

## مساله: قیمت خانه

در دیتاستی که ضمیمه شده، تعدادی ویژگی که احتمالاً بر قیمت نهایی خانه تاثیرگذار است لیست شده است. دو فایل به نام‌های `data` و `to_submit` قرار داده شده است. شما باید به کمک محتویات `data`، مدل را آموزش داده و سپس از محتویات `to_submit` برای ارسال خروجی مدل خود استفاده کنید. برای خروجی یک فایل دوستونی به فرمت `csv` را آپلود نمایید که ستون اول، `id` خانه‌هایی که در فایل `to_submit` آمده (به همان ترتیب) و ستون دوم، قیمت پیشنهادی مدل شماست.

پیشنهاد می‌شود مدل‌های مختلفی را پیاده‌سازی کنید و هر یک از این مدل‌ها را بصورت یک نوت بوک جداگانه به همراه توضیحات روش انتخابی و توجیه نتایج آن، آماده سازید و تحویل دهید.

مراحل:

### 1) پیش‌پردازش

- a. مدیریت کردن داده‌های پرت و مقادیر گمشده (`missing`)
- b. مدیریت داده‌های ترتیبی و غیرعددی (`categorical`)
- c. نرمال‌سازی و اسکیل کردن
- d. برخی تحلیل‌های آماری روی متغیرها
- i. هیستوگرام
- ii. چولگی (و مثلاً استدلال برای تشخیص داده‌های پرت)
- iii. واریانس روی ستون‌ها (یافتن و تصمیم‌گیری برای ستون‌های با واریانس خیلی کم)

### 2) تقسیم‌بندی داده‌ها و ساخت مدل

- a. می‌توانید نسبت‌های مختلفی را برای تقسیم داده‌ها استفاده کنید.
- b. مدل‌هایی که انتخاب می‌کنید حتماً توجیه داشته باشند.

### 3) تولید خروجی به کمک هر مدل از داده‌های فایل `to_submit`

- a. می‌توانید خروجی مدل‌های مختلف خود را در یک فایل نیز ارسال کنید که یک ستون برای `id` و برای هر مدل یک ستون برای قیمت خانه قرار داشته باشد.

## برخی نکات:

- ✓ استفاده از کتابخانه‌های پایتون آزاد است.
- ✓ حتما بخشی از نتایج و دستاوردهای خود را بصورت نمودارهای مختلف نمایش دهید.
- ابتدا ویژگی‌های همبسته را بیابید و سپس چندتا از ویژگی‌های با همبستگی بالا با قیمت خانه را بصری نمایید.
- سوالاتی مانند موارد زیر را بصورت گرافیکی و متنی پاسخ دهید.
  - آیا کیفیت ساخت بر قیمت تاثیر دارد؟
  - آیا اندازه (تعداد) اتاق با قیمت ارتباطی دارد؟
  - آیا فاصله از ریل راه‌آهن تاثیری بر قیمت دارد؟
  - ...
- خروجی مدل‌های مختلف را بصورت بصری با هم مقایسه کنید.
- ✓ انتخاب ویژگی را بصورت جدی لحاظ کنید و تاثیر آن را مورد بررسی قرار دهید.
- خروجی مدل پیش و پس از انتخاب ویژگی را بررسی کنید.
- ✓ الگوریتم‌های خود را از نظر زمان محاسبه (روی ماشین یکسان) نیز با هم مقایسه کنید.

## توجه:

- \*\* باید پروژه خود را ارائه دهید که زمان آن پس از پایان مهلت ارسال جواب‌ها اطلاع رسانی خواهد شد.
- \*\* از زمان آپلود فایل شما یکماه فرصت دارید روی این موضوع کار کنید.
- \*\* لااقل دو مدل مجزا را برای مساله پیاده‌سازی نمایید تا نتایج قابل مقایسه باشند. مدل‌ها و تحلیل‌های بیشتر نمره اضافی در پی خواهد داشت.
- \*\* گزارش شما در نوت بوک باید به زبان فارسی و کامل باشد.
- \*\* از تحویل جواب‌های مشابه و یا پاسخ‌هایی که قادر به توضیح یا تغییر در آن نیستید پرهیز کنید.