

بخش اول:

تصور کنید در شرکتی کار می‌کنید که می‌خواهد به مشتریان خود در زمینه‌ی سرمایه‌گذاری املاک مشاوره دهد. شرکت شما قصد دارد از داده‌های موجود در بازار املاک برای ساخت یک مدل پیش‌بینی‌کننده استفاده کند که قیمت خانه‌ها را بر اساس ویژگی‌های مختلف مانند موقعیت مکانی، سال ساخت، متراژ و غیره پیش‌بینی کند. به عنوان بخشی از تیم داده، شما و همکارانتان مسئول جمع‌آوری داده‌ها، پاک‌سازی آنها و ساخت مدل‌های پیش‌بینی هستید تا به بهترین شکل ممکن به مشتریان کمک کنید.

یک فایل CSV در اختیار شما قرار داده شده است که شامل داده‌های مورد نیاز برای ساخت مدل‌های مورد نظر است. همچنین یک فایل Jupyter Notebook در به شما داده شده است. این فایل شامل مراحل مورد نیاز برای انجام کامل پروژه است. شما باید بخش‌های مختلف این فایل را تکمیل کنید.

مدلهایی که باید آموزش دهید به ترتیب:

1. Linear Regression
2. Polynomial Regression
3. Ridge Regression
4. Lasso Regression
5. Elastic Net Regression
6. XGBoost Regression

بخش مربوطه به هر کدام از مدل‌های زیر در داخل فایل Notebook مشخص شده است. همچنین باید میزان دقت هرکدام از مدل‌های بالا بر روی داده‌های مورد نظر را محاسبه کنید.

درباره روند کار هرکدام از مدل‌های بالا تحقیق کنید و در فایل گزارش خود هرکدام را توضیح دهید. همچنین نتایج حاصل از هر مدل را بررسی کنید و بیان کنید که هر مدل چرا چنین نتیجه‌ای را خروجی می‌دهد و نتایج حاصله را مقایسه کنید.

بخش دوم:

تیم داده شما پس از موفقیت در پیش‌بینی قیمت املاک، حالا قصد دارد با استفاده از داده‌های خرید مشتریان در فروشگاه برای استخراج قواعد انجمنی، الگوها و روابط میان محصولات را بررسی کند.

هدف این بخش، کشف الگوهای خرید مشتریان و تحلیل ارتباطات بین محصولات است. با این اطلاعات، ما می‌توانیم به فروشگاه توصیه‌هایی ارائه دهیم که به بهبود استراتژی‌های فروش و افزایش درآمد کمک کند.

در زیر مراحل انجام این بخش گفته شده است که باید به ترتیب در Notebook داده‌شده تکمیل شوند.

۱. بارگیری و پیش‌پردازش داده‌ها:

- بارگیری داده‌های خرید مشتریان از فروشگاه به صورت یک فایل CSV.
- تبدیل داده‌های ورودی به فرمتی که برای استخراج قواعد انجمنی مناسب باشد.
- حذف داده‌های تکراری یا ناقص و تمیز کردن داده‌ها برای آماده‌سازی آنها برای استفاده در الگوریتم‌های استخراج قواعد انجمنی.

۲. استخراج مجموعه‌های پرتکرار:

- استفاده از الگوریتم Apriori یا FP-Growth برای شناسایی مجموعه‌های پرتکرار از محصولات مختلف.

- تعیین آستانه حداقل حمایت برای شناسایی مجموعه‌های پرتکرار.

۳. تولید قواعد انجمنی:

- استفاده از مجموعه‌های پرتکرار برای تولید قواعد انجمنی با استفاده از الگوریتم Apriori یا FP-Growth.

- تنظیم آستانه حداقل اطمینان برای انتخاب قواعد با قدرت پیش‌بینی مناسب.

در انتها در فایل گزارش خود نتایج استخراج قواعد انجمنی به همراه توضیحات مرحله به مرحله فرایند را به صورت کامل قرار دهید.