

בחלק א הגדרנו 3 מסווגים , decision tree , random forest , adaboost , את כולם הגדרנו עם פרמטרים דיפולטיבים ללא הגדרות מיוחדות מלבד random על מנת שנקבל הגרלה קבועה תמיד.
בהמשך אבצע 6 ניסויים (בכל ניסוי שיניתי הגדרות לכל מסווג ובחנתי את התוצאות שלו)
סה"כ 18 בדיקות (6 לכל מסווג) בכל ניסוי ניתן לראות גם מי המסווג עם אחוז הדיוק הגבוה ביותר

```
#decision tree
model_decision_tree = DecisionTreeClassifier(random_state=42)
#random forest
model_random_forest = RandomForestClassifier(random_state=42)
#adaboost
model_adaboost = AdaBoostClassifier(random_state=42)
```

תוצאה:

results for decision_tree:	results for RandomForest	results for AdaBoost:
accuracy:	accuracy:	accuracy:
0.951048951048951	0.965034965034965	0.951048951048951
confusion_matrix:	confusion_matrix:	confusion_matrix:
[[51 3]	[[51 3]	[[53 1]
[4 85]]	[2 87]]	[6 83]]
precision:	precision:	precision:
0.9659090909090909	0.9666666666666667	0.9880952380952381
recall:	recall:	recall:
0.9550561797752809	0.9775280898876404	0.9325842696629213
f-score:	f-score:	f-score:
0.96045197740113	0.9720670391061452	0.9595375722543352

the best classifier based on f-score is: RandomForest

ניסוי מספר 1:

ב Tree בחרתי לשנות את Gini ל Entropy לשנות את הערך הדיפולטי של Splitter , לשנות את ה Split הדיפולטיבי מ 2 ל - 3 ולהגדיר מקסימום עומק ל - 8

ב - Forest בחרתי לשנות את Gini ל Entropy , לשנות את n_estimators מהדיפולט ל - 200 וגם פה להוסיף את אותו העומק וה Split ששמתי בעץ למעלה.

ב - Adaboost שיניתי את ה n_estimators ל 100 ואת learning rate ל 2.5

```
#decision tree
model_decision_tree = DecisionTreeClassifier(random_state=42,
criterion='entropy', splitter = 'random', max_depth=8,
min_samples_split=3)

#random forest
model_random_forest = RandomForestClassifier(random_state=42,
n_estimators=200, criterion='entropy', max_depth=8,
min_samples_split=3)

#adaboost
model_adaboost = AdaBoostClassifier(random_state=42, n_estimators=100,
learning_rate=2.5,)
```

תוצאה :

results for decision_tree:	results for RandomForest:	results for AdaBoost:
accuracy:	accuracy:	accuracy:
0.9300699300699301	0.972027972027972	0.951048951048951
confusion_matrix:	confusion_matrix:	confusion_matrix:
[[50 4]	[[51 3]	[[53 1]
[6 83]]	[1 88]]	[6 83]]
precision:	precision:	precision:
0.9540229885057471	0.967032967032967	0.9880952380952381
recall:	recall:	recall:
0.9325842696629213	0.9887640449438202	0.9325842696629213
f-score:	f-score:	f-score:
0.9431818181818182	0.9777777777777777	0.9595375722543352

the best classifier based on f-score is: RandomForest

ניסוי מספר 2:

ב Tree ו ב Forest בחרתי להחזיר ל Gini עם אותם הפרמטרים מניסוי 1 ולבדוק את ההבדל.
ב - Adaboost שיניתי את ה n_estimators ל 120 ואת learning rate ל 2.0

```
#decision tree
model_decision_tree = DecisionTreeClassifier(random_state=42,
criterion='gini', splitter = 'random', max_depth=8, min_samples_split=3)

#random forest
model_random_forest = RandomForestClassifier(random_state=42,
n_estimators=200, criterion='gini', max_depth=8, min_samples_split=3)

#adaboost
model_adaboost = AdaBoostClassifier(random_state=42, n_estimators=120,
learning_rate=2.0,)
```

תוצאה:

results for decision_tree:	results for RandomForest:	results for AdaBoost:
accuracy: 0.9300699300699301	accuracy: 0.965034965034965	accuracy: 0.916083916083916
confusion_matrix: [[50 4] [6 83]]	confusion_matrix: [[51 3] [2 87]]	confusion_matrix: [[47 7] [5 84]]
precision: 0.9540229885057471	precision: 0.9666666666666667	precision: 0.9230769230769231
recall: 0.9325842696629213	recall: 0.9775280898876404	recall: 0.9438202247191011
f-score: 0.9431818181818182	f-score: 0.9720670391061452	f-score: 0.9333333333333333

the best classifier based on f-score is: RandomForest

ניסוי מספר 3 :

ב Tree בחרתי להגדיל את העומק ל-12 ולהחזיר את ה Split ל-2 (לדיפולט).
ב - Forest בחרתי להגדיל את העומק ל-12 ולהחזיר את ה Split ל-2 (לדיפולט) ו n_estimators ל-100
ב - Adaboost שיניתי את ה n_estimators ל-120 ואת learning rate ל-2.0

```
#decision tree
model_decision_tree = DecisionTreeClassifier(random_state=42,
criterion='gini',splitter = 'random', max_depth=12, min_samples_split=2)

#random forest
model_random_forest = RandomForestClassifier(random_state=42,
n_estimators=100,criterion='gini', max_depth=12, min_samples_split=2)

#adaboost
model_adaboost = AdaBoostClassifier(random_state=42, n_estimators=150,
learning_rate=1.5,)
```

תוצאה:

results for decision_tree:	results for RandomForest:	results for AdaBoost:
accuracy:	accuracy:	accuracy:
0.916083916083916	0.965034965034965	0.965034965034965
confusion_matrix:	confusion_matrix:	confusion_matrix:
[[51 3]	[[51 3]	[[53 1]
[9 80]]	[2 87]]	[4 85]]
precision:	precision:	precision:
0.963855421686747	0.9666666666666667	0.9883720930232558
recall:	recall:	recall:
0.898876404494382	0.9775280898876404	0.9550561797752809
f-score:	f-score:	f-score:
0.9302325581395349	0.9720670391061452	0.9714285714285714

the best classifier based on f-score is: RandomForest

ניסוי מספר 4:

ב Tree בחרתי להקטין את העומק ל-4 את ה-Split לשנות ל-5 ולהחזיר ל Entropy ולהוסיף 2 פרמטרים חדשים
(ccp_alpha, max_features)

ב - Forest בחרתי להקטין את העומק ל-5 את ה-Split ל-5 ו n_estimators ל-150 ולהוסיף פרמטרים חדשים
(ccp_alpha, max_features, min_samples_leaf)

ב - Adaboost שיניתי את ה n_estimators ל-150 ואת learning rate ל-1.6 והוספתי פרמטר חדש (algorithm)

```
#decision tree
model_decision_tree = DecisionTreeClassifier(random_state=42,
criterion='entropy', max_depth=4,
min_samples_split=5, ccp_alpha=0.0199, max_features = 3)

#random forest
model_random_forest = RandomForestClassifier(random_state=42,
n_estimators=150, criterion='entropy', max_depth=5,
min_samples_split=5, min_samples_leaf = 5, ccp_alpha=0.01, max_features=2)

#adaboost
model_adaboost = AdaBoostClassifier(random_state=42, n_estimators=150,
learning_rate=1.6, algorithm='SAMME' )
```

תוצאה:

results for decision_tree:	results for RandomForest:	results for AdaBoost:
accuracy:	accuracy:	accuracy:
0.965034965034965	0.972027972027972	0.9790209790209791
confusion_matrix:	confusion_matrix:	confusion_matrix:
[[51 3]	[[51 3]	[[53 1]
[2 87]]	[1 88]]	[2 87]]
precision:	precision:	precision:
0.9666666666666667	0.967032967032967	0.9886363636363636
recall:	recall:	recall:
0.9775280898876404	0.9887640449438202	0.9775280898876404
f-score:	f-score:	f-score:
0.9720670391061452	0.9777777777777777	0.9830508474576272

the best classifier based on f-score is: AdaBoost

ניסוי מספר 5:

ב Tree בחרתי לשנות ל log loss ולשחק עם שני הפרמטרים חדשים (ccp_alpha , max_features).
ב - Forest בחרתי לשנות ל log loss ו n_estimators ל 100 ולשחק עם הפרמטרים חדשים ccp.
(alpha, max_features ,min_samples_leaf).
ב - Adaboost שיניתי רק את learning rate ל 1.5

```
#decision tree
model_decision_tree = DecisionTreeClassifier(random_state=42,
criterion='log_loss', max_depth=4,
min_samples_split=5, ccp_alpha=0.0199, max_features = 4)

#random forest
model_random_forest = RandomForestClassifier(random_state=42,
n_estimators=100, criterion='log_loss', max_depth=5,
min_samples_split=2, min_samples_leaf = 3, ccp_alpha=0.018)

#adaboost
model_adaboost = AdaBoostClassifier(random_state=42, n_estimators=150,
learning_rate=1.5, algorithm='SAMME' )
```

תוצאה:

results for decision_tree:	results for RandomForest:	results for AdaBoost:
accuracy:	accuracy:	accuracy:
0.972027972027972	0.972027972027972	0.965034965034965
confusion_matrix:	confusion_matrix:	confusion_matrix:
[[54 0]	[[51 3]	[[53 1]
[4 85]]	[1 88]]	[4 85]]
precision:	precision:	precision:
1.0	0.967032967032967	0.9883720930232558
recall:	recall:	recall:
0.9550561797752809	0.9887640449438202	0.9550561797752809
f-score:	f-score:	f-score:
0.9770114942528736	0.9777777777777777	0.9714285714285714
the best classifier based on f-score is: RandomForest		

ניסוי מספר 6:

ב Tree בחרתי לשחק עם שני הפרמטרים (splitter , min_samples_split).
ב - Forest בחרתי לשנות את max_depth להוסיף (min_impurity_decrease , min_weight_fraction_leaf , min_samples_leaf , max_features , alpha)
ב - Adaboost שיניתי רק את learning rate ל 1.0

```
#decision tree
model_decision_tree = DecisionTreeClassifier(random_state=42,
criterion='log_loss', max_depth=4,
min_samples_split=7, ccp_alpha=0.0199, max_features = 4 , splitter = 'random')

#random forest
model_random_forest = RandomForestClassifier(random_state=42,
n_estimators=100, criterion='log_loss', max_depth=None,
min_samples_split=2, ccp_alpha=0.01, max_features=5,
min_weight_fraction_leaf = 0.5, min_impurity_decrease=0.1)

#adaboost
model_adaboost = AdaBoostClassifier(random_state=42, n_estimators=150,
learning_rate=1.0, algorithm='SAMME' )
```

תוצאה:

results for decision_tree:	results for RandomForest:	results for AdaBoost:
accuracy:	accuracy:	accuracy:
0.9230769230769231	0.951048951048951	0.9790209790209791
confusion_matrix:	confusion_matrix:	confusion_matrix:
[[53 1]	[[52 2]	[[53 1]
[10 79]]	[5 84]]	[2 87]]
precision:	precision:	precision:
0.9875	0.9767441860465116	0.9886363636363636
recall:	recall:	recall:
0.8876404494382022	0.9438202247191011	0.9775280898876404
f-score:	f-score:	f-score:
0.9349112426035503	0.96	0.9830508474576272

the best classifier based on f-score is: AdaBoost

ב- Tree בחרתי את ניסוי מספר 5 כי בו ה F-score היה הכי גבוה , ב Forest ו Adaboost
התוצאה הכי גבוהה הייתה זהה בכמה ניסויים ולכן בחרתי את ניסוי מספר 6 אך היה אפשר להגיע
לאותו אחוז גם עם חלק מהניסויים האחרים.