

گزارش کار: تشخیص دیابت با استفاده از مجموعه داده Pima Indians Diabetes

هدف

هدف این پروژه آشنایی با نحوه استفاده از SVM برای دسته‌بندی داده‌های پزشکی و تشخیص دیابت است.

توضیحات پروژه

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

مراحل انجام کار

بارگذاری مجموعه داده

- مجموعه داده از فایل data.csv بارگذاری شد.

بررسی مجموعه داده

- چند سطر ابتدایی مجموعه داده برای بررسی ساختار آن نمایش داده شد.

آماده‌سازی ویژگی‌ها و متغیر هدف

- ویژگی‌ها (X) و متغیر هدف (y) جدا شدند، با فرض اینکه ستون آخر متغیر هدف است و سایر ستون‌ها ویژگی‌ها هستند.

تقسیم داده‌ها

- داده‌ها به دو بخش 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

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

پیش‌بینی و ارزیابی مدل

- عملکرد مدل با استفاده از معیارهایی مانند دقت، فراخوانی و F1-Score ارزیابی 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 به صورت بصری ترسیم شد.

نتایج

- دقت: مدل دقتی در حدود 0.73 به دست آورد.
- دقت: **(Precision)** دقت مدل حدود 0.65 بود.
- فراخوانی: **(Recall)** فراخوانی مدل حدود 0.56 بود.
- **F1-Score**: امتیاز F1 مدل حدود 0.60 بود.

نتیجه‌گیری

در این پروژه، از مدل SVM با کرنل RBF برای تشخیص دیابت بر اساس مجموعه داده Pima Indians Diabetes استفاده شد. عملکرد مدل با استفاده از معیارهای مختلف مانند دقت، دقت (Precision)، فراخوانی و F1-Score ارزیابی شد. ماتریس سردرگمی و نمودار ROC نیز نمایشی بصری از عملکرد مدل ارائه دادند. نتایج نشان دادند که مدل SVM در تشخیص دیابت عملکرد قابل قبولی دارد، اگرچه هنوز امکان بهبود وجود دارد.