



به نام خدا

## پروژه جایگزین امتحان درس هوش مصنوعی



تاریخ تحویل: ساعت ۲۱ روز ۶ مرداد

نیمسال دوم ۹۸-۹۹

سلام

در این پروژه با یک مسئله رگرسیون روبرو هستید. هدف از این پروژه تخمین قیمت گوشی تلفن همراه با توجه به مجموعه‌ای از آگهی‌ها است.

### مجموعه داده:

مجموعه داده‌ای که در اختیار شما قرار داده شده است زیر مجموعه‌ای کوچک از آگهی‌های دیوار است که تاریخ انتشار آنها قبل از سال ۲۰۱۷ است و آزادسازی شده است<sup>۱</sup>.

ویژگی‌های ارائه شده در این مجموعه داده به شرح زیر است:

**Brand:** برند گوشی

**City:** شهری که آگهی در آن منتشر شده است.

**Title:** عنوان آگهی را نمایش می‌دهد. هر آگهی شامل یک عنوان و یک شرح است.

**Desc:** شرح آگهی. این فیلد معمولاً طولانی‌تر از عنوان آگهی است و شامل توضیحاتی در مورد کالای مورد نظر است. توجه داشته باشید که عنوان و شرح آگهی توسط کاربر ارائه می‌شوند و اشتباهات نوشتاری زیادی از جمله اشتباهات املایی و حذف فاصله‌ها در آنها دیده می‌شود. همچنین بسیاری از آگهی‌ها به زبان محاوره ارائه شده‌اند.

**Image\_count:** کاربر ممکن است به همراه هر آگهی تعدادی تصویر از کالا را هم آپلود کرده باشد. این ویژگی تعداد این تصاویر را نشان می‌دهد.

**Created\_at:** روز هفته و ساعت انتشار آگهی را نشان می‌دهد.

**Price:** قیمت تعیین شده توسط کاربر برای کالا را نمایش می‌دهد. این فیلد عددی است و مقدار ۱- برای آن یعنی کاربر قیمتی برای کالای خود تعیین نکرده است. هدف از این پروژه تخمین این مقدار است. توجه داشته باشید که آگهی‌ها

<sup>1</sup> [https://research.cafebazaar.ir/visage/divar\\_datasets/](https://research.cafebazaar.ir/visage/divar_datasets/)

مربوط به چند سال پیش هستند و تفاوت رنج قیمت گوشی‌ها در مقایسه با این روزها طبیعی است. ولی در نظر داشته باشید که اشتباهاتی در ورود قیمت توسط کاربر هم اتفاق می‌افتد، مانند حذف یک یا چند صفر.

### روش انجام پروژه:

در انجام این پروژه انتظار داریم که شما قدم‌های اصلی لازم در یک مسئله یادگیری ماشین را انجام دهید. در مورد این قدم‌ها در کلاس صحبت کرده‌ایم و در پروژه‌های طول ترم هم با آنها روبرو شده‌اید.

انتظار داریم که در نهایت گزارشی به همراه کدهای خود آپلود کنید که در آن موارد زیر مشخص باشد:

- ۱- مراحل انجام کار
  - ۲- تصمیمات گرفته شده در هر مرحله و دلیل اتخاذ آنها
  - ۳- نتایج و تحلیل‌های انجام شده
  - ۴- ایده‌هایی برای بهبود مدل. ممکن است ایده‌هایی داشته باشید که فرصت یا امکانات کافی برای اجرای آنها در بازه تعیین شده وجود نداشته باشد. در بخش انتهایی گزارش خود می‌توانید آنها را مطرح و درباره‌شان بحث کنید.
- توجه داشته باشید که گزارش باید روال منطقی و قابل فهم داشته باشد ولی لازم نیست وقت زیادی برای مرتب کردن فرمت آن بگذارید. حتی می‌توانید توضیحاتتان را در همان نوت‌بوکی که کدها قرار دارند، ارائه دهید. (البته ارائه کدها در فرم نوت بوک الزامی نیست.)

### قواعد انجام پروژه:

- شما می‌توانید از کتابخانه‌های یادگیری ماشین موجود استفاده کنید و نیازی به پیاده‌سازی مجدد ماژول‌های موجود نیست. ولی باید برای انتخاب‌های خود دلیل کافی داشته باشید.
- جستجو در اینترنت مجاز است.
- هرگونه همفکری و مشورت با افراد دیگر غیر مجاز است. این شامل دانشجویان کلاس، افراد خارج از کلاس و ارائه پرسش در فضای مجازی است.
- هر گونه به اشتراک گذاری کدها، ایده‌ها و گزارش غیر مجاز است. پس توجه داشته باشید که دستاوردهای کارتان را در فضاهای عمومی یا هر فضایی که دیگران ممکن است به آن دسترسی داشته باشند ارائه ندهید.
- تا ساعت ۲۱ روز ۶ مرداد فرصت دارید تا کدها و گزارش خود را در محل تعیین شده در cecm آپلود کنید. اگر به هر دلیلی برای اتصال به سامانه مشکل داشتید می‌توانید خروجی کار خود را از طریق ایمیل برای من ارسال کنید. ([h.fadaei@ut.ac.ir](mailto:h.fadaei@ut.ac.ir))

## توصیه‌ها:

- زمان خود را مدیریت کنید. برای ما انجام تمام قدم‌های کار از کیفیت خروجی مهم‌تر است. پس مطمئن شوید برای همه قدم‌ها فرصت کافی اختصاص می‌دهید.
- در صورتی که نیاز به استفاده از ابزارهای پایه پردازش متن داشتید می‌توانید از کتابخانه هضم<sup>۲</sup> که برای زبان فارسی ارائه شده است استفاده کنید.
- توجه داشته باشید که در این پروژه شما با یک مسئله واقعی که تا حدی ساده‌سازی شده است روبرو هستید. پس اگر در کوتاه مدت نتایج خوبی نگرفتید ناامید نشوید و به کارتان ادامه دهید.
- حداقل سه سال از انتشار این آگهی‌ها می‌گذرد. پس از قدیمی بودن مدل گوشی‌ها و کم بودن نسبی قیمت‌ها تعجب نکنید.

## ملاحظات:

- تمامی نتایج باید در یک فایل فشرده با عنوان `AI-Final-#STDID.zip` تحویل داده شود. این فایل باید شامل موارد زیر باشد:
  - یک پوشه به نام `Code` شامل کدهای تمام قسمت‌هایی که پیاده‌سازی نموده‌اید.
  - گزارش پروژه با فرمت `PDF`
  - در صورتی که از `Notebook Jupyter` استفاده می‌کنید نیازی به ارسال جداگانه کدها و گزارش نیست و هر دو را می‌توانید در یک فایل `Notebook` ارائه دهید. حتما خروجی `html` فایل `Notebook` خود را نیز همراه فایل `Notebook` ارسال کنید
- توجه داشته باشید که علاوه بر ارسال فایل‌های پروژه، این پروژه به صورت حضوری نیز تحویل گرفته خواهد شد. بنابراین تمام بخش‌های پروژه باید قابلیت اجرای مجدد در زمان تحویل حضوری را داشته باشند. همچنین در صورت عدم حضور در تحویل حضوری نمرهای دریافت نخواهید کرد.
- اگر در طول انجام پروژه ابهامی داشتید می‌توانید از طریق ایمیل یا پیام صوتی یا متنی در تلگرام سوالتان را از من بپرسید. البته توجه داشته باشید که ممکن است من پیام شما را با تاخیر ببینم. پس در انتظار جواب کار خود را متوقف نکنید.
- همچنین در صورتی که نیاز دیدم توضیح اضافه‌ای ارائه دهم، این کار را از طریق فروم درس انجام خواهیم داد. پس لطفا توجه کافی به فروم داشته باشید.

موفق باشید

فدائی

---

<sup>2</sup> <https://www.sobhe.ir/hazm/docs/index.html>  
<https://pypi.org/project/hazm/>