

## דף עבודה – מדדי הערכה

לא רק בלימוד מכונה אלא גם בחיים הכלליים, ובמיוחד בחיי העסקים, תוכלו לשמוע שאלות כמו:

"עד כמה המוצר שלך מדויק?" או "עד כמה המכונה שלך מדויקת?"

אם כן, איך נענה על שאלות אלו?

בדף עבודה זה נגדיר מדדים המשמשים להערכת תוצאות הסיווגים ונדון ביתרונות ובחסרונות של כל

אחד מהמדדים. המדדים הם:

- Accuracy
- Precision
- Recall
- F1-Score

בשביל לחשב את כל אחד מארבעת המדדים נציג את טבלת הבילבול (Confusion Matrix) ונדגים אותה על ערכים אמיתיים.

נניח שיש לנו מכונה שמקבלת כקלט תמונות של כלבים וחתולים ומחזירה את הסיווג – האם בתמונה יש תצלום של כלב או חתול.

העברנו לדגימה סט של 100 תמונות (50 תמונות של כלבים ו 50 תמונות של חתולים)

אבל המודל שלנו סיווג – 60 תמונות כחתולים, ו 40 תמונות ככלבים

נייצג את הנתונים בטבלה שהכותרות שלהם הן: בעמודות – הערכים החזויים predicted (לאחר הסיווג), בשורות – הערכים האמיתיים actual.

שלב ראשון - נגדיר מה מבחינתנו תשובה חיובית <= תמונה של חתול

שלב שני - נשאל את השאלות הבאות:

- 1) כמה דגימות המודל שלנו סיווג כ"חתול" ובאמת הן תמונות של "חתול" –  $TP = 42$
- 2) כמה דגימות המודל שלנו סיווג כ"כלב" ובאמת הן תמונות של "כלב" –  $TN = 32$
- 3) כמה דגימות המודל שלנו סיווג כ"חתול" והן לא תמונות של "חתול" –  $FP = 18$
- 4) כמה דגימות המודל שלנו סיווג כ"כלב" והן לא תמונות של "כלב" –  $FN = 8$

נארגן את הנתונים בטבלה:

Confusion Matrix		Predicted classes	
		cat	dog
Actual classes	cat	42	8
	dog	18	32

ככלל:

		חזוי	
		TRUE	FALSE
אמיתי	TRUE	TP	FN
	FALSE	FP	TN

## תרגול ב PYTHON

1. כתוב פונקציה שמחשבת את המדד הראשון: Accuracy

הפונקציה מקבלת את הפרמטרים הבאים: TP, TN, FP, FN

ומחזירה את דיוק המכונה, כמה דגימות המכונה שלנו זיהתה נכון מסך כל הדגימות שלנו לפי החישוב:

$$accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

זמן את הפונקציה עם סט ערכים הבאים:

TP = 42

TN = 32

FP = 18

FN = 8

הדפס את התוצאה על המסך

אם חישובתם נכון הפונקציה צריכה להחזיר 0.74, מה דעתך? האם המודל מוצלח? האם המכונה שלנו מדויקת?

כעת הרץ את הפונקציה שלך עם הנתונים הבאים:

		חזוי	
		חתול	כלב
אמיתי	חתול	0	8
	כלב	18	74

$$TP = 0$$

$$TN = 74$$

$$FP = 18$$

$$FN = 8$$

### הדפס את התוצאה על המסך

עפ"י הנתונים החדשים המודל לא הצליח לזהות אפילו תמונה אחת של חתול בצורה נכונה, אבל הפונקציה Accuracy עדיין מחזירה את הערך 0.74 !!

**מסקנה: צריך לחפש מדד נוסף שיעריך את המודל שלנו בצורה טובה יותר**

2. שנה את שם הפונקציה מסעיף הקודם ל model\_evaluate

הפונקציה מקבלת את הערכים הבאים: TP, TN, FP, FN

ומחזירה שני ערכים: **Precision, Accuracy**

הוסף לחישוב מדד נוסף: **Precision**

המדד Precision, משמעו מספר הדגימות החיוביות שהמודל גם חזה אותן כחיוביות לחלק

לכל הדגימות החיוביות **שהמודל חזה**. מחושב עפ"י החישוב הבא:

$$precision = \frac{TP}{TP + FP}$$

זמן את הפונקציה עם סט ערכים הבאים:

$$TP = 42$$

$$TN = 32$$

$$FP = 18$$

$$FN = 8$$

### הדפס את התוצאה על המסך

אם חישובתם נכון הפונקציה צריכה להחזיר **Precision=0.70**, מה דעתך? האם המודל מוצלח? האם המכונה שלנו מדויקת?

כעת הרץ את הפונקציה שלך עם הנתונים הבאים:

$$TP = 0$$

$$TN = 74$$

$$FP = 18$$

FN = 8

### הדפס את התוצאה על המסך

אם חישבתם נכון הפונקציה צריכה להחזיר  $Precision=0$ .

זה כבר משקף נכון יותר האם המודל שלנו מוצלח או לא. במקרה שבו המודל לא הצליח לחזות ערכים חיוביים בצורה נכונה קיבלנו שהמודל שלנו אינו מדויק

תארו לכם שבמקום סיווג תמונות יש לפניכם תוצאות בדיקת קורונה, עם הנתונים הבאים:

		חזוי	
		חיובי	שלילי
מציאות	חיובי	74	24
	שלילי	2	0

במצב הזה:  $Precision=0.97$ ,  $Accuracy=0.74$ . נראה שהמודל ממש טוב. אבל, האם נראה לכם הגיוני להודיע ל 24 אנשים שהם שליליים לקורונה למרות שבפועל חולים???

**מסקנה: צריך לחפש מדד נוסף שיעריך את המודל שלנו בצורה טובה יותר**

3. השתמש בפונקציה מסעיף הקודם: `model_evaluate`

הפונקציה מקבלת את הערכים הבאים: TP, TN, FP, FN

ומחזירה שלושה ערכים: **Precision, Accuracy, Recall**

הוסף לחישוב מדד נוסף: **Recall**

המדד Recall, משמעו מספר הדגימות החיוביות שהמודל גם חזה אותן כחיוביות לחלק לכל הדגימות החיוביות ~~שהמודל חזה שהיו בנתונים~~. מחושב עפ"י החישוב הבא:

$$recall = \frac{TP}{TP + FN}$$

זמן את הפונקציה עם סט ערכים הבאים:

TP = 74

TN = 0

FP = 2

FN = 24

### הדפס את התוצאה על המסך

במצב הזה: Accuracy=0.74 ,Precision=0.97,Recall=0.75.

התבלבלתם?

לפי איזה מדד נחשב? בכל בעיה נתאים את המדד לסוג הבעיה? אולי יש מדד שיספק לנו את הדיוק בכל סוגי הבעיות?

יש מדד שמשקלל בתוכו את כל סוגי הבעיות: F1-score

4. השתמש בפונקציה מסעיף הקודם: model\_evaluate  
הפונקציה מקבלת את הערכים הבאים: TP, TN, FP, FN  
ומחזירה ארבעה ערכים: Precision, Accuracy, Recall, f1  
הוסף לחישוב מדד נוסף: f1  
ככל השערך המקבל במדד f1, מתקרב לערך 1, המודל שלנו מדויק יותר  
המדד f1 מחושב עפ"י החישוב הבא:

$$F_1 = \frac{2}{\frac{1}{recall} + \frac{1}{precision}} = 2 \cdot \frac{precision \cdot recall}{precision + recall}$$

זמן את הפונקציה עם סט ערכים הבאים:

TP = 74

TN = 0

FP = 2

FN = 24

הדפס את התוצאה על המסך

במצב הזה: Accuracy=0.74 ,Precision=0.97,Recall=0.75, f1=0.84.