

باسمه تعالی



دانشگاه صنعتی اصفهان

دانشکده برق و کامپیوتر

## گزارش میانی پروژه درس داده کاوی

موضوع پروژه:

پیش بینی اختلال خواب افراد بر اساس سبک زندگی آنها

اعضای تیم:

فاطمه جان نثاری (۴۰۱۱۴۳۰۳)

حوری دهش (۹۸۲۱۴۱۳)

اردیبهشت ۱۴۰۳

موضوع پروژه ما پیش‌بینی اختلال خواب افراد بر اساس سبک‌زندگی است و کارهایی که تاکنون انجام دادیم شامل موارد زیر است:

- (۱) ابتدا دیتاست را لود کردیم و از آن داده‌ها را خواندیم.
- (۲) به اطلاعات آماری دیتاست پرداختیم مثلاً اینکه شامل چند سطر و ستون است و نیز نوع ویژگی‌ها، تعداد داده‌های یکتا یا مواردی مثل میانه، چارک‌ها و... را به دست آوردیم.
- (۳) سپس به طور مفصل به تجزیه و تحلیل داده‌ها پرداختیم به این صورت که نموداری برای پراکندگی داده‌ها رسم کردیم، تعداد افرادی که هر نوع از اختلال خواب دارند را روی نمودار هیستوگرام نمایش دادیم، و بعد رابطه هر یک از ویژگی‌ها را با اختلال خواب نمایش دادیم. مثلاً رابطه جنسیت، شغل، کیفیت خواب، فعالیت‌های فیزیکی، سن، مدت زمان خواب، فشار خون و BMI، سطح استرس با اختلال خواب (برای نمایش هر یک از این رابطه‌ها نمودار مناسب را رسم کردیم).

چالشی که در این بخش با آن روبه‌رو شدیم این بود که درون دیتاست ویژگی اختلال خواب شامل یکسری فیلدهای **nan** بود و نمودارهایی که می‌کشیدیم به صورت **none** بودند به همین خاطر نمودارها غلط نشان داده می‌شد در نهایت با تبدیل **nan** به **none** مشکل حل شد.

- (۴) در گام بعد روی داده‌ها پیش‌پردازش انجام دادیم. به این صورت که چک کردیم داده **nan** نداشته باشیم، برای فشار خون به ازای نرمال ۰ و به ازای غیر نرمال ۱ قرار دادیم، آیدی‌ها را حذف کردیم، مقادیر عددی بعضی ستون‌ها را به بازه‌های مشخص تبدیل کردیم، داده‌های غیر عددی (مثلاً رشته) را به داده عددی تبدیل کردیم.

- (۵) در مرحله بعدی، تقسیم‌بندی داده‌ها را انجام دادیم و بخش آموزشی و تست را مشخص کردیم.

- (۶) سپس به سراغ ایجاد کردن مدل روی داده‌ها رفتیم. برای مدل‌سازی ما قصد داریم ۶ مدل بسازیم و آنها را **tune** کنیم. مدل‌های درخت تصمیم، جنگل تصادفی و **gradient boosting classifier** را ساختیم و با استفاده از **grid search** و **bayes search** آنها را **tune** کردیم.

چالش کنونی ما این است که دقت مدل قبل از **tuning** بیشتر از بعد از **tuning** است و گمان می‌کنم به دنبال علت این مشکل و رفع آن هستیم.

کارهای باقیمانده شامل ساخت بقیه مدل‌ها و **tune** کردن، سپس تست کردن مدل‌ها و بررسی دقت آنها و در نهایت انتخاب بهترین مدل و استفاده از آن است.