

لازم نیست همه‌ی ۳۴ ویژگی رو وارد مدل فازی کنید تا از پیچیدگی مدل و قوانین پرهیز کنید، در صورتی که ویژگی‌های دیتاست را بررسی و ویژگی‌های مهم را اکتشاف کنید. در نتیجه:

باید از انتخاب ویژگی (**Feature Selection**) استفاده کنید و فقط مهم‌ترین ورودی‌ها رو نگه دارید.

(نکته: حداقل ۱۰ ویژگی را برای مدل سازی استفاده کنید).

یکی از این راهکارها استفاده از الگوریتم ژنتیک که در تمرین قبل یاد گرفتید هست.

یکی دیگر از راهکارهای ساده تحلیل همبستگی (**Correlation Analysis**) می‌باشد. ویژگی‌هایی که با متغیر خروجی (Target) همبستگی بالا دارن رو نگه و ویژگی‌های مشابه با هم (multicollinear) رو حذف کنید.

راهکارهای دیگر:

تحلیل اطلاعاتی (**Information Gain, Mutual Information**): از کتابخانه‌هایی مثل

`sklearn.feature_selection.mutual_info_classif` استفاده کنید تا ببینید هر ویژگی چقدر اطلاعات به پیش‌بینی خروجی اضافه می‌کند.

روش‌های آماری یا مدل پایه‌ای (مثلاً درخت تصمیم): یک مدل ساده مثل درخت تصمیم اجرا کنید و اهمیت ویژگی‌ها رو ببینید و فقط ویژگی‌های با اهمیت بالا رو نگه دارید.