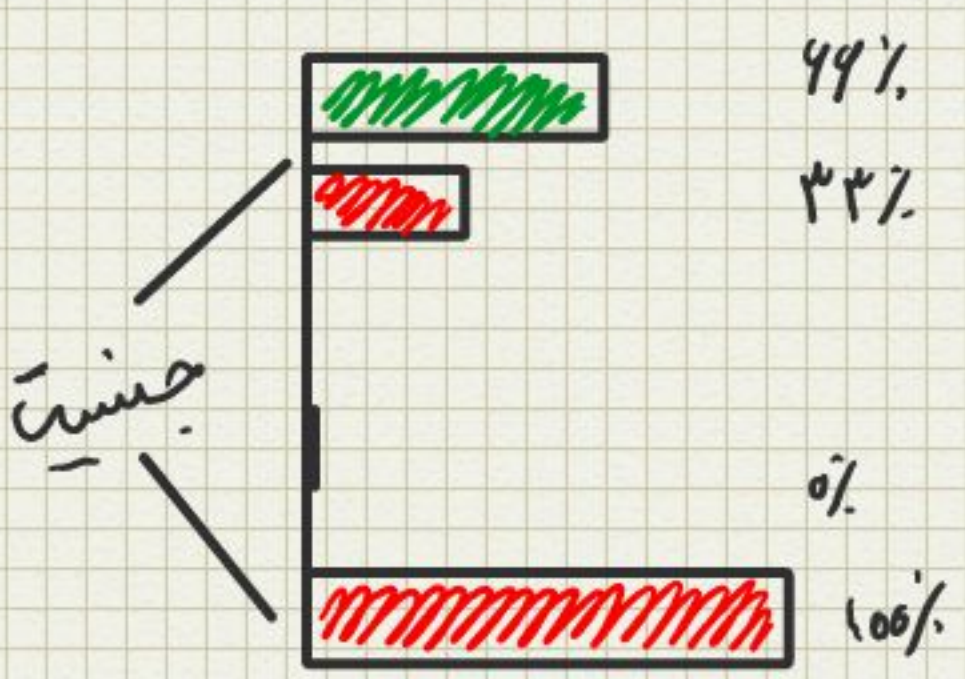
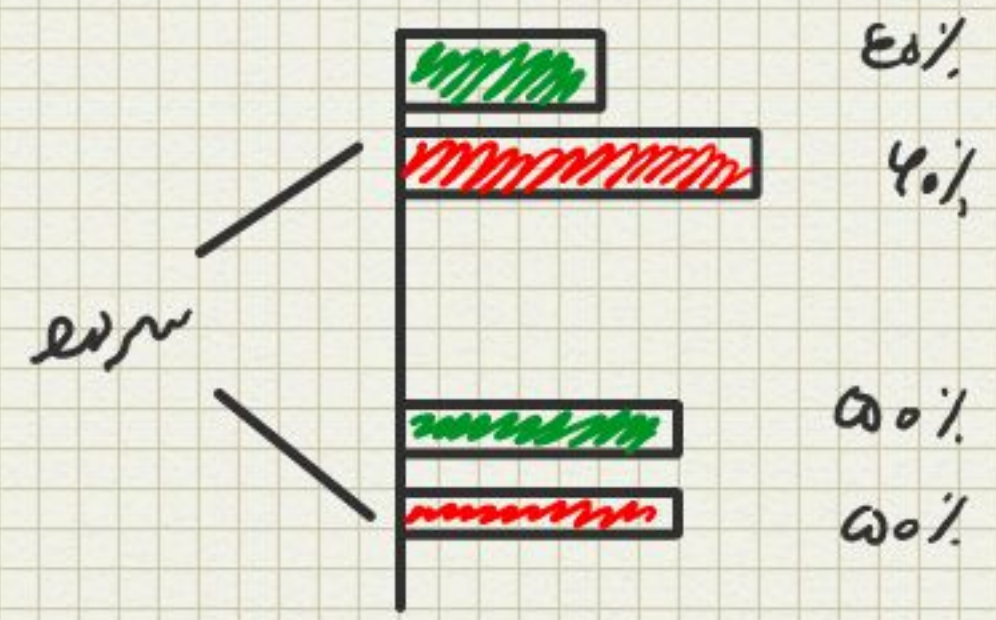
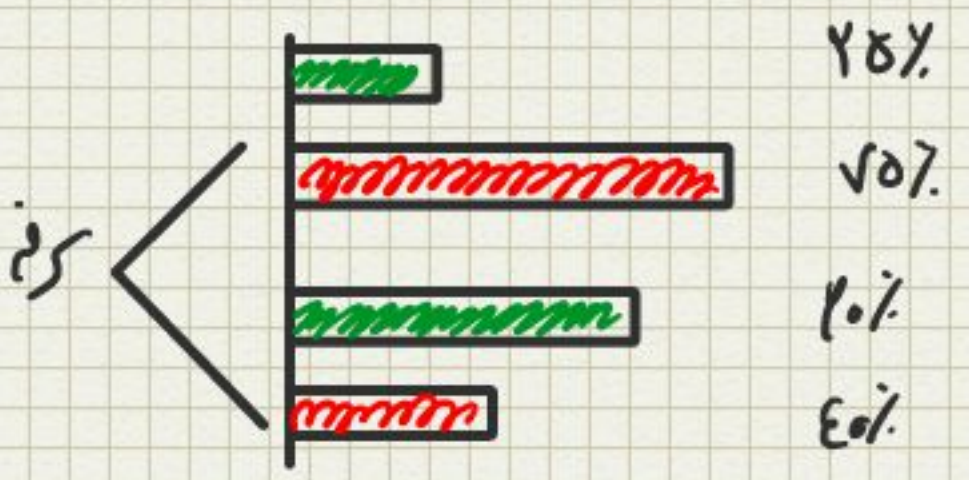
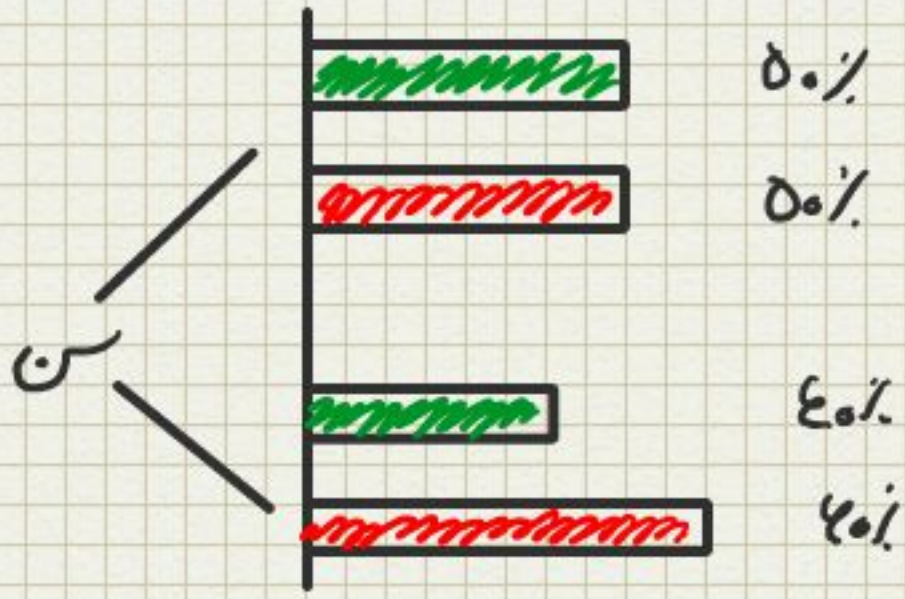
محمدرضا شمس اشکذری	95100156	دوم اسفند 1396

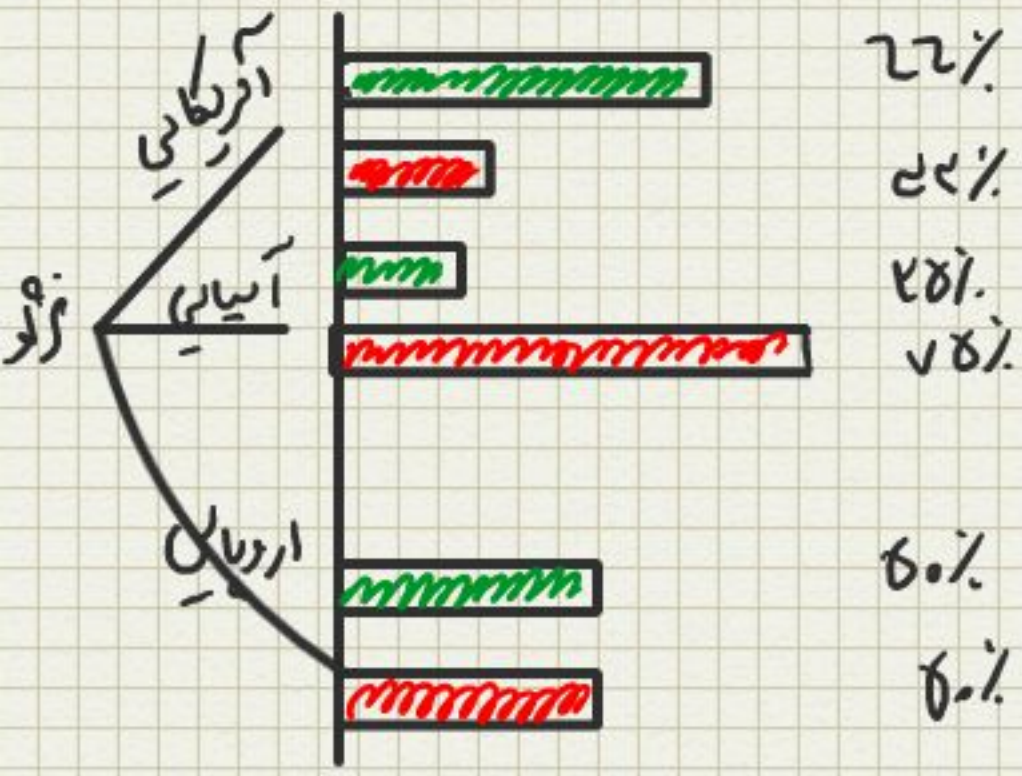
سوال یک) برای هر کدام از ویژگی ها، نمودار جدا کننده را رسم می کنیم تا بهترین آنها را برای شروع ترسیم درخت تصمیم گیری بیابیم.

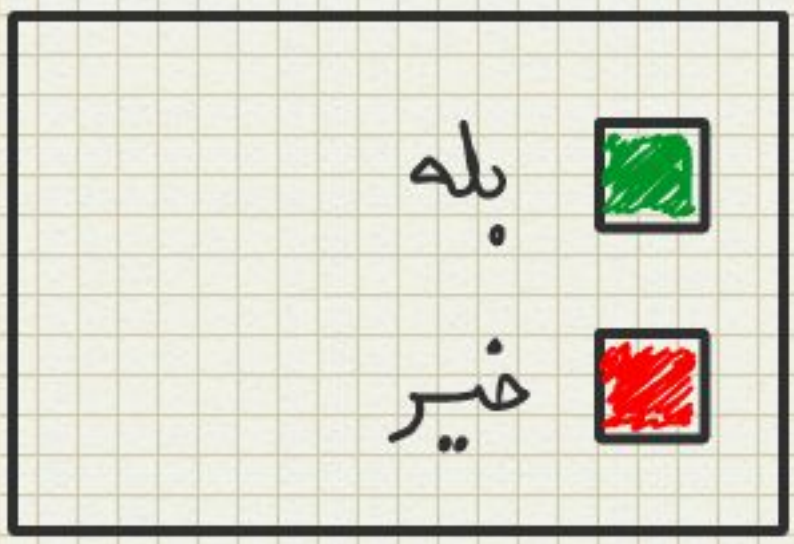




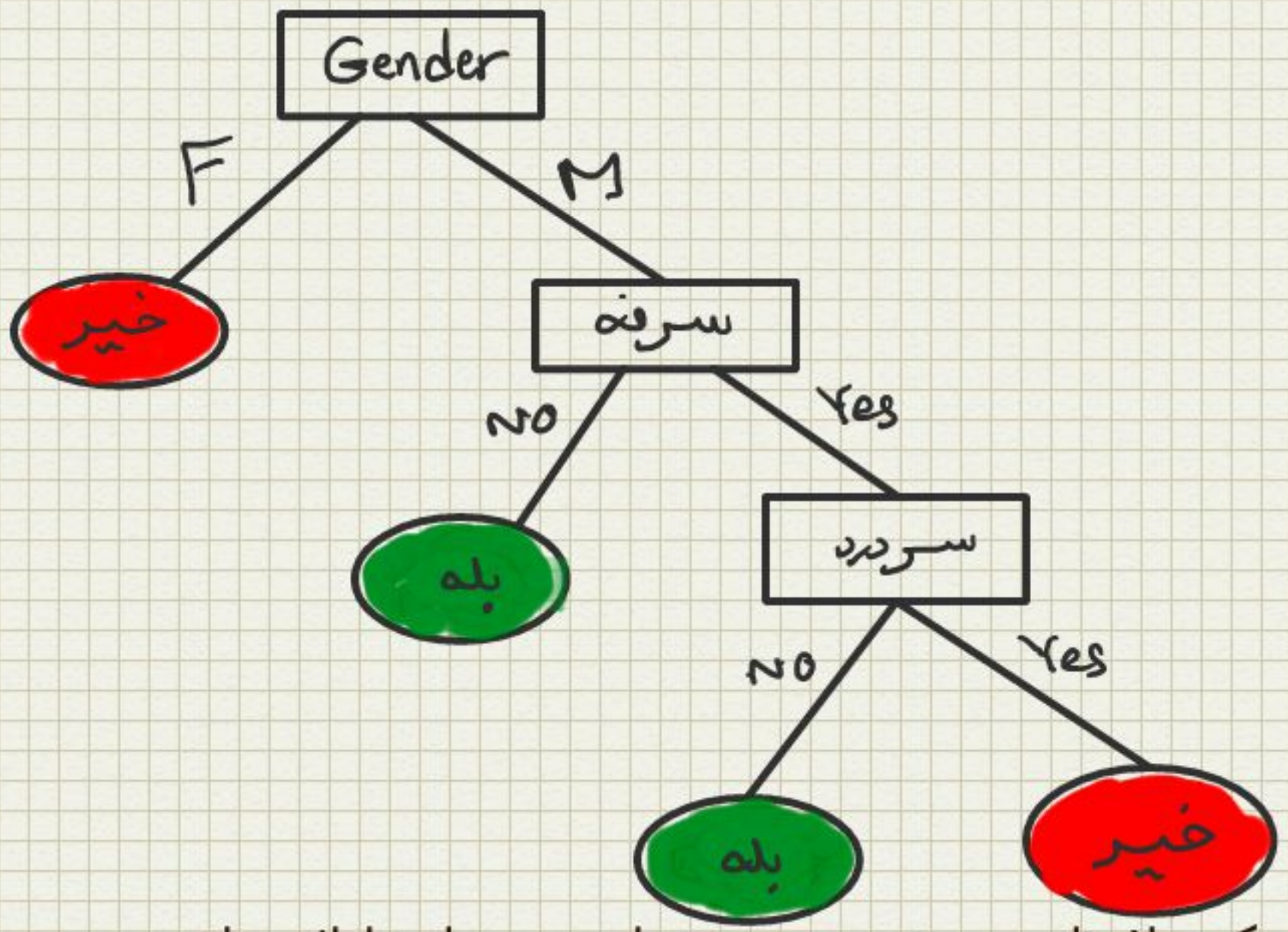








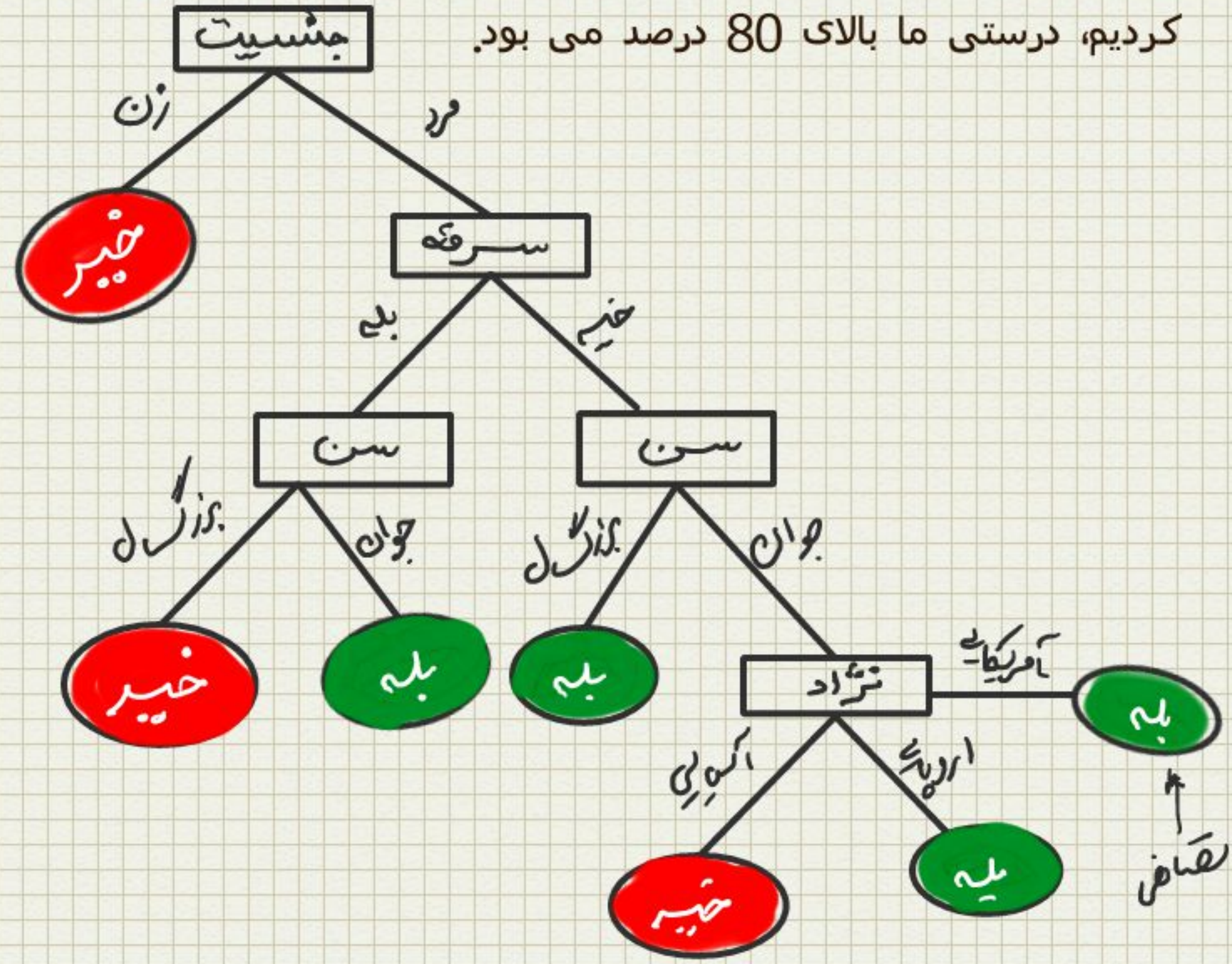
با توجه به نمودارهایی که رسم شد، به راحتی می فهمیم که به‌ترین ویژگی در ابتدای کار باید استفاده کنیم، جنسیت است. لذا درخت ما با توجه به داده ها اگر زنی آزمایش دهد، بیمار بودنش را رد می کند. در ادامه هم در هر مرحله برای داده های باقیمانده باز ویژگی بهتر را انتخاب می کنیم و به همین شیوه پیش می رویم تا نهایتا درخت ما به شکل زیر در می آید:



همان طور که ملاحظه می شود، سن و نژاد در جدول ما اثر نداشتند و ما آنها را اضافه نکردیم. البته به جای سرفه می توانستیم از سن استفاده کنیم ولی خب همین کافی است. یعنی با در نظر گرفتن آنها شانسی بهتر از پنجاه پنجاه نخواهیم داشت و نهایتا دقتمان روی تست به شکل زیر خواهد بود:

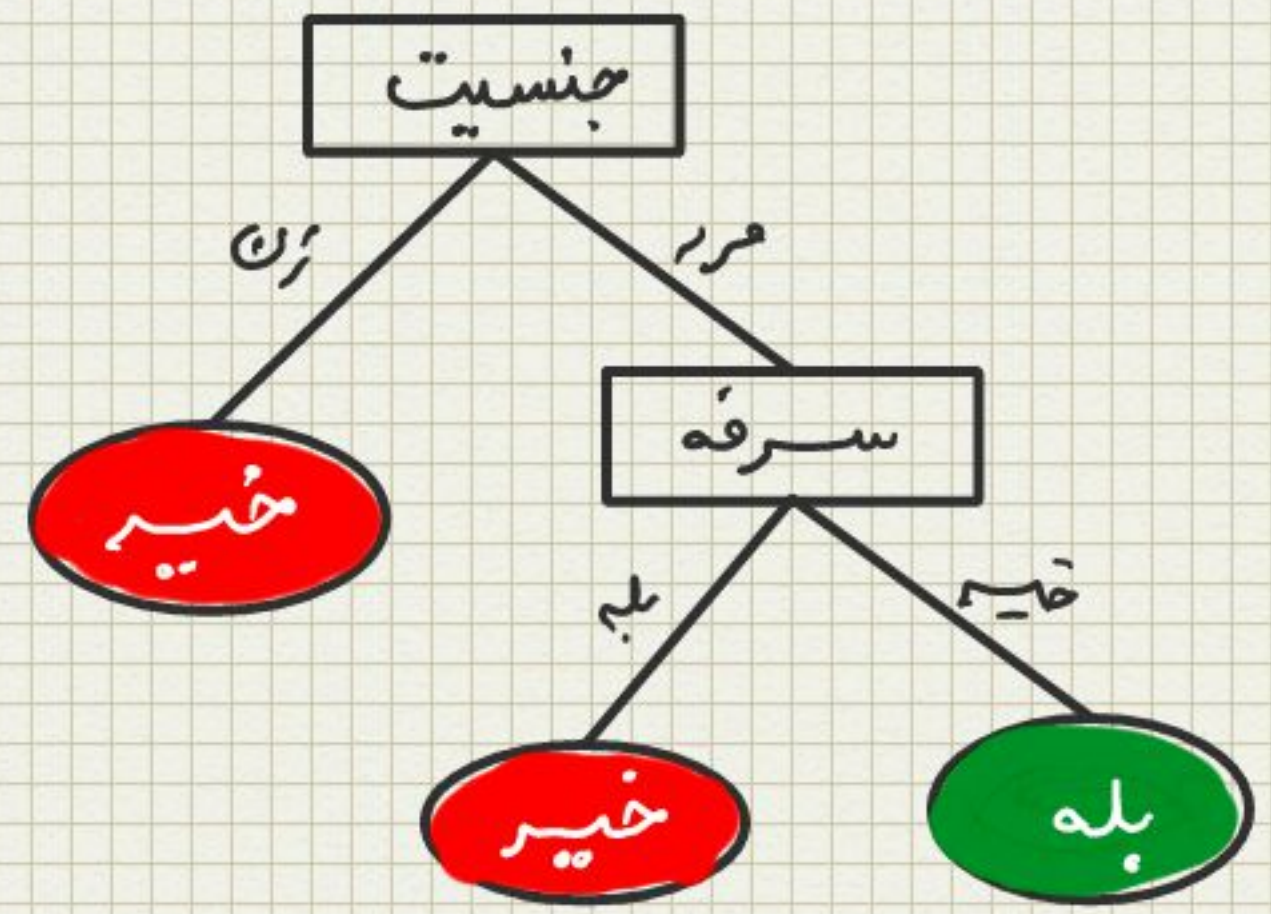
$$accuracy = \frac{4}{6} = 0.66$$

1اگر ویژگی سردرد وجود نداشت، این بار با روش بالا به درخت زیر می رسیدیم و درستی در بدترین حالت به اندازه سری قبل خواهد بود. و اگر از شانس خوب، برای نژاد آمریکایی هم درست پیش بینی می کردیم، درستی ما بالای 80 درصد می بود.



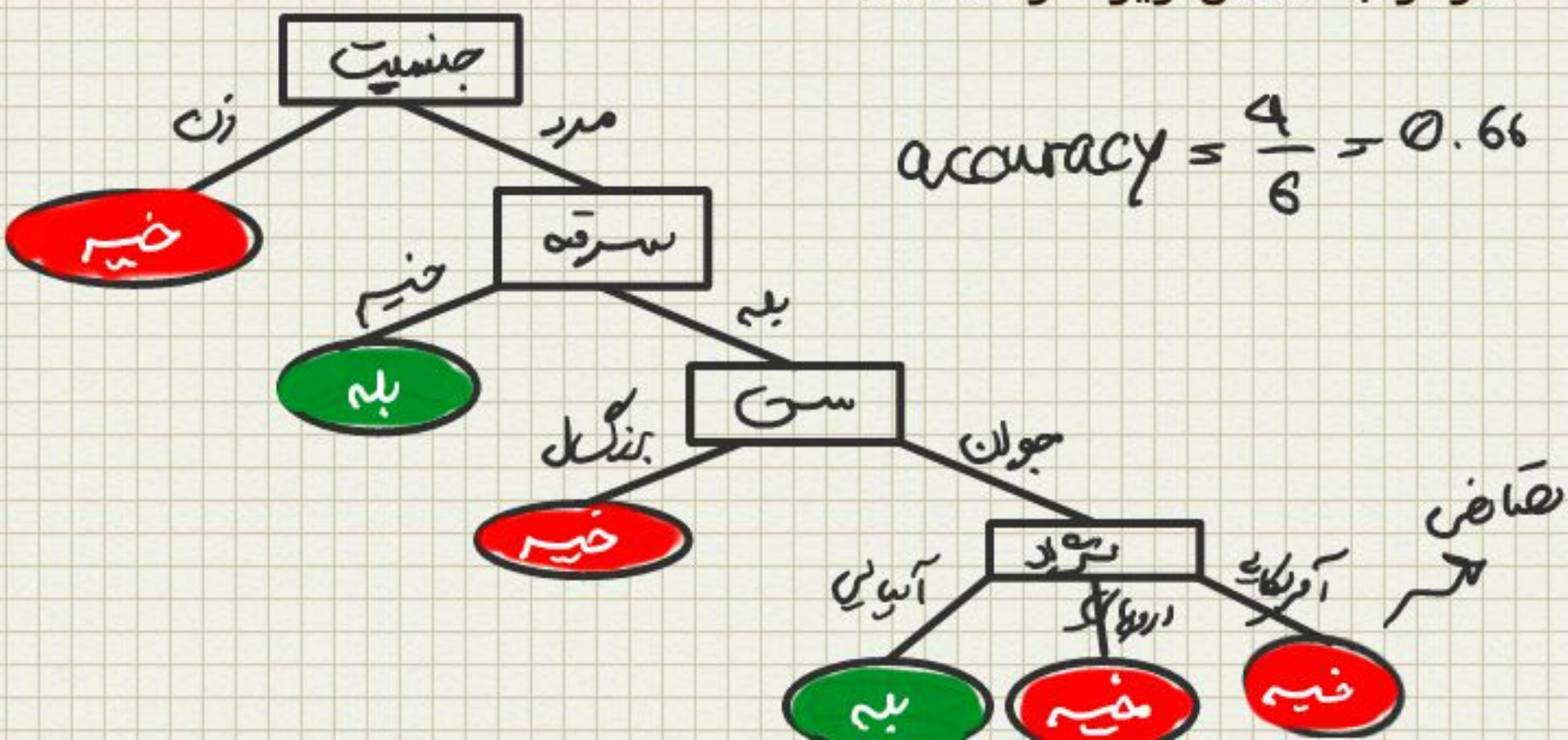
2 اگر ویژگی نژاد وجود نداشت، هیچ فرقی نمی کرد زیرا در درخت نخست هم همان طور که ملاحظه شد اثری نداشت. اما اگر هم ویژگی سردرد را نداشتیم و هم سردرد، می بایست در جدول بالا به جای ویژگی نژاد پیش بینی می کردیم که این فرد مبتلا نیست. چون 66 درصد داده ها چنین می گفتند.

3اگر پزشک برای تام سرماخوردگی تشخیص نمی داد، به خاطر کم بودن حجم داده ها نتیجه ی عجیبی می گرفتیم و جدول به شکل زیر می شد:



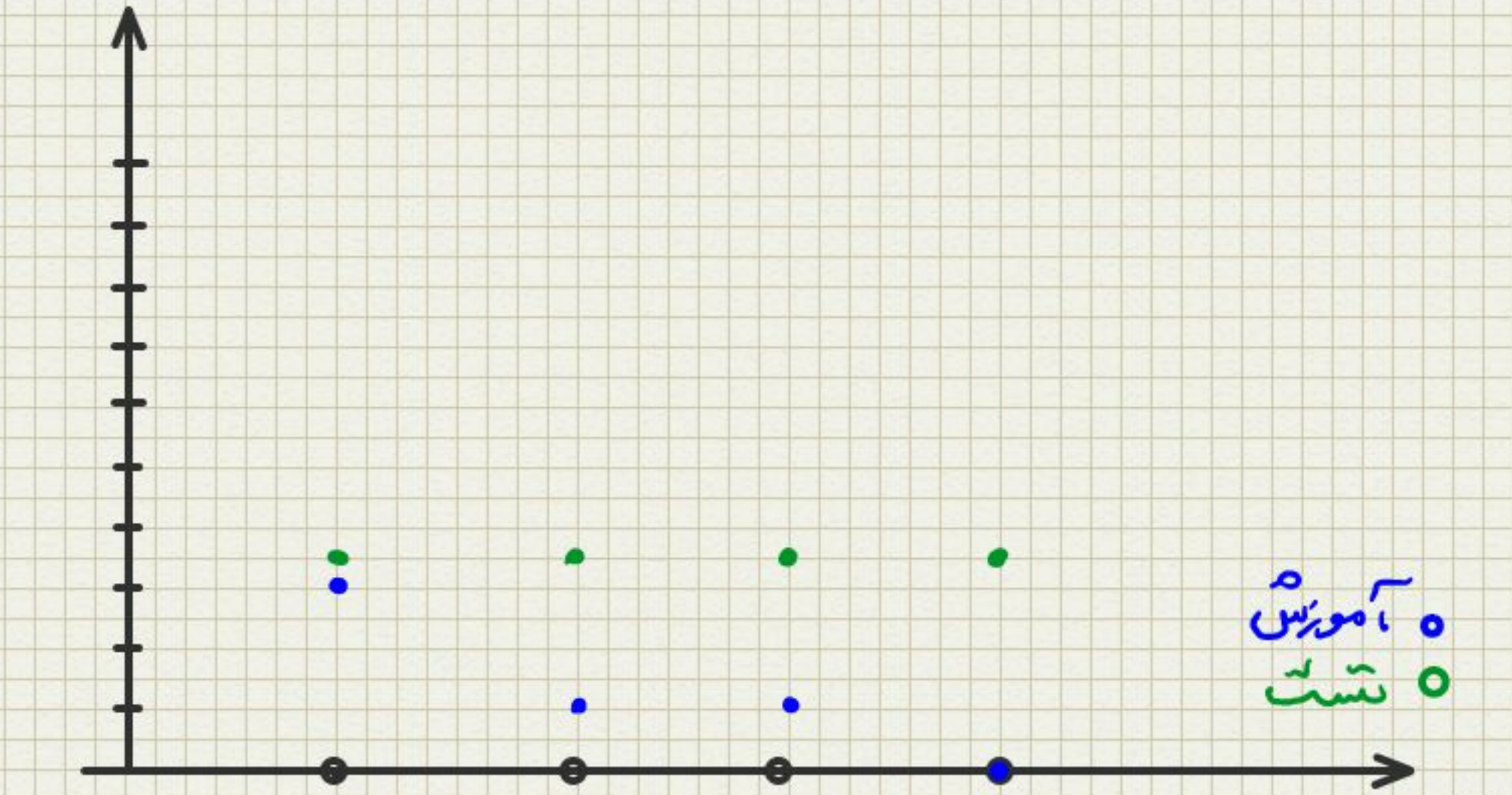
این یعنی ما هر مردی که سرفه نکند را بیمار می دانیم!! عجیب تر آن است که همچنان درستی مدلمان ثابت می ماند و این دیگر به خاطر کم بودن داده های تست است.

4 اضافه کردن ستون های با احتمال برابر که کمکی نمی کند، اما از انجا که تیلور مرد جوانی است که بیمار نیست کمک کننده است و نمودار به شکل زیر خواهد شد:



$$accuracy = \frac{4}{6} = 0.66$$

5نمودار توابع خطای مرحله آموزش و خطای تست برای یک درخت بر حسب حداکثر عمق محدود شده:



6 نمودار توابع خطای آموزش و تست برای یک درخت بر حسب سایز مجموعه آموزش: