



مینی پروژه سوم

۱ SVM

۱.۱ ریاضیات: بررسی معادله صفحه جداکننده پیشنهاد شده توسط SVM

داده‌های زیر را در نظر بگیرید:

$$\mathbf{X}^{(1)} = [1, -1, 1] \rightarrow \text{کلاس} = -1$$

$$\mathbf{X}^{(2)} = [-3, 1, 1] \rightarrow \text{کلاس} = 1$$

$$\mathbf{X}^{(3)} = [-3, 1, -1] \rightarrow \text{کلاس} = -1$$

$$\mathbf{X}^{(4)} = [1, 2, 1] \rightarrow \text{کلاس} = -1$$

$$\mathbf{X}^{(5)} = [-1, -1, 2] \rightarrow \text{کلاس} = 1$$

این داده‌ها به‌طور خطی جداسازی‌پذیر هستند. معادله صفحه جداکننده‌ای که SVM پیشنهاد می‌کند، باید یافت شود. راهنمایی: برای پیدا کردن معادله این صفحه، باید از تابع هدف زیر استفاده کنید:

$$\phi(\alpha) = \sum_{i=1}^N \alpha_i - \frac{1}{2} \sum_{i,j=1}^N \alpha_i \alpha_j y_i y_j K(x_i, x_j)$$

شکل ۱: تابع هدف برای پیدا کردن مقادیر α

در این تابع:

- α ها پارامترهای لاگرانژ هستند که برای هر داده در مسئله بهینه‌سازی استفاده می‌شوند.
- Y ها برچسب‌های کلاس‌ها هستند که معمولاً به‌صورت ۱ و -۱ نمایش داده می‌شوند.
- K مربوط به کرنل است.
- N تعداد داده‌های دیتاست است.

برای حل معادله و پیدا کردن مقادیر α ، می‌توانید از روش حل مسائل برنامه‌ریزی درجه دوم (Quadratic Programming) استفاده کنید.

شما می‌توانید کد مربوط به qp_solver را خودتان پیاده‌سازی کنید یا از بسترهای آنلاین برای حل معادلات و پیدا کردن مجهولات استفاده کنید.

۲.۱ دیتاست - بررسی و تحلیل و دسته‌بندی دیتاست کیفیت هوای پکن (Beijing PM2.5 Data Set)

بررسی اولیه دیتاست

۱.۲.۱ دانلود و بررسی ویژگی‌های دیتاست:

- برای حل این قسمت این مجموعه داده را [دانلود](#) کنید.
- این دیتاست شامل ۱۳ ستون است که به ترتیب شامل موارد زیر می‌باشند:

No	year	month	day	hour	pm2.5	DEWP	TEMP	PRES	cbwd	lws	ls	lr
----	------	-------	-----	------	-------	------	------	------	------	-----	----	----

شکل ۲: ستون‌های دیتاست کیفیت هوای پکن

- ویژگی‌ها را بررسی کنید و برای هر ویژگی توضیح دهید که نشان‌دهنده چه نوع اطلاعاتی هستند و واحد اندازه‌گیری آن‌ها چیست.
- بررسی کنید که کدام ویژگی‌ها می‌توانند با آلودگی هوا ارتباط داشته باشند و چگونگی ارتباط آن‌ها را بیان کنید.
- آیا همین ویژگی‌ها برای بررسی آلودگی هوا کافی هستند؟ چه ویژگی‌های دیگری می‌توانند مفید باشند؟
- ویژگی‌هایی که می‌توان بر اساس ویژگی‌های فعلی مطرح کرد و دیتاست را بهبود بخشید، مطرح کنید.

۲.۲.۱ خواندن دیتاست و تحلیل اولیه:

- فایل CSV را بارگذاری کرده و در قالب DataFrame نمایش دهید.
- چند داده اول را مشاهده کرده و تحلیل اولیه‌ای از ساختار دیتاست ارائه دهید.
- با بررسی نمونه‌های اولیه و ترسیم ویژگی‌های مختلف، دیدگاه کلی خود را در مورد کیفیت، ساختار و تراکم داده‌ها بیان کنید.

پیش‌پردازش و تمیز کردن داده‌ها

۳.۲.۱ بررسی و مدیریت مقادیر گمشده (Missing Values):

- وجود missing value را بررسی کنید.
- حضور این نمونه‌ها و مقادیر چه تاثیری در ادامه کار و مدل ما خواهد داشت؟
- اگر در این دیتاست مقادیر گمشده‌ای وجود دارد، با روشی مناسب این مشکل را حل کنید و توضیح دهید چرا از آن روش استفاده کردید.

۴.۲.۱ داده‌های categorical:

- ویژگی‌های categorical در دیتاست وجود دارند؟
- در صورت وجود، چه روشی برای تبدیل کردن این ویژگی‌ها به دسته‌بندی‌های عددی وجود دارد؟
- یک ستون جدید ایجاد کرده و ویژگی مربوطه را به صورت عددی بنویسید و توضیح دهید که چگونه دسته‌بندی را انجام داده‌اید.

۵.۲.۱ داده‌های پرت:

- حضور این نوع داده‌ها چه تاثیری در مدل و روند کار خواهد داشت؟
- بررسی کنید که آیا ویژگی‌های ['pm2.5', 'TEMP', 'DEWP', 'lws'] دارای داده‌های پرت هستند؟
- این مشکل داده‌های پرت دیتاست را با روشی متناسب با دیتاست حل کنید.

۶.۲.۱ دسته‌بندی استاندارد غلظت PM2.5:

- استانداردهای کیفیتی هوا (AQI) را برای سطح غلظت ذرات معلق PM2.5 بررسی کنید.
- مقادیر PM2.5 را بر اساس این استاندارد در دسته‌های مختلف طبقه‌بندی کرده و در یک ستون جدید قرار دهید.

تحلیل اکتشافی داده‌ها (Exploratory Data Analysis - EDA)

۷.۲.۱ ایجاد ویژگی‌های جدید:

ویژگی Lag:

- Lag Feature به معنای استفاده از مقادیر گذشته یک ویژگی به عنوان ویژگی جدید است. در تحلیل سری‌های زمانی، این ویژگی به ما کمک می‌کند تا تأثیر مقادیر قبلی بر مقدار فعلی را بررسی کنیم.
- ویژگی‌های lag معمولاً برای سه دسته ایجاد میشوند:
 - ۱ الی ۳ ساعت گذشته
 - ۲۴ ساعت گذشته
 - ۱۶۸ ساعت گذشته
- بررسی کنید کاربرد هریک برای چه مواردی هست و در چه شرایطی میتوان ازشون استفاده کرد.
- این ویژگی را ایجاد کرده و مقادیر ۲ ساعت قبل PM2.5 و ۲۴ ساعت قبل PM2.5 را در ستون جدید ایجاد کنید.
- بعد از ایجاد ستون مربوط به این ویژگی، آیا چند سطر اول دیتا مقادیر NaN خواهند داشت؟ دلیل این موضوع را بررسی کنید.
- مشکل مقادیر NaN را به روشی مناسب حل کنید.
- با رسم نمودار مناسب برای lag_24 و lag_2 بررسی کنید چه الگویی را نمایش میدهند و نتیجه رو با توجه به کاربرد هر یک بررسی کنید.
- Correlation بین pm2.5 و ستون‌های مربوط به ویژگی lag را بررسی کرده، heatmap آن را رسم کرده و نتایج رو تحلیل کنید.

ویژگی Rolling Statistics:

- Rolling Statistics شامل محاسبه آماره‌هایی مانند میانگین، واریانس و انحراف معیار روی یک بازه متحرک است.
- بررسی کنید از این ویژگی در چه شرایطی میتوان استفاده کرد و نتایج آن چه کمکی به ما خواهد کرد.
- یک ویژگی جدید ایجاد کرده و Rolling Statistics را برای PM2.5 محاسبه کنید. (از پنجره ۲۴ ساعته استفاده کنید).
- آیا برای چند سطر اول داده‌ها این مقدار ثابت است؟ چرا؟
- نمودار pm2.5 بر حسب ستون مربوط به pm2.5-rolling statistics رسم کنید. نتیجه این نمودار رو تحلیل کنید.

انکود کردن ویژگی‌های تناوبی:

- در این دیتاست ویژگی‌های تناوبی وجود دارد؟
- برای انکود کردن این ویژگی‌ها، چه روشی پیشنهاد می‌شود؟ دلیل انکود کردن ویژگی‌های تناوبی را نیز توضیح دهید.
- در صورت وجود ویژگی‌های تناوبی در این دیتاست، آن‌ها را با روش مناسب انکود کرده و توضیح دهید چرا از آن روش استفاده کرده‌اید.
- ویژگی ساعت را انکود کرده با روشی مناسب و سپس نمودار scatter مربوط به آن را رسم نمایید. این نمودار چه نتیجه‌ای را به ما نشان میدهد؟ چطور میتوان توسط این نمودار بررسی کرد چه زمان‌هایی به صورت مکرر و متناوب آلودگی بالایی داریم؟

ویژگی‌های زمانی پیشرفته:

- یک ستون جدید مربوط به شاخص فصلی در دیتاست ایجاد کنید.
- بررسی کنید در چه فصلی از سال میزان آلودگی بیشتر هست! دلیل این موضوع را نیز بررسی نمایید.
- همین موضوع را برای روزهای هفته نیز بررسی کنید و مشخص کنید چه روزهایی از هفته دارای آلودگی بیشتری هست و علت آن چه می‌باشد. (با استفاده از متغیرهای day میتوان شاخص هفتگی ایجاد کرد).
- تاثیرات فشار هوا، دما و باد شدید را بر میزان آلودگی بررسی کنید.
- اگر الگوی ساعتی خاصی برای آلودگی هوا و تغییرات PM2.5 وجود دارد، آن را بررسی کنید.

۸.۲.۱ مصورسازی داده‌ها:

- تاریخ و زمان داخل دیتاست به صورت year, month, day, hour قرار گرفته‌اند. برای نمایش PM2.5 نمودار را بر حسب تاریخ و زمان رسم کنیم، چگونه باید از این ویژگی‌ها استفاده کنید؟
- با روشی مناسب یک ستون جدید مخصوص زمان و تاریخ ایجاد کنید که اطلاعات هر ۴ ویژگی، year, month, day, hour را در بر بگیرد.
- مقدار PM2.5 را بر اساس تاریخ و زمان نمایش دهید.
- بررسی کنید که معمولاً در چه تاریخی آلودگی هوا بیشینه بوده است.
- بین ماه‌های January و December کدام یک PM2.5 کمتری دارد؟ با اینکه هر دوی این ماه‌ها زمستان هستند، چرا بین PM2.5 آنها تفاوت وجود دارد؟

بررسی توازن داده‌ها (Data Balancing)

۹.۲.۱ بررسی توزیع داده‌ها:

- آیا داده‌های این دیتاست نامتوازن هست؟
- در صورتی که دیتاست بررسی شده نامتوازن هست، آن را متوازن کرده و در دیتافریم جدیدی ذخیره کنید. (لزوماً نیازی نیست از تمام ویژگی‌هایی که در دیتاست وجود دارد و طی مراحل قبل ایجاد کردیم استفاده کنید)

نرمال‌سازی داده‌ها و ایجاد داده‌های train, validation, test

۱۰.۲.۱ نرمال‌سازی داده‌ها:

- با روشی مناسب داده‌ها را نرمال‌سازی نمایید. دلیل انتخاب روش استفاده‌شده برای نرمال‌سازی داده‌ها را توضیح دهید.
- داده‌های نرمال‌سازی‌شده را در یک دیتافریم جدید ذخیره کنید.

۱۱.۲.۱ تقسیم داده‌ها به مجموعه‌های آموزش و آزمون و ارزیابی:

- داده‌ها را به دو دیتافریم ویژگی‌ها (X) و هدف (y) تقسیم کنید. (لزوماً نیازی نیست X شامل تمامی ویژگی‌های موجود در دیتاست باشد).
- داده‌ها را به سه مجموعه (train (70%), validation (15%), test (15%) تقسیم‌بندی کنید.

۱۲.۲.۱ مدل‌سازی با استفاده از SVM (Support Vector Machine) برای دسته‌بندی

- چه هایپرپارامترهایی وجود دارد که در الگوریتم SVM حائز اهمیت هست و می‌توان آن‌ها را بهینه کرد؟
- آیا در کرنل‌ها نیز هایپرپارامترهای حائز اهمیت وجود دارند؟
- SVM-linear، SVM-polynomial و SVM-rbf را یک‌بار با استفاده از توابع از پیش آماده‌شده و کتابخانه‌ها مدل کنید، و مدل‌ها را ارزیابی کنید. در انتها نتایج این مدل‌ها را با هم بررسی کنید.

- SVM-linear، SVM-polynomial و SVM-rbf را یکبار دیگر بدون استفاده از توابع از پیش آماده‌شده و کتابخانه‌ها مدل کنید و تمامی توابع را خودتان بنویسید. مجدد ارزیابی را برای این مدل‌ها نیز انجام داده و در انتها نتایج این مدل‌ها را با هم بررسی کنید.
- چه تفاوتی در نتایج مدل‌هایی که با استفاده از توابع از پیش تعریف‌شده ایجاد شدند، با نتایج مدل‌هایی که خودتان نوشتید مشاهده می‌کنید؟ علت آن را بررسی کنید.
- با استفاده از Random Search و Grid Search به ترتیب برای بازه دلخواه و مقادیر دلخواه برای دو کرنل Linear و RBF بهترین طبقه‌بندی را پیدا کنید و نتایج را با مدل‌های قبلی بررسی کنید.
- نمودار loss داده‌های validation و داده‌های train را رسم کنید برای مدل‌های مختلف. (یکبار قبل از پیش‌پردازش داده و یکبار پس از پیش‌پردازش داده) بررسی کنید که هر مورد از مراحل پیش‌پردازش چه تأثیری در روند loss شما داشته است.

۱۳.۲.۱ مدل‌سازی با استفاده از SVM (Support Vector Machine) برای پیش‌بینی

- نحوه عملکرد الگوریتم SVM برای پیش‌بینی چگونه است؟
- هایپرپارامترهای این مدل برای پیش‌بینی با هایپرپارامترهای مدل دسته‌بندی یکسان هست؟
- از نظر عملکرد چه تفاوت‌هایی وجود دارد؟
- همان داده‌های هوای شهر پکن را در نوت‌بوکی دیگر باز کرده و با انجام مراحل پیش‌پردازش، تا مرحله متوازن‌سازی دادگان مطابق قبل پیش‌روی کنید.
- برای متوازن‌سازی دادگان از روشی مناسب استفاده نمایید و دادگان را متوازن سازید.
- در ادامه یک مدل SVM ایجاد کرده و برای پیش‌بینی میزان PM2.5 از آن استفاده نمایید.
- برای ارزیابی این مدل می‌توان مطابق قبل عمل کرد و از روش‌های ارزیابی که برای دسته‌بندی استفاده شده بود مجدد استفاده کرد؟ چرا؟
- در صورتی که نمی‌توان از روش‌های قبلی استفاده مرد، از تکنیک‌های مناسب ارزیابی استفاده نمایید و نتایج را گزارش دهید.

بهبودسازی مدل SVM با PSO (Particle Swarm Optimization) - امتیازی

۱۴.۲.۱ معرفی الگوریتم PSO:

- توضیح دهید که PSO چیست و چگونه می‌توان از آن برای بهبودسازی پارامترهای SVM استفاده کرد.

۱۵.۲.۱ پیاده‌سازی SVM با PSO:

- پارامترهای مدل SVM با کرنل RBF را با PSO بهبود کنید.
- مدل بهبودشده را اجرا کرده و ارزیابی کنید.
- نتایج مدل PSO-Optimized RBF SVM را با مدل معمولی SVM-rbf مقایسه کنید.

۳.۱ بررسی و پیاده‌سازی مقاله

۱.۳.۱ عنوان مقاله

Innovative SVM optimization with differential gravitational fireworks for superior air pollution classification

نویسندگان: Bian Chao & Huang Guangqiu

سال انتشار: 2024

۲.۳.۱ دیتاست استفاده شده:

در این بخش، توضیح دهید که دیتاست استفاده شده در مقاله چیست، ویژگی‌های آن چیست و چه نوع داده‌هایی شامل می‌شود.

۳.۳.۱ مدل استفاده شده:

مدل SVM با یک روش بهینه‌سازی نوآورانه به نام Differential Gravitational Fireworks برای دسته‌بندی کیفیت هوا استفاده شده است. هدف اصلی، افزایش دقت مدل دسته‌بندی با استفاده از بهینه‌سازی هوشمند هایپرپارامترها بوده است. در صورتی که تغییرات نوآورانه‌ای در مدل ایجاد شده، باید این تغییرات و اهداف آن‌ها به طور کامل توضیح داده شود.

۴.۳.۱ نتایج:

در این قسمت نتایج مدل‌ها بررسی و تحلیل می‌شوند. مدل‌های ارائه شده با روش پیشنهادی، از نظر دقت، سرعت همگرایی، و پیچیدگی محاسباتی مقایسه شده‌اند. بهبود قابل توجهی نسبت به مدل‌های پایه SVM گزارش شده است.

۵.۳.۱ پیاده‌سازی:

- هدف از این بخش، پیاده‌سازی الگوریتم مقاله است.
- ابتدا دیتاست کیفیت هوای پکن را بارگذاری کرده و مراحل آماده‌سازی شامل بررسی مقادیر گم‌شده، داده‌های پرت، و توازن داده‌ها را انجام دهید.
- با توجه به رویکرد مقاله، یک مدل SVM طراحی و پیاده‌سازی کنید که با استفاده از تکنیک Differential Evolution (DE) یا الگوریتم پیشنهادی مقاله، هایپرپارامترهای آن بهینه شوند.
- مدل طراحی شده را روی دیتاست اجرا کرده و عملکرد آن را ارزیابی کنید.
- نتایج مدل پیشنهادی را با یک SVM استاندارد مقایسه کنید.
- تحلیل دقیقی از تفاوت‌ها از نظر دقت، سرعت همگرایی، و پیچیدگی محاسباتی ارائه دهید. در کدام ویژگی‌ها مدل پیشنهادی برتری دارد؟
- بررسی کنید چگونه الگوریتم Differential Evolution به global search در فضای هایپرپارامترها کمک کرده است. همچنین مزایا و محدودیت‌های استفاده از DE را نسبت به الگوریتم‌های بهینه‌سازی مانند Genetic Algorithm (GA) یا Particle Swarm Optimization (PSO) توضیح دهید.

۶.۳.۱ محصول نهایی

در این بخش، هدف ایجاد یک مدل نهایی و کاربردی بر اساس تحلیل‌های پیشین است:

- بعد از بررسی تمام مدل‌ها و روش‌های بهینه‌سازی SVM، مدلی که بهترین عملکرد را داشته انتخاب کرده و آن را کامل آموزش دهید.
- سپس داده‌های واقعی را که دارای ویژگی‌هایی مطابق با واحدهای دیتاست اولیه هستند، به مدل وارد کرده و خروجی را بررسی کنید.
- حداقل ۵ ورودی واقعی به مدل ارائه دهید، خروجی مدل را تحلیل کرده و گزارش کامل ارائه دهید.
- برای تهیه داده‌های واقعی، می‌توانید از سرویس‌هایی نظیر [OpenWeather](#) استفاده کنید.

۷.۳.۱ بخش امتیازی (تحقیقاتی)

- جدیدترین روش‌های ارزیابی عملکرد مدل‌های SVM را برای هر دو حالت دسته‌بندی و پیش‌بینی بررسی کنید.
- این روش‌ها از چه معیارهایی مانند Accuracy, Recall, AUC-ROC, F1-Score و ... استفاده می‌کنند؟ در چه شرایطی هر معیار مناسب‌تر است؟
- بررسی کنید که آیا ترکیب مدل‌های SVM با روش‌های مختلف بهینه‌سازی می‌تواند باعث بهبود عملکرد شود؟ آیا در شرایط مختلف مانند داده‌های نامتوازن، نویزی یا پیچیده کاربردی هستند؟

- در پایان ایده‌ها و پیشنهادهای خود را برای بهبود عملکرد مدل‌ها و استفاده از روش‌های نوآورانه ارائه دهید.
- تمامی مقالات مطالعه‌شده به صورت منابع در انتهای گزارش آورده شوند.
- در هر قسمت که به مقاله‌ای اشاره می‌شود، از روش صحیح ارجاع‌دهی استفاده شود.

۲ کاهش ابعاد

مجموعه داده‌ی ^۱MNIST Fashion شامل ۷۰,۰۰۰ تصویر سیاه‌وسفید با اندازه‌ی ۲۸×۲۸ پیکسل از انواع پوشاک در ۱۰ دسته‌ی مختلف است.

۱.۲ سوالات

(الف) مجموعه داده را بارگذاری کرده، از هر یک از ده دسته‌ی موجود یک تصویر انتخاب کرده و آن‌ها را به صورت یک ردیف نمایش دهید. سپس به تمام داده‌ها نویز گاوسی^۲ با میانگین صفر و انحراف معیار ۰/۲ اضافه کرده و مقادیر پیکسل‌ها را در بازه‌ی [0, 1] محدود نمایید. در پایان، نسخه‌های نویزی همان تصاویر را نیز به صورت جداگانه نمایش دهید.

(ب) الگوریتم PCA را بدون استفاده از هیچ کتابخانه‌ی آماده پیاده‌سازی نمایید. سپس نسبت واریانس توضیح داده‌شده برای مؤلفه‌های اصلی را محاسبه کرده و آن را به صورت نمودار میله‌ای ترسیم نمایید. چند مؤلفه اصلی کافی است تا بخش عمده‌ای از اطلاعات داده (مثلاً ۹۰٪ یا بیشتر از واریانس کل) حفظ شود؟

(ج) الگوریتم PCA را با استفاده از کتابخانه‌ی ^۳scikit-learn اجرا کرده، داده‌ها را به دو مؤلفه‌ی اصلی کاهش دهید و آن‌ها را در قالب نمودار پراکندگی دوبعدی نمایش دهید، به گونه‌ای که هر دسته با رنگی متفاوت مشخص باشد.

(د) از PCA برای بازسازی تصاویر نویزی استفاده نمایید. تصاویر نویزی را به زیرفضای مؤلفه‌های اصلی با تعداد مشخصی مؤلفه (مثلاً ۱۰۰) فشرده‌سازی کرده و آن‌ها را بازسازی نمایید. سپس سه ردیف تصویر شامل نسخه‌ی اصلی، نسخه‌ی نویزی و نسخه‌ی بازسازی‌شده را برای همان ۱۰ تصویر نمایش دهید.

(ه) الگوریتم LDA را با استفاده از کتابخانه‌ی ^۳scikit-learn پیاده‌سازی نمایید. سپس داده‌ها را به دو مؤلفه کاهش داده و نتایج را در قالب نمودار پراکندگی دوبعدی نمایش دهید به گونه‌ای که کلاس‌ها با رنگ متفاوت مشخص باشند. نتایج LDA را با خروجی PCA مقایسه و تحلیل نمایید.

(و) ماتریس‌های پراکندگی درون‌کلاسی S_w و بین‌کلاسی S_b را محاسبه کرده و سپس ماتریس جدایی‌پذیری $S_w^{-1}S_b$ را تشکیل دهید. مقادیر ویژه‌ی این ماتریس را به صورت نزولی محاسبه و در قالب نمودار رسم نمایید.

(ز) مقدار $\text{trace}(S_w^{-1}S_b)$ را نسبت به تعداد ویژگی‌های انتخاب‌شده محاسبه کرده و آن را در قالب یک نمودار ترسیم نمایید. سپس تأثیر افزایش تعداد ویژگی‌ها بر مقدار این ردیابی را تحلیل و تفسیر نمایید. تعداد بهینه ویژگی‌ها را بیابید.

۲.۲ امتیازی

روش t-SNE^۵ را به صورت مفهومی توضیح دهید. سپس این روش را بدون استفاده از هیچ‌گونه کتابخانه‌ی آماده پیاده‌سازی نمایید و نتایج آن را برای داده‌های انتخاب‌شده نمایش دهید. مزایا و محدودیت‌های آن را نیز با روش‌های دیگر کاهش بُعد مانند PCA و LDA مقایسه نمایید.

^۱MNIST Fashion

^۲Gaussian noise

^۳scikit-learn

^۴trace: مجموع قطر اصلی یک ماتریس مربعی

^۵t-distributed Stochastic Neighbor Embedding

در انجام این مینی پروژه حتماً به نکات زیر توجه کنید:

- موعد تحویل این تمرین، ساعت ۱۸:۰۰ روز سه شنبه ۶ خردادماه ۱۴۰۴ است.
- برای گزارش لازم است که پاسخ هر سوال و زیربخش هایش به ترتیب و به صورت مشخص نوشته شده باشند. بخش زیادی از نمره به توضیحات دقیق و تحلیل های کافی شما روی نتایج بستگی خواهد داشت.
- لازم است که در صفحه اول گزارش خود لینک مخزن گیت هاب را و گوگل کولب مربوط به مینی پروژه خود را درج کنید. درخصوص گیت هاب، یک مخزن خصوصی درست کنید و آی دی های MJAHMADEE و AliBagheriNejad را به عنوان Collaborator به مخزن اضافه کنید. پروژه های گیت هاب می بایست در انتهای ترم پابلیک شوند. درمقابل، لینک گوگل کولب را در حالتی که دسترسی عمومی دارد به اشتراک بگذارید. دفترچه کد گوگل کولب باید به صورت منظم و با بخش بندی مشخص تنظیم شده باشد و خروجی سلول های اجرا شده قابل مشاهده باشد. در گیت هاب نیز یک مخزن برای درس و یک پوشه مجزا برای هر مینی پروژه ایجاد کنید.
- (آموزش پرایوت کردن مخزن گیت هاب و آموزش افزودن Collaborator به مخزن گیت هاب)
- هر جا از دفترچه کد گوگل کولب شما نیاز به فراخوانی فایلی خارج از محیط داشت، مطابق آموزش های ارائه شده ملزم هستید از دستور gdown استفاده کنید و مسیرهای فایل ها را طوری تنظیم کنید که صرفاً با اجرای سلول های کد، امکان فراخوانی و خواندن فایل ها توسط هر کاربری وجود داشته باشد.
- در تمامی مراحل تعریف داده و مدل و هر جای دیگری که مطابق آموزش های ویدیویی و به لحاظ منطقی نیاز است، Random State را برابر با دو رقم آخر شماره دانشجویی خود در نظر بگیرید.
- استفاده از ابزارهای هوشمند (مانند ChatGPT) در کمک گرفتن برای بهبود کدها مجاز است؛ اما لازم است تمام جزئیات مواردی که در خروجی های مختلف گزارش خود عنوان می کنید را به خوبی خوانده، درک و تحلیل کرده باشید. استفاده از این ابزارهای هوشمند در نوشتن گزارش و تحلیل ها ممنوع است.
- در جاهایی که با توجه به دو رقم آخر شماره دانشجویی خود محدود به انتخاب عدد، متغیر و یا داده ای خاص شده اید، برای تست های اضافه تر و نمایش بهبود در نتایج خود، مجاز هستید از مقادیر دیگر هم استفاده کنید.
- رعایت نکات بالا به حرفه ای تر شدن شما کمک خواهد کرد و اهمیتی معادل مطالب درسی فراگرفته شده دارد؛ بنابراین، در صورت عدم رعایت هریک از این نکات، از نمره تمرین شما کسر خواهد شد.
- آی دی پرسش هرگونه سوال درخصوص پرسش شماره 1
- آی دی پرسش هرگونه سوال درخصوص پرسش شماره 2