

درخواست پیشنهاد (RFP) برای پروژه‌های تحلیل موضوعی (Topic Modeling)

در حوزه هوش مصنوعی

کد مسئله : SI- 6

۱. مقدمه

این پروژه به منظور توسعه یک سیستم تحلیل موضوعی طراحی شده است که قادر به شناسایی و تحلیل موضوعات موجود در مجموعه‌های داده متنی، به صورت خودکار و کارآمد باشد. هدف از این سیستم، استخراج الگوهای موضوعی از داده‌های متنی و تسهیل فرآیندهای تحلیل داده برای بهبود تصمیم‌گیری و کسب اطلاعات عمیق‌تر است.

۲. اهداف پروژه

هدف اصلی این پروژه توسعه یک سیستم تحلیل موضوعی است که بتواند به صورت دقیق و موثر موضوعات مختلف را از مجموعه‌های داده متنی، استخراج کند. اهداف فرعی عبارتند از:

- a. **شناسایی موضوعات کلیدی** : توسعه سیستمی که قادر به شناسایی و طبقه‌بندی موضوعات کلیدی در داده‌های متنی باشد.
- b. **پشتیبانی از چندین زبان** : سیستم باید توانایی تحلیل متون به زبان‌های مختلف (به‌ویژه زبان‌های مورد استفاده در بازار هدف) را داشته باشد.
- c. **قابلیت کار با داده‌های غیرساختاریافته** : سیستم باید بتواند داده‌های متنی غیرساختاریافته را تحلیل کند و موضوعات مرتبط را استخراج نماید.
- d. **ارائه بصری داده‌ها** : سیستم باید قابلیت ارائه بصری موضوعات شناسایی شده به صورت نمودار و گزارش‌های گرافیکی را داشته باشد.
- e. **یکپارچگی با سیستم‌های موجود** : سیستم باید قابلیت یکپارچگی با سیستم‌های داخلی و پایگاه‌های داده موجود در شرکت را داشته باشد.

f. قابلیت یادگیری و بهبود مستمر: سیستم باید بتواند با استفاده از داده‌های جدید و بازخورد کاربران، به مرور زمان بهبود یابد.

۳. دامنه کار

پروژه شامل مراحل زیر خواهد بود:

a. تحلیل و بررسی نیازها: تحلیل نیازهای شرکت، شناسایی موارد استفاده اصلی و تعیین مشخصات سیستم.

b. طراحی معماری سیستم: طراحی معماری سیستم تحلیل موضوعی، شامل انتخاب مدل‌های یادگیری، پردازش زبان طبیعی و تعیین منابع داده.

c. توسعه و پیاده‌سازی: توسعه الگوریتم‌ها و مدل‌های تحلیل موضوعی با استفاده از تکنیک‌های پیشرفته هوش مصنوعی.

d. آزمون و ارزیابی: ارزیابی عملکرد سیستم از نظر دقت، کارایی و توانایی تحلیل موضوعات به‌درستی.

e. آموزش و مستندسازی: آموزش تیم داخلی برای استفاده و مدیریت سیستم و ارائه مستندات فنی و کاربری.

f. پشتیبانی و بهبود مستمر: ارائه خدمات پشتیبانی و بهبود مستمر پس از تحویل سیستم.

۴. الزامات فنی

a. الگوریتم‌ها و مدل‌ها: استفاده از مدل‌های پیشرفته مانند Latent Dirichlet Allocation (LDA)، Non-Negative Matrix Factorization (NMF)، و مدل‌های مبتنی بر Transformers برای تحلیل موضوعی.

b. پایگاه داده‌های آموزشی: استفاده از مجموعه داده‌های متنی با کیفیت بالا و تنوع زبانی گسترده برای آموزش و ارزیابی مدل‌ها.

c. پلتفرم‌های توسعه: سیستم باید قابلیت اجرا بر روی سرورهای محلی و همچنین در محیط‌های ابری (Cloud) را داشته باشد.

- d. استانداردهای امنیتی: تضمین امنیت داده‌ها و حفاظت از اطلاعات حساس متنی.
- e. مقیاس‌پذیری و کارایی: سیستم باید توانایی پردازش و تحلیل مقادیر زیادی از داده‌های متنی را با کارایی بالا داشته باشد.
- f. API یکپارچه‌سازی: ارائه API های استاندارد برای یکپارچگی با سایر سیستم‌های نرم‌افزاری موجود.

۵. الزامات غیر فنی

- a. تجربه و توانمندی‌ها: ارائه‌دهندگان باید تجربه کافی در توسعه سیستم‌های تحلیل موضوعی و کار با داده‌های متنی بزرگ را داشته باشند.
- b. کیفیت و پشتیبانی: ارائه‌دهندگان باید متعهد به ارائه محصولات با کیفیت بالا و خدمات پشتیبانی پس از تحویل باشند.
- c. مدیریت پروژه: استفاده از روش‌های مدیریت پروژه استاندارد برای تضمین تحویل به‌موقع و در چارچوب بودجه.

۶. زمانبندی و تحویل‌ها

پیشنهاد می‌شود که زمانبندی پروژه به صورت زیر تنظیم شود:

- a. تحلیل و بررسی نیازها 4 هفته
- b. طراحی معماری سیستم 6 هفته
- c. توسعه و پیاده‌سازی 14 هفته
- d. آزمون و ارزیابی 4 هفته
- e. آموزش و مستندسازی 2 هفته
- f. پشتیبانی و بهبود مستمر 6 ماه پس از تحویل

۷. نحوه ارائه پیشنهاد

ارائه‌دهندگان باید پیشنهادات خود را حداکثر تا [تاریخ] ارسال نمایند. پیشنهادها باید شامل موارد زیر باشند:

- a. توضیحات شرکت و تجربه‌های مرتبط: اطلاعاتی درباره شرکت و نمونه‌های پروژه‌های مشابه انجام شده.
- b. پیشنهاد فنی: توضیحات کامل درباره روش‌ها، مدل‌ها و تکنولوژی‌هایی که قرار است به کار گرفته شوند.
- c. پیشنهاد مالی: جزئیات هزینه‌ها و بودجه مورد نیاز برای اجرای پروژه.
- d. برنامه زمانی: جدول زمانی برای اجرای هر یک از مراحل پروژه.
- e. پشتیبانی و نگهداری: توضیحات درباره خدمات پشتیبانی و نگهداری پس از تحویل سیستم.