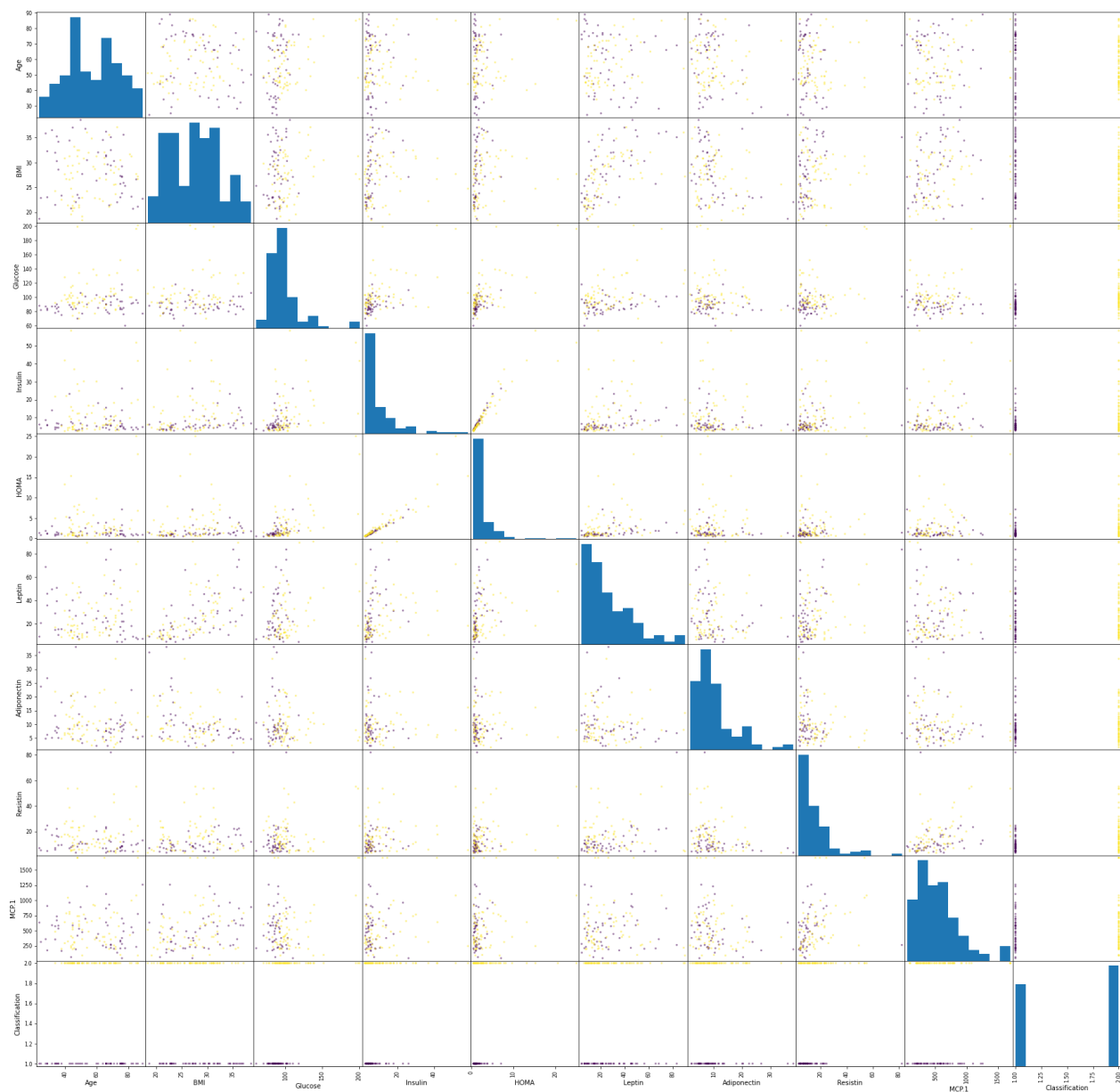


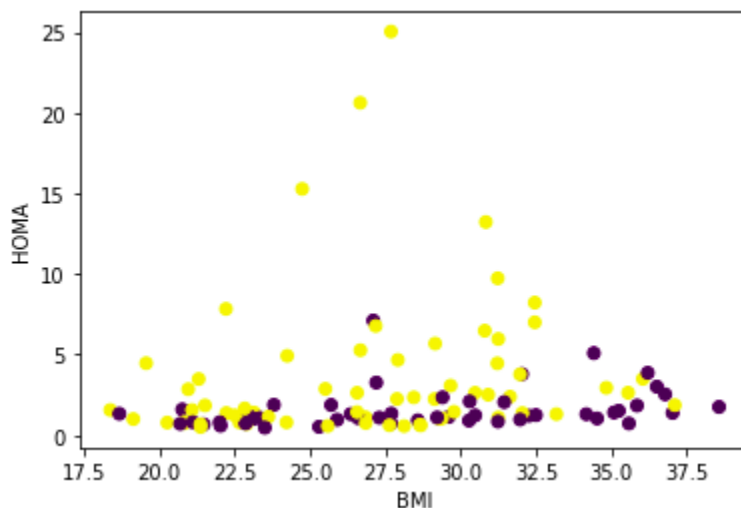


قسمت اول: تحلیل نمودار دادگان

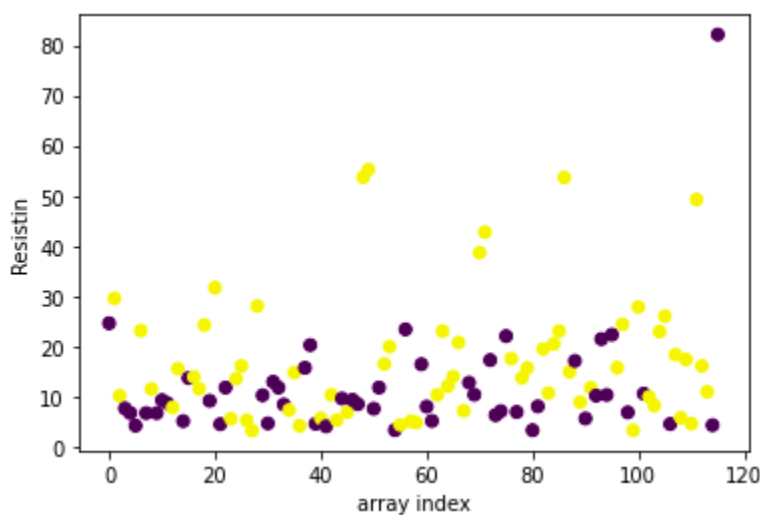
در این قسمت نکته نهایی و خلاصه شده این است که داده های ما فقط با دو یا یک ویژگی و feature تفکیک پذیر نیستند. و با دیدن نمودار های زیر کامل واضح است:



یک نمونه دو فیچر: که جدا کننده نیست و هیچکدام از نمودار های دو فیچر ای جدا کننده نبودند.



یک نمونه تک فیچر: که خط افقی جدا کننده نداشت و من با مقایسه ی نمودار های تک فیچر هم به رابطه جداکنندگی خاصی بین نمودار ها پی نبردم.



با افزایش بعد برای نمونه از تک فیچر به دو فیچر باید تفکیک پذیری ما بیشتر شود که در اینجا به مقدار کم این اتفاق می افتد و جدا پذیری با افزایش بعد افزایش می یابد.
و در کل مهم ترین کاربرد این نمودار ها تشخیص تفکیک پذیر بودن داده ها به صورت شهودی میباشد.

قسمت اول: کد

در این قسمت به دلیل سادگی کد ها خیلی نیاز به توضیح ندارند.
در کل به نظرم با توجه به نوع توزیع داده ها بهترین راه حل **logistic regression** یا **SVM linear** بوده و **Decision Tree** هم بد نبود
اما مدل **SVM** با کرنل **RBF** و خود **KNN** مدل های خوبی برای این توزیع نبودند و عملکرد **KNN** سخت
بودن توزیع داده برای کلاستر کردن را نشان میدهد و عملکرد بد **RBF** به نظر به دلیل انتخاب تابعی نامناسب
برای تغییر فضا است.
اگر بخواهیم دقتی بالاتر از این ها دریافت کنیم باید از شبکه های عصبی **MLP** استفاده میکنیم.

فقط برای فهم روند کد زنی
برای استفاده از هر مدل چهار قسمت کد درون فایل **ipynb** موجود است
اولی : فقط یک کد ساده و بدون **k-fold**
دومی : **k-fold** با استفاده از **k=5** و استفاده از **cross_val_score** برای **k-fold**
سومی : **k-fold** با استفاده از **k=5** و با فقط با استفاده از **KFold**
چهارمی : یافتن بهترین **k** برای **k-fold-cross-validation**
تنها مدلی که با **k-fold** در آن به **overfit** ختم شد **Decision Tree** بود.

در قسمت **KNN** هم هدف ما مقایسه ی **k** های مختلف برای **KNN** بوده است و به همین دلیل خروجی ما یک نمودار از دقت است.