**תרגיל מספר 3- נדב וילדר 312173776**

**חלק 1: הכנת הנתונים:**

**משימה 1:**

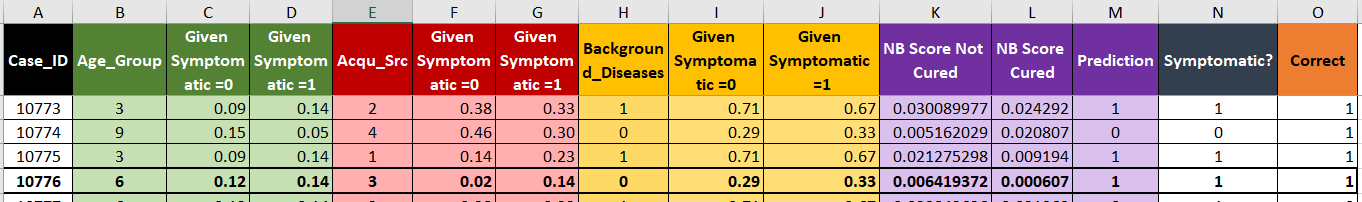
1. א. לא כל המשתנים הם אינפורמטיביים, ישנו משתנה שבין אם נשאיר אותו או נסיר לא תשתנה התוצאה.

ב. רוב המשתנים אלו בקנה אחד עם ההשערות שלי למעט מחלות רקע שדווקא חשבתי שלבעלי מחלות רקע יש יותר סיכוי לסימפטומים אך נראה כי אין זה משפיע כל כך.

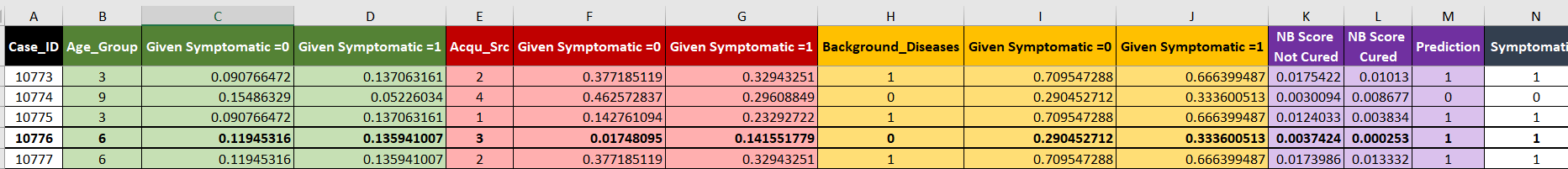
1. הפיצ'ר מין לא אינפורמטיבי. הוא לא תורם לנו כי ההסתברות להיות זכר בהינתן סימפטומטי וא-סימפטומטי זהה. כנ"ל גם עבור לנקבה.

**חלק 2: חיזוי:**

**משימה 4:**

7.

בחרתי את מקרה מספר 10776. המודל חזה שהמקרה יהיה סימפטומטי כיוון ש לפי החישוב: 0.14\*0.14\*0.33=0.006>0.12\*0.02\*0.29=0.0006

8.

בחרתי את מקרה מספר 10776. המודל חזה שהמקרה יהיה סימפטומטי כיוון ש לפי החישוב: 0.14\*0.14\*0.33\*0.58=0.003>0.12\*0.02\*0.29\*0.42=0.0002

9. המשמעות היא שגם אם מתחשבים ב 𝑝(𝐻=ℎ) וגם אם לא המודל מזהה את מקרה 10776 כסימפטומטי.

10.ניתן לראות את מקרה מספר 10702 שב1NSB נחזה שהוא א-סימפטומטי ואילו בNSB2 נחזה שהוא סימפטומטי. הדבר קרה מהיות שהסיכוי הכללי להיות סימפטומטי גדול יותר מהסיכוי להיות לא סימפטומטי. לכן מקרים בהם ההפרש בין ציון NB CURED גבוה מNOTCURED, מכפלה בסיכוי שהוא כן סימפטומטי יכול לשנות את החיזוי.

**חלק 3: ערכת תוצאות המודל:**

**משימה 6:**

19. מהיות שבמדד הaccuracy אנחנו נותנים אותו משקל TP וTN יכול להיגרם מצב ש"טיפלנו" בבן אדם שהוגדר כחולה למרות שהוא בכלל בריא. יש לזה משמעות מאוד כבדה.

20. פירוט על המודולים:

**NSB1**: עבור מודל זה ניתן לשים לב כי בין שלושת המדדים שלו , אין פער גדול. לכן קשה לצאת במסקנות עבור מודל זה בנפרד. בנוסף התוצאות לא מאוד חד משמעיות (62 אחוז) מה שמחזק את חולשתו. בהשוואה שלו לBM1 ניתן לראות שהוא יותר מדויק ב Specificity. כלומר סבירות גבוה יחסית שמבין הא-סימפטומטיים יחזה א-סימפטומטי.

**NSB2**: עבור מודל זה ניתן לשים לב כי שקיבלנו תוצאה מאוד טובה ב Recall / Sensitiviy. כלומר, המודל יזהה בסבירות גבוהה אדם סימפטומטי אם אכן כזה. ניתן לראות גם שבהשוואה לBM2 התוצאות של Recall / Sensitiviy טובות יותר, מה שמחזק את הטענה.

**NBS-1 Benchmark:** מודל זה מבוסס על פרופורציות הסימפטומטיים בNSB1 , כלומר בסביבות ה54% . לפי מודל arbitrary זה ניתן להסיק שאין ממש מה להסיק עבור מודל זה.

**NBS-2 Benchmark:** מודל זה מבוסס על פרופורציות הסימפטומטיים בNSB2 , כלומר בסביבות ה70% . לפי מודל arbitrary זה ניתן להסיק שבינו לבין עצמו אין יותר מידי מה להסיק אך בינו לבין NSB2 ניתן לראות פער בSensitivity.

המודל בו הייתי משתמש עבור הבעיה שפתרנו בהתאם לתוצאות שקיבלנו ובהנחה שכל הנבדקים הסימפטומטיים הם בעלי תופעות לוואי בלתי הפיכות, הוא NSB2 מכיוון ש ל Recall / Sensitiviy שלו יש תוצאה טובה.