رتبه)	عامل کلیدی در تعیین کیفیت خواب بر اساس الگوریتم‌ها									
		XGBoost					BaggingRegressor				
		RRSE	RAE	RMSE	MAE	CC	RRSE	RAE	RMSE	MAE	CC
۲۷ الی ۳۷ سال	۶/۸۲ (۴)	مدت زمان خواب					مدت زمان خواب				
		۰/۹۹۹۹	۰/۰۰۰۸	۰/۰۰۲۴	۰/۰۰۱۴	۰/۰۰۳۸	۰/۹۲۱۳	۰/۱۶۱۵	۰/۳۵۸۴	۰/۱۴۵۲	۰/۲۸۰۴
۳۸ الی ۴۸ سال	۷/۲۳ (۳)	مدت زمان خواب					مدت زمان خواب				
		۱/۰	۰/۰۰۰۱	۰/۰۰۰۱	۰/۰۰۰۱	۰/۰۰۰۱	۰/۹۵۵۲	۰/۰۹۳۳	۰/۱۵۰۵	۰/۱۵۲۱	۰/۲۱۱۵
۴۹ الی ۵۹ سال	۸ (۱)	مدت زمان خواب					مدت زمان خواب				
		۱/۰	۰/۰۰۰۱	۰/۰۰۰۱	۰/۰۰۰۱	۰/۰۰۰۱	۰/۹۹۹۵	۰/۰۰۹	۰/۰۳۰۱	۰/۰۰۷۱	۰/۰۲۲۳
همه سنین	۷/۳۱ (۲)	مدت زمان خواب					مدت زمان خواب				
		۰/۹۹۹۹	۰/۰۰۰۲	۰/۰۰۰۴	۰/۰۰۰۲	۰/۰۰۰۳	۱/۰	۰/۰	۰/۰	۰/۰	۰/۰

چقدر سن اشخاص افزایش پیدا می‌کند، کیفیت خواب آنها نیز بهتر می‌شود. بر اساس الگوریتم‌های XGBoost و BaggingRegressor، عامل کلیدی در تعیین کیفیت خواب اشخاص با سنین مختلف، مدت زمان خواب است. مطابق با معیارهای ارزیابی، هر دو الگوریتم XGBoost و BaggingRegressor توانستند با دقت بسیار بالایی این نکته را تشخیص دهند. با این حال، الگوریتم XGBoost به صورت جزئی، دقت بالاتری داشته است.

● اشخاص در سنین ۲۷ الی ۳۷: سطح استرس و ضربان قلب بیشتری داشتند.

● اشخاص در سنین ۳۸ الی ۴۸: اختلال خواب بیشتری داشتند.

● اشخاص در سنین ۴۹ الی ۵۹: مدت زمان خواب، کیفیت خواب، سطح فعالیت بدنی، شاخص توده بدنی، تعداد قدم‌های روزانه، و فشار خون بیشتری داشتند.

همان‌طور که در جدول ۶ مشاهده می‌شود، هر

جدول ۷. حداقل، حداکثر و میانگین مقادیر ویژگی‌های اشخاص با مشاغل مختلف

ویژگی‌ها	حسابدار			دکتر			مهندس			وکیل			پرستار			معلم		
	حداقل	میانگین	حداکثر	حداقل	میانگین	حداکثر	حداقل	میانگین	حداکثر	حداقل	میانگین	حداکثر	حداقل	میانگین	حداکثر	حداقل	میانگین	حداکثر
مدت زمان خواب	۶/۵	۷/۱۱	۸	۶	۶/۹۷	۸/۲	۶/۵	۷/۹۹	۸/۵	۶/۹	۷/۴۱	۷/۹	۵/۹	۷/۰۶	۸/۲	۶/۳	۶/۶۹	۷/۲
کیفیت خواب	۷	۷/۸۹	۹	۶	۶/۶۵	۹	۵	۸/۴۱	۹	۷	۷/۸۹	۸	۵	۷/۳۷	۹	۵	۶/۹۷	۸
سطح فعالیت بدنی	۴۵	۵۸/۱۱	۸۰	۳۰	۵۵/۳۵	۹۰	۳۰	۵۱/۸۶	۹۰	۵۰	۷۰/۴۲	۹۰	۳۵	۷۸/۵۹	۹۰	۳۰	۶۲۴۵	۶۰
سطح استرس	۳	۴/۵۹	۷	۳	۶/۷۳	۸	۳	۳/۸۸	۷	۵	۵/۰۶	۶	۳	۵/۵۵	۸	۴	۴/۵۲	۷
شاخص توده بدنی	۱	۱/۱۶	۲	۱	۱/۱۱	۳	۱	۱/۰۴	۲	۱	۱/۱۳	۳	۱	۱/۹۰	۲	۱	۱/۸۷	۳
ضربان قلب	۶۷	۶۸/۸۶	۷۳	۶۵	۷۱/۵۲	۸۶	۶۵	۶۷/۱۹	۸۰	۶۸	۶۹/۶۴	۸۴	۶۸	۷۲	۸۰	۶۵	۶۷/۲۲	۸۲
تعداد قدم‌های روزانه	۶۰۰۰	۶۸۸۱/۰۸	۷۵۰۰	۳۵۰۰	۶۸۰۸/۴	۱۰۰۰۰	۴۰۰۰	۵۹۸۰/۹	۸۰۰۰	۳۳۰۰	۷۶۶۱/۷	۸۰۰۰	۴۰۰۰	۸۰۵۷/۵	۱۰۰۰۰	۳۵۰۰	۵۹۵۷/۵	۷۰۰۰
اختلال خواب	۰	۰/۳۸	۲	۰	۰/۱۴	۲	۰	۰/۱۷	۲	۰	۰/۱۵	۲	۰	۰/۹۲	۲	۰	۱/۴۵	۲
فشار خون سیستولیک	۱۱۵	۱۱۷/۷۳	۱۳۰	۱۱۸	۱۲۳	۱۴۲	۱۱۷	۱۲۵/۹۰	۱۳۲	۱۲۵	۱۲۹/۹۶	۱۳۵	۱۱۷	۱۳۸/۵	۱۴۰	۱۱۵	۱۳۱/۲۲	۱۴۰
فشار خون دیاستولیک	۷۵	۷۶/۹۲	۸۵	۷۵	۸۰/۵	۹۲	۷۶	۸۱/۳۸	۸۷	۸۲	۸۵	۸۸	۷۶	۹۳/۷۳	۹۵	۷۵	۸۶/۹	۹۰

۴-۴- عوامل کلیدی در کیفیت خواب اشخاص با مشاغل مختلف

در مجموعه داده سلامت خواب و سبک زندگی، ۱۳ ویژگی مرتبط با کیفیت خواب و عادات روزانه مورد بررسی قرار گرفته است [۲۱]. از ویژگی‌های مورد بررسی، شغل اشخاص است. در این مجموعه داده، اشخاص با مشاغل حسابدار (۳۷ نفر)، دکتر (۷۱ نفر)، مهندس (۶۳ نفر)، وکیل (۴۷ نفر)، مدیر (۱ نفر)، پرستار (۷۳ نفر)، نماینده فروش (۲ نفر)، فروشنده (۳۲ نفر)، دانشمند (۴ نفر)، مهندس نرم‌افزار (۴ نفر) و معلم (۴۰ نفر) وجود داشتند. از آنجایی که تعداد اشخاصی که مشاغل مدیر، نماینده فروش، دانشمند، و مهندس نرم‌افزار داشتند، کمتر از ۱۰ نفر بود و همچنین شغل فروشنده، به دلیل این که مقادیر ویژگی‌های این اشخاص، تا حد قابل توجهی مشابه یکدیگر بود؛ از مجموعه داده حذف شدند؛ چون امکان ارائه تحلیل درست برای آنها فراهم نبود. در جدول ۷، حداقل، حداکثر و میانگین مقادیر ویژگی‌های اشخاص به تفکیک مشاغل آنها

ارائه شده است. همان طور که مشاهده می‌شود، مقادیر حداقل و حداکثر ویژگی‌ها در مشاغل مختلف تقریباً مشابه

یکدیگر است ولی به طور میانگین:

- **دکترها:** سطح استرس بیشتری داشتند.
 - **مهندس‌ها:** مدت زمان خواب و کیفیت خواب بیشتری داشتند.
 - **پرستارها:** سطح فعالیت بدنی، شاخص توده بدنی، ضربان قلب، تعداد قدم‌های روزانه، و فشار خون بیشتری داشتند.
 - **معلم‌ها:** اختلال خواب بیشتری داشتند.
- همان طور که در جدول ۸ مشاهده می‌شود، مهندس‌ها بیشترین کیفیت خواب را دارند و وکیل‌ها در رتبه بعدی هستند. همچنین دکترها، کمترین کیفیت خواب را در میان همه مشاغل دارا هستند. عامل کلیدی در تعیین کیفیت خواب اشخاص بر اساس الگوریتم‌های XGBoost و Bagging Regressor، برای حسابدارها، سن و سطح استرس، برای دکترها، مدت زمان خواب و سطح استرس، برای مهندس‌ها،

جدول ۸. عامل کلیدی در کیفیت خواب اشخاص با مشاغل مختلف

عامل کلیدی در تعیین کیفیت خواب بر اساس الگوریتم‌ها										میانگین کیفیت خواب (رتبه)	شغل
BaggingRegressor					XGBoost						
RRSE	RAE	RMSE	MAE	CC	RRSE	RAE	RMSE	MAE	CC		
سطح استرس					سن					۷/۸۹ (۲)	حسابدار
۰/۰	۰/۰	۰/۰	۰/۰	۱/۰	۰/۰۰۰۳	۰/۰۰۰۲	۰/۰۰۰۱	۵/۳۵۸۴	۰/۹۹۹۹		
سطح استرس					مدت زمان خواب					۶/۶۵ (۶)	دکتر
۰/۰	۰/۰	۰/۰	۰/۰	۱/۰	۰/۰۰۰۲	۰/۰۰۰۱	۰/۰۰۰۲	۰/۰۰۰۱	۰/۹۹۹۹		
سن					جنسیت					۸/۴۱ (۱)	مهندس
۰/۰	۰/۰	۰/۰	۰/۰	۱/۰	۰/۰۰۰۱	۰/۰۰۰۱	۷/۳۵۲۴	۶/۵۱۱۲	۱/۰		
تعداد قدم‌های روزانه					شاخص توده بدنی					۷/۸۹ (۲)	وکیل
۰/۰	۰/۰	۰/۰	۰/۰	۱/۰	۰/۰۰۰۵	۰/۰۰۰۵	۰/۰۰۰۱	۰/۰۰۰۱	۰/۹۹۹۹		
سطح استرس					مدت زمان خواب					۷/۳۷ (۳)	پرستار
۰/۰	۰/۰	۰/۰	۰/۰	۱/۰	۵/۵۰۲۰	۵/۰۸۶۲	۸/۲۵۳۰	۷/۶۲۹۳	۱/۰		
ضربان قلب					سطح استرس					۶/۹۷ (۵)	معلم
۰/۰	۰/۰	۰/۰	۰/۰	۱/۰	۰/۰۰۰۲	۰/۰۰۰۱	۰/۰۰۰۱	۶/۵۴۴۵	۰/۹۹۹۹		
مدت زمان خواب					مدت زمان خواب					۷/۳۱ (۴)	همه مشاغل
۰/۰	۰/۰	۰/۰	۰/۰	۱/۰	۰/۰۰۰۳	۰/۰۰۰۲	۰/۰۰۰۴	۰/۰۰۰۲	۰/۹۹۹۹		

منظور شناسایی مهم‌ترین عوامل موثر بر کیفیت خواب اشخاص به تفکیک ویژگی‌های مختلف استفاده شد. برای این منظور از الگوریتم‌های XGBoost و BaggingRegressor استفاده شد که معیارهای ارزیابی به دست آمده از این الگوریتم‌ها نشان داد که این دو الگوریتم با دقت بالایی می‌توانند این عوامل را شناسایی کنند. نتایج نشان دادند که مهم‌ترین عامل موثر بر کیفیت خواب اشخاص با جنسیت و سنین مختلف، مدت زمان خواب است؛ ولی مهم‌ترین عامل در تعیین کیفیت خواب اشخاص با مشاغل مختلف، سطح استرس آنها است.

آخرین نتایج تحقیقاتی به‌دست آمده از سلامت روان و کیفیت خواب اشخاص بیان می‌کند که عوامل موثر بر کیفیت خواب افراد بسیار متنوع هستند. به عبارت دیگر، طیف وسیعی از عوامل شامل وراثت، موقعیت اجتماعی، موقعیت مکانی، تجربه‌ها و عادات، سبک زندگی، وضعیت جسمانی و غیره می‌توانند بر سلامت روان یک شخص تاثیرگذار باشند. در این مقاله، با محدود کردن مسئله، صرفاً ۱۳ ویژگی مرتبط با کیفیت خواب و عادات روزانه اشخاص مورد بررسی قرار گرفت. به عنوان کار آینده

جنسیت و سن، برای وکیل‌ها، شاخص توده بدنی و تعداد قدم‌های روزانه، برای پرستارها، مدت زمان خواب و سطح استرس، و برای معلم‌ها، سطح استرس و ضربان قلب است. همان‌طور که مشاهده می‌شود، سطح استرس، عامل کلیدی پرتکرار در تعیین کیفیت خواب اشخاص با مشاغل مختلف است. مطابق با معیارهای ارزیابی، هر دو الگوریتم XGBoost و Bagging Regressor توانستند با دقت بسیار بالایی این نکته را تشخیص دهند. با این حال، الگوریتم Bagging Regressor به صورت جزئی، دقت بالاتری داشته است.

۵- نتیجه‌گیری

در این مقاله از مجموعه داده سلامت خواب و سبک زندگی که ۱۳ ویژگی مرتبط با کیفیت خواب و عادات روزانه شامل جنسیت، سن، شغل، مدت زمان خواب، کیفیت خواب، سطح فعالیت بدنی، سطح استرس، شاخص توده بدنی، ضربان قلب، تعداد قدم‌های روزانه، مشکلات خواب و شاخص‌های مربوط به فشار خون (فشار خون سیستولیک و دیاستولیک) است را پوشش می‌دهد، به

Cyberspace and Sleep Quality among Students of University of Medical Sciences. *Pajouhan Scientific Journal*, 18(3), 24-33.

[12] Majidaei, M., Pireinaladin, S., & Kasaei, A. (2015). The Role of Cell-Phones over Use in Anticipation of Sleep Quality, Anxiety and Depression among University Students. *Journal of Education and Community Health Year*, 2(2), 38-46.

[13] Moebus, M., & Holz, C. (2024). Personalized interpretable prediction of perceived sleep quality: Models with meaningful cardiovascular and behavioral features. *PLOS One*, 19(7), 1-25.

[14] Credico, A. D., Perpetuini, D., Izzicupo, P., Gaggi, G., Mammarella, N., Domenico, A. D., Palumbo, R., Malva, P. L., Cardone, D., Merla, A., Ghinassi, B., & Baldassarre, A. D. (2024). Predicting Sleep Quality through Biofeedback: A Machine Learning Approach Using Heart Rate Variability and Skin Temperature. *Clocks & Sleep*, 6(3), 322-337.

[15] Zhang, L., Zhao, S., Yang, W., Yang, Z., Wu, Z. a., Zheng, H., & Lei, M. (2024). Utilizing machine learning techniques to identify severe sleep disturbances in Chinese adolescents: an analysis of lifestyle, physical activity, and psychological factors. *Front Psychiatry*, 15, 1447281.

[16] Satapathy, S. K., Patel, V., Gandhi, M., & Mohapatra, R. K. (14-16 March, 2024, pp. 1-6). Comparative Study of Brain Signals for Early Detection of Sleep Disorder Using Machine and Deep Learning Algorithm. *IEEE International Conference on Interdisciplinary Approaches in Technology and Management for Social Innovation (IATMSI)*, Gwalior, India.

[17] Kumar, M., Ahmed, B., Mishra, H. M., & Jha, A. K. (May 2024, pp. 1-7). Predictive Sleep Disorder Modelling: Using Machine Learning with Lifestyle and Sleep Health Data. *International Conference on Advances in Computing, Communication and Applied Informatics (ACCAI)*,

[18] Alshammari, T. S. (2024). Applying Machine Learning Algorithms for the Classification of Sleep Disorders. *IEEE Access*, 12, 36110-36121.

[19] Airlangga, G. (2024). Evaluating Machine Learning Models for Predicting Sleep Disorders in a Lifestyle and Health Data Context. *JIKO*, 7(1), 51-57.

[20] Malik, N., Gupta, V., & Sharma, V. (2024, pp. 1-4). Predictive Analysis of Sleep Disorders Using Machine Learning: A Comprehensive Analysis. *International Conference on Trends in Quantum Computing and Emerging Business Technologies (TQCEBT)*,

[21] Tharmalingam, L.). Sleep Health and Lifestyle Dataset. Kaggle. <https://www.kaggle.com/datasets/uom190346a/sleep-health-and-lifestyle-dataset>

می‌توان عوامل بیشتری را مورد بررسی قرار داد و با ارائه یک پرسش‌نامه و جمع‌آوری داده از طیف گسترده‌تری از اشخاص، تحلیل‌های دقیق‌تری با در نظر گرفتن عواملی مانند سلامت جسمانی، عادات غذایی، محیط خواب، میزان استفاده از فناوری و غیره، ارائه کرد.

مراجع

[1] Schuh, G., Reinhart, G., Prote, J.-P., Sauermann, F., Horsthofer, J., Oppolzer, F., & Knoll, D. (2019). Data mining definitions and applications for the management of production complexity. *Procedia Cirp*, 81, 874-879.

[2] Akulwar, P., Pardeshi, S., & Kamble, A. (30-31 Aug. 2018, pp. 513-519). Survey on Different Data Mining Techniques for Prediction. *2nd International Conference on I-SMAC*, Palladam, India.

[3] Papakyriakou, D., & Barbounakis, I. (2022). Data Mining Methods: A Review. *International Journal of Computer Applications*, 183(49), 5-19.

[4] Yaman, M. A., Rattay, F., & Subasi, A. (2021). Comparison of Bagging and Boosting Ensemble Machine Learning Methods for Face Recognition. *Procedia Computer Science*, 194, 202-209.

[5] Deng, X., Ye, A., Zhong, J., Xu, D., Yang, W., Song, Z., Zhang, Z., Guo, J., Wang, T., Tian, Y., Pan, H., Zhang, Z., Wang, H., Wu, C., Shao, J., & Chen, X. (2022). Bagging-XGBoost algorithm based extreme weather identification and short-term load forecasting model. *Energy Reports*, 8, 8661-8674.

[6] González, S., García, S., Ser, J. D., Rokach, L., & Herrera, F. (2020). A practical tutorial on bagging and boosting based ensembles for machine learning: Algorithms, software tools, performance study, practical perspectives and opportunities. *Information Fusion*, 64, 205-237.

[7] Naghash Asadi, A., Mirzaei, Z., & Nikookar, S. (2023). Providing predictive models of financial indicators using regression and neural networks (in Persian). *Computing Science Journal*, 8(3), 26-38.

[8] Farooq, F., Ahmed, W., Akbar, A., Aslam, F., & Alyousef, R. (2021). Predictive modeling for sustainable high-performance concrete from industrial wastes: A comparison and optimization of models using ensemble learners. *Journal of Cleaner Production*, 292, 126032.

[9] Baharloo, S., Moosazadeh, M., & Setareh, J. (2021). Relationship between Personality Characteristics and Sleep Quality, Daily Sleepiness and Quality of Life in University Students. *Journal of Mazandaran University of Medical Sciences*, 31(198), 144-158.

[10] Karimi, R., Abedini, R., & Arshadi, N. (2019). The relationship of sleep quality with mental health and job burnout. *Contemporary Psychology*, 13(2), 121-129.

[11] Barati, M., Rashidi Darab, S., Bakhshian, P., Afshari, M., & Doosti Irani, A. (2020). Study Relationship the Overuse of