به نام فروهر

«پروژه سوم»

بوت‌کمپ هوش مصنوعی کوئرا

بهار ۱۴۰۵

مهلت ارسال پاسخ: جمعه ۶ شهریورماه ساعت ۱۱:۵۹ شب

زمان ارائه‌ی گروهی: دوشنبه و سه شنبه ۹ و ۱۰ شهریورماه

بخش اول: معرفی داده

در این پروژه با مجموعه‌داده‌ای واقعی از محصولات و نظرات کاربران فروشگاه اینترنتی دیجی‌کالا کار خواهید کرد. مجموعه‌داده‌ی در اختیار شما شامل بیش از یک میلیون رکورد مربوط به محصولات و بیش از شش میلیون نظر فارسی ثبت‌شده توسط کاربران است.
اطلاعات محصولات و نظرات در دو فایل جداگانه ارائه شده‌اند و می‌توان آن‌ها را با استفاده از شناسه‌ی محصول به یکدیگر مرتبط کرد.

جهت دریافت مجموعه‌داده از نسخه‌ی ثابت زیر استفاده کنید:
https://huggingface.co/datasets/RadeAI/Digikala_comments_products/tree/89c3133b169c8d3793db8834f56f32fee33d9db0

فایل‌های اصلی مجموعه‌داده عبارت‌اند از:
digikala-products.csv
digikala-comments.csv

این مجموعه‌داده در قالب خام در اختیار شما قرار می‌گیرد. بنابراین تضمینی وجود ندارد که همه‌ی مقادیر کامل، یکتا، متوازن، بدون داده‌ی پرت یا مستقیماً آماده‌ی استفاده در مدل باشند. بررسی کیفیت داده، تصمیم‌گیری درباره‌ی نحوه‌ی استفاده از مقادیر گمشده، داده‌های تکراری، مقادیر غیرعادی، پیش‌پردازش متن فارسی و هر تصمیم دیگری که برای ساخت سیستم خود لازم می‌دانید، بخشی از پروژه است. انتظار می‌رود هر تصمیمی که در این مسیر می‌گیرید قابل توضیح و مبتنی بر تحلیل داده باشد.

ستون‌های فایل محصولات به شرح زیر هستند:

id: شناسه‌ی محصول
title_fa: عنوان فارسی محصول
Rate: امتیاز ثبت‌شده برای محصول
Rate_cnt: تعداد امتیازهای ثبت‌شده برای محصول
Category1: دسته‌بندی سطح اول محصول
Category2: دسته‌بندی سطح دوم محصول
Brand: برند محصول
Price: قیمت محصول
Seller: فروشنده‌ی محصول
Is_Fake: نشانگر مربوط به غیراصل بودن کالا
min_price_last_month: کمترین قیمت ثبت‌شده در ماه گذشته
sub_category: دسته‌بندی کلی محصول

ستون‌های فایل نظرات به شرح زیر هستند:

id: شناسه‌ی نظر
title: عنوان نظر
body: متن اصلی نظر
created_at: زمان ثبت نظر

rate: امتیاز ثبت‌شده توسط کاربر
recommendation_status: وضعیت پیشنهاد خرید محصول توسط کاربر
is_buyer: مشخص می‌کند که کاربر خریدار محصول بوده است یا خیر
product_id: شناسه‌ی محصول مربوط به نظر
advantages: نکات مثبت ثبت‌شده توسط کاربر
disadvantages: نکات منفی ثبت‌شده توسط کاربر
likes: تعداد رأی مثبت به نظر
dislikes: تعداد رأی منفی به نظر
seller_title: نام فروشنده‌ی محصول در زمان ثبت نظر
seller_code: شناسه‌ی فروشنده
true_to_size_rate: نظر کاربر درباره‌ی تناسب اندازه‌ی محصول در موارد مرتبط

بخش دوم: دستیار هوشمند خرید و تحلیل محصولات

یک فروشگاه اینترنتی بزرگ با صدها هزار محصول و میلیون‌ها نظر کاربری روبه‌رو است. بررسی دستی این حجم از اطلاعات برای مشتریان و حتی تیم‌های داخلی فروشگاه دشوار است. مشتری هنگام انتخاب محصول ممکن است مجبور باشد میان تعداد زیادی گزینه جست‌وجو کند و صدها نظر را بخواند؛ از طرف دیگر مدیران دسته‌بندی نیز برای فهمیدن مشکلات پرتکرار محصولات، مقایسه‌ی برندها و شناخت بازخورد کاربران به ابزاری مقیاس‌پذیر نیاز دارند.

هدف این پروژه طراحی و پیاده‌سازی یک سیستم هوشمند مبتنی بر مدل‌های زبانی است که بتواند با استفاده از اطلاعات واقعی محصولات و نظرات واقعی کاربران، به جست‌وجو، مقایسه، تحلیل و پاسخ‌گویی درباره‌ی محصولات کمک کند.

نحوه‌ی حل مسئله بر عهده‌ی شماست. استفاده از روش‌هایی مانند Prompting، Retrieval-Augmented Generation، Embedding، مدل‌های محلی، مدل‌های Fine-tuning، API، ابزارها، Agentها یا هر معماری دیگری اجباری نیست. شما باید بر اساس نیاز مسئله، محدودیت منابع و نتایج آزمایش‌های خود تصمیم بگیرید چه معماری‌ای مناسب‌تر است. استفاده از یک روش پیچیده‌تر به خودی خود امتیاز بیشتری ندارد؛ مهم این است که بتوانید نشان دهید انتخاب شما به حل بهتر مسئله منجر شده است.

سیستم نهایی شما باید حداقل قابلیت‌های زیر را پوشش دهد.

بخش ۱) جست‌وجو و کشف محصول

کاربر باید بتواند نیاز خود را به زبان طبیعی فارسی بیان کند و سیستم بر اساس اطلاعات موجود، محصولات مناسب را پیدا و معرفی کند. درخواست کاربر ممکن است ترکیبی از ویژگی‌هایی مانند نوع محصول، بازه‌ی قیمت، برند، میزان رضایت کاربران یا سایر معیارهایی باشد که از داده قابل استخراج هستند.

برای مثال کاربر ممکن است بپرسد:
«یک کیف برای استفاده‌ی روزمره می‌خوام که خیلی گرون نباشه و خریدارها هم ازش راضی باشن.»

یا:
«چند وسیله‌ی کاربردی برای سفر معرفی کن که نظر کاربران دربارشون خوب باشه.»

انتخاب روش جست‌وجو، نحوه‌ی ترکیب اطلاعات ساخت‌یافته و متنی و چگونگی رتبه‌بندی محصولات بر عهده‌ی شماست.

بخش ۲) پرسش و پاسخ مبتنی بر نظرات کاربران

سیستم باید بتواند درباره‌ی یک محصول مشخص، بر اساس نظرات واقعی کاربران پاسخ دهد. برای نمونه:
«مردم بیشتر از چه چیزی در این محصول راضی بودند؟»
«ایرادهای پرتکرار این محصول چیست؟»
«خریداران درباره‌ی کیفیت این محصول چه گفته‌اند؟»
«آیا با توجه به تجربه‌ی کاربران ارزش خرید دارد؟»

پاسخ‌هایی که درباره‌ی تجربه و نظر کاربران داده می‌شوند باید به داده‌ی واقعی متکی باشند. سیستم شما باید بتواند شناسه‌ی نظرات یا شواهدی را که برای تولید پاسخ استفاده کرده است ارائه کند. در ارزیابی، صرفاً روان و قانع‌کننده بودن متن کافی نیست و ادعاهای سیستم باید با داده‌های موجود سازگار باشند.

بخش ۳) مقایسه‌ی محصولات

سیستم باید امکان مقایسه‌ی دو یا چند محصول را نیز فراهم کند. این مقایسه می‌تواند بر اساس قیمت، اطلاعات محصول، میزان رضایت کاربران، نقاط قوت و ضعف پرتکرار و سایر شواهد موجود در داده انجام شود.

برای نمونه:
«این دو محصول را با هم مقایسه کن. از نظر کاربران کدام انتخاب بهتری است و چرا؟»

در پاسخ‌های خود به تفاوت میان اطلاعات مستقیم موجود در داده، شواهد حاصل از نظرات کاربران و استنتاج یا پیشنهاد نهایی مدل توجه کنید. سیستم نباید اطلاعاتی را که در داده یا شواهد قابل پشتیبانی نیستند به عنوان واقعیت قطعی بیان کند.

بخش ۴) تحلیل برای مدیران فروشگاه

کاربر سیستم شما همیشه یک خریدار نیست. فرض کنید مدیر یک دسته‌ی محصول نیز می‌خواهد از همین داده برای شناخت بهتر بازار و تجربه‌ی مشتریان استفاده کند. سیستم باید بتواند حداقل چند نوع پرسش تحلیلی سطح دسته یا مجموعه‌ای از محصولات را پاسخ دهد.

برای مثال:
«پرتکرارترین شکایت کاربران در این دسته چیست؟»
«کاربران بیشتر درباره‌ی چه ویژگی‌هایی از محصولات این دسته ناراضی هستند؟»

«کدام محصولات نظر زیادی دارند اما درصد پیشنهاد خرید آن‌ها پایین است؟»
«بازخورد کاربران درباره‌ی چند برند اصلی این دسته چه تفاوتی دارد؟»

انتخاب دامنه و روش اجرای این بخش بر عهده‌ی گروه است؛ اما انتظار می‌رود خروجی سیستم فراتر از خلاصه‌سازی چند نظر محدود باشد و بتواند اطلاعات مفیدی برای تصمیم‌گیری ایجاد کند.

بخش سوم: پیش‌بینی وضعیت پیشنهاد خرید

علاوه بر سیستم اصلی، می‌خواهیم یک مسئله‌ی قابل اندازه‌گیری نیز از دل همین داده حل شود. ستون recommendation_status نشان می‌دهد که کاربر محصول را پیشنهاد کرده، پیشنهاد نکرده یا نظر مشخصی در این مورد نداشته است.

یک مدل یا روش طراحی کنید که با دریافت اطلاعات متنی یک نظر، وضعیت پیشنهاد خرید آن را در یکی از سه کلاس زیر پیش‌بینی کند:

recommended
not_recommended
no_idea

انتخاب داده‌های ورودی، نحوه‌ی پاک‌سازی و نمونه‌گیری، چگونگی تقسیم داده‌ها، مدل مورد استفاده و روش آموزش بر عهده‌ی شماست. توجه کنید که باید از نشت اطلاعات میان داده‌های آموزش، اعتبارسنجی و آزمون جلوگیری شود و انتخاب‌های شما در این بخش نیز قابل دفاع باشند.

معیار اصلی ارزیابی این بخش Macro F1 خواهد بود. علت استفاده از این معیار آن است که عملکرد مدل روی هر سه کلاس اهمیت دارد و صرفاً عملکرد خوب روی کلاس پرتعداد کافی نیست.

بخش چهارم: ارزیابی سیستم نهایی

بخش مهمی از پروژه، نحوه‌ی ارزیابی سیستم شماست. نمایش چند نمونه‌ی موفق برای ارزیابی یک سیستم مبتنی بر مدل زبانی کافی نیست. انتظار می‌رود گروه شما یک مجموعه‌ی ارزیابی مناسب طراحی کرده و عملکرد سیستم را از چند جهت بررسی کند.

حداقل موارد زیر باید در گزارش و ارائه‌ی شما بررسی شوند:

کیفیت پاسخ: آیا پاسخ به نیاز کاربر مربوط و مفید است؟
Grounding: آیا ادعاهای سیستم توسط اطلاعات محصول یا نظرات بازیابی‌شده پشتیبانی می‌شوند؟
Retrieval Quality: آیا شواهد و نظرات انتخاب‌شده واقعاً به سؤال مرتبط هستند؟
Recommendation Prediction: عملکرد مدل روی مسئله‌ی پیش‌بینی recommendation_status با معیار Macro F1.
Latency: زمان پاسخ‌گویی سیستم و نحوه‌ی تغییر آن در سناریوهای مختلف.
Cost: هزینه‌ی استفاده از APIها و تعداد Tokenهای مصرف‌شده.

Failure Analysis: نمونه‌های واقعی از شکست سیستم، علت احتمالی آن‌ها و تلاش شما برای بهبودشان.

در صورتی که از LLM-as-a-Judge یا مدل زبانی دیگری برای ارزیابی استفاده می‌کنید، باید درباره‌ی معیار ارزیابی، Prompt مورد استفاده، محدودیت‌های این روش و در صورت امکان مقایسه‌ی آن با ارزیابی انسانی توضیح دهید.

محدودیت هزینه‌ی API

برای هر گروه حداکثر ۵ دلار اعتبار API در نظر گرفته شده است. این اعتبار شامل مرحله‌ی توسعه، آزمایش و اجرای نهایی پروژه است. مدیریت این بودجه بخشی از مسئله‌ی مهندسی شماست.

استفاده از مدل‌های متن‌باز و محلی، Embeddingهای محلی و سایر ابزارهای رایگان محدودیتی از نظر این بودجه ندارد. در گزارش نهایی لازم است حداقل تعداد درخواست‌های API، مجموع هزینه‌ی مصرف‌شده و در صورت امکان تعداد Tokenهای ورودی و خروجی را گزارش کنید.

سیستمی که برای دستیابی به کیفیت مشابه هزینه‌ی کمتر یا زمان پاسخ‌گویی مناسب‌تری داشته باشد، از نظر مهندسی راه‌حل بهتری محسوب می‌شود. بنابراین از ارسال بی‌دلیل حجم بزرگی از داده به مدل‌های گران‌قیمت خودداری کنید و تصمیم‌های خود را اندازه‌گیری کنید.

نکته‌های کلی

- مجموعه‌داده به شکل خام در اختیار شما قرار گرفته است و کیفیت داده بخشی از مسئله است. هر نوع پاک‌سازی، حذف، نگه‌داشتن، نمونه‌گیری، نرمال‌سازی یا تبدیل داده باید قابل توضیح باشد.

- الزامی به استفاده از تمام رکوردهای مجموعه‌داده وجود ندارد. در صورت محدودیت منابع می‌توانید از بخشی از داده استفاده کنید؛ اما روش انتخاب این بخش نباید دلخواه و بدون استدلال باشد.
- در آزمایش‌های خود باید یک Baseline ساده داشته باشید تا بتوان نشان داد تغییرات بعدی واقعاً باعث بهبود شده‌اند.
- استفاده از RAG، Fine-tuning، LoRA، Agent، Tool Calling، Reranking، Query Rewriting یا هر تکنیک دیگری اجباری نیست. استفاده از هر کدام باید بر اساس نیاز مسئله و شواهد تجربی توجیه شود.
- در صورت استفاده از مدل‌های پیش‌آموخته یا APIها، تمام اعضای فعال گروه باید درک مناسبی از نحوه‌ی کار و نقش آن‌ها در سیستم داشته باشند.
- کدهای خود را خوانا، ماژولار و قابل اجرا بنویسید. اجرای سیستم و ارزیابی آن باید قابل بازتولید باشد.
- فایل‌های کلیدی پروژه، نحوه‌ی نصب وابستگی‌ها، نحوه‌ی اجرای سیستم و نحوه‌ی اجرای ارزیابی باید در README توضیح داده شوند.
- مهم‌ترین بخش این پروژه تحلیل و تفسیر شما از مسئله، انتخاب‌ها، نتایج و شکست‌های سیستم است. در ارائه‌ی نهایی روی روند حل مسئله و شواهد تجربی تمرکز کنید، نه توضیح خطبه‌خط کد.
- مدیریت منابع محاسباتی نیز بخشی از پروژه است. مجموعه‌داده بزرگ است و انتظار می‌رود راه‌حلی متناسب با منابع موجود طراحی کنید.

بخش امتیازی (بیشینه: ۲۵ نمره)

امتیاز اضافه برای استفاده‌ی صرف از یک تکنولوژی یا مدل پیچیده در نظر گرفته نمی‌شود. بخش امتیازی زمانی ارزش دارد که بتوانید با آزمایش نشان دهید قابلیت جدید یا روش پیشنهادی شما به بهبود قابل اندازه‌گیری یا ایجاد ارزش جدید منجر شده است.

موارد زیر نمونه‌هایی از کارهای امتیازی هستند:

- بهبود Retrieval با روش‌هایی مانند Hybrid Search یا Reranking و ارائه‌ی شواهد کمی از اثر آن. (۴ نمره)
- Fine-tuning یا LoRA روی یک مدل و نشان دادن بهبود نسبت به Baseline مناسب. (۴ نمره)
- طراحی Router یا معماری چندمدلی برای کاهش هزینه یا افزایش کیفیت. (۳ نمره)
- طراحی روش ارزیابی پیشرفته‌تر و اعتبارسنجی آن با ارزیابی انسانی. (۳ نمره)
- طراحی رابط کاربری یا داشبورد تعاملی کاربردی برای مشتری یا مدیر فروشگاه. (۳ نمره)
- استفاده‌ی مؤثر از Quantization، Caching یا سایر روش‌های بهینه‌سازی برای کاهش هزینه یا زمان پاسخ‌گویی. (۳ نمره)
- طرح یک مسئله‌ی جدید و مرتبط با داده، با تأیید منتور، و ارائه‌ی راه‌حل و ارزیابی قابل قبول. (۳ نمره)
- ارائه‌ای جذاب و منسجم با خط داستانی مناسب و نمایش روشن تصمیم‌ها، آزمایش‌ها، نتایج و Demo نهایی. (۲ نمره)

موفق باشید 🤩