

سیستم هوشمند خدمات مشاوره‌ای خرید خانه



تعریف مسئله:

پروژه اول مربوط به مجموعه داده قیمت نهایی فروش منازل مسکونی می‌باشد. این مجموعه داده یکی از مسابقه‌های وبسایت Kaggle می‌باشد که در حال حاضر در جریان است (<https://www.kaggle.com/c/house-prices-advanced-regression-techniques>). هدف از اجرای این پروژه ارائه مدلی جهت پیش‌بینی هرچه بهتر قیمت فروش منازل مسکونی می‌باشد. در اختیار داشتن چنین مدل‌هایی می‌تواند به عنوان معیاری در فرآیند قیمت گذاری، برآورد سود پروژه‌های ساختمانی، تعیین استراتژی معاملاتی مشاورین مسکن و ارائه خدمات مشاوره ای به صورت خودکار مفید واقع گردد. این پروژه دارای دو قسمت اصلی می‌باشد که قسمت اول شامل سه مرحله است:

قسمت اول (۴۰٪ نمره)

ارائه مدل پیش‌بینی

مرحله اول: بررسی متغیرها، مطالعه مقادیر گمشده و ارائه راهکاری جهت حل مشکل مقادیر گمشده (توجه در صورت حذف نقاط).
مرحله دوم: ارائه مدلی جهت پیش‌بینی قیمت. در این مرحله حداقل دو روش متفاوت جهت پیش‌بینی باید مورد استفاده قرار گیرد. در این مرحله می‌توان متغیرهای جدیدی را از میان متغیرهای موجود استخراج کرده و در مدل استفاده نمود (این گام که feature engineering نام دارد کاملاً اختیاری می‌باشد)

مرحله سوم: تعریف یک شاخص مناسب و مقایسه روش‌های ارائه شده در مرحله دوم.

مدل‌های مورد نظر خود را با استفاده از داده‌های **train_data** برازش نمایید. از بهترین مدل خود جهت پیش‌بینی داده‌های **test_data** استفاده نموده و مقادیر پیش‌بینی شده را در یک ستون جدید با نام **price prediction** اضافه کنید. فایل **test_data** را به همراه گزارش پایانی خود تحویل دهید.

قسمت دوم (۶۰٪ نمره)

برای این قسمت در صورت نیاز می‌توان از مدل برازش شده در قسمت قبل استفاده نمود و نیاز به برازش هیچ مدل جدیدی وجود نخواهد داشت)

ارائه سیستم خدماتی

به عنوان کسی که هم اکنون قسمتی از یک پایگاه داده با مجموعه اطلاعات مربوط به منازل مسکونی را در اختیار دارید، اگر بخواهید وبسایتی طراحی نمایید که در آن شخص با در اختیار داشتن میزان مشخصی از سرمایه، متراژ، منطقه مورد انتخاب، تعداد اتاق خواب مورد نظر، قدمت منزل، کاربری زمین (تجاری، مسکونی، کشاورزی...) و تعداد طبقات مشخص بتواند بهترین گزینه‌های ممکن را مورد بررسی قرار دهد، چه روشی جهت مرتب کردن و نمایش گزینه‌ها انتخاب می‌نمایید؟

در این جا ممکن است دو حالت رخ دهد. در حالت اول شما بدون توجه به مدل برازش شده در قسمت قبلی سوال، باید به گونه ای بتوانید بهترین گزینه‌ها را برای مشتری انتخاب نموده و آن‌ها را نمایش دهید. تفسیر معنا و مفهوم بهترین بر عهده شما می‌باشد. پس در این دیدگاه ما به مدل برازش شده هیچ توجهی نخواهیم داشت و تنها نگاهمان را به خواسته مشتری معطوف می‌نماییم.

روش خود را جهت یافتن ۳ عدد از مناسب ترین گزینه‌ها برای یک مشتری با معیارهای زیر بکار گیرید:

سرمایه: ۱۹۸۷۰۰

متراژ: حدود ۹۶۰

منطقه مورد انتخاب: بی اهمیت

تعداد اتاق خواب: ۴

قدمت منزل: حداکثر ۲۰ سال

کاربری زمین: کشاورزی نباشد

تعداد طبقات: ۲

پرواضح است که مشخصه‌های مد نظر مشتری بسیار کمتر از متغیرهای مسئله می‌باشد. پیدا کردن راه حل برای این مشکل بر عهده شما می‌باشد.

در حالت دوم، ما با مسئله ای مواجه هستیم که مشتری دیدگاهی غیرواقعی نسبت به قیمت منزل مسکونی مورد نظر دارد. در این حالت می توان از مدل برازش شده در قسمت قبل استفاده نموده و قیمت منزل پیش بینی شود. در صورتی که میزان سرمایه فرد با قیمت پیش بینی شده (با توجه به معیارهای مورد نظر) متفاوت باشد، میزان سرمایه واقعی مورد نیاز باید به صورت خودکار پیشنهاد داده شود. می توان تفاوت های بالای ۵۰۰۰ هزار دلار را به عنوان تفاوت های غیرقابل چشم پوشی در نظر گرفت. به فرض مثال اگر هزینه برآورد شده با توجه به معیارهای منتخب مشتری برابر با ۲۰۶۰۰۰ هزار دلار بوده و فرد قیمت ۱۹۸۰۰۰ هزار دلار را انتخاب نموده باشد، مقدار واقعی سرمایه مورد نیاز به عنوان پیشنهاد باید به مشتری عرضه شود.

برای مشتری فوق، آیا نیازی به نمایش مقدار واقعی سرمایه مورد نیاز وجود دارد؟ تخمین شما از قیمت منزلی با شرایط ذکر شده چیست؟

اگر فردی با در اختیار داشتن سرمایه مشخصی نتواند منزل مورد نظر خود را بر روی وبسایت شما بیابد، چگونه می توانید بدون درخواست افزایش سرمایه و با تغییر حداقلی در معیارهای مورد نظر مشتری بهترین گزینه ها را برای او نمایش دهید؟

شیوه گزارش نویسی:

کلیه گزارش ها باید دارای شش قسمت چکیده (یک پاراگراف)، مقدمه و تعریف مسئله (حداکثر یک صفحه)، رویکرد، نتایج، نتیجه گیری و ضمیمه (کدهای استفاده شده)، باشند. رعایت نکات زیر در نگارش گزارش می تواند مفید واقع گردد:

- زبان مورد استفاده می تواند فارسی یا انگلیسی باشد.
- هیچ گونه حداقلی بر تعداد صفحات، تعداد نمودارها و تعداد جداول وجود نداشته و تنها محتوای علمی و منطقی گزارش مورد بررسی قرار خواهد گرفت.
- خلاقیت در بکارگیری روش ها، نمودارها و تفسیر و تحلیل نتایج ملاک های اصلی نمره دهی محسوب خواهند شد.
- نیازی به تشریح تفصیلی روش های استفاده شده (قسمت رویکرد) نمی باشد. تنها شرح مختصری از روش به همراه ذکر دلیل استفاده از آن کفایت می کند.
- نکته اساسی در نگارش گزارش ها بیان مختصر و مفید تمامی مراحل کار می باشد.
- متن گزارش باید پیوسته و دارای انسجام باشد. رعایت توالی منطقی در گزارش نویسی از معیارهای اساسی نگارش می باشد.
- تمامی نمودارها و جداول باید دارای شماره و عنوان مختصر بوده و هیچ نمودار یا جدولی بدون تحلیل و بررسی در گزارش قرار داده نشود.
- محورهای افقی و عمودی نمودارها باید نام گذاری شده باشند.
- در صورتی که روش مورد استفاده شما مستلزم بررسی پاره ای از مفروضات می باشد، این مرحله در قسمت نتایج ارائه گردد.
- در صورتی که تکنیک cross-validation جهت تنظیم پارامترها مورد استفاده قرار گیرد، جزئیات خروجی های این فرآیند در قسمت نتایج تشریح گردد.