



دانشگاه صنعتی امیرکبیر
(پلی تکنیک تهران)

دانشکده ریاضی و علوم کامپیوتر

استاد درس: دکتر مهدی قطعی

استاد کارگاه: بهنام یوسفی مهر

بهار ۱۴۰۳

راهنمای تمرین ۵

درس هوش مصنوعی و کارگاه



مقدمه

در این تمرین هدف تحلیل یکی از سه دیتاست زیر می باشد. یکی از آنها را به دلخواه انتخاب کرده و سپس باتوجه به گام های خواسته شده در ادامه، آن را تحلیل کنید.

• ADHD | Mental Health

• Bank Account Fraud Dataset Suite (NeurIPS 2022)

• ADHD Diagnosis Data

مراحل اجرای تمرین

برای اجرا مراحل از کتابخانه های آماده و دستورات مربوطه استفاده کنید. نیازی به پیاده سازی الگوریتم ها نمی باشد.

پیش پردازش داده

- ساختار داده و اطلاعات موجود را بررسی کنید
- پیش پردازش های مورد نیاز برای اجرای مدل ها را انجام دهید.
- در صورتی که برای classification ابتدا داده ها را متوازن کنید، تاثیر مثبت در ارزیابی شما خواهد داشت.

پیاده سازی مدل

الف) خوشه بندی

- در این مرحله میتوانید برای تحلیل بهتر نتیجه و بررسی خوشه ها از الگوریتم های کاهش بعد استفاده کنید.
- از یک الگوریتم خوشه بندی مانند K-Means استفاده کنید.
- نتایج را بررسی کرده و اگر خوشه هایی به دست آمد، این خوشه ها را تحلیل کنید.

**ب) طبقه بندی****استفاده از کتابخانه predict lazy**

در ابتدا با استفاده از این کتابخانه مدل های مختلف طبقه بندی را مقایسه کرده و براین اساس مدلی را که عملکرد بهتری دارد انتخاب کنید و مراحل بعد را انجام دهید. (در صورتی که این مدل در کلاس تدریس نشده بود میتوانید از بهترین مدل تدریس شده استفاده کنید)

- داده ها را به نسبت ۷۰ به ۳۰ برای train و test جدا کنید
- از یک الگوریتم طبقه بندی مانند KNN یا شبکه عصبی برای طبقه بندی استفاده کنید. (محدودیتی در انتخاب مدل نیست و شما میتوانید برای افزایش دقت مدل های مختلف را ارزیابی کنید)
- معیار های مختلف مانند F1, Accuracy و ... برای بررسی مدل را محاسبه کنید.
- با استفاده از تکنیک cross validation داده ها را جدا کرده و مراحل دوم و سوم را اجرا کنید.
- دقت و سایر معیارهای مربوط به طبقه بندی را تحلیل کنید.

نکته مهمی که باید به آن توجه کنید نشت داده یا DATA LEAKAGE می باشد، به همین علت در صورتی که در هنگام پیش پردازش داده ها از الگوریتم های کاهش بعد استفاده کرده اید؛ برای داده های train از دستور fit transform و برای داده های تست تنها از دستور transform استفاده کنید.

با توجه به جدید بودن و پتانسیل دیتاست های داده شده در صورتی که برای هریک از دیتاست ها بتوانید مدلی با نوآوری یا F1 بالا کسب کنید پس از بررسی آن با آقای یوسفی مهر در صورت تایید مدل شما توسط ایشان به شما برای چاپ مقاله آن کمک خواهند کرد.



معیارهای سنجش تمرین

- توانایی به کار بردن الگوریتم خوشه‌بندی
- توانایی تحلیل و توصیف خوشه‌ها
- توانایی به کار بردن الگوریتم طبقه‌بندی و تقسیم داده‌ها به train و test
- ارزیابی دقت و معیارهای دیگر طبقه‌بندی
- توضیحات کامل در قالب پایان نامه

مهلت تحویل

مهلت تحویل تمرین تا ساعت ۲۳:۵۹ روز ۲۲ اردیبهشت می باشد.
لطفا فایل pdf گزارش کار و فایل پیاده سازی کد را در قالب یک فایل zip در سامانه کورسز و کوئرا بارگذاری نمایید.