1. فیچر سلکشن

برای قسمت اول ابتدا accuracy هر فیچر را با الگوریتم طبقه بندی ID3 محاسبه میکنیم که در اینجا با الگوریتم K-fold برای 5 حالت accuracy را محاسبه میکنیم سپس میانگین این5 مقدار را برای هر فیچر ست میکنیم.

در مرحله بعد ماتریس correlation را به دست می اوریم و برای همه مقادیر قدر مطلق را حساب میکنیم.

سپس میانگین correlation را برای هر فیچر حساب میکنیم :

در اینجا با توجه به نحوه پیاده سازی ، برای به دست اوردن f1، من میانگین را به جای max برای هر فیچر حساب کردم.

حالا دو متریک accuracy , mean\_correlation برای هر فیچر داریم که من 17 فیچر در نظر گرفته ام.

برای به دست اوردن f1 ابتدا correlation را عکس میکنیم تا بیشترین مقدار بهترین باشد مانند accuracy و برای هر فیچر مقدار f1 را حساب میکنیم.

حالا این مقادیر را سورت میکنیم و بر اساس این مقدار و تعداد فیچری که میخواهیم، دیتا اصلی را برمیگردانیم.

2.خوشه بندی

در مرحله بعد با استفاده از الگوریتم kmeans داده ها را خوشه بندی میکنیم و هر خوشه را با استفاده از یک classifier آموزش می دهیم و سپس یک حلقه بر روی دیتا تست میزنیم و predict\_label هر داده تست را به دست می اوریم و در نهایت دقت ان را محاسبه میکنیم که البته برای این دیتا به دلیل اینکه دیتا تمیزی هست به طور واضح نمیتوانیم بهتر شدن دقت را ببینیم و در چند حالت خاص با تغییر تعداد کلاستر ها و فیچر ها این اتفاق رخ می دهد.

3. مسائل مختلف

در مرحله اخر هم برای یک حالت سه کلاسه(AB-CD-E) و سه حالت دو کلاسه

(AB-CD , AB-E, CD-E) دقت را اندازه گیری میکنیم.

برای این کار دیتایی که همه فیچر ها را دارد را به different\_issue پاس میدهیم و در انجا بعد از تغییر مسئله feature\_selection را صدا میزنیم و دلیلش این است که بعد از تغییر مسئله دقت و f1 برای هر فیچر تغییر می کند پس دوباره باید فییچر سلکشن انجام شود.