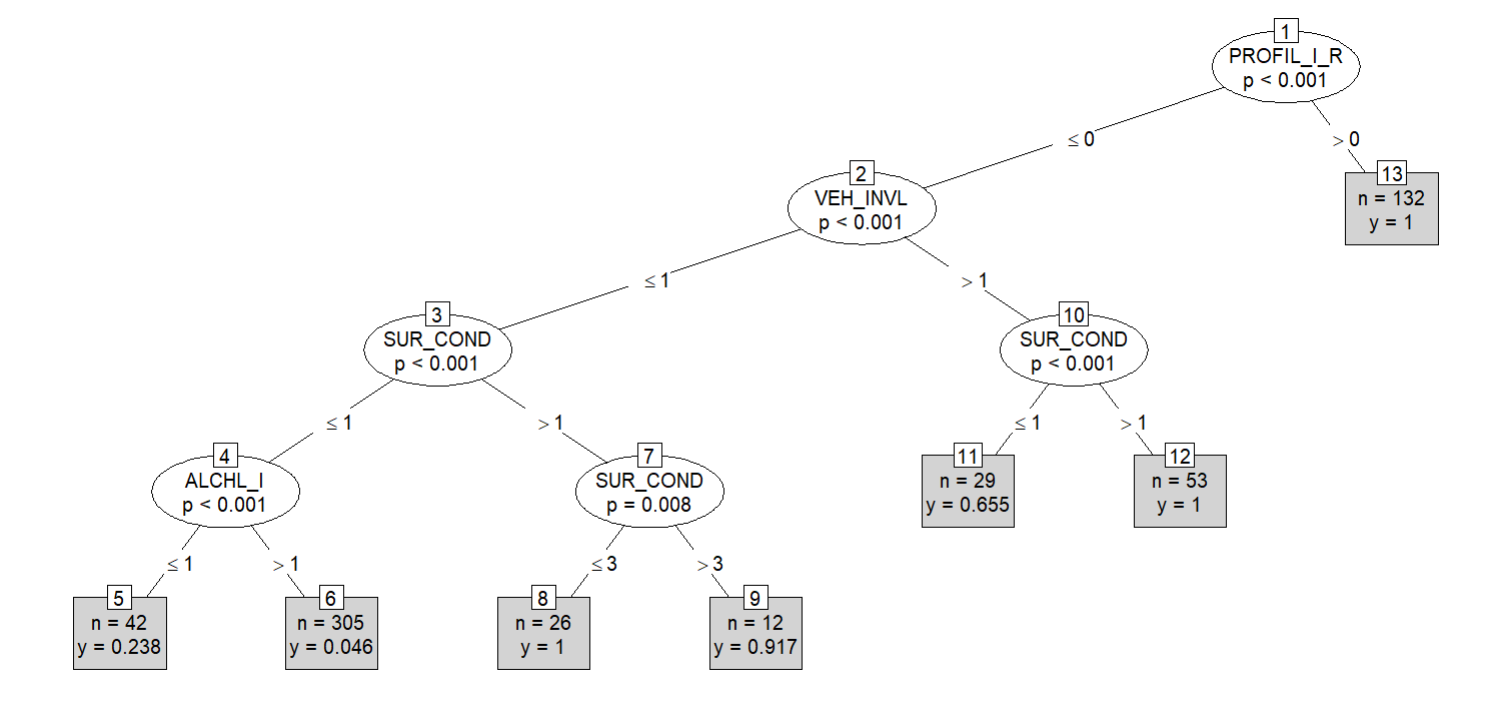
**ניב קוטק (ת.ז: 208236315) קובץ פלט לתרגיל מספר 4**

|  |  |
| --- | --- |
| **מטלה** | **פתרון** |
| שמות המודלים שהורצו | עץ החלטה, רשת נוירונים, רגרסיה לוגיסטית, יער אקראי ,( מכונת וקטורים תומכים) |
| שם המודל הטוב ביותר | רשת נוירונים וגם רגרסיה לוגיסטית |
| AUC של המודל הטוב ביותר | 0.972048 |

עץ ההחלטה שהתקבל עבור מודל עץ החלטה ותאור שלו (y, n, המספרים על הקשתות):

**הסבר על העץ**: שורש אחד, 5 צמתים פנימיים (13 צמתים) , 7 עלים, 7 דרכים. 12 קשתות   
**עבור כל עלה** (7 עלים , מספרי עלים – 5,6,8,9,11,12,(13).  
**n**: מספר התצפיות שנכנסו (42,305,26,12,29,53,132)  
**y**: ההסתברות לקבלת הערך החזוי – מספר נפגעים גדול מאפס (1) או לא (0). כאשר ההסתברות היא 1 אז זה בוודאות שמספר הנפגעים גדול מאפס. אם זה 0 אז זה בוודאות מספר הנפגעים לא גדול מאפס. אם זה לא 1 או לא 0 אז צריך להחליט מה הערך סף שלנו (CUTOFF).  
**מספרים על הקשתות**:   
 כאשר פרופיל הכביש קטן שווה לאפס אז זה ילך לצומת של מספר כלי הרכב המעורבים, אחרת זה ילך לעלה מספר 13.  
כאפשר מספר כלי הרכב המעורבים גדול מ1 : זה ילך לצומת (10) של מצב פני הכביש. אם צומת פני הכביש קטן או שווה 1 זה ילך לעלה מספר 11 אחרת זה ילך לעלה מספר 12.

כאפשר מספר כלי הרכב המעורבים קטן או שווה ל1 : זה ילך לצומת (3) של מצב פני הכביש. אם צומת פני הכביש קטן או שווה 1 זה ילך לצומת 4 של נוכחות אלכוהול אחרת זה ילך לצומת 7 של מצב פני הכביש. אם מצב פני הכביש קטן או שווה ל 3 זה ילך לעלה מספר 8 אחרת זה ילך לעלה מספר 9. או נוכחות אלכוהול זה קטן או שווה ל1 זה ילך לעלה מספר 5 אחרת זה ילך לעלה מספר 6.  
**לסיכום:** מספרים על הקשתות מתייחסים לערכים שמייצגים את תוצאות ההחלטות שהעץ מקבל באותה צומת (כמו התוצאות שהסברתי קודם). מספרים אלה עוזרים להבין את חלוקת הנתונים ותורמים למעקב אחר תהליך קבלת ההחלטות בעץ, כולל הכיצד הנתונים מנותחים ומסווגים בכל שלב של התהליך. לבסוף הוא מגיע לעלה שבו מתקבלת החלטה הסופית של ערך החזוי.

**מדדי הערכה עבור המודל הטוב ביותר (תוצאה + הסבר):**

|  |  |  |  |
| --- | --- | --- | --- |
| **מדד** | **תוצאה** | **תוצאה לאחר שינוי cutoff** | **משמעות המדד כאשר positive class=0** |
| Accuracy | 0.945 | 0.9575 | סך כל הפעמים שהמודל מזהה נכון את מספר הנפגעים שלא גדול מ0 וכן גדול 0 חלקי סה"כ כל התצפיות של המודל. |
| Sensitivity | 0.9585 | 1.0000 | היכולת לזהות נכון את ה **class** החשוב שהוא (0) כלומר שמספר הנפגעים לא גדול מ0. הוא מחושב ע"י מספר הפעמים שהוא מזהה נכון מתוך סה"כ הפעמים שמספר הנפעים לא גדול מ0. כלומר סך כל העמודה 0. |
| Specificity | 0.9290 | 0.9071 | היכולת לזהות מי כן בתאונה מספר הנפעים גדול מ 0.  הוא מחושב ע"י י מספר הפעמים שהוא מזהה נכון מתוך סה"כ הפעמים שמספר הנפעים גדול מ 0 כלומר סך כל העמודה 1. |

**הסבר להשפעת שינוי ה cutoff על המדדים -**אם cutoff=0.5 אז כל מי שגדול מהסתברות של 0.5 אסווג אותו שמספר הנפגעים גדול מ0 , אחרת אסווג אותו שלא.

|  |  |  |
| --- | --- | --- |
| **1** | **0** |  |
| 13 | 208 | **0** |
| 170 | 9 | **1** |

אם cutoff=0.6 אז כל מי שגדול מהסתברות של 0.6 אסווג אותו שמספר הנפגעים גדול מ0 , אחרת אסווג אותו שלא.

|  |  |  |
| --- | --- | --- |
| **1** | **0** |  |
| 17 | 217 | **0** |
| 166 | 0 | **1** |

**מה קורה לטבלה**:  
כאשר העליתי את סף ההחלטה שלי, מקרים שבעבר ניבאתי שמספר הנפגעים בהם יהיה גדול מאפס, עברו כעת לקבוצה בה הניבוי הוא שמספר הנפגעים יהיה אפס. כלומר, חלק מהמקרים שנמצאו קודם בשורה התחתונה של המטריצה עברו כעת לשורה העליונה .

המשמעות של שינוי זה היא שעל אף שהרגישות (Sensitivity) של המודל שלי תשתפר – כלומר, הוא יזהה יותר נכון את המקרים החיוביים – הדיוק הספציפי (Specificity) יפחת, כי המודל יתחיל לפספס יותר מקרים של תוצאות שליליות אמיתיות.

**שדות**:  
  
ALCOHOL= נוכחות (1) או היעדר (2) של אלכוהול

PROFIL\_I\_R= פרופיל הכביש; רמה (1),אחר(0)

SUR\_COND = מצב פני הכביש של הכביש; יבש (1), רטוב (2), מראה/ רפש (3), קרח (4), rn ידוע (9)

VEH\_INVL= מספר כלי הרכב המעורבים

MAX\_SEV\_IR= נוכחות של פציעות / הרוגים: אין פצועים (0), פציעה (1), מוות (2)  
  
Casualties = האם מספר הנפגעים גדול מ 0 (1) או לא (0)