چالش ورود به بوت‌کمپ مهندس داده۲

از اینکه بوت‌کمپ مهندس داده۲ آکادمی همراه اول را برای رشـــد و توســـعه مهارت‌های خود انتخاب کرده‌اید، خوشـحالیم! برای تکمیل فرآیند ثبت‌نام، علاوه بر ارسـال رزومه، لازم است به چالش مطرح شـده پاسـخ دهید. این مرحله به ما کمک می‌کند تا شـما را بهتر بشـناسـیم و همراهی موثرتری در مسـیر یادگیریتان داشـته باشـیم. لطفاً فایل پاسـخ چالش به همراه رزومه، طبق دسـتورالعمل‌های پایانی بارگذاری نمائید .

- ## هدف از این چالش

سنجش اولیه دانش پایه و توان تحلیلی داوطلبان پیش از شروع بوت‌کمپ است.

- ## بخش اول - چالش ورودی و بررسی اولیه

در این چالش، یک محیط شبیه‌سازی‌شده مدیریت هوشمند شبکه آب ارائه شده که در آن سه مدل هوش مصنوعی به‌صورت مداوم داده‌های پایش را پردازش کرده و خروجی خود را به شکل Data Stream ارسال می‌کنند.
وظیفه شما به عنوان مهندس داده، طراحی سامانه‌ای برای دریافت جریان داده، اعتبارسنجی و تفکیک رکوردهای سالم و خراب، ذخیره‌سازی داده‌ها و تولید گزارش است.
در این چالش نیازی به طراحی یا آموزش مدل هوش مصنوعی نیست و تمرکز اصلی بر پیاده‌سازی بخش‌های مهندسی داده شامل پردازش، ذخیره‌سازی و پایش داده‌هاست.

سه مدل فعال روی شبکه

مدل	خروجی	توضیح
مدل تشخیص نشتی leak_detector	leak / normal	بررسی فشار و جریان ایستگاه و تشخیص وضعیت نشتی (leak) یا عادی (normal).
مدل پیش‌بینی افت فشار pressure_drop_predictor	drop / stable	پیش‌بینی افت فشار آینده با استفاده از روند داده‌ها؛ خروجی افت (drop) یا پایدار.
مدل پیش‌بینی تقاضای مصرف demand_forecaster	عدد صحیح ۱ تا ۵۰۰۰	پیش‌بینی میزان مصرف آب هر ایستگاه در یک ساعت آینده

نحوه اجرای سرور

هر بار که یکی از مدل‌ها پیش‌بینی جدیدی انجام میدهد، یک رکورد لاگ تولید می‌شود. این رکوردها توسط اسکریپت server_stream_ai_water.py ساخته می‌شوند. سرور هر ۰.۰۲۵ ثانیه یک رکورد جدید را به‌صورت رشته ی JSON، از طریق سوکت TCP روی پورت ۹۰۳٤ ارسال می‌کند؛ یعنی در شرایط عادی حدود ۴۰ رکورد در ثانیه وارد برنامه‌ی شما میشود.

مدیریت جریان داده

فایل water_ai_stream_server.py را دانلود و در مسیر دلخواه ذخیره کنید.

سپس دستور زیر را در ترمینال اجرا کنید:

```
python3 water_ai_stream_server.py
```

اجرای دستور بالا، سرور روی آدرس localhost:9034 شروع به کار میکند و نرخ خرابی انتخاب شده برای همان اجرا را در ترمینال چاپ میکند. برای متوقف کردن آن میتوانید از کلیدهای C+Ctrl استفاده کنید.

ساختار داده ها

فیلد	توضیح	شرط اعتبار
reading_id	شناسه رکورد	عدد صحیح مثبت
station_id	شناسه ایستگاه	یکی از ST-01 تا ST-10
model_name	نام مدل	leak_detector, pressure_drop_predictor, demand_forecaster
predicted_label	خروجی مدل	وابسته به نوع مدل
confidence_score	میزان اطمینان	عدد بین ۰ و ۱
response_time_ms	زمان پاسخ	عدد مثبت
timestamp	زمان ثبت	Timestamp معتبر

مقادیر معتبر predicted_label

Model	Label
leak_detector	leak / normal
pressure_drop_predictor	drop / stable
demand_forecaster	عدد صحیح ۱ تا ۵۰۰۰

ماموریت های شما

دریافت جریان داده

برنامه‌ای با پایتون بنویسید که به سوکت TCP روی پورت ۹۰۳٤ متصل شده و داده‌ها را به‌صورت Stream دریافت کند. رکوردها پس از جدا شدن بر اساس \n و تکمیل شدن، به دیکشنری پایتون تبدیل شوند؛ زیرا TCP مرز رکوردها را حفظ نمی‌کند و نیاز به مدیریت بافر دارد.

اعتبارسنجی و ذخیره سازی داده‌ها

رکوردهای سالم و خراب را از یکدیگر جدا کنید. توجه داشته باشید که یک رکورد ممکن است همزمان چند مشکل داشته باشد. رکوردهای سالم را در clean_readings.csv و رکوردهای خراب را دربad_readings.csv ذخیره کنید.

گزارش لحظه ای

هر ۲۰ ثانیه یک ردیف جدید در فایل real_time_reports.csv ثبت کنید. گزارش باید شامل موارد
تعداد کل رکوردها، تعداد رکوردهای سالم، تعداد رکوردهای خراب و تعداد ایستگاه‌های فعال باشد.
پس از انجام وظایف ذکرشده، به سوالات پاسخ دهید. برای هر سوال، راهکار خود را همراه با دلیل توضیح دهید.

سوالات

افزایش حجم داده
فرض کنید تعداد ایستگاه‌های پایش در سطح شهر به‌شدت افزایش یابد و نرخ ورودی از ۵۰ به ۵۰,۰۰۰ رکورد در ثانیه برسد. آیا پایتون و CSVپاسخگو هستند؟ گلوگاه‌ها را مشخص کنید و برای دریافت، پردازش و ذخیره‌سازی راهکار مقیاس‌پذیرتری پیشنهاد دهید.

چند جریان هم‌زمان
فرض کنید هر مدل از طریق یک اتصال سوکت مستقل داده ارسال کند. چگونه چند اتصال را همزمان مدیریت میکنید

جلوگیری از هدر رفت داده

فرض کنید هیچ رکوردی نباید در اثر قطع اتصال، کرش برنامه یا خطای ذخیره سازی از بین برود چرا که همین رکوردها ممکن است نشانه ی یک نشتی واقعی باشند. چه مکانیزمی برای صف، بافر، تأیید دریافت و بازیابی پیشنهاد می‌کنید؟

برای پاسخ چالش بوت‌کمپ از این لینک استفاده کنید.

پس از ورود به لینک چالش، خواهشمنداست سوالات، رزومه و پاسخ خود را بارگذاری کنید.
پاسخ چالش را در یک فایل فشرده zip یا rar (شامل نام و نام‌خانوادگی) قرار داده و در لینک مربوطه بارگزاری نمایید.

پاسخ‌ها به ترتیب زمان ارسال بررسی می‌شوند. با توجه به ظرفیت محدود بوت‌کمپ، پس از تکمیل ظرفیت، مهلت ارسال به پایان خواهد رسید.

لطفاً رزومه به‌روز خود را در بخش مربوط به بارگذاری رزومه در فرم ارسال کنید.
رزومه برای بررسی سوابق کاری، تجربه‌های مرتبط و میزان تناسب مسیر حرفه‌ای شما با بوت‌کمپ مهندس داده۲ استفاده می‌شود.

آکادمی همراه‌اول

مشتاق دیدار شما هستیم