



دانشگاه صنعتی امیرکبیر
(پلی تکنیک تهران)

دانشکده ریاضی و علوم کامپیوتر

استاد درس: دکتر مهدی قطعی

استاد کارگاه: بهنام یوسفی مهر

آبان ۱۴۰۴

راهنمای پروژه ۳

درس هوش مصنوعی و کارگاه



توضیح تمرین

یادگیری ماشین و تکنیک‌های آن کاربردهای مهمی در حوزه‌های متنوع دارند. طبقه‌بندی یا همان (classification) و خوشه‌بندی یا همان (clustering) از معروف‌ترین و پرکاربردترین متدها و تکنیک‌های یادگیری ماشین هستند که در پروژه‌های مختلف از آن‌ها استفاده می‌شود، و یکی از این حوزه‌ها، حوزه ی پزشکی است.

یکی از کاربردهای مهم طبقه‌بندی در رشته ی پزشکی، این است که تعیین کنیم آیا یک فرد بیمار هست یا خیر. همچنین برای تعیین این که کدام افراد، براساس ویژگی‌های داده شده، به هم شبیه هستند، از خوشه‌بندی استفاده می‌کنیم تا بتوانیم دقیق‌تر به بررسی داده‌ها بپردازیم و بتوانیم به نتایج قابل اعتماد برسیم.

مجموعه داده‌ی مورد استفاده در این پروژه شامل اطلاعاتی کاربردی از افراد از جمله سن، جنسیت، نژاد، قند خون و اطلاعاتی از این قبیل است.

این پروژه شامل ۴ فاز اصلی است که عبارتند از:

۱. پیش پردازش

۲. خوشه‌بندی

۳. طبقه‌بندی در هر خوشه

۴. طبقه‌بندی روی کل داده (بدون خوشه‌بندی) و تحلیل و مقایسه‌ی کل نتایج

توضیحات پروژه و شرح وظایف

بخش اول: پیش پردازش

در این بخش باید با توجه به دانش خود و بررسی مجموعه داده، آن را پاکسازی کرده تا بتوانید از آن استفاده کرده و بهترین نتایج ممکن را به دست آورید.

بخش دوم: خوشه‌بندی

در این بخش با الگوریتم‌های خوشه بندی آشنا شده و آن‌ها را روی این مجموعه داده پیاده‌سازی می‌کنیم. در ادامه، با معیارهای ارزیابی خوشه‌بندی آشنا شده و آن‌ها را برای خوشه‌بندی‌هایی که انجام دادیم محاسبه می‌کنیم تا بتوانیم شهود بهتری از نتایجمان داشته باشیم و بتوانیم مقایسه کنیم.



بخش سوم: طبقه‌بندی در هر خوشه

در این مرحله، با الگوریتم‌های طبقه‌بندی آشنا شده و سپس در هر خوشه، این الگوریتم‌های را اجرا کرده و سپس ارزیابی می‌کنیم

بخش چهارم: طبقه‌بندی روی کل داده (بدون خوشه‌بندی) و تحلیل و مقایسه‌ی کل نتایج

در این بخش، الگوریتم‌های طبقه‌بندی را روی کل مجموعه داده پیاده‌سازی کرده و نتایج را با حالت قبلی، یعنی پیاده‌سازی طبقه‌بندی روی نتایج خوشه‌بندی، مقایسه کرده تا به بهترین راه ممکن برسیم.

معیار های ارزیابی

ارزیابی این پروژه بر اساس موارد زیر انجام می شود:

- پیاده سازی کامل تسک های خواسته شده در فایل تمپلیت
 - اجرای صحیح تمام مراحل کد و رسیدن به نتایج مطلوب
 - آموزش مناسب مدل ها و ارزیابی آن ها با استفاده از معیارهای داده شده
 - تحلیل و مقایسه مدل های متفاوت و استدلال در مورد نحوه عملکرد آن ها
 - پاسخ به سوالات داده شده در فایل نوت بوک و تسلط به پاسخ ها و مفاهیم آن
- توجه شود در این پروژه، حداقل یکی از مدل‌های اجرا شده (هم در انتهای بخش ۳ هم در بخش ۴) امتیاز F۱ آن باید حداقل مقدار ۵۰ درصد داشته باشد و مقدار کمتر از ۵۰ درصد قابل قبول نیست.
- گزارش پروژه باید شامل موارد زیر باشد:
- شرح مراحل انجام شده در پروژه
 - بیان مفاهیم و استدلال مناسب برای پاسخ‌های ارائه شده برای سوال‌های داخل نوت بوک
 - استدلال و دلیل برای انتخاب روش های متفاوت
 - خلاصه ای از عملکرد مدل های استفاده شده
 - نتایج به دست آمده از هر مدل و مقایسه آن ها



فایل ارسالی

در نهایت باید یک فایل زیپ در سامانه بارگذاری شود که شامل موارد زیر است:

- فایل نوت بوک پروژه در قالب Jupyter Notebook (.ipynb)
- لینک اجرای پروژه در گوگل کولب، ذخیره شده در یک فایل متنی (.txt)
- گزارش کار پروژه در قالب یک فایل PDF

مهلت تحویل

مهلت تحویل تمرین تا ساعت ۲۳:۵۹ روز ۱۶ آبان می باشد.

با استفاده از این **لینک** میتوانید به درایو پروژه سوم دسترسی داشته باشید. تمپلیت کد در این درایو می باشد.
توجه: دقت کنید که برای پیاده سازی پروژه عملی باید از طریق لینک کولب در درایو، یک کپی از آن را در کولب خود ایجاد کنید یا اینکه از طریق بخش Files، فایل تمپلیت را دانلود کرده و در سیستم خود به آن دسترسی داشته باشید.

در صورت وجود هر گونه ابهام در ارتباط با پروژه عملی میتوانید با آیدی های زیر در ارتباط باشید:

- ایلیا اسدی
- محمدرضا شیخ الاسلامی
- آراین جعفری
- سید سینا نگهبان