



نکته مهم درباره محیط اجرا

این تمرین را حتماً در یکی از محیط‌های تعاملی زیر انجام دهید:

گزینه ۱: Google Colab

- مزیت: نیازی به نصب چیزی ندارید، فقط با مرورگر کار کنید

لینک: colab.research.google.com

- راهنمای یک Notebook جدید بسازید و کدها را در سلول‌های جداگانه اجرا کنید

نمودارها مستقیماً زیر هر سلول نمایش داده می‌شوند

گزینه ۲: Jupyter Notebook

سوال ۱: تحلیل داده‌های اول‌چری با افزودن ویژگی جدید

هدف پروژه: در این مینی‌پروژه، شما با استفاده از دیتاست اول‌چری، به تحلیل ویژگی‌های اول‌چری و افزودن ویژگی‌های جدید می‌پردازید. این پروژه به شما کمک می‌کند تا با تکنیک‌های تحلیل داده، مصورسازی و مفهوم افزودن ویژگی جدید آشنا شوید.

مراحل پروژه:

۱. جمع‌آوری و آماده‌سازی داده‌ها:

- دیتاست اول‌چری را از لینک [Iris Dataset](#) دانلود کرده و در پروژه خود بارگذاری کنید.
- ساختار داده‌ها را بررسی کرده و نوع داده‌ها و مقادیر گمشده را تحلیل کنید. در صورت وجود مقادیر گمشده، آنها را مدیریت کنید.
- همبستگی بین ویژگی‌های مختلف (مانند طول و عرض گلبرگ) را محاسبه کنید.

۲. افزودن ویژگی جدید:

- نسبت طول گلبرگ به عرض گلبرگ را محاسبه کرده و به دیتاست اضافه کنید. این ویژگی جدید به نام "petal_ratio" نامگذاری می‌شود.

۳. مصورسازی داده‌ها:

- از نمودارهای مناسب برای مصورسازی داده‌ها استفاده کنید. این شامل:
 - نمودار Heatmap برای مشاهده همبستگی ویژگی‌ها
 - نمودار جفتی (Pairplot) برای بررسی روابط بین ویژگی‌ها
 - نمودار Boxplot برای بررسی توزیع ویژگی جدید "petal_ratio" بر اساس گونه‌ها
 - نمودار پراکنده‌گی (Scatter Plot) برای بررسی رابطه بین طول گلبرگ و نسبت طول به عرض گلبرگ

۴. تحلیل نهایی:

- یافته‌های خود را تحلیل کنید و به سوالات زیر پاسخ دهید:
 - چگونه ویژگی جدید (نسبت) می‌تواند به تفکیک گونه‌ها کمک کند؟
 - آیا رابطه‌ای بین نسبت و طول گلبرگ وجود دارد؟
 - کدام گونه‌ها بیشترین و کمترین نسبت طول به عرض گلبرگ را دارند؟

سوال ۲: چالش داده‌های اقلیمی

تحلیل و پیش‌بینی گرمایش جهانی با استفاده از کتابخانه‌های Pandas، NumPy و Matplotlib

مراحل پروژه

۱- آماده‌سازی داده

دیتاست را از لینک [datasets/global-temp](#) دانلود کرده و در پروژه خود بارگذاری کنید.

وظایف:

۱. حذف مقادیر NaN:

.NaN شناسایی و حذف مقادیر.

. پر کردن مقادیر NaN با میانگین ۵ ساله اطراف.

۲. افزودن ستون Decade:

تبديل سال‌ها به دهه‌ها. به عنوان مثال، سال ۱۹۸۰ به "۱۹۸۰" و سال ۱۹۹۵ به "۱۹۹۵".

۳. فیلتر کردن کشورها:

نگهداری فقط کشورهایی که حداقل ۱۰۰ سال داده دارند.

اطمینان حاصل کنید که داده‌ها به صورت صحیح برای تحلیل‌های بعدی آماده شده‌اند.

۲- تحلیل عددی (NumPy)

۱. محاسبه میانگین دمای هر دهه:

.decade × temperature anomaly محاسبه میانگین دما برای هر دهه و ساخت ماتریس

اطمینان حاصل کنید که داده‌ها به درستی گروه‌بندی و میانگین‌گیری شده‌اند.

۲. محاسبه Trend خطی دما:

استفاده از np.polyfit() برای محاسبه خط روند دما.

محاسبه نرخ گرمایش بر حسب $^{\circ}\text{C}$ بر دهه.

۳. پیش‌بینی دما تا سال ۲۰۵۰:

پیش‌بینی دما با سه سناریو:

خوش‌بینانه: کاهش ۳۰٪ نرخ

پایه: نرخ فعلی

بدینانه: افزایش ۵۰٪ نرخ

۳- مصورسازی

۱. نمودار دما + خط روند:

رسم نمودار دما همراه با خط روند برای تجزیه و تحلیل تغییرات دما در طول زمان.

۲. Heatmap تغییرات دهه‌ای:

ایجاد Heatmap برای نمایش تغییرات دما در دهه‌های مختلف.

استفاده از رنگ‌ها برای نمایش شدت تغییرات.

۳. نمودار سه سناریو تا ۲۰۵۰:

- رسم نمودار پیش‌بینی دما برای سه سناریو تا سال ۲۰۵۰.

- استفاده از رنگ‌های مختلف برای هر سناریو و نمایش واضح آن‌ها.

۴. ترکیب در یک تصویر:

- ترکیب تمام نمودارها در یک تصویر (۳ زیرنمودار) با استفاده از Subplots.

- اطمینان از اینکه نمودارها به خوبی قابل مشاهده و خواندن هستند.

۴- مرحله چهارم

۱. محاسبه همبستگی:

- با استفاده از np.corrcoef، همبستگی دمای جهانی و داده‌های CO₂ را بسنجید.

- رسم نمودار همبستگی و تحلیل نتیجه.

محدودیت‌ها

- فقط از کتابخانه‌های matplotlib, numpy, pandas و استفاده شود.

:

سؤال ۳: تحلیل متنی ساده با استفاده از pandas و مقدمات NLP

آشنایی با پردازش متن (Text Processing) در پایتون با استفاده از ابزارهای پایه مانند pandas و مفاهیم ابتدایی NLP

مرحله ۱: خواندن داده و آماده‌سازی اولیه

۱. فایل reviews.csv را با استفاده از `pandas.read_csv()` بخوانید و در یک DataFrame ذخیره کنید.

۲. بررسی کنید که فایل شامل ستون زیر باشد:

• ReviewText: شامل متن نظرات کاربران درباره یک محصول یا سرویس

۳. از وجود مقادیر خالی (NaN) در ستون ReviewText مطمئن شوید و در صورت وجود، آنها را حذف یا با مقدار مناسب جایگزین کنید.

مرحله ۲: استخراج ویژگی‌های متنی اولیه

۱. محاسبه طول هر متن (بر حسب تعداد کاراکتر):

• یک ستون جدید به نام TextLength ایجاد کنید که تعداد کاراکترهای هر متن را نشان می‌دهد.

۲. شمارش تعداد واژه‌ها در هر متن:

• یک ستون دیگر با نام WordCount اضافه کنید که تعداد کلمات هر متن را محاسبه می‌کند.

مرحله ۳: تحلیل فراوانی کلمات در کل داده

۱. تمام متن‌های موجود در ستون ReviewText را به یک متن بزرگ ترکیب کنید.

۲. با استفاده از توابع رشته‌ای پایتون (مثل `.lower()`, `.split()`, `.replace()`) یا کتابخانه‌های کمکی:

• متن را به لیستی از کلمات (tokens) تبدیل کنید.

• کلمات رایج مانند نقطه، کاما، و ... را حذف کنید (اختیاری).

۳. فراوانی (تعداد تکرار) هر کلمه را محاسبه کنید.

۴. ۱۰ کلمه پر تکرار را در قالب یک جدول نمایش دهید.

مرحله ۴: مصورسازی فراوانی کلمات

۱. از کتابخانه‌های matplotlib یا seaborn استفاده کنید تا یک نمودار میله‌ای (barplot) از ۱۰ کلمه پر تکرار رسم کنید.

• محور X: کلمات

• محور Y: تعداد تکرار

• عنوان نمودار: "Top 10 Most Frequent Words in Reviews"

مرحله ۵: تحلیل و نتیجه‌گیری

به صورت نوشتاری پاسخ دهید:

- کدام کلمات بیشترین استفاده را دارند؟ آیا این کلمات معنای خاصی دارند؟
- آیا تکرار زیاد برخی کلمات نشان از رضایت یا نارضایتی کاربران دارد؟
- چه پیش‌پردازش‌هایی می‌توان انجام داد تا تحلیل دقیق‌تری داشته باشیم؟ (مثلاً حذف stopwords، ریشه‌یابی کلمات و ...)

۴: شبیه‌سازی ساده‌ی سیستم تشخیص چهره (Face Unlock) با مقایسه‌ی تصویری

هدف: آشنایی با ایده‌ی پایه‌ی تشخیص چهره از طریق مقایسه‌ی تصویری با استفاده از `Matplotlib`, `NumPy`, و تحلیل عددی ساده.

مرحله ۱: بارگذاری و نمایش تصاویر چهره

۱. دو تصویر با فرمت `.jpg` یا `.png` را آماده کنید:

• تصویر مرجع (مثلاً صاحب دستگاه) → `face_reference.png`

• تصویر کاربر فعلی برای مقایسه → `face_input.png`

۲. با استفاده از `matplotlib.pyplot.imread()` این دو تصویر را بخوانید.

۳. اگر تصویر رنگی است، آن را به `grayscale` تبدیل کنید:

با استفاده از `plt.imshow()` دو تصویر را کنار هم نمایش دهید (با `.subplot()`).

مرحله ۲: تحلیل عددی تصویر با NumPy

۱. اندازه‌ی تصاویر را یکسان کنید (برش یا `resize` در صورت نیاز با `numpy`).

۲. دو تصویر را به آرایه‌ی عددی تبدیل کرده و با هم مقایسه کنید.

۳. محاسبه‌ی معیارهای عددی برای تشخیص شباهت:

• میانگین اختلاف پیکسل‌ها (Mean Absolute Error)

• میانگین مربعات اختلاف (Mean Squared Error)

مرحله ۳: نمایش نتیجه تشخیص

۱. اگر مقدار `MSE` یا `MAE` کمتر از مقدار مشخصی بود، بنویسید:

• چهره مشابه است (Face Matched)

• چهره متفاوت است (Face Not Matched)

۲. از `plt.title()` یا `plt.text()` برای نمایش نتیجه روی نمودار استفاده کنید.

مرحله ۴: ذخیره نتایج در DataFrame (تحلیل مقایسه‌ای)

اگر چند تصویر مقایسه‌ای دارید (مثلاً ۵ تصویر ورودی را با تصویر مرجع مقایسه می‌کنید):

۱. برای هر تصویر ورودی، مقدار `MSE` را محاسبه کنید.

۲. نتایج را در یک جدول `pandas.DataFrame` ذخیره کنید:

۳. نمودار میله‌ای از مقدار `MSE` برای هر تصویر رسم کنید.

مرحله ۵: تحلیل و نتیجه‌گیری

پاسخ به این سؤالات:

- چقدر تفاوت بین تصاویر هست؟
- آیا سیستم ساده‌ی شما در تشخیص چهره‌ها خوب عمل می‌کنه؟
- فکر می‌کنید چه عواملی باعث خطا در تشخیص می‌شوند؟ (مثالاً نور، زاویه، چرخش؟)

نحوه تحويل:

تمرین را در **Git** آپلود کنید در هنگام تحويل تمرین فایل های شما در گیت بررسی میشود.