

שלב ב :

DecisionTreeClassifier .1

הפרמטרים המרכזיים של DecisionTreeClassifier הם:

פרמטר	תיאור	ערכים נפוצים לדוגמה
criterion	הקריטריון לקביעת איכות הפיצול:	'gini' (ברירת מחדל), 'entropy'
splitter	אסטרטגיית בחירת הפיצול בצומת:	'best' (ברירת מחדל), 'random'
max_depth	העומק המקסימלי של העץ (מונע התאמת יתר).	None (ברירת מחדל), מספרים שלמים
min_samples_split	מספר הדוגמאות המינימלי הדרוש כדי לפצל צומת.	2 (ברירת מחדל), 5, 10
min_samples_leaf	מספר הדוגמאות המינימלי בכל עלה (leaf).	1 (ברירת מחדל), 2, 5
max_features	מספר המאפיינים המרבי שניתן לבדוק בכל צומת.	'None', 'sqrt', 'log2'

<pre>criterion='entropy' max_depth=5 min_samples_split=10 min_samples_leaf=5</pre>	<pre>Train Accuracy: 0.97 Test Accuracy: 0.90  Classification Report (Test):       precision    recall  f1-score   support        0:       0.86       0.87       0.87        166       1:       0.93       0.92       0.92        290   accuracy:       0.90 macro avg:       0.90       0.90       0.90        456 weighted avg:    0.90       0.90       0.90        456  Confusion Matrix (Test): [[145  21]  [ 23 267]]</pre>	<p>דיוק נמוך יותר על סט הבדיקה, ייתכן שעדיין יש התאמת יתר (Overfitting) וטעויות משמעותיות בזיהוי שתי הקטגוריות, במיוחד בקטגוריה 0</p>
<pre>criterion='entropy' max_depth=7 min_samples_split=10 min_samples_leaf=5</pre>	<pre>Train Accuracy: 0.97 Test Accuracy: 0.90  Classification Report (Test):       precision    recall  f1-score   support        0:       0.86       0.87       0.87        166       1:       0.93       0.92       0.92        290   accuracy:       0.90 macro avg:       0.90       0.90       0.90        456 weighted avg:    0.90       0.90       0.90        456  Confusion Matrix (Test): [[145  21]  [ 23 267]]</pre>	<p>שינוי הפרמטרים לא השפיע באופן מהותי על ביצועי המודל. ייתכן ש-DecisionTreeClassifier הגיע למגבלת היכולת שלו על סט הנתונים.</p>
<pre>criterion='entropy' max_depth=10 min_samples_split=10 min_samples_leaf=2</pre>	<pre>Train Accuracy: 0.97 Test Accuracy: 0.93  Classification Report (Test):       precision    recall  f1-score   support        0:       0.85       0.96       0.90        166       1:       0.98       0.90       0.94        290   accuracy:       0.93 macro avg:       0.91       0.93       0.92        456 weighted avg:    0.93       0.93       0.93        456  Confusion Matrix (Test): [[160   6]  [ 28 262]]</pre>	<p>Test Accuracy: 0.93 שיפור לעומת 0.90 בהרצה הקודמת שיפור עומק העלים והמינימום לדוגמאות לפיצול תרם לשיפור ביצועים בקטגוריה 1</p>
<pre>criterion='gini' max_depth=15 min_samples_split=5 min_samples_leaf=1</pre>	<pre>Train Accuracy: 0.98 Test Accuracy: 0.89  Classification Report (Test):       precision    recall  f1-score   support        0:       0.81       0.91       0.86        166       1:       0.94       0.88       0.91        290   accuracy:       0.89 macro avg:       0.88       0.89       0.88        456 weighted avg:    0.90       0.89       0.89        456  Confusion Matrix (Test): [[151  15]  [ 35 255]]</pre>	<p>שינוי הפרמטרים לא שיפר את יכולת ההכללה. המודל סובל מתאמת יתר (Overfitting) ומביצועים בינוניים על סט הבדיקה.</p>

<pre>criterion='entropy' max_depth=10 min_samples_split=10 min_samples_leaf=4 random_state=42</pre>	<pre>Train Accuracy: 0.97 Test Accuracy: 0.93  Classification Report (Test):       precision    recall  f1-score   support       0       0.85       0.96       0.90         166      1       0.98       0.90       0.94         290     accuracy         0.91         0.93         0.93         456   macro avg         0.91         0.93         0.92         456  weighted avg         0.93         0.93         0.93         456  Confusion Matrix (Test): [[160   6]  [ 28 262]]</pre>	<p>שינוי הפרמטרים (עלים גדולים יותר ופיצולים מינימליים) שיפר את הביצועים ושמר על איזון טוב.</p>
---	--	---

2. RandomForestClassifier

RandomForestClassifier מבוסס על עצי החלטה רבים. הנה הפרמטרים המרכזיים:

פרמטר	תיאור	ערכים נפוצים לדוגמה
n_estimators	מספר העצים ביער.	100 (ברירת מחדל), 50, 200
criterion	הקריטריון לקביעת איכות הפיצול (בדומה ל-DecisionTreeClassifier).	'gini', 'entropy'
max_depth	העומק המקסימלי של כל עץ.	None, 5, 10
min_samples_split	מספר הדוגמאות המינימלי הדרוש לפיצול בצומת.	10, 5, 2
min_samples_leaf	מספר הדוגמאות המינימלי בכל עלה.	5, 2, 1
max_features	מספר המאפיינים המרבי לבחירה בכל עץ.	'sqrt' (ברירת מחדל), 'log2', None
bootstrap	אם להשתמש בדגימה עם החזרה לבחירת דוגמאות עבור כל עץ.	True (ברירת מחדל), False

<p>n_estimators=100 max_depth=10 min_samples_split=5 max_features='sqrt'</p>	<p>Train Accuracy: 0.99 Test Accuracy: 0.95</p> <p>Classification Report (Test):</p> <table><thead><tr><th></th><th>precision</th><th>recall</th><th>f1-score</th><th>support</th></tr></thead><tbody><tr><td>0</td><td>0.92</td><td>0.93</td><td>0.93</td><td>166</td></tr><tr><td>1</td><td>0.96</td><td>0.95</td><td>0.96</td><td>290</td></tr><tr><td>accuracy</td><td></td><td></td><td>0.95</td><td>456</td></tr><tr><td>macro avg</td><td>0.94</td><td>0.94</td><td>0.94</td><td>456</td></tr><tr><td>weighted avg</td><td>0.95</td><td>0.95</td><td>0.95</td><td>456</td></tr></tbody></table> <p>Confusion Matrix (Test):</p> <pre>[[155  11]  [ 14 276]]</pre>		precision	recall	f1-score	support	0	0.92	0.93	0.93	166	1	0.96	0.95	0.96	290	accuracy			0.95	456	macro avg	0.94	0.94	0.94	456	weighted avg	0.95	0.95	0.95	456	<p>המודל מתאים כמעט לחלוטין את סט האימון, ביצועים טובים מאוד על סט הבדיקה.</p> <p>פחות טעויות לעומת עץ החלטה. המודל מזהה היטב את שתי הקטגוריות</p>
	precision	recall	f1-score	support																												
0	0.92	0.93	0.93	166																												
1	0.96	0.95	0.96	290																												
accuracy			0.95	456																												
macro avg	0.94	0.94	0.94	456																												
weighted avg	0.95	0.95	0.95	456																												
<p>n_estimators=200 max_depth=15 min_samples_split=5 max_features='log2' criterion='entropy'</p>	<p>Train Accuracy: 0.99 Test Accuracy: 0.96</p> <p>Classification Report (Test):</p> <table><thead><tr><th></th><th>precision</th><th>recall</th><th>f1-score</th><th>support</th></tr></thead><tbody><tr><td>0</td><td>0.94</td><td>0.96</td><td>0.95</td><td>166</td></tr><tr><td>1</td><td>0.98</td><td>0.97</td><td>0.97</td><td>290</td></tr><tr><td>accuracy</td><td></td><td></td><td>0.96</td><td>456</td></tr><tr><td>macro avg</td><td>0.96</td><td>0.96</td><td>0.96</td><td>456</td></tr><tr><td>weighted avg</td><td>0.96</td><td>0.96</td><td>0.96</td><td>456</td></tr></tbody></table> <p>Confusion Matrix (Test):</p> <pre>[[159   7]  [ 10 280]]</pre>		precision	recall	f1-score	support	0	0.94	0.96	0.95	166	1	0.98	0.97	0.97	290	accuracy			0.96	456	macro avg	0.96	0.96	0.96	456	weighted avg	0.96	0.96	0.96	456	<p>Test Accuracy: 0.96 (שיפור קל לעומת 0.95).</p> <p>הגדלת מספר העצים ושינוי עומק העצים עזרו לשפר את ביצועי המודל</p>
	precision	recall	f1-score	support																												
0	0.94	0.96	0.95	166																												
1	0.98	0.97	0.97	290																												
accuracy			0.96	456																												
macro avg	0.96	0.96	0.96	456																												
weighted avg	0.96	0.96	0.96	456																												
<p>n_estimators=300 max_depth=20 min_samples_split=10 min_samples_leaf=5 max_features='log2'</p>	<p>Train Accuracy: 0.96 Test Accuracy: 0.94</p> <p>Classification Report (Test):</p> <table><thead><tr><th></th><th>precision</th><th>recall</th><th>f1-score</th><th>support</th></tr></thead><tbody><tr><td>0</td><td>0.91</td><td>0.93</td><td>0.92</td><td>166</td></tr><tr><td>1</td><td>0.96</td><td>0.94</td><td>0.95</td><td>290</td></tr><tr><td>accuracy</td><td></td><td></td><td>0.94</td><td>456</td></tr><tr><td>macro avg</td><td>0.93</td><td>0.94</td><td>0.94</td><td>456</td></tr><tr><td>weighted avg</td><td>0.94</td><td>0.94</td><td>0.94</td><td>456</td></tr></tbody></table> <p>Confusion Matrix (Test):</p> <pre>[[155  11]  [ 16 274]]</pre>		precision	recall	f1-score	support	0	0.91	0.93	0.92	166	1	0.96	0.94	0.95	290	accuracy			0.94	456	macro avg	0.93	0.94	0.94	456	weighted avg	0.94	0.94	0.94	456	<p>הגבלת העומק (20) ומספר העצים הגבוה תרמו ליציבות המודל ולביצועים מאוזנים.</p>
	precision	recall	f1-score	support																												
0	0.91	0.93	0.92	166																												
1	0.96	0.94	0.95	290																												
accuracy			0.94	456																												
macro avg	0.93	0.94	0.94	456																												
weighted avg	0.94	0.94	0.94	456																												

n_estimators=500 max_depth=None min_samples_split=5 max_features='sqrt'	Train Accuracy: 0.98 Test Accuracy: 0.95  Classification Report (Test): <pre> precision    recall  f1-score   support  0           0.93      0.93      0.93       166 1           0.96      0.96      0.96       290  accuracy          0.95          0.95          0.95       456 macro avg         0.95          0.95          0.95       456 weighted avg      0.95          0.95          0.95       456 </pre> Confusion Matrix (Test): <pre> [[155  11]  [ 12 278]] </pre>	ביצועים טובים על סט האימון, עם התאמת יתר מינימלית. <b>Test Accuracy: 0.95 →</b> הביצועים הגבוהים ביותר מבין המסווגים. המודל היציב והמצטיין ביותר בהרצות הנוכחיות. השימוש במספר רב של עצים עם בחירת מאפיינים רנדומלית תרם לדיוק גבוה.
n_estimators=300 max_depth=15 min_samples_split=10 max_samples=0.8 max_features=0.8	Train Accuracy: 0.99 Test Accuracy: 0.94  Classification Report (Test): <pre> precision    recall  f1-score   support  0           0.93      0.92      0.92       166 1           0.95      0.96      0.96       290  accuracy          0.94          0.94          0.94       456 macro avg         0.94          0.94          0.94       456 weighted avg      0.94          0.94          0.94       456 </pre> Confusion Matrix (Test): <pre> [[152  14]  [ 12 278]] </pre>	<b>Test Accuracy: 0.94 →</b> תוצאה חזקה ויציבה. המודל שמר על יציבות עם ביצועים גבוהים מאוד ואיזון מצוין בין דיוק לשליפות

3. AdaBoostClassifier

AdaBoostClassifier עובד עם מספר לומדים חלשים (weak learners), ברירת המחדל היא DecisionTreeClassifier.

פרמטר	תיאור	ערכים נפוצים לדוגמה
n_estimators	מספר הלומדים החלשים (weak learners) שיוכשרו.	50 (ברירת מחדל), 100, 200
learning_rate	קצב הלמידה שמשפיע על תרומת כל לומד חלש.	1.0 (ברירת מחדל), 0.1, 0.01
algorithm	האלגוריתם: 'SAMME.R' (ברירת מחדל) או 'SAMME'.	'SAMME.R', 'SAMME'
base_estimator	לומד חלש מותאם אישית (ברירת המחדל הוא DecisionTreeClassifier עם max_depth=1).	DecisionTreeClassifier

<pre>estimator= DecisionTreeClassifier (max_depth=1) n_estimators=50 learning_rate=0.5 algorithm='SAMME'</pre>	<pre>Train Accuracy: 1.00 Test Accuracy: 0.96  Classification Report (Test):       precision    recall  f1-score   support       0       0.93      0.96      0.95       166      1       0.98      0.96      0.97       290   accuracy         0.96       456  macro avg       0.95      0.96      0.96       456 weighted avg       0.96      0.96      0.96       456  Confusion Matrix (Test): [[160   6]  [ 12 278]]</pre>	המודל מתאים בצורה מושלמת את סט האימון הביצועים הטובים ביותר על סט הבדיקה המודל משיג את התוצאות הטובות ביותר, עם הכי פחות טעויות בשתי הקטגוריות
<pre>estimator= DecisionTreeClassifier (max_depth=1) n_estimators=100 earning_rate=0.8 algorithm='SAMME'</pre>	<pre>Train Accuracy: 1.00 Test Accuracy: 0.95  Classification Report (Test):       precision    recall  f1-score   support       0       0.92      0.95      0.93       166      1       0.97      0.95      0.96       290   accuracy         0.95       456  macro avg       0.95      0.95      0.95       456 weighted avg       0.95      0.95      0.95       456  Confusion Matrix (Test): [[158   8]  [ 14 276]]</pre>	הגדלת מספר הלומדים החלשים (weak learners) ושינוי קצב הלמידה לא שיפרו את הביצועים.

<div>estimator= DecisionTreeClassifier (max_depth=2) n_estimators=150 learning_rate=0.3 algorithm='SAMME.R'</div>	<div>Train Accuracy: 1.00 Test Accuracy: 0.90</div> <div>Classification Report (Test):</div> <table><thead><tr><th></th><th>precision</th><th>recall</th><th>f1-score</th><th>support</th></tr></thead><tbody><tr><td>0</td><td>0.81</td><td>0.95</td><td>0.87</td><td>166</td></tr><tr><td>1</td><td>0.97</td><td>0.88</td><td>0.92</td><td>290</td></tr><tr><td>accuracy</td><td></td><td></td><td>0.90</td><td>456</td></tr><tr><td>macro avg</td><td>0.89</td><td>0.91</td><td>0.90</td><td>456</td></tr><tr><td>weighted avg</td><td>0.91</td><td>0.90</td><td>0.90</td><td>456</td></tr></tbody></table> <div>Confusion Matrix (Test):</div> <div>[[157 9] [ 36 254]]</div>		precision	recall	f1-score	support	0	0.81	0.95	0.87	166	1	0.97	0.88	0.92	290	accuracy			0.90	456	macro avg	0.89	0.91	0.90	456	weighted avg	0.91	0.90	0.90	456	<div>Test Accuracy: 0.90 ירידה לעומת 0.95</div> <div>בהרצה הקודמת העלאת מספר הלומדים החלשים ושינוי קצב הלמידה לא הובילו לשיפור, אלא אולי להתאמת יותר</div>
	precision	recall	f1-score	support																												
0	0.81	0.95	0.87	166																												
1	0.97	0.88	0.92	290																												
accuracy			0.90	456																												
macro avg	0.89	0.91	0.90	456																												
weighted avg	0.91	0.90	0.90	456																												
<div>estimator= DecisionTreeClassifier (max_depth=3) n_estimators=80 learning_rate=1.0 algorithm='SAMME.R'</div>	<div>Train Accuracy: 1.00 Test Accuracy: 0.93</div> <div>Classification Report (Test):</div> <table><thead><tr><th></th><th>precision</th><th>recall</th><th>f1-score</th><th>support</th></tr></thead><tbody><tr><td>0</td><td>0.90</td><td>0.91</td><td>0.90</td><td>166</td></tr><tr><td>1</td><td>0.95</td><td>0.94</td><td>0.94</td><td>290</td></tr><tr><td>accuracy</td><td></td><td></td><td>0.93</td><td>456</td></tr><tr><td>macro avg</td><td>0.92</td><td>0.93</td><td>0.92</td><td>456</td></tr><tr><td>weighted avg</td><td>0.93</td><td>0.93</td><td>0.93</td><td>456</td></tr></tbody></table> <div>Confusion Matrix (Test):</div> <div>[[151 15] [ 17 273]]</div>		precision	recall	f1-score	support	0	0.90	0.91	0.90	166	1	0.95	0.94	0.94	290	accuracy			0.93	456	macro avg	0.92	0.93	0.92	456	weighted avg	0.93	0.93	0.93	456	<div>שיפורים קלים בזיהוי, אך עדיין מאחור ביחס ל- RandomForest</div>
	precision	recall	f1-score	support																												
0	0.90	0.91	0.90	166																												
1	0.95	0.94	0.94	290																												
accuracy			0.93	456																												
macro avg	0.92	0.93	0.92	456																												
weighted avg	0.93	0.93	0.93	456																												
<div>estimator= DecisionTreeClassifier (max_depth=2) n_estimators=200 learning_rate=0.1 algorithm='SAMME.R'</div>	<div>Train Accuracy: 1.00 Test Accuracy: 0.93</div> <div>Classification Report (Test):</div> <table><thead><tr><th></th><th>precision</th><th>recall</th><th>f1-score</th><th>support</th></tr></thead><tbody><tr><td>0</td><td>0.88</td><td>0.93</td><td>0.90</td><td>166</td></tr><tr><td>1</td><td>0.96</td><td>0.93</td><td>0.94</td><td>290</td></tr><tr><td>accuracy</td><td></td><td></td><td>0.93</td><td>456</td></tr><tr><td>macro avg</td><td>0.92</td><td>0.93</td><td>0.92</td><td>456</td></tr><tr><td>weighted avg</td><td>0.93</td><td>0.93</td><td>0.93</td><td>456</td></tr></tbody></table> <div>Confusion Matrix (Test):</div> <div>[[154 12] [ 21 269]]</div>		precision	recall	f1-score	support	0	0.88	0.93	0.90	166	1	0.96	0.93	0.94	290	accuracy			0.93	456	macro avg	0.92	0.93	0.92	456	weighted avg	0.93	0.93	0.93	456	<div>הקטנת learning_rate ושינוי מספר הלומדים החלשים שיפרו את היציבות והאיזון בביצועים</div>
	precision	recall	f1-score	support																												
0	0.88	0.93	0.90	166																												
1	0.96	0.93	0.94	290																												
accuracy			0.93	456																												
macro avg	0.92	0.93	0.92	456																												
weighted avg	0.93	0.93	0.93	456																												