

به نام خدا

مسابقه شماره 1 دوره هوش مصنوعی فناوری

بهار 1400

یک وبسایت ارائه دهنده آگهی املاک (مثل دیوار یا کلید) در شهر میلان ایتالیا قصد دارد فرآیند تایید آگهی ها را اتوماتیک کند. به این صورت که اگر قیمت پیشنهاد شده برای آگهی توسط مالک، از نرم بازار خیلی بیشتر یا کمتر بود، اجازه انتشار این آگهی را ندهد. در این مسابقه قصد داریم الگوریتم هوش مصنوعی ای طراحی کنیم که قیمت هر آگهی را تخمین بزند.

به این منظور دو دیتاست در اختیار شما قرار گرفته است که از آگهی های فروش ملک بر روی سایت های آگهی خدمات و کالا (مانند دیوار) استخراج شده اند.

دیتاست اول با نام `milan_houses_train.csv` شامل فیچر های هر خانه آگهی شده، مانند مساحت، تعداد اتاق خواب و ... و همچنین قیمت پیشنهاد داده شده است.

دیتاست دوم با نام `milan_houses_test.csv` شامل بخش دیگری از آگهی ها به همراه فیچر های مربوطه است با این تفاوت که از این فایل قیمت پیشنهاد داده شده حذف شده است.

شما در ابتدا باید با استفاده از فایل `train` به یک الگوریتم دقیق و بهینه دست یابید و زمانی که از مدل خود مطمئن شدید این الگوریتم را روی دیتای `test` اجرا کرده و نتایج را در یک فایل `csv` با دو ستون که ستون اول همان `id` هر آگهی است که در فایل `test` موجود است و ستون دوم قیمت پیشبینی شده الگوریتم شما است را برای بررسی و رده بنده مسابقه ارسال کنید.

معیار ارزیابی:

برای ارزیابی الگوریتم ها ما از معیار `RMSE` با فرمول زیر استفاده میکنیم:

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_{true}^i - y_{pred}^i)^2}$$

نکته شماره 1: شما مختارید برای رسیدن به بهترین الگوریتم از تمامی روش هایی که در کلاس ذکر شده و یا نشده است استفاده کنید. بنابراین توصیه میشود با جستجو در اینترنت روش های متفاوت را یاد بگیرید و در کار خود بکار بگیرید.

نکته شماره 2: توصیه میشود به مرحله تمیز کردن داده ها و فیچر ها توجه ویژه ای کنید.

نکته شماره 3: شما مختارید در طول برگزاری مسابقه سوالات خود را از همکلاسی ها یا مدرس کلاس بپرسید. (توصیه میشود این سوالات در گروه عمومی اسلک مطرح شود تا هم دیگر همکلاسی ها از بحث ها استفاده کنند و هم امتیاز شما برای مشارکت عمومی در کلاس بالا رود).

نکته شماره 4: با در نظر گرفتن نکته شماره 3 لطفا این پروژه را به صورت شخصی انجام دهید و اخلاق حرفه ای و آکادمیک را در استفاده از کار دیگران در نظر داشته باشید.

نکته شماره 5: به برنده این مسابقه در پایان زمان برگزاری جایزه ای تقدیم خواهد شد.

نکته شماره 6: برنده مسابقه در اولین جلسه پس از مسابقه ارائه کوتاهی از روش های به کار برده شده ارائه خواهد کرد.

نکته شماره 7: ارزیابی الگوریتم ها و ارائه جدول رده بندی در دو نوبت یکی روز جمعه و دیگری در روز دوشنبه هفته آینده انجام خواهد شد، بنابراین توصیه میشود تا روز جمعه به یک راه حل اولیه ای برسید و جواب را ارسال کنید تا جایگاه خود را در جدول بدانید و تا روز دوشنبه هفته بعد روش خود را بهبود ببخشید.