



دانشگاه صنعتی شریف
دانشکده مهندسی صنایع

پیشنهاد پروژه‌ی کلاسی

درس مدل‌سازی و تصمیم‌گیری داده‌محور

استاد درس

سرکار خانم دکتر نفیسه صدقی

اعضای گروه

اشکان فرهودی زارع ۴۰۱۲۱۱۹۱۷

محمدحسین محمودی ۴۰۱۲۱۱۹۷۴





فهرست مطالب

پیشنهاد اول ۳

معرفی مقاله ۳

معرفی مجموعه داده ۳

پیشنهاد دوم ۵

معرفی مقاله ۵

معرفی مجموعه داده ۵



پیشنهاد اول

معرفی مقاله

فینخال و همکاران در پژوهشی تحت عنوان *پیش‌بینی احتمال ابتلا به دیابت در مراحل اولیه با استفاده از تکنیک‌های داده کاوی* [۱] به مسئله‌ی تشخیص زودهنگام بیماری دیابت پرداختند. با توجه به رشد سریع معضل‌های سلامت بدون علامت، آن‌ها به پیش‌بینی احتمال توسعه‌ی بیماری دیابت در افراد با استفاده از متدهای دسته‌بندی پرداختند. برای حل مسئله، تیم فینخال از یک مجموعه داده‌ی پرسشنامه‌ی بیمارستان سیلپت^۱ بنگلادش شامل ۵۲۰ رکورد و ۱۶ ویژگی استفاده کردند. در طول تحلیل و پیش‌بینی، تیم محققین از متدهای مختلف شامل دسته‌بندی کننده‌ی بیز ساده^۲، رگرسیون لجستیک^۳ و جنگل تصادفی^۴ استفاده کردند. علاوه بر این، برای صحت‌سنجی عملکرد مدل‌ها از اعتباردهی متقابل^۵ و تقسیم‌بندی درصدی^۶ استفاده نمودند. در نهایت هم پژوهشگران یک ابزار کامل برای افراد حقیقی جهت بررسی ریسک بیماری دیابت در بین آن‌ها ارائه کردند.

معرفی مجموعه داده

مجموعه داده‌ی مورد استفاده در این پژوهش شامل گزارش‌هایی از ۵۲۰ نفر است که علائم مرتبط با دیابت، از جمله علائم دیابت بالقوه را نشان می‌دهند. این مجموعه داده از بیمارستان دیابت سیلپت در شهر سیلپت بنگلادش، از طریق پرسشنامه‌های مستقیمی که هم برای افراد مبتلا به دیابت تازه تشخیص داده شده و هم برای افراد غیر دیابتی که علائم مرتبط را تجربه می‌کنند، جمع‌آوری شده است. پیش‌پردازش داده‌ها با مدیریت مقادیر نامعلوم از طریق حذف ورودی‌های ناقص انجام می‌شود که منجر به ایجاد مجموعه داده‌ی نهایی با اندازه‌ی ۵۰۰ می‌شود که ۳۱۴ مورد آن مثبت (نشان‌دهنده‌ی وجود خطر ابتلا به دیابت) و ۱۸۶ مورد منفی (نشان‌دهنده‌ی عدم وجود خطر ابتلا به دیابت) هستند. متغیرهای این مجموعه داده و مقادیری که اختیار می‌کنند در جدول صفحه‌ی بعد ارائه شده است.

- Sylhet¹
- Naïve Bayes²
- Logistic Regression³
- Random Forest⁴
- Cross Validation⁵
- Percentage Split⁶



ویژگی	مقادیر اتخاذ شده
سن	۲۰-۳۵
	۳۶-۴۵
	۴۶-۵۵
	۵۵-۵۶
	بیشتر از ۶۵
جنسیت	مرد، زن
پرادراری ^۷	بله، خیر
عطش بیش از حد ^۸	بله، خیر
کاهش وزن ناگهانی	بله، خیر
ضعف	بله، خیر
پراشتهایی ^۹	بله، خیر
برفک تناسلی ^{۱۰}	بله، خیر
تاری دید	بله، خیر
خارش	بله، خیر
بی‌قراری	بله، خیر
بهبودی همراه با تاخیر	بله، خیر
کاهش توانایی تحرک در ناحیه‌ی پا ^{۱۱}	بله، خیر
سفتی عضلاتی	بله، خیر
طاسی منطقه‌ای ^{۱۲}	بله، خیر
چاقی	بله، خیر
دسته	مثبت، منفی

⁷ Polyuria
⁸ Polydispia
⁹ Polyphagia
¹⁰ Genital Thrush
¹¹ Partial Paresis
¹² Alopecia



پیشنهاد دوم

معرفی مقاله

ونگیپورام و همکاران در پژوهشی تحت عنوان پیش‌بینی خطر دیابت جوانان با استفاده از داده‌های انهنس و یادگیری ماشین [۲] با استفاده از یک مجموعه داده‌ی بزرگ از نظرسنجی ملی سلامت و تغذیه^{۱۳} شامل ۲۹۷۰ رکورد از نوجوانان ۱۲ تا ۱۹ سال به ارزیابی عملکرد یک دستورالعمل غربالگری بالینی کودکان در شناسایی بیماری پیش‌دیابت و دیابت شیرین^{۱۴} مابین جوانان مبتلا به این بیماری با استفاده از نشانگرهای زیستی تشخیصی انجمن دیابت آمریکا^{۱۵} پرداختند. آن‌ها در این پژوهش با استفاده از شاخص‌های ارزیابی استاندارد، دستورالعمل‌ها را با استفاده از طبقه‌بندی‌کننده‌های یادگیری ماشین مورد ارزیابی قرار دادند. با توجه به تفاوت عملکرد الگوریتم‌ها در بین زیرگروه‌های مختلف جمعیتی، یافته‌های آنان مهر تاییدی بر ناکافی بودن دستورالعمل غربالگری بالینی کودکان است.

معرفی مجموعه داده

مجموعه داده‌ی مورد استفاده در این پژوهش داده‌های نظرسنجی ملی سلامت و تغذیه است که یک نظرسنجی جامع مقطعی شامل اطلاعات مربوط به سلامت از طریق مصاحبه، معاینات پزشکی و آزمایشگاهی می‌باشد. این مجموعه داده شامل اطلاعات مربوط به ۲۹۷۰ نوجوان ۱۲ تا ۱۹ ساله است که سال‌های ۲۰۰۵ تا ۲۰۱۶ را پوشش می‌دهد و آن‌ها به طور خاص افرادی را انتخاب کردند که نشانگرهای زیستی تشخیص مربوط به پیش‌دیابت و دیابت شیرین برای آن‌ها در دسترس باشد.

متغیرهای این مجموعه داده و مقادیری که اختیار می‌کنند در جدول صفحه‌ی بعد ارائه شده است.

National Health and Nutritional Examination Survey (NHANES)¹³

PreDM/DM¹⁴

American Diabetes Association Diagnostic Biomarkers¹⁵



ویژگی	مقادیر اتخاذ شده
جنسیت	مرد، زن
نژاد/قومیت	سفید غیر اسپانیایی، سیاه غیر اسپانیایی، اسپانیایی، دیگر
بازه‌ی سنی	۱۲-۱۴ ۱۵-۱۷ ۱۸-۱۹
دسته‌بندی صدک شاخص توده‌ی بدنی ^{۱۶}	کمتر از ۸۵ ۸۵-۹۵ ۹۵-۹۹ بیشتر از ۹۹
در معرض خطر ابتلا به پیش‌دیابت و دیابت شیرین بر اساس دستورالعمل‌های غربالگری بالینی کودکان ^{۱۷}	بله، خیر

BMI Percentile Categories¹⁶

At risk for PreDM/DM based on AAP/ADA pediatric clinical screening guidelines¹⁷



منابع و مراجع

- [1] M. M. F. Islam, R. Ferdousi, S. Rahman, and H. Y. Bushra, "Likelihood Prediction of Diabetes at Early Stage Using Data Mining Techniques," *Advances in Intelligent Systems and Computing*, vol. 992, pp. 113–125, 2020, doi: 10.1007/978-981-13-8798-2_12/COVER.
- [2] N. Vangeepuram, B. Liu, P. hsiang Chiu, L. Wang, and G. Pandey, "Predicting youth diabetes risk using NHANES data and machine learning," *Scientific Reports* 2021 11:1, vol. 11, no. 1, pp. 1–9, May 2021, doi: 10.1038/s41598-021-90406-0.