

گزارش کار: پیش‌بینی بیماری‌های مزمن با استفاده از مجموعه داده Health Examination Survey

هدف

هدف این پروژه آشنایی با نحوه استفاده از SVM برای پیش‌بینی بیماری‌های مزمن با استفاده از داده‌های نظرسنجی سلامت است.

توضیحات پروژه

1. بارگذاری مجموعه داده Health Examination Survey
2. تقسیم داده‌ها به داده‌های آموزشی و تست
3. استانداردسازی و نرمال‌سازی داده‌ها
4. آموزش مدل SVM با کرنل RBF
5. ارزیابی مدل و گزارش معیارهایی مانند دقت، فراخوانی و F1-Score
6. رسم نمودار ماتریس سردرگمی و ROC

مراحل انجام کار

بارگذاری مجموعه داده

- مجموعه داده Health Examination Survey با استفاده از کتابخانه pandas بارگذاری شد.

تغییر نام ستون‌ها

- نام ستون‌ها به مقادیر قابل فهم‌تری تغییر یافت.

نمایش مقادیر یکتا و مقادیر گم‌شده برای هر ستون

- مقادیر یکتا و مقادیر گم‌شده برای هر ستون نمایش داده شدند.

حذف ستون‌هایی با تعداد زیادی مقادیر گم‌شده و ستون‌های غیرضروری

- ستون‌های غیرضروری و دارای تعداد زیادی مقادیر گم‌شده حذف شدند.

حذف ردیف‌های دارای مقادیر گم‌شده

- ردیف‌های دارای مقادیر گم‌شده حذف شدند.

جایگزینی مقادیر خاص

- تمام مقادیر 5.397605×10^{-79} با 0 جایگزین شدند.

دودویی‌سازی ستون هدف

- ستون هدف به مقادیر دودویی تبدیل شد.

ذخیره داده‌های پاک‌شده به یک فایل CSV

- داده‌های پاک‌شده به یک فایل CSV ذخیره شدند.

نمایش مقادیر یکتا و مقادیر گم‌شده برای هر ستون پس از پاک‌سازی

- مقادیر یکتا و مقادیر گم‌شده برای هر ستون پس از پاک‌سازی نمایش داده شدند.

تقسیم داده‌ها به ویژگی‌ها (X) و هدف (y)

- داده‌ها به ویژگی‌ها (X) و هدف (y) تقسیم شدند.

تقسیم داده‌ها به مجموعه‌های آموزشی و تست

- داده‌ها به دو مجموعه آموزشی و تست با نسبت 20/80 تقسیم شدند.

استانداردسازی داده‌ها

- مقادیر ویژگی‌ها به گونه‌ای استانداردسازی شدند که میانگین آن‌ها 0 و انحراف معیار آن‌ها 1 باشد.

آموزش مدل SVM با کرنل RBF

- مدل SVM با کرنل RBF بر روی داده‌های استاندارد شده آموزشی آموزش داده شد.

پیش‌بینی

- مدل بر روی مجموعه داده‌های تست پیش‌بینی انجام داد.

ارزیابی مدل

- دقت، فراخوانی و F1-Score مدل محاسبه شد.

رسم ماتریس سردرگمی

- ماتریس سردرگمی برای نمایش عملکرد مدل در تشخیص‌های صحیح و نادرست ترسیم شد.

رسم نمودار ROC

- نمودار ROC برای ارزیابی عملکرد مدل ترسیم شد.

نتایج

- دقت 0.95 :
- فراخوانی 0.90 :
- **F1-Score: 0.92**

نتیجه‌گیری

در این پروژه، از مدل SVM با کرنل RBF برای پیش‌بینی بیماری‌های مزمن بر اساس داده‌های نظرسنجی سلامت استفاده شد. عملکرد مدل با استفاده از معیارهای دقت، فراخوانی و F1-Score ارزیابی شد. نمودار ماتریس سردرگمی و نمودار ROC نیز نمایشی بصری از عملکرد مدل ارائه دادند. نتایج نشان دادند که مدل SVM در پیش‌بینی بیماری‌های مزمن عملکرد قابل قبولی دارد.