תאריך הגשה: ‏06/02/2024

מגישים: רתם קשאני 209073352 ודויד קופלב 208870279

תרגיל מספר 2 מערכות לומדות

**שאלה 13**

**כדי להבין אילו נתונים נדרשים כדי לחשב את הערכים בכל תא של מטריצת הבלבול, נגדיר:  
  
- True Positive (TP): מספר המקרים שנחזו נכון כחיוביים.**

**- True Negative (TN): מספר המקרים שנחזו נכון כשליליים.**

**- False Positive (FP): מספר המקרים שנחזו באופן שגוי כחיוביים.**

**- False Negative (FN): מספר המקרים שנחזו באופן שגוי כשליליים.**

**כל תא במטריצת הבלבול מייצג את מספר המקרים שנכללים בקטגוריה ספציפית. השורות מתאימות למחלקות בפועל, והעמודות מתאימות למחלקות החזויות.**

**להלן הנוסחאות לחישוב הערכים עבור כל תא בשורה ה-**i**:**

**1. True Positive (TP) (אלמנטים אלכסוניים): נוסחה: TP = מספר המקרים שבהם מחלקה בפועל = מחלקה חזויה**

**2. (FP) False Positive (סכום לאורך העמודות, לא כולל האלמנט האלכסוני): נוסחה: FP = סכום הערכים בעמודה - TP**

**3. False Negative (FN) (סכום לאורך השורות, לא כולל האלמנט האלכסוני): נוסחה: FN = סכום הערכים בשורה - TP**

**4.** **True Negative (TN) (סכום של כל הערכים מלבד אלה שבאותה שורה ועמודה כמו TP): נוסחה: TN = סכום כולל - (TP + FP + FN)**

**שאלה 14**

**כדי להבין אילו נתונים נדרשים כדי לחשב את הערכים בכל תא של מטריצת הבלבול, נגדיר:  
  
- True Positive (TP): מספר המקרים שנחזו נכון כחיוביים.**

**- True Negative (TN): מספר המקרים שנחזו נכון כשליליים.**

**- False Positive (FP): מספר המקרים שנחזו באופן שגוי כחיוביים.**

**- False Negative (FN): מספר המקרים שנחזו באופן שגוי כשליליים.**

**כל תא במטריצת הבלבול מייצג את מספר המקרים שנכללים בקטגוריה ספציפית. השורות מתאימות למחלקות בפועל, והעמודות מתאימות למחלקות החזויות.**

**להלן הנוסחאות לחישוב הערכים עבור כל תא בשורה ה-**j**:**

1. **True Positive (TP)** (אלמנט אלכסוני):   
נוסחה: TP = confusion\_matrix[j, j] (כאשר j הוא אינדקס העמודה)

2. **False Positive (FP)** (סכום לאורך השורות, לא כולל האלמנט האלכסוני):   
נוסחה: FP = sum(confusion\_matrix[:, j]) - TP

3. **False Negative (FN)** (סכום לאורך העמודות, לא כולל האלמנט האלכסוני):   
נוסחה: FN = sum(confusion\_matrix[j, :]) - TP

4. **True Negative (TN)** (סכום של כל הערכים מלבד הערכים באותה שורה ועמודה כמו TP): נוסחה: TN = sum(confusion\_matrix) - (sum(confusion\_matrix[:, j]) + sum(confusion\_matrix[j] ,:]) - TP)

**שאלה 15  
מדדים בהקשר של classification report:**

**1. Precision (דיוק): מדד לחישוב כמה מהפיסולים החיוביים שמסומנים כחיוביים באופן תקן (True Positives) ביחס לסך כל התוצאות שמסומנות כחיוביות (True Positives + False Positives).  
  
Precision (פורמולה):**

**לדוג' עבור ספרה 9:**

**2. Recall (אחיזה): מדד לחישוב כמה מהפיסולים החיוביים שמסומנים כחיוביים באופן תקן (True Positives) ביחס לסך כל התוצאות האמיתיות החיוביות (True Positives + False Negatives).  
  
Recall (פורמולה):**

**לדוג' עבור ספרה 3:**

**3. F1-Score (סקור F1): מדד המשקף באופן משוקלל בין Precision ו-Recall. הערכים של F1-Score יתקבלו כפל הנגזרת של הדיוק והאחיזה, מתחלקת בסכום הדיוק והאחיזה.**

**F1-Score (פורמולה):**

**זהו חישוב ייצוגי, ועשוי להשתנות בהתאם לשימוש ספציפי במודל.**

**שאלה 16**

**מקרה אחד בעולם האמיתי שבו**  precision**מרבי הוא חיוני הוא בתחום האבחון הרפואי, במיוחד בגילוי סרטן. שקול שימוש בטכניקות הדמיה מתקדמות, כגון ממוגרפיה או** MRI**, לצורך בדיקת סרטן השד.**

**בבדיקת סרטן השד, המטרה היא לזהות נגעים סרטניים פוטנציאליים בשלב המוקדם ביותר האפשרי כדי לשפר את תוצאות הטיפול. תוצאות חיוביות כוזבות (זיהוי שגוי של מצב שאינו סרטני כסרטן) עלולות להוביל ללחץ מיותר, הליכי אבחון פולשניים נוספים, ולעיתים אף טיפולים מיותרים כמו ניתוח או כימותרפיה. מצד שני, שלילי שווא (אי זיהוי נגע סרטני ממשי) עלולים לעכב את הטיפול, ולאפשר לסרטן להתקדם לשלבים מתקדמים יותר.**

**בהקשר זה, השגת** precision **מרבי היא חיונית מכיוון שהיא ממזערת תוצאות חיוביות שגויות.** precision **באבחון רפואי מתייחס ליחס בין חיוביות אמיתיות לסכום של חיוביות אמיתיות וחיוביות שגויות.** precision **גבוה יותר פירושו שכאשר המערכת מסמנת נגע סרטני פוטנציאלי, סביר יותר שמדובר בסרטן ממשי, מה שמפחית את הסיכויים להתערבויות מיותרות.**

**על ידי אופטימיזציה ל** precision **מירבי בבדיקת סרטן השד, אנשי מקצוע בתחום הבריאות יכולים להבטיח שמטופלות שזוהו בסיכון אכן נמצאות בסבירות גבוהה יותר ללקות בסרטן, מה שמוביל לאבחנות מדויקות יותר וטיפולים בזמן וממוקד יותר. זה מדגים את החשיבות של** precision **במצבים שבהם ההשלכות של חיוביות כוזבות יכולות להיות השפעות משמעותיות על חייהם של אנשים.**

**שאלה 17**

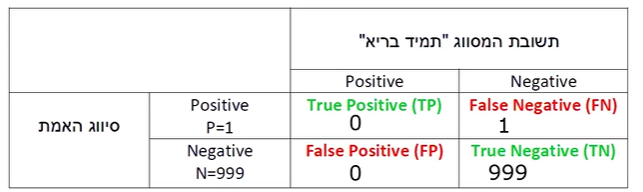
דוגמה שבה recall מקסימלי הוא חיוני ניתן למצוא בהקשר של מיון אבטחה בשדה התעופה. באבטחת שדות תעופה, המטרה העיקרית היא לזהות ולמנוע כניסה של כל איומים פוטנציאליים, כגון נשק או מטעני חבלה, אל כלי טיס.

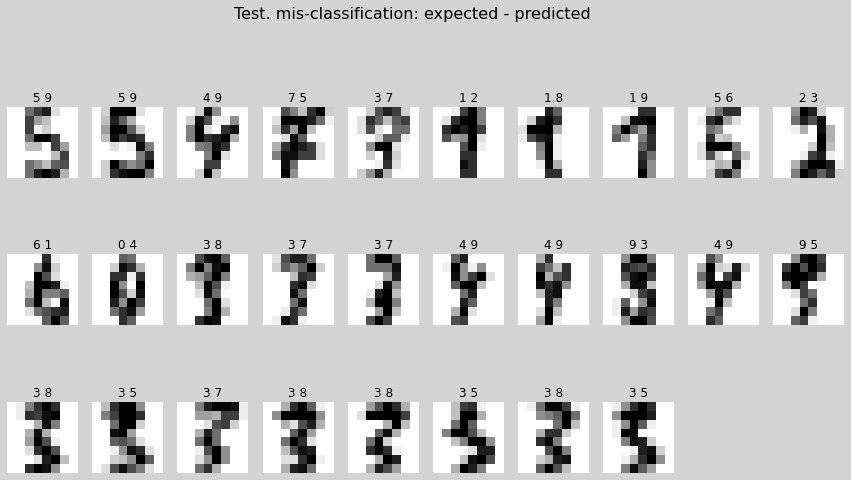
recall, בהקשר זה, מתייחס ליכולתה של מערכת האבטחה לזהות נכון את כל האיומים הממשיים, כולל הן חיוביות אמיתיות (איומים שזוהו נכון) והן שליליות כוזבות (איומים שהוחמצו). השגת recall מקסימלי חיונית בתרחיש זה מכיוון שהחמצת אפילו איום בודד עלולה להיות בעלת השלכות קטסטרופליות על בטיחות הנוסעים והצוות.

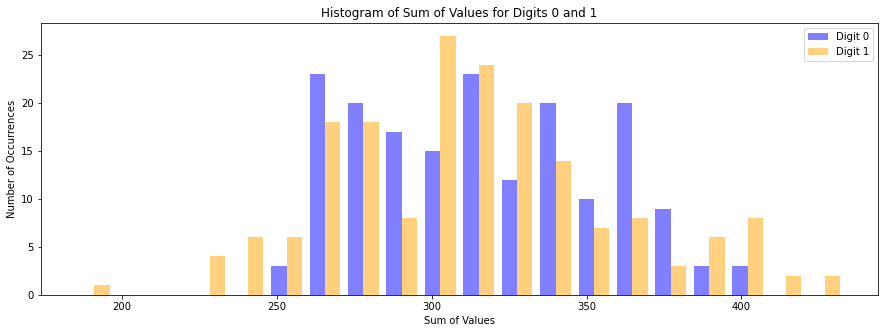
אם למערכת בדיקת אבטחה יש recall נמוך, פירוש הדבר שהיא עלולה להיכשל בזיהוי פריטים מסוכנים מסוימים במהלך תהליך הבדיקה, מה שיאפשר להם לעבור בנקודות בידוק ביטחוניות ללא זיהוי. הדבר מהווה סיכון רציני לבטיחות התעופה ולביטחון הלאומי.

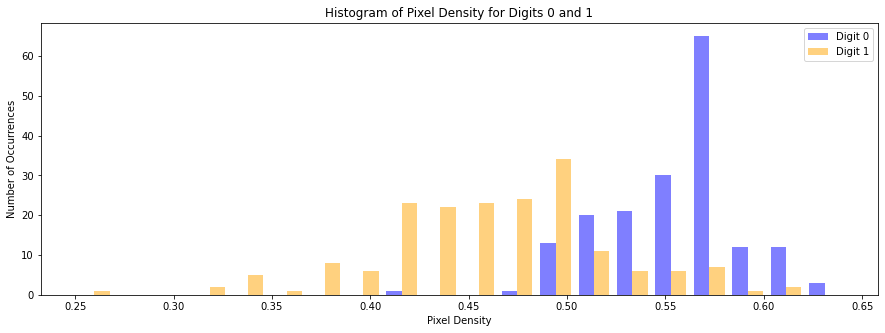
כדי להפחית את הסיכון הזה, מערכות האבטחה בשדות תעופה מתוכננות לעתים קרובות לתת עדיפות ל recall מקסימלי, גם אם זה אומר לקבל שיעור גבוה יותר של תוצאות חיוביות שגויות (פריטים שאינם מאיומים שזוהו באופן שגוי כאיומים). בעוד אזעקות שווא יכולות להיות לא נוחות ולהוביל לאמצעי סינון נוספים, העדיפות היא למזער את הסיכוי להחמיץ איום ממשי.

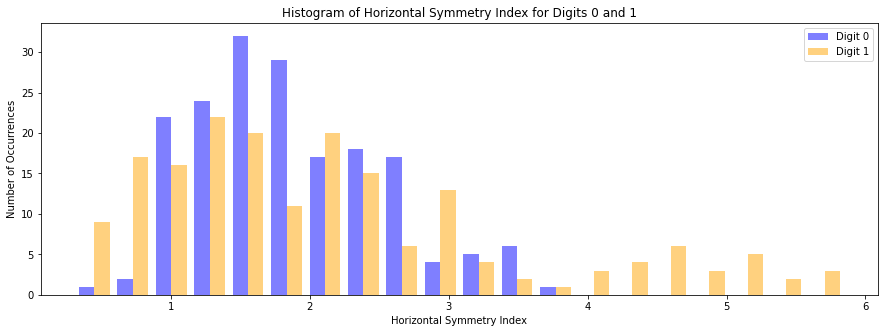
באבטחת שדות התעופה, השגת recall מקסימלי מסייעת להבטיח תהליך סינון יסודי ויעיל יותר, תורם לבטיחות הכוללת של נסיעות אוויריות ומניעת פרצות אבטחה פוטנציאליות. דוגמה זו ממחישה את החשיבות של ה recallבמצבים שבהם ההשלכות של החמצת חיובי אמיתי הן גבוהות ביותר.

**שאלה 18**

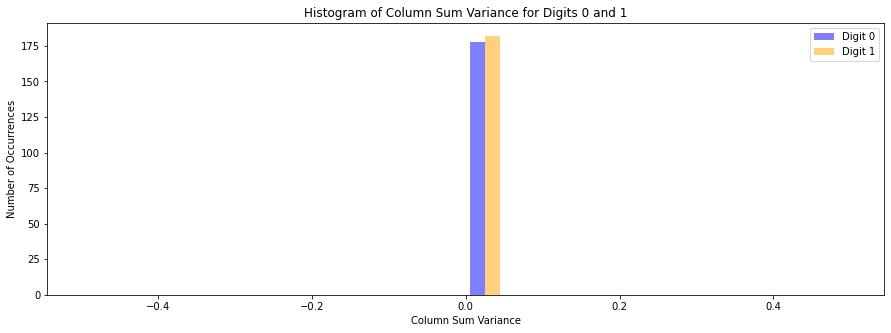
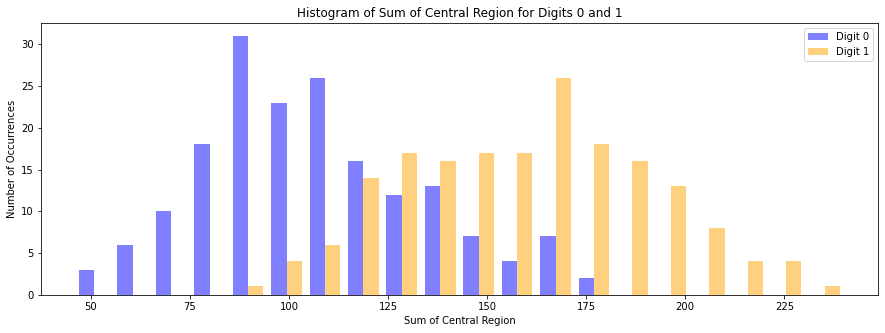
**שאלה 19**

**שאלה 20**

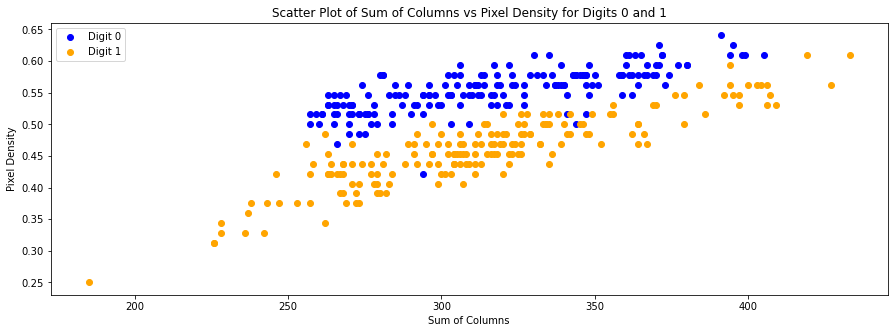
הסבר על פעולת המאפיין: מחשב את הסכום של כל ערכי הפיקסלים במטריצת התמונה בגודל 8x8 ומייצג את העוצמה או הכהות הכוללת של התמונה. ערכים גבוהים יותר עשויים להצביע על תמונה כהה יותר.

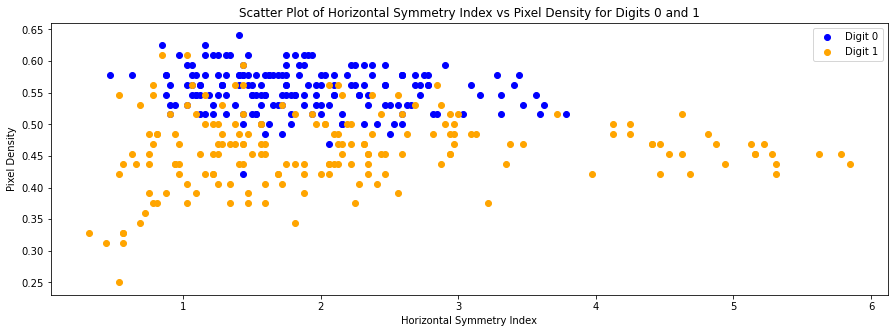
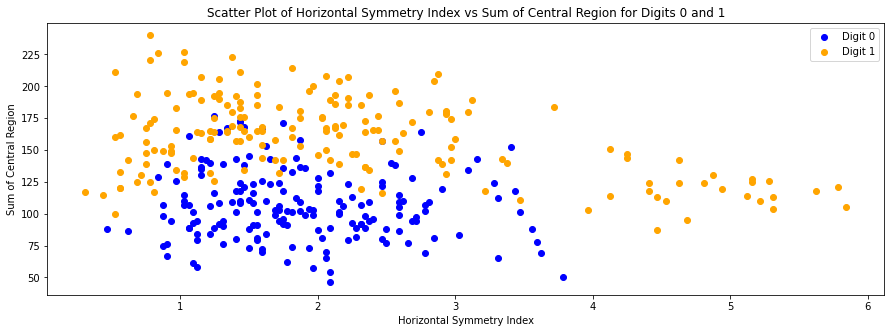
****הסבר על פעולת המאפיין: מחשב את היחס בין פיקסלים שאינם אפס למספר הכולל של פיקסלים במטריצת התמונה של 8x8 ומשקף את הפרופורציה של התמונה שתפוסה במידע. צפיפות גבוהה יותר עשויה להצביע על תמונה עם מילוי צפוף יותר.

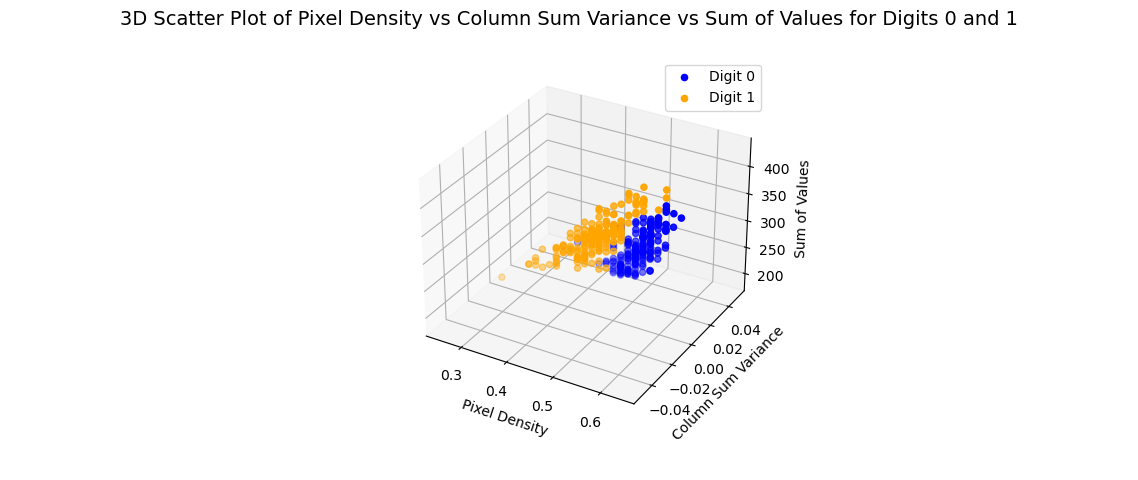
הסבר על פעולת המאפיין: **מדד הסימטריה האופקית מודד עד כמה תמונה סימטרית לאורך הציר האופקי שלה. הוא מחשב את ההפרש המוחלט בין כל פיקסל למקבילו המשתקף אופקית על פני מרכז התמונה. סכום ההבדלים הללו, מנורמל לפי מספר הפיקסלים הכולל, נותן ערך המציין את מידת הסימטריה האופקית: קרוב יותר ל-1 פירושו סימטריה גבוהה, בעוד שקרוב יותר ל-0 פירושו סימטריה נמוכה.**

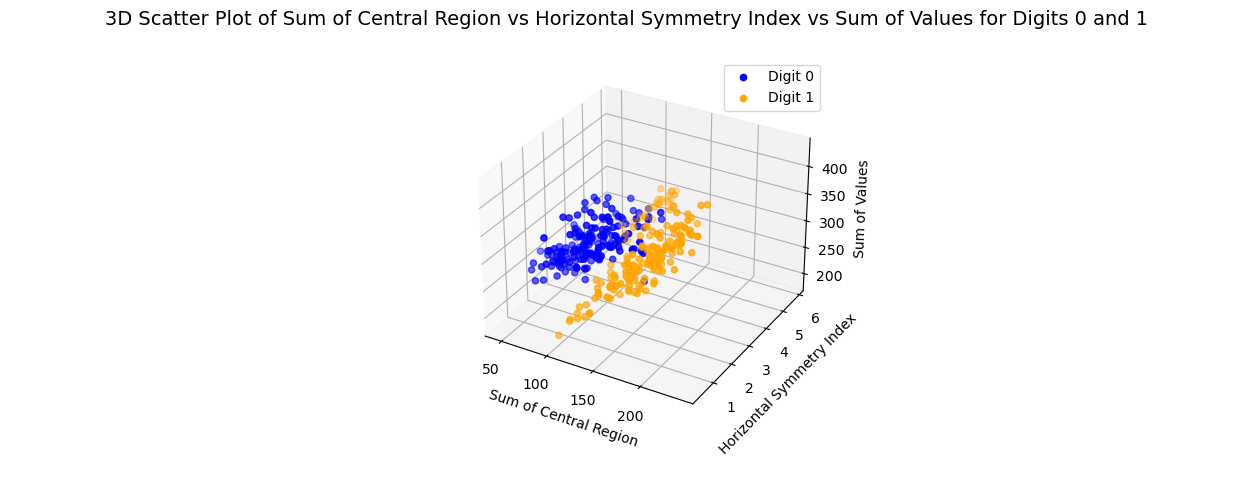
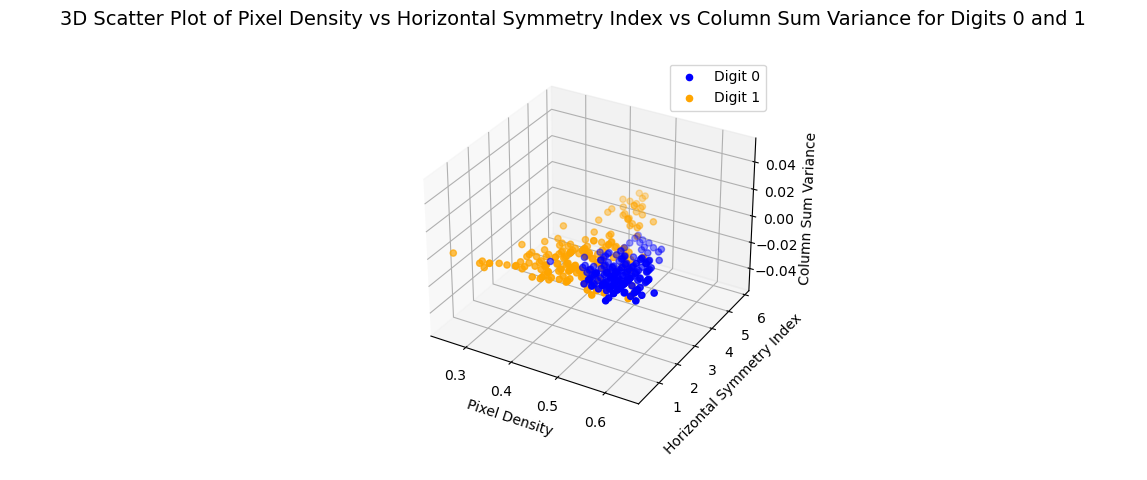
****הסבר על פעולת המאפיין: **פעולת שונות סכום העמודה מחשבת את השונות של סכומי הערכים בכל עמודה של מטריצה. זה כרוך בסיכום הערכים בכל עמודה, חישוב השונות של הסכומים הללו ומתן מדד עד כמה סכומי העמודה חורגים מערכם הממוצע. פעולה זו שימושית לניתוח ההתפלגות והשונות של נתונים לפי עמודה במטריצות, המשמשות לעתים קרובות בתחומים כמו סטטיסטיקה, עיבוד תמונה וניתוח נתונים.**

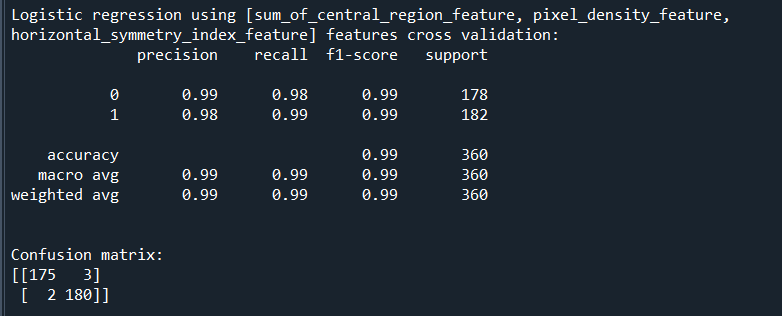
הסבר על פעולת המאפיין: פעולת סיכום השטח המרכזי במטריצה כוללת חיבור ערכי אזור ספציפי הנמצא במרכז המטריצה. אזור זה הוא בדרך כלל תת-קבוצה מרובעת או מלבנית של המטריצה המקורית.

**הצגת זוגות מאפיינים:**

****

**הצגת שלשות של מאפיינים:**

****



המאפיינים שבחרנו הם:   
סכום האזור המרכזי במטריצה שפעולת סיכום השטח המרכזי במטריצה כוללת חיבור ערכי אזור ספציפי הנמצא במרכז המטריצה. אזור זה הוא בדרך כלל תת-קבוצה מרובעת או מלבנית של המטריצה המקורית.  
צפיפות פיקסלים שמחשב את היחס בין פיקסלים שאינם אפס למספר הכולל של פיקסלים במטריצת התמונה של 8x8 ומשקף את הפרופורציה של התמונה שתפוסה במידע. צפיפות גבוהה יותר עשויה להצביע על תמונה עם מילוי צפוף יותר.  
מדד סימטריה אופקית ש**מודד עד כמה תמונה סימטרית לאורך הציר האופקי שלה. הוא מחשב את ההפרש המוחלט בין כל פיקסל למקבילו המשתקף אופקית על פני מרכז התמונה. סכום ההבדלים הללו, מנורמל לפי מספר הפיקסלים הכולל, נותן ערך המציין את מידת הסימטריה האופקית: קרוב יותר ל-1 פירושו סימטריה גבוהה, בעוד שקרוב יותר ל-0 פירושו סימטריה נמוכה.**