

۱. تولید دیتاست فارسی با کیفیت

استفاده از ابزار `gpt4free` و مدل `gpt4.1` برای تولید دیتاست مصنوعی با کیفیت. برای مدیریت چالش هایی مثل رفتن برق بنده کد تولید دیتاست را با لپتاپ در کولب پیاده سازی کرده و سپس با گوشی به کولب متصل شده و کد را اجرا کردم البته یک پاور بانک ۳۰ هزار هم در صورت نیاز برای پشتیبانی باطری موبایل وجود داشت.

`scripts/generate_data.py`

۲. آموزش دو مدل Naive Bayes و SVM

نتایج ارزیابی دقت بالایی داشت و بعد از این دو مدل سراغ پیاده سازی مدل `bert` برای فاین تیون رفتیم ولی سیستمم نکشید و وقت زیادی هم از من برد ولی نتیجه نداد با این سیستم. "[HooshvareLab/bert-fa-base-uncased](#)"
نتایج ارزیابی را میتوان در `models/model_evaluation.csv` مشاهده کرد.

۳. API با FastAPI

`uvicorn app.main:app --reload --port 8000`

<http://localhost:8000/docs>

۴. داکرایز کردن

<http://localhost:8000/docs>

برای آماده سازی محیط `Anaconda` از شل اسکریپت `conda_install.sh` استفاده کنید و برای بالا آوردن داکر از `build_and_run.sh` استفاده کنید.

تولید دیتاست و داکرایز کردن وقت زیادی از من برد.