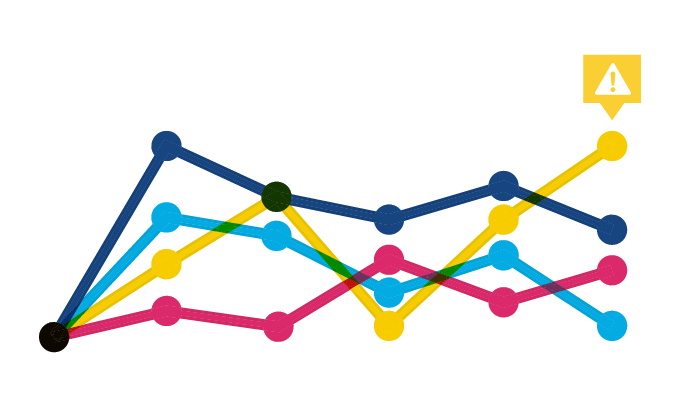
****

**دانشکده‌ی مهندسی صنایع**

**دانشگاه صنعتی شریف**

**پیشنهاد پروژه‌ی کلاسی**

**درس مدل‌سازی و تصمیم‌گیری داده‌محور**



**استاد درس**

سرکار خانم دکتر نفیسه صدقی

**اعضای گروه**

اشکان فرهودی زارع ۴۰۱۲۱۱۹۱۷

محمدحسین محمودی ۴۰۱۲۱۱۹۷۴

**آبان ۱۴۰۲**

# فهرست مطالب

**پیشنهاد اول.........................................................................................................................۳۱**

معرفی مقاله.........................................................................................................................................................۳

معرفی مجموعه‌داده.............................................................................................................................................۳

**پیشنهاد دوم........................................................................................................................۵۱**

معرفی مقاله.........................................................................................................................................................۵

معرفی مجموعه‌داده.............................................................................................................................................۵

# پیشنهاد اول

## معرفی مقاله

فنیخال و همکاران در پژوهشی تحت عنوان *پیش‌بینی احتمال ابتلا به دیابت در مراحل اولیه با استفاده از تکنیک‌های داده کاوی[1]* 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 مجموعه‌داده‌ی پرسشنامه‌ی بیمارستان سیلهت[[1]](#footnote-1) بنگلادش شامل ۵۲۰ رکورد و ۱۶ ویژگی استفاده کردند. در طول تحلیل و پیش‌بینی، تیم محققین از متد‌های مختلف شامل دسته‌بندی‌کننده‌ی بیز ساده[[2]](#footnote-2)، رگرسیون لجستیک[[3]](#footnote-3) و جنگل تصادفی[[4]](#footnote-4) استقاده کردند. علاوه‌بر این، برای صحت‌سنجی عملکرد مدل‌ها از اعتباردهی متقابل[[5]](#footnote-5) و تقسیم‌بندی درصدی[[6]](#footnote-6) استفاده نمودند. در نهایت هم پژوهشگران یک ابزار کامل برای افراد حقیقی جهت بررسی ریسک بیماری دیابت در بین آن‌ها ارائه کردند.

## معرفی مجموعه‌داده

مجموعه‌داده‌ی 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 مجموعه‌داده از بیمارستان دیابت سیلهت در شهر سیلهت 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 شده‌است. پیش‌پردازش داده‌ها با مدیریت مقادیر نامعلوم از طریق حذف ورودی‌های ناقص انجام می‌شود که منجر به ایجاد مجموعه‌داده‌ی 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.

|  |  |
| --- | --- |
| **ویژگی** | **مقادیر اتخاذ شده** |
| سن | ۲۰-۳۵  ۳۶-۴۵  ۴۶-۵۵  ۵۵-۵۶  بیشتر از ۶۵ |
| جنسیت | مرد، زن |
| پر‌ادراری[[7]](#footnote-7) | بله، خیر |
| عطش بیش از حد[[8]](#footnote-8) | بله، خیر |
| کاهش وزن ناگهانی | بله، خیر |
| ضعف | بله، خیر |
| پراشتهایی[[9]](#footnote-9) | بله، خیر |
| برفک تناسلی[[10]](#footnote-10) | بله، خیر |
| تاری دید | بله، خیر |
| خارش | بله، خیر |
| بی‌قراری | بله، خیر |
| بهبودی همراه با تاخیر | بله، خیر |
| کاهش توانایی تحرک در ناحیه‌ی پا[[11]](#footnote-11) | بله، خیر |
| سفتی عضلاتی | بله، خیر |
| طاسی منطقه‌ای[[12]](#footnote-12) | بله، خیر |
| چاقی | بله، خیر |
| **دسته** | **مثبت، منفی** |

# پیشنهاد دوم

## معرفی مقاله

ونگیپورام و همکاران در پژوهشی تحت عنوان *پیش‌بینی خطر دیابت جوانان با استفاده از داده‌های انهنس و یادگیری ماشین[2]* با استفاده از یک مجموعه‌داده‌ی بزرگ از نظرسنجی ملی سلامت و تغذیه[[13]](#footnote-13) شامل ۲۹۷۰ رکورد از نوجوانان ۱۲ تا ۱۹ سال به ارزیابی عملکرد یک دستورالعمل غربالگری بالینی کودکان در شناسایی بیماری پیش‌دیابت و دیابت شیرین[[14]](#footnote-14) مابین جوانان مبتلا به این بیماری با استفاده از نشانگرهای زیستی تشخیصی انجمن دیابت آمریکا[[15]](#footnote-15) 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 مجموعه‌داده

مجموعه‌داده‌ی 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 مجموعه‌داده 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.

|  |  |
| --- | --- |
| **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[[16]](#footnote-16) | 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[[17]](#footnote-17)** | **بله، خیر** |

# 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.

1. Sylhet [↑](#footnote-ref-1)
2. Naïve Bayes [↑](#footnote-ref-2)
3. Logistic Regression [↑](#footnote-ref-3)
4. Random Forest [↑](#footnote-ref-4)
5. Cross Validation [↑](#footnote-ref-5)
6. Percentage Split [↑](#footnote-ref-6)
7. Polyuria [↑](#footnote-ref-7)
8. Polydispia [↑](#footnote-ref-8)
9. Polyphagia [↑](#footnote-ref-9)
10. Genital Thrush [↑](#footnote-ref-10)
11. Partial Paresis [↑](#footnote-ref-11)
12. Alopecia [↑](#footnote-ref-12)
13. National Health and Nutritional Examination Survey (NHANES) [↑](#footnote-ref-13)
14. PreDM/DM [↑](#footnote-ref-14)
15. American Diabetes Association Diagnostic Biomarkers [↑](#footnote-ref-15)
16. BMI Percentile Categories [↑](#footnote-ref-16)
17. At risk for PreDM/DM based on AAP/ADA pediatric clinical screening guidelines [↑](#footnote-ref-17)