

מבוא לראייה חישובית ממ"ן 12

מאת: יאיר חריט 207282955

התוכניות תחילה טוענות את הקובץ `config.py` המכיל את כל הנדרש על מנת להפעיל את האלגוריתמים ולסכם את הנתונים שייצרו- ייבוא מסדי הנתונים, סידורם, אתחול משתנים, יתר על כן, קובץ זה שומר על אחידות ומונע כפל קוד.

לאחר אתחול המשתנים, נפעיל את האלגוריתם הנבחר על 80% מהנתונים, האלגוריתם יחשב ויסווג את הנתונים. את התוצאות שהתקבלו נשווה מול המידע הקיים במסד הנתונים על מנת לקבוע את טיב האלגוריתם.

עבור אלגוריתם kNN נאסוף את התוצאות עבור ערכי ה-k השונים עד `max_k` ולבסוף נציג את המידע הנאסף בגרף לשם השוואה.

עבור אלגוריתם (י) SVM נאסוף את התוצאות עבור כל מסווג בערכי C שונים ולבסוף נשווה את הנתונים הללו בעזרת גרפים.

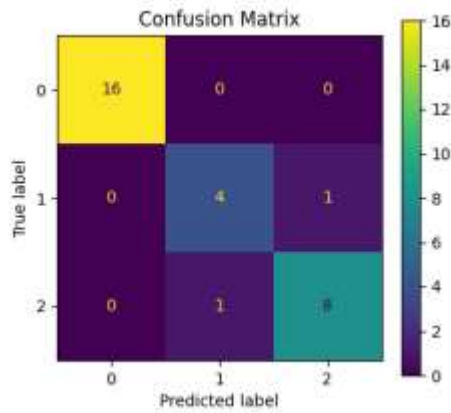
על מנת לבדוק את הקוד ובכדי לשמור על זמן עיבוד הגיוני צמצמתי את MNIST ל-10,000 דוגמיות, המידע הנתון מטה חושב עבור חצי ממסד הנתונים- 30,000 דוגמיות.

להלן חלק מהנתונים שנאספו, שאר הנתונים נמצאים בתיקיית `figures`.

IRIS Dataset

kNN

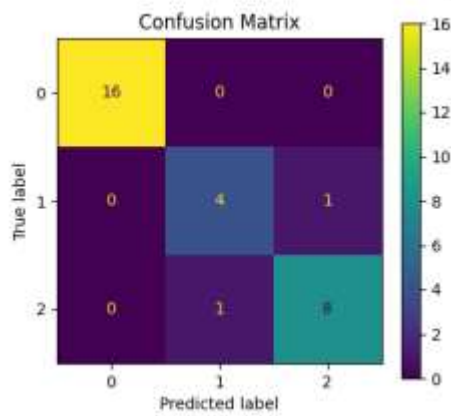
iris 1NN



Classification Report

	precision	recall	f1-score	support
0	1.00	1.00	1.00	16
1	0.80	0.80	0.80	5
2	0.89	0.89	0.89	9
accuracy	0.93			30
macro avg	0.90	0.90	0.90	30
weighted avg	0.93	0.93	0.93	30

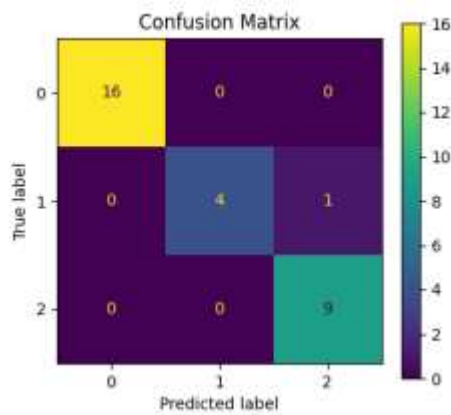
iris 2NN



Classification Report

	precision	recall	f1-score	support
0	1.00	1.00	1.00	16
1	0.80	0.80	0.80	5
2	0.89	0.89	0.89	9
accuracy	0.93			30
macro avg	0.90	0.90	0.90	30
weighted avg	0.93	0.93	0.93	30

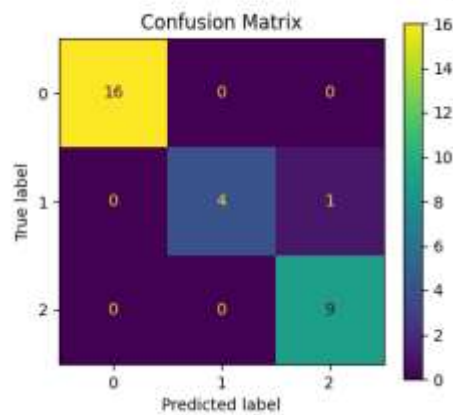
iris 5NN



Classification Report

	precision	recall	f1-score	support
0	1.00	1.00	1.00	16
1	1.00	0.80	0.89	5
2	0.90	1.00	0.95	9
accuracy	0.97			30
macro avg	0.97	0.93	0.95	30
weighted avg	0.97	0.97	0.97	30

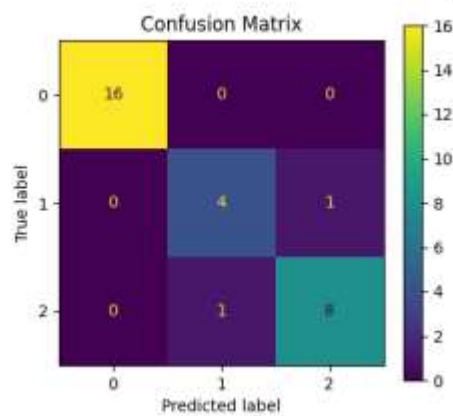
iris 21NN



Classification Report

	precision	recall	f1-score	support
0	1.00	1.00	1.00	16
1	1.00	0.80	0.89	5
2	0.90	1.00	0.95	9
accuracy	0.97			30
macro avg	0.97	0.93	0.95	30
weighted avg	0.97	0.97	0.97	30

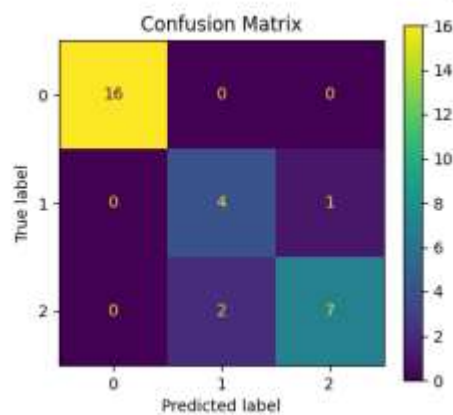
iris 30NN



Classification Report

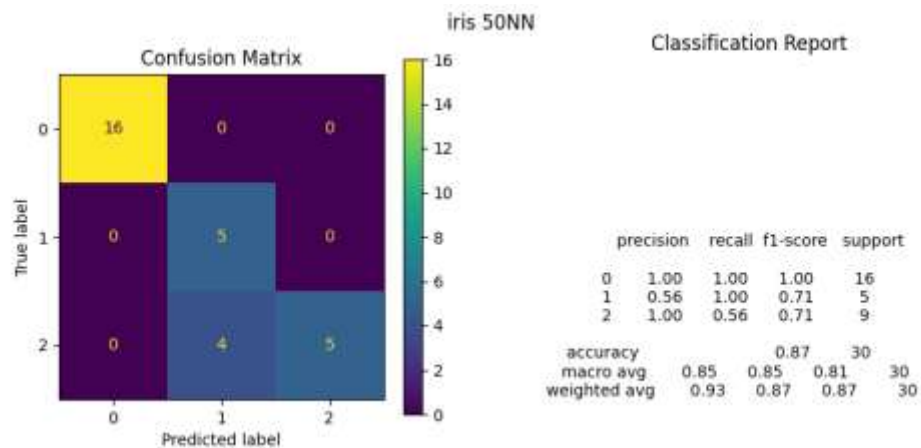
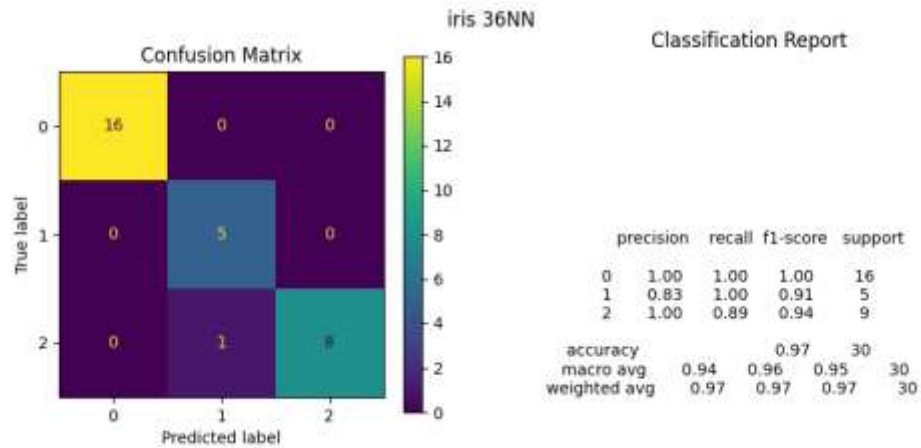
	precision	recall	f1-score	support
0	1.00	1.00	1.00	16
1	0.80	0.80	0.80	5
2	0.89	0.89	0.89	9
accuracy	0.93			30
macro avg	0.90	0.90	0.90	30
weighted avg	0.93	0.93	0.93	30

iris 32NN

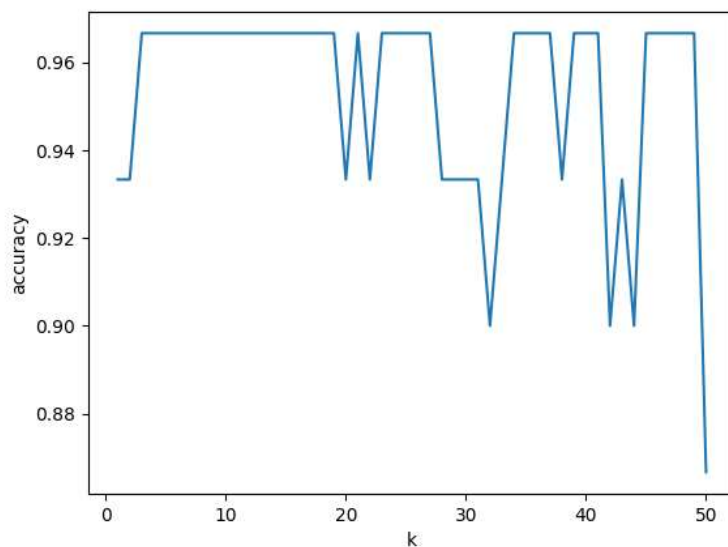


Classification Report

	precision	recall	f1-score	support
0	1.00	1.00	1.00	16
1	0.67	0.80	0.73	5
2	0.88	0.78	0.82	9
accuracy	0.90			30
macro avg	0.85	0.86	0.85	30
weighted avg	0.91	0.90	0.90	30

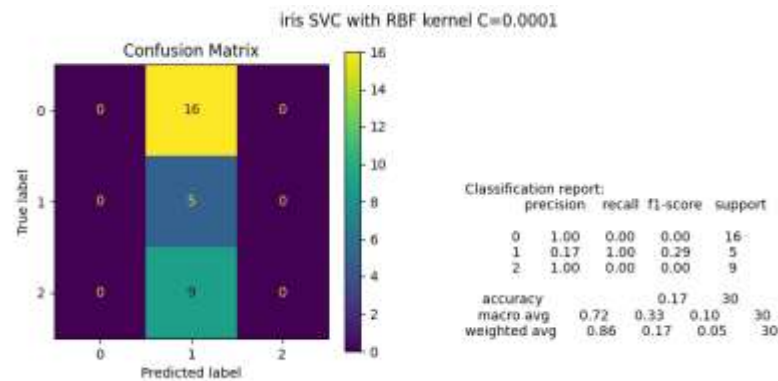
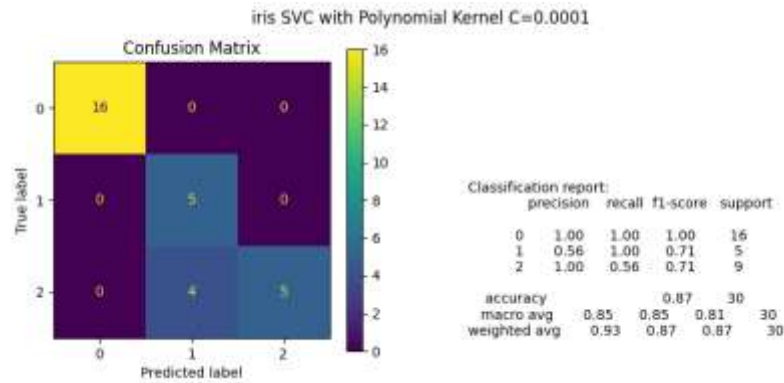
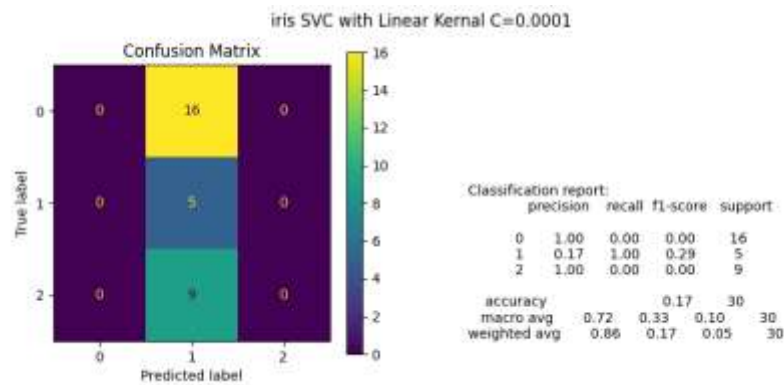
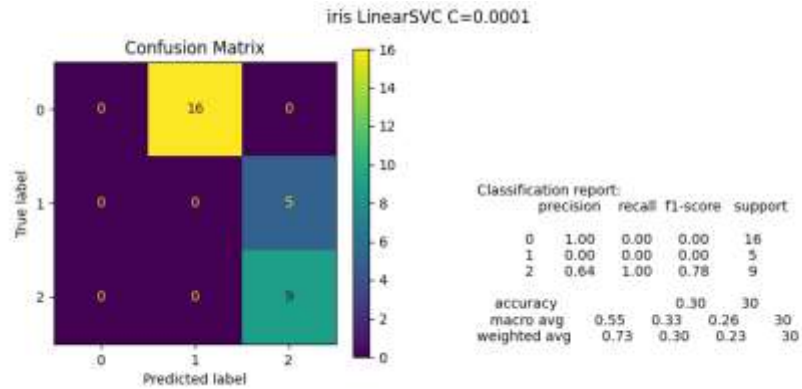


iris kNN 50 summery

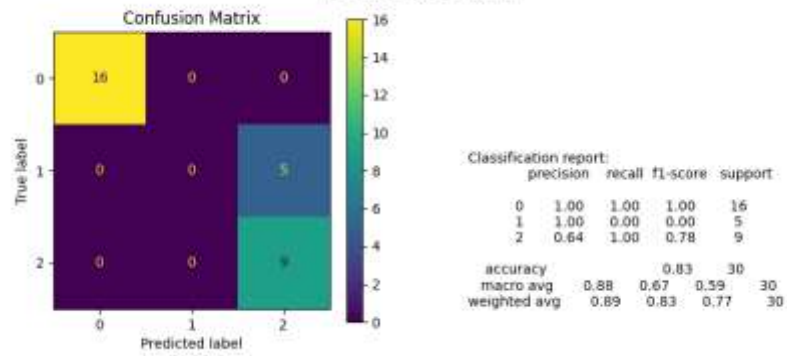


מן הנתונים הללו ניתן להסיק כי מספר השכנים האופטימלי נמוך, שכן כבר כאשר מסווגים בעזרת $k=1$ רמת הדיוק היא 0.93. יתר על כן, ניתן להגיד בפשטנות כי ככל שמספר השכנים גדל רמת הדיוק יורדת.

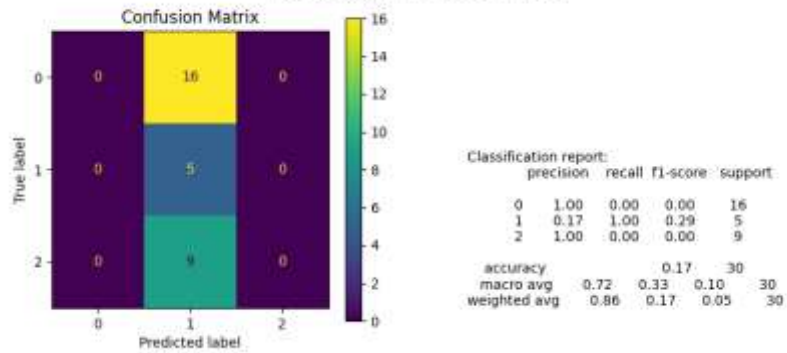
:SVMs



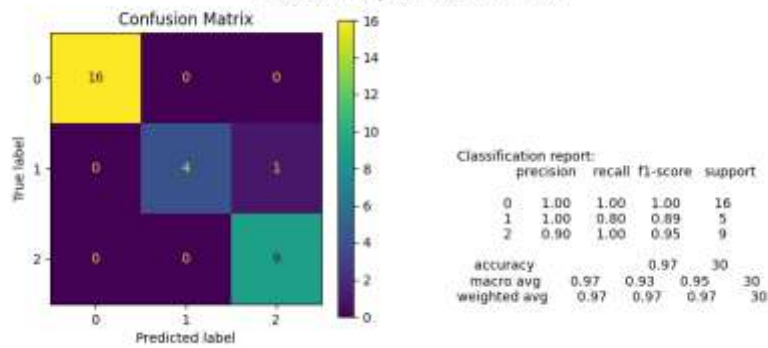
iris LinearSVC C=0.001



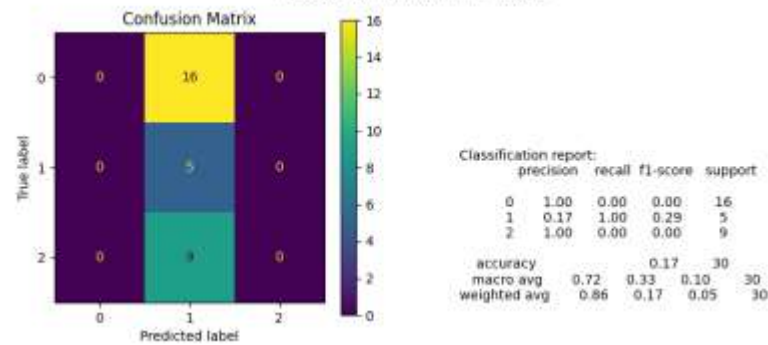
iris SVC with Linear Kernel C=0.001



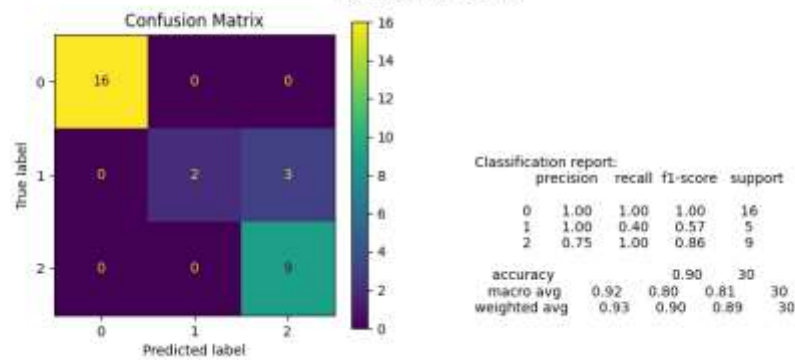
iris SVC with Polynomial Kernel C=0.001



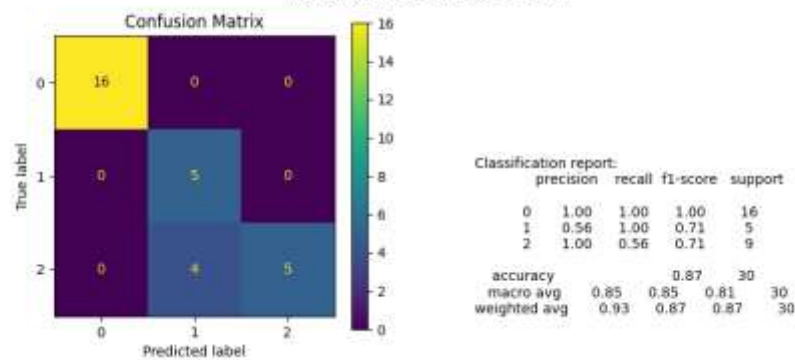
iris SVC with RBF kernel C=0.001



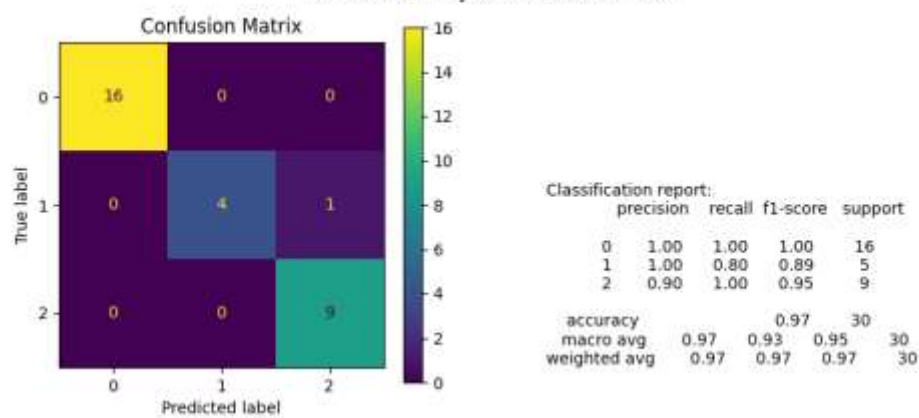
iris LinearSVC C=0.01



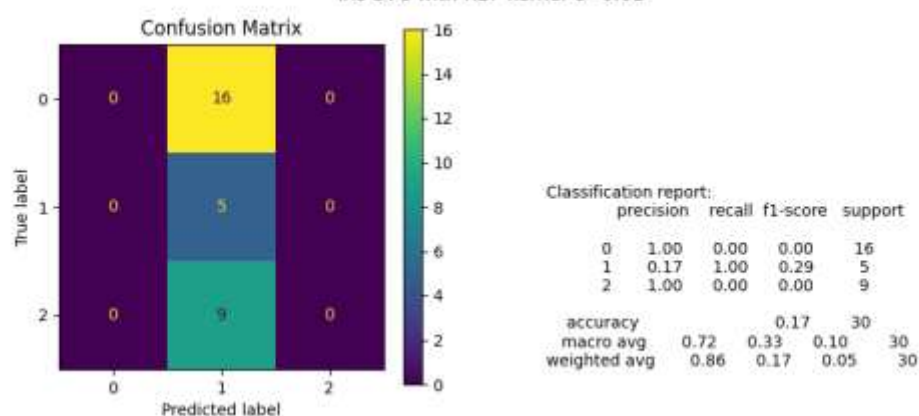
iris SVC with Linear Kernel C=0.01



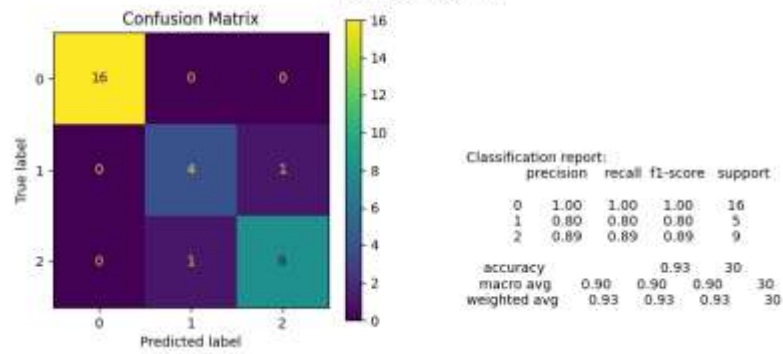
iris SVC with Polynomial Kernel C=0.01



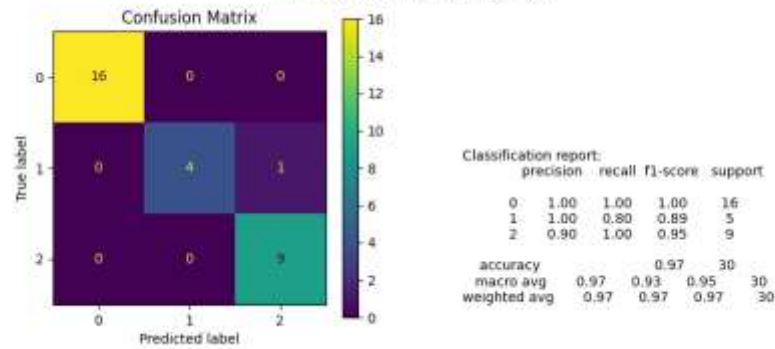
iris SVC with RBF kernel C=0.01



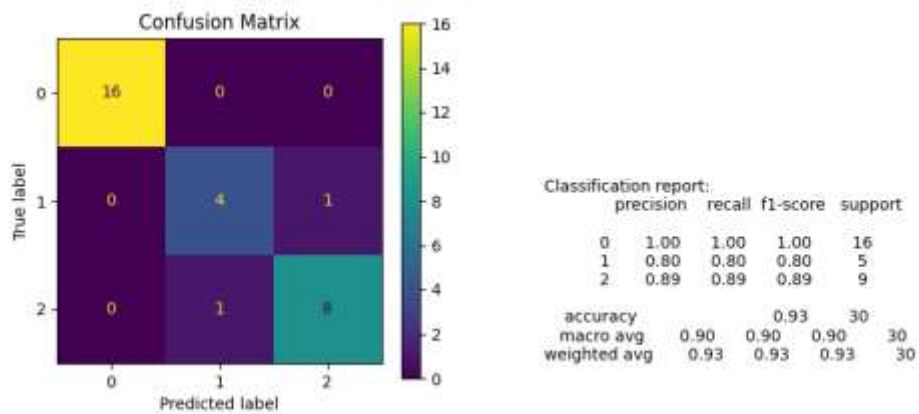
iris LinearSVC C=0.1



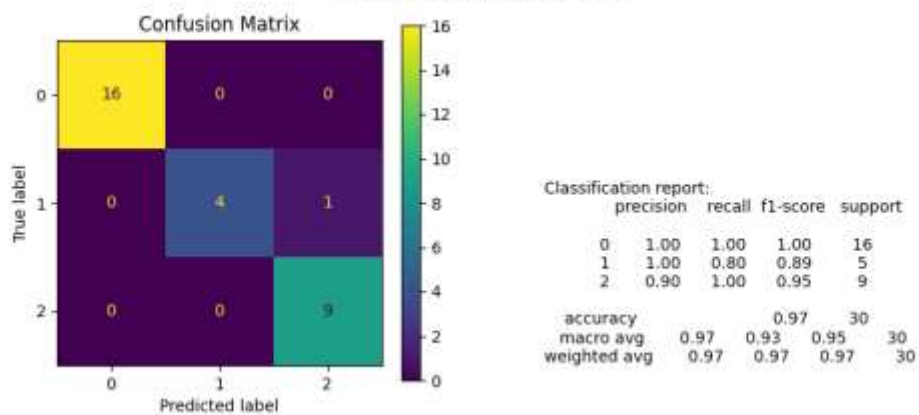
iris SVC with Linear Kernel C=0.1



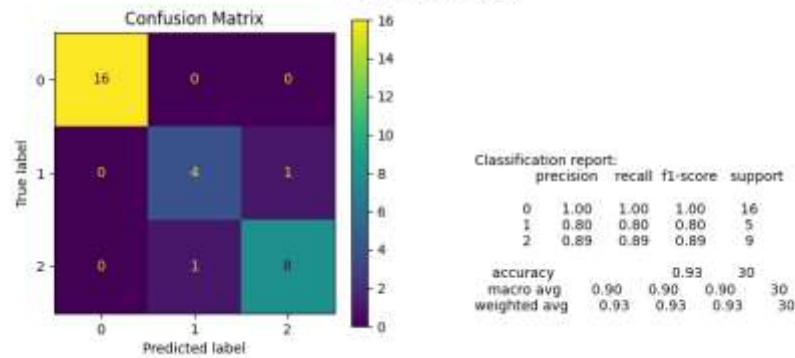
iris SVC with Polynomial Kernel C=0.1



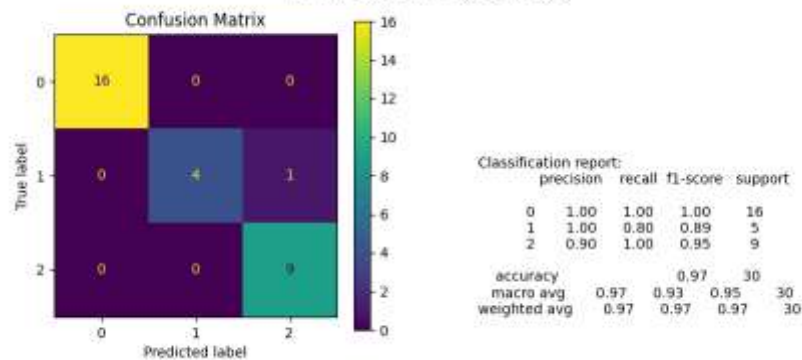
iris SVC with RBF kernel C=0.1



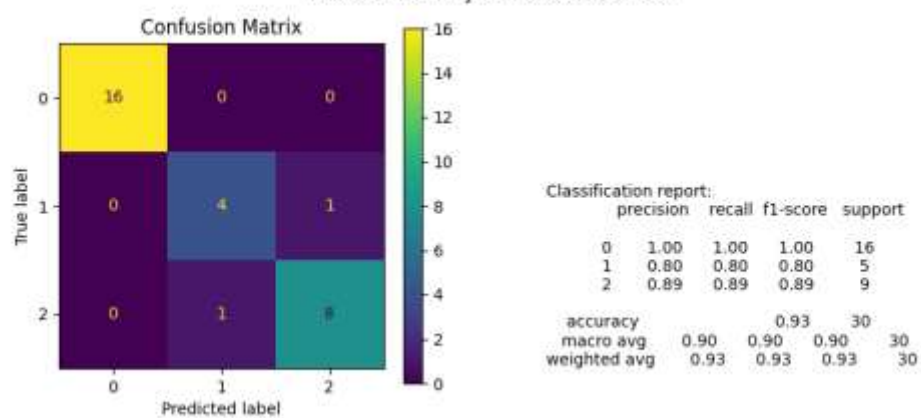
iris LinearSVC C=0.5



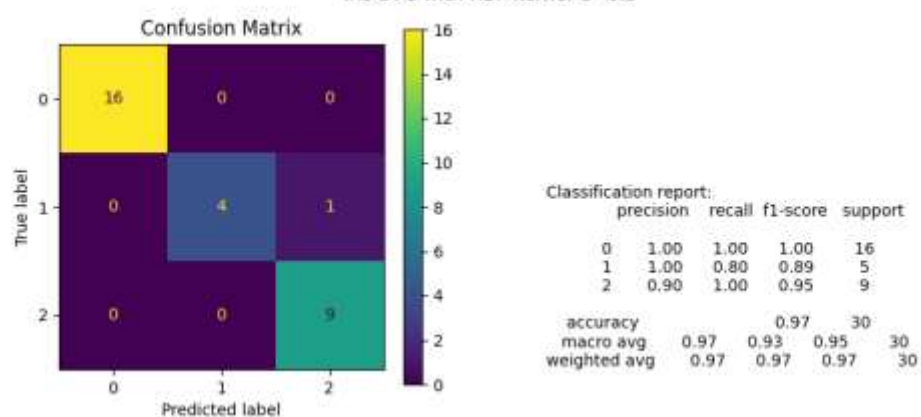
iris SVC with Linear Kernel C=0.5

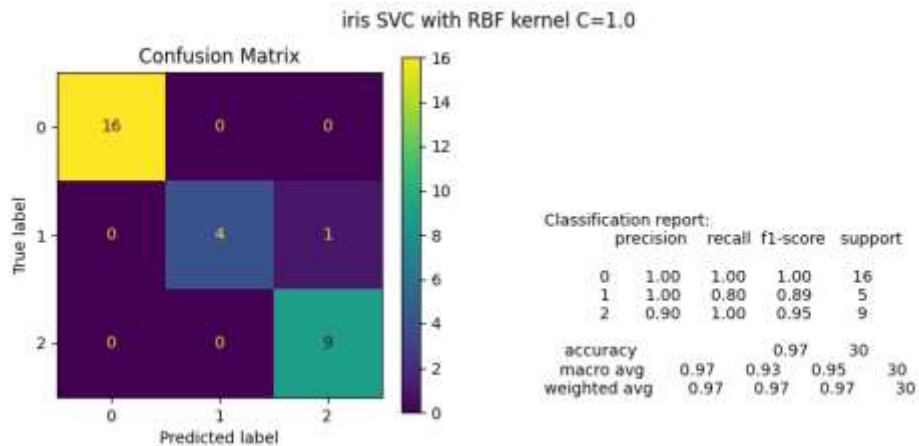
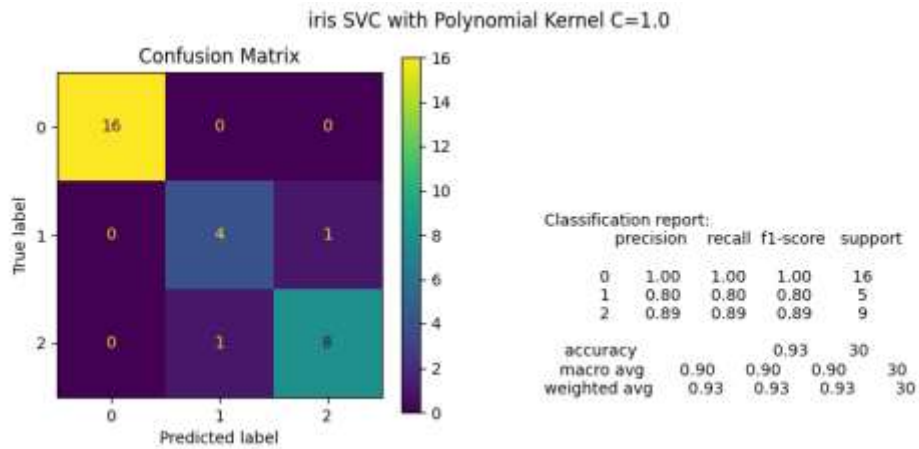
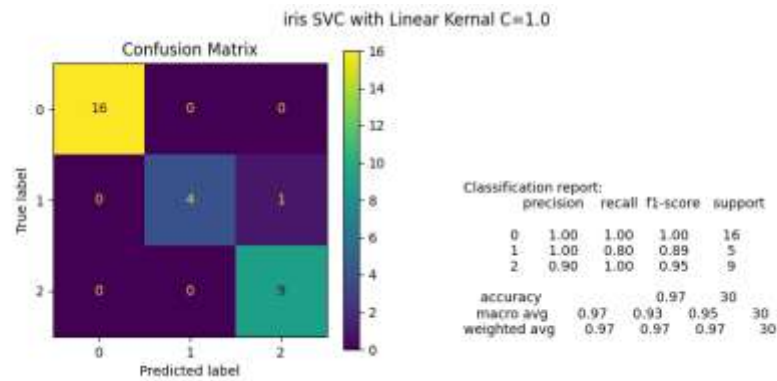
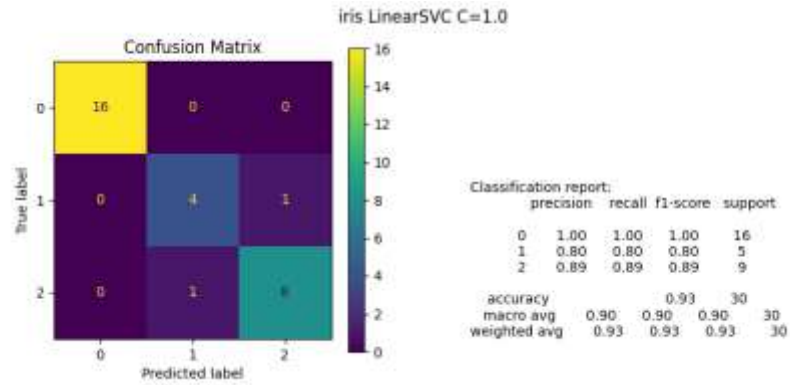


iris SVC with Polynomial Kernel C=0.5

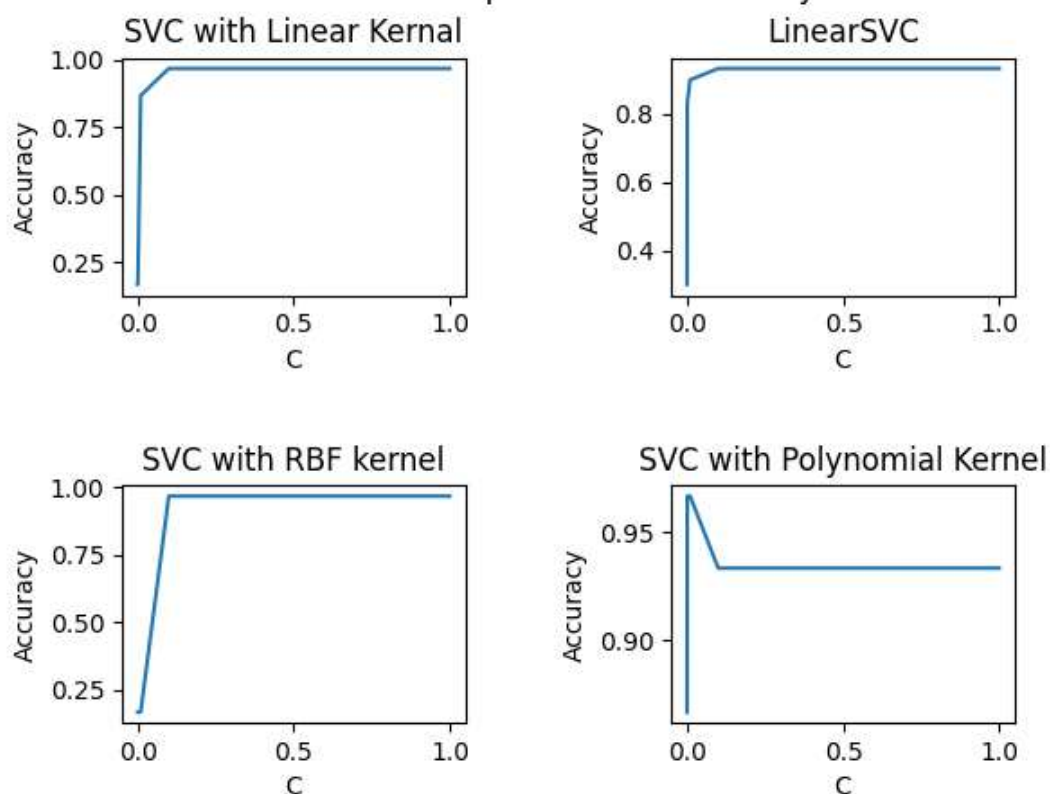


iris SVC with RBF kernel C=0.5





iris SVM sample size 150 summary



מן נתונים אלו ניתן להסיק כי מעבר ל-0.1, במקרה זה, המשקל של הערך של C אינו משפיע רבות על רמת דיוק האלגוריתמים.

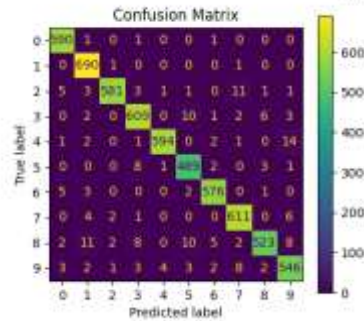
יתר על כן, ניתן לראות כי כאשר $C=1$, 4 האלגוריתמים השונים נותנים תוצאות ברמת דיוק דומה.

לסיכום, ניתן להסיק כי האלגוריתמים שנבדקו יסווגו את מסד הנתונים של IRIS ברמת דיוק זהה ל-KNN, מאחר וזמן העיבוד של KNN טוב יותר- הוא האלגוריתם המתאים יותר במקרה זה.

MNIST

kNN

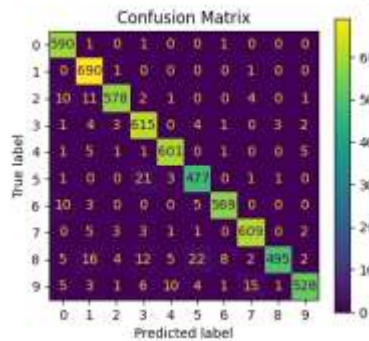
mnist 1NN



Classification Report

	precision	recall	f1-score	support
0	0.97	0.99	0.98	593
1	0.96	1.00	0.98	692
2	0.99	0.96	0.97	607
3	0.96	0.96	0.96	633
4	0.99	0.97	0.98	615
5	0.95	0.97	0.96	504
6	0.98	0.98	0.98	587
7	0.96	0.98	0.97	624
8	0.98	0.92	0.94	571
9	0.94	0.95	0.95	574
accuracy	0.97			6000
macro avg	0.97	0.97	0.97	6000
weighted avg	0.97	0.97	0.97	6000

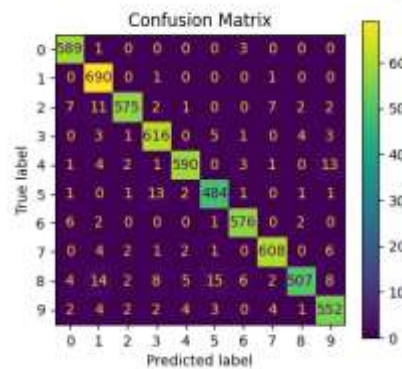
mnist 2NN



Classification Report

	precision	recall	f1-score	support
0	0.95	0.99	0.97	593
1	0.93	1.00	0.97	692
2	0.98	0.95	0.96	607
3	0.93	0.97	0.95	633
4	0.97	0.98	0.97	615
5	0.93	0.95	0.94	504
6	0.98	0.97	0.97	587
7	0.96	0.98	0.97	624
8	0.99	0.87	0.92	571
9	0.98	0.92	0.95	574
accuracy	0.96			6000
macro avg	0.96	0.96	0.96	6000
weighted avg	0.96	0.96	0.96	6000

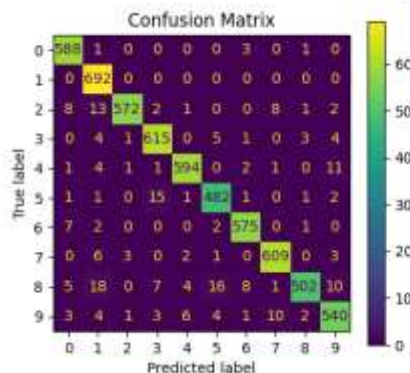
mnist 3NN



Classification Report

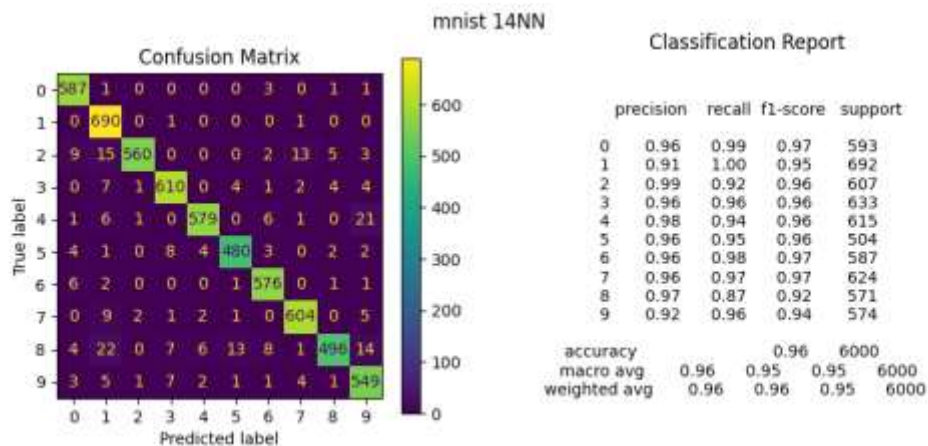
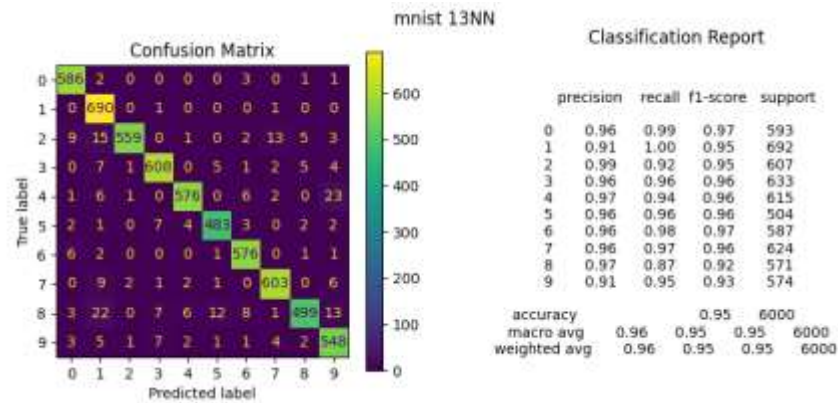
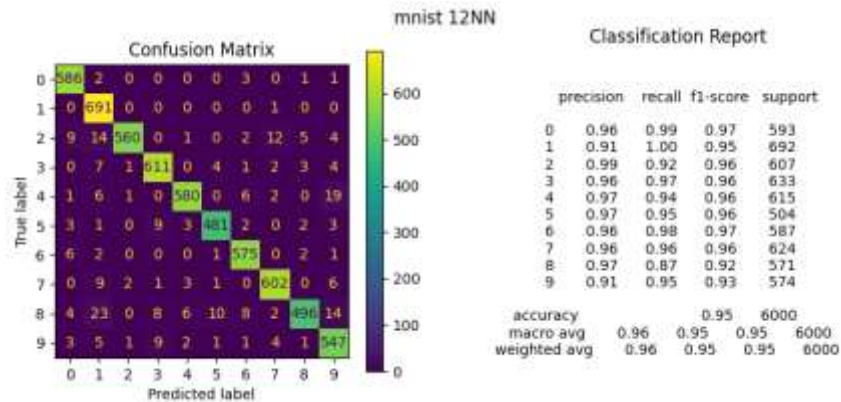
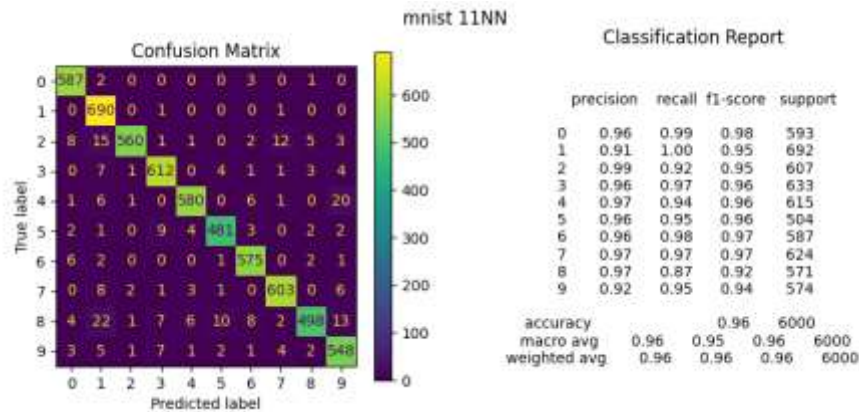
	precision	recall	f1-score	support
0	0.97	0.99	0.98	593
1	0.94	1.00	0.97	692
2	0.98	0.95	0.96	607
3	0.96	0.97	0.96	633
4	0.98	0.96	0.97	615
5	0.95	0.96	0.96	504
6	0.98	0.98	0.98	587
7	0.98	0.97	0.98	624
8	0.98	0.89	0.93	571
9	0.94	0.96	0.95	574
accuracy	0.96			6000
macro avg	0.97	0.96	0.96	6000
weighted avg	0.96	0.96	0.96	6000

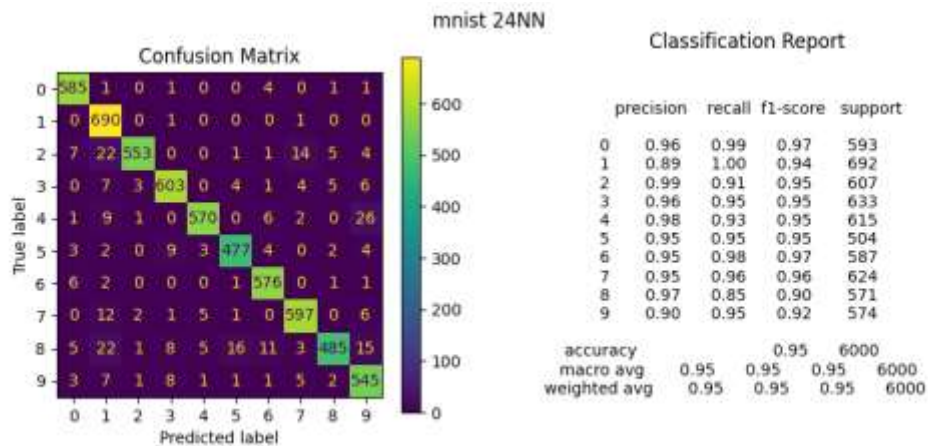
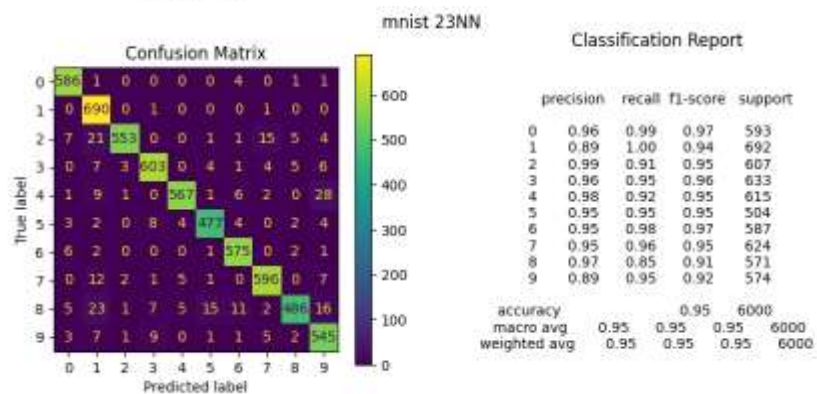
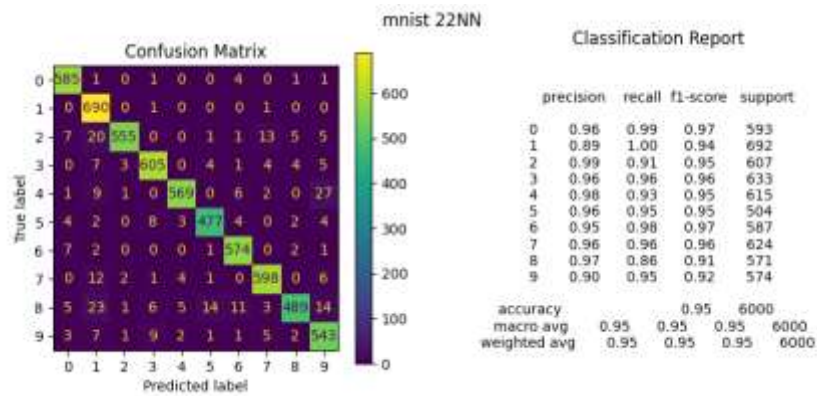
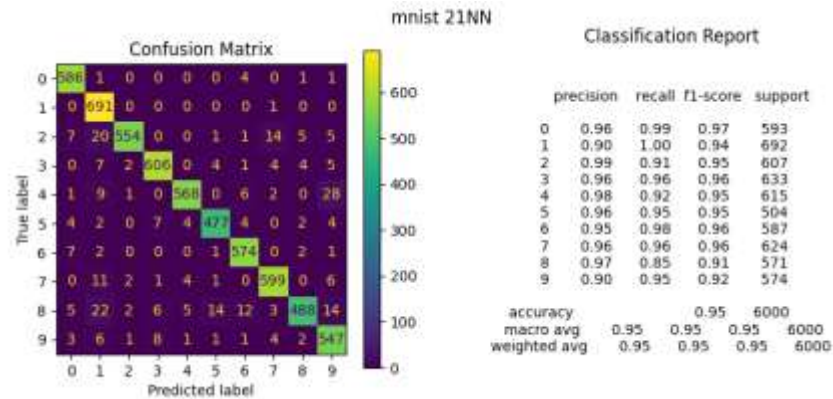
mnist 4NN

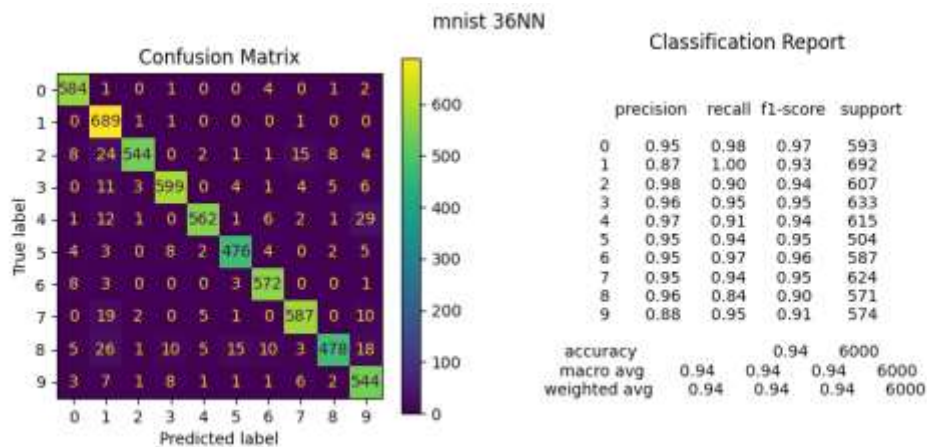
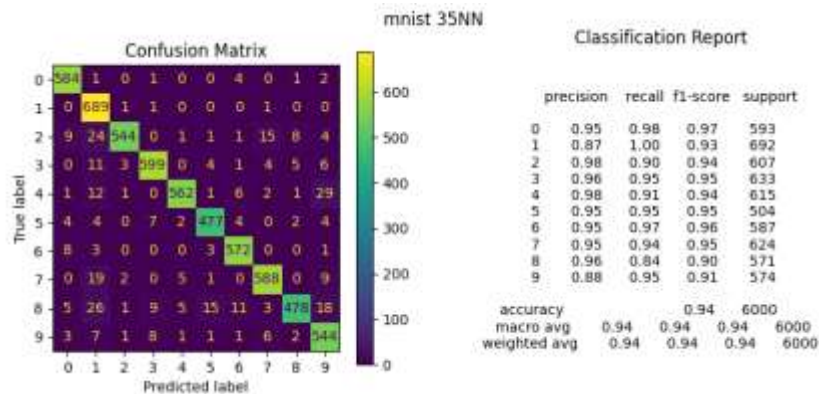
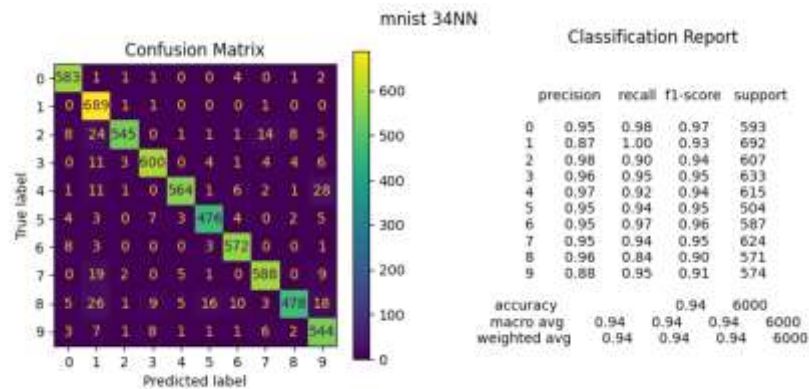
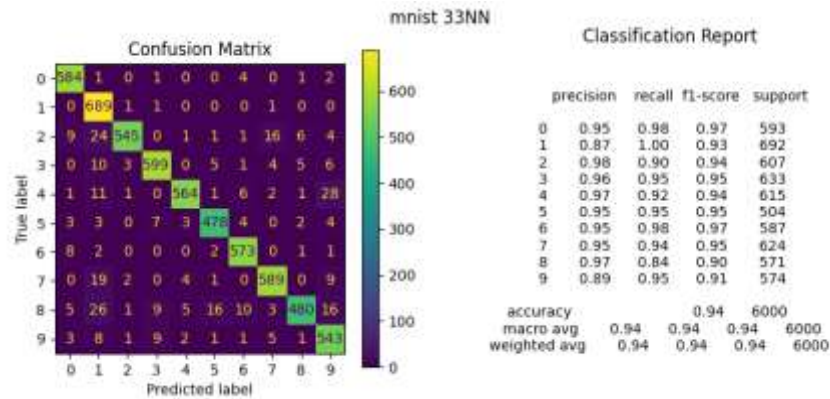


Classification Report

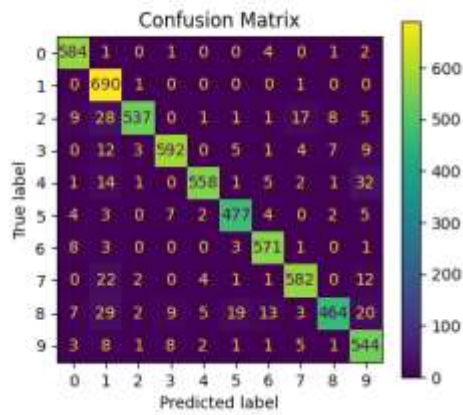
	precision	recall	f1-score	support
0	0.96	0.99	0.98	593
1	0.93	1.00	0.96	692
2	0.99	0.94	0.97	607
3	0.96	0.97	0.96	633
4	0.98	0.97	0.97	615
5	0.95	0.96	0.95	504
6	0.97	0.98	0.98	587
7	0.97	0.98	0.97	624
8	0.98	0.88	0.93	571
9	0.94	0.94	0.94	574
accuracy	0.96			6000
macro avg	0.96	0.96	0.96	6000
weighted avg	0.96	0.96	0.96	6000







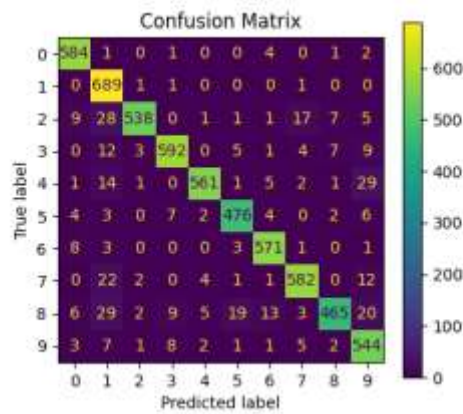
mnist 49NN



Classification Report

	precision	recall	f1-score	support
0	0.95	0.98	0.97	593
1	0.85	1.00	0.92	692
2	0.98	0.88	0.93	607
3	0.96	0.94	0.95	633
4	0.98	0.91	0.94	615
5	0.94	0.95	0.94	504
6	0.95	0.97	0.96	587
7	0.95	0.93	0.94	624
8	0.96	0.81	0.88	571
9	0.86	0.95	0.90	574
accuracy	0.93			6000
macro avg	0.94	0.93	0.93	6000
weighted avg	0.94	0.93	0.93	6000

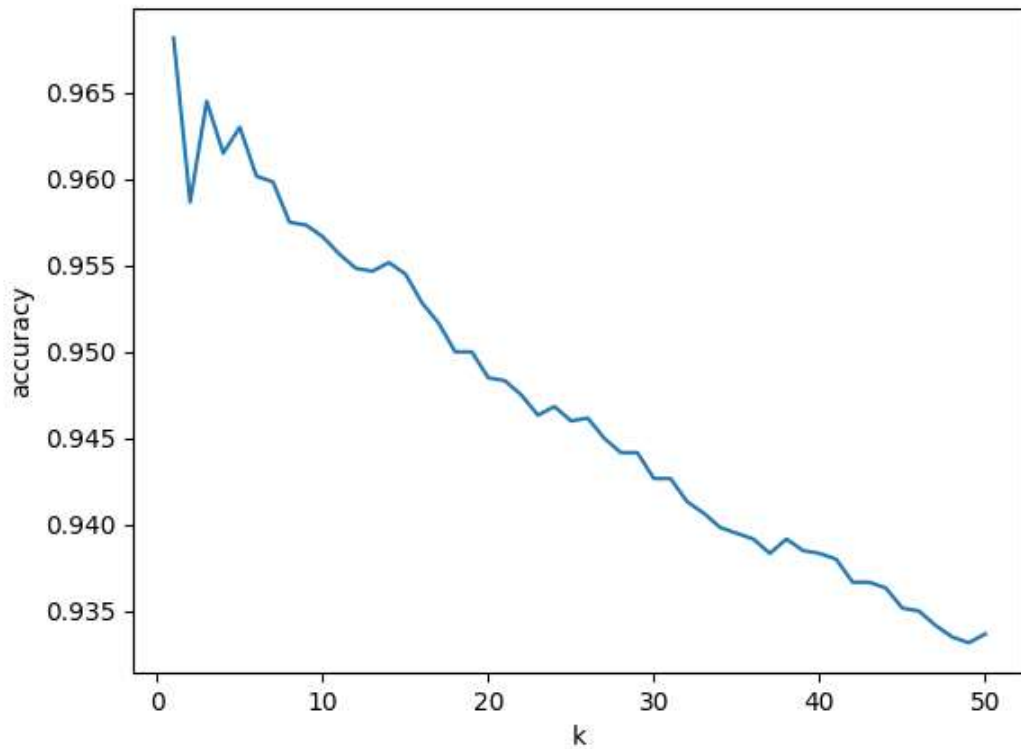
mnist 50NN



Classification Report

	precision	recall	f1-score	support
0	0.95	0.98	0.97	593
1	0.85	1.00	0.92	692
2	0.98	0.89	0.93	607
3	0.96	0.94	0.95	633
4	0.98	0.91	0.94	615
5	0.94	0.94	0.94	504
6	0.95	0.97	0.96	587
7	0.95	0.93	0.94	624
8	0.96	0.81	0.88	571
9	0.87	0.95	0.91	574
accuracy	0.93			6000
macro avg	0.94	0.93	0.93	6000
weighted avg	0.94	0.93	0.93	6000

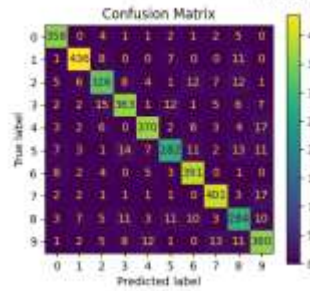
mnist kNN 50 summery



מן הנתונים הללו ניתן להסיק כי מספר השכנים האופטימלי נמוך, כלומר, ככל שנעזרים ביותר שכנים כך רמת הדיוק יורדת (במקרה זה ב-0.03%). ניתן לראות שהשיפוע מתון יותר ביחס ל-kNN עבור IRIS, ייתכן וזאת משום שהערכים (השכנים) אותם נבדוק די דומים האחד לשני-כולם בצבע דומה, עם צפיפות דומה וכו'.

:SVM

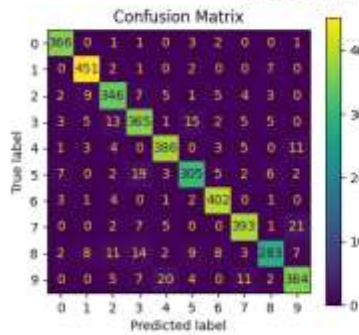
mnist LinearSVC C=0.0001



Classification report:

	precision	recall	f1-score	support
0	0.92	0.96	0.94	374
1	0.94	0.94	0.94	463
2	0.87	0.85	0.86	382
3	0.89	0.88	0.89	414
4	0.92	0.90	0.91	413
5	0.88	0.80	0.84	351
6	0.91	0.94	0.92	414
7	0.92	0.93	0.93	429
8	0.81	0.82	0.81	347
9	0.85	0.87	0.86	413
accuracy			0.89	4000
macro avg	0.89	0.89	0.89	4000
weighted avg	0.89	0.89	0.89	4000

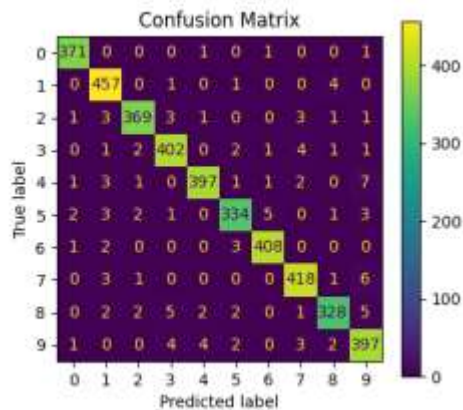
mnist SVC with Linear Kernel C=0.0001



Classification report:

	precision	recall	f1-score	support
0	0.95	0.98	0.97	374
1	0.95	0.97	0.96	463
2	0.89	0.91	0.90	382
3	0.87	0.88	0.87	414
4	0.91	0.93	0.92	413
5	0.89	0.87	0.88	351
6	0.94	0.97	0.96	414
7	0.93	0.92	0.92	429
8	0.92	0.82	0.86	347
9	0.90	0.88	0.89	413
accuracy			0.92	4000
macro avg	0.91	0.91	0.91	4000
weighted avg	0.92	0.92	0.91	4000

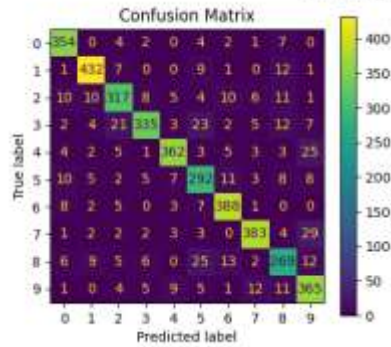
mnist SVC with Polynomial Kernel C=0.0001



Classification report:

	precision	recall	f1-score	support
0	0.98	0.99	0.99	374
1	0.96	0.99	0.98	463
2	0.98	0.97	0.97	382
3	0.97	0.97	0.97	414
4	0.98	0.96	0.97	413
5	0.97	0.95	0.96	351
6	0.98	0.99	0.98	414
7	0.97	0.97	0.97	429
8	0.97	0.95	0.96	347
9	0.94	0.96	0.95	413
accuracy			0.97	4000
macro avg	0.97	0.97	0.97	4000
weighted avg	0.97	0.97	0.97	4000

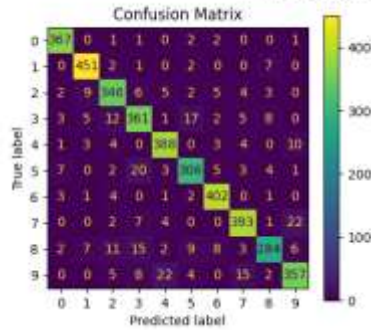
mnist LinearSVC C=0.001



Classification report:

	precision	recall	f1-score	support
0	0.89	0.95	0.92	374
1	0.93	0.93	0.93	463
2	0.85	0.83	0.84	382
3	0.92	0.81	0.86	414
4	0.92	0.88	0.90	413
5	0.78	0.83	0.80	351
6	0.90	0.94	0.92	414
7	0.92	0.89	0.91	429
8	0.80	0.78	0.79	347
9	0.81	0.88	0.85	413
accuracy			0.87	4000
macro avg	0.87	0.87	0.87	4000
weighted avg	0.88	0.87	0.87	4000

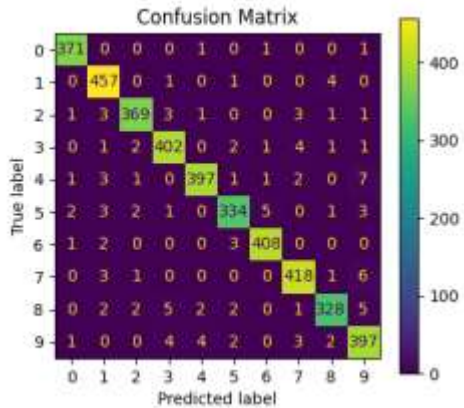
mnist SVC with Linear Kernel C=0.001



Classification report:

	precision	recall	f1-score	support
0	0.95	0.98	0.97	374
1	0.95	0.97	0.96	463
2	0.89	0.91	0.90	382
3	0.86	0.87	0.87	414
4	0.91	0.94	0.92	413
5	0.89	0.87	0.88	351
6	0.94	0.97	0.96	414
7	0.92	0.92	0.92	429
8	0.92	0.82	0.86	347
9	0.90	0.86	0.88	413
accuracy			0.91	4000
macro avg	0.91	0.91	0.91	4000
weighted avg	0.91	0.91	0.91	4000

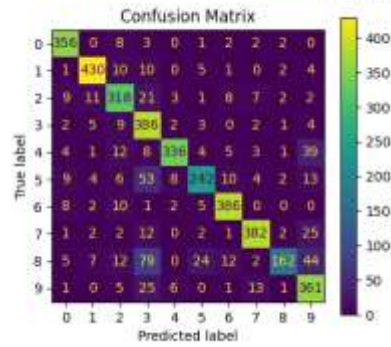
mnist SVC with Polynomial Kernel C=0.001



Classification report:

	precision	recall	f1-score	support
0	0.98	0.99	0.99	374
1	0.96	0.99	0.98	463
2	0.98	0.97	0.97	382
3	0.97	0.97	0.97	414
4	0.98	0.96	0.97	413
5	0.97	0.95	0.96	351
6	0.98	0.99	0.98	414
7	0.97	0.97	0.97	429
8	0.97	0.95	0.96	347
9	0.94	0.96	0.95	413
accuracy			0.97	4000
macro avg	0.97	0.97	0.97	4000
weighted avg	0.97	0.97	0.97	4000

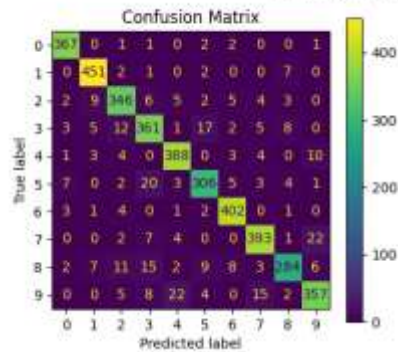
mnist LinearSVC C=0.01



Classification report:

	precision	recall	f1-score	support
0	0.90	0.95	0.92	374
1	0.93	0.93	0.93	463
2	0.81	0.83	0.82	382
3	0.65	0.93	0.76	414
4	0.94	0.81	0.87	413
5	0.84	0.69	0.76	351
6	0.91	0.93	0.92	414
7	0.92	0.89	0.91	429
8	0.93	0.47	0.62	347
9	0.73	0.87	0.80	413
accuracy			0.84	4000
macro avg	0.86	0.83	0.83	4000
weighted avg	0.86	0.84	0.84	4000

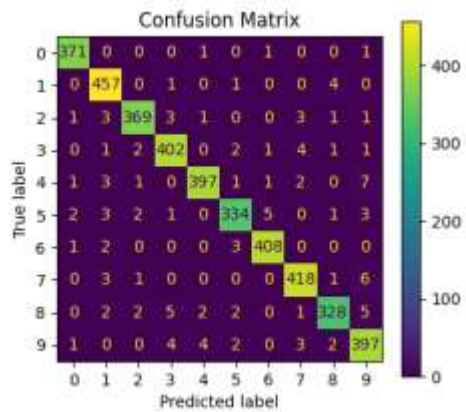
mnist SVC with Linear Kernel C=0.01



Classification report:

	precision	recall	f1-score	support
0	0.95	0.98	0.97	374
1	0.95	0.97	0.96	463
2	0.89	0.91	0.90	382
3	0.86	0.87	0.87	414
4	0.91	0.94	0.92	413
5	0.89	0.87	0.88	351
6	0.94	0.97	0.96	414
7	0.92	0.92	0.92	429
8	0.92	0.82	0.86	347
9	0.90	0.86	0.88	413
accuracy			0.91	4000
macro avg	0.91	0.91	0.91	4000
weighted avg	0.91	0.91	0.91	4000

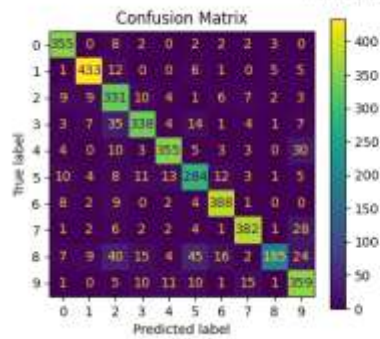
mnist SVC with Polynomial Kernel C=0.01



Classification report:

	precision	recall	f1-score	support
0	0.98	0.99	0.99	374
1	0.96	0.99	0.98	463
2	0.98	0.97	0.97	382
3	0.97	0.97	0.97	414
4	0.98	0.96	0.97	413
5	0.97	0.95	0.96	351
6	0.98	0.99	0.98	414
7	0.97	0.97	0.97	429
8	0.97	0.95	0.96	347
9	0.94	0.96	0.95	413
accuracy			0.97	4000
macro avg	0.97	0.97	0.97	4000
weighted avg	0.97	0.97	0.97	4000

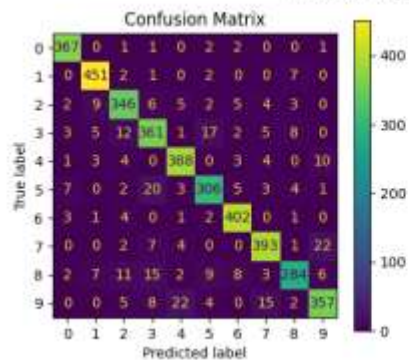
mnist LinearSVC C=0.1



Classification report:

	precision	recall	f1-score	support
0	0.89	0.95	0.92	374
1	0.93	0.94	0.93	463
2	0.71	0.87	0.78	382
3	0.86	0.82	0.84	414
4	0.90	0.86	0.88	413
5	0.76	0.81	0.78	351
6	0.90	0.94	0.92	414
7	0.91	0.89	0.90	429
8	0.93	0.53	0.68	347
9	0.78	0.87	0.82	413
accuracy			0.85	4000
macro avg	0.86	0.85	0.85	4000
weighted avg	0.86	0.85	0.85	4000

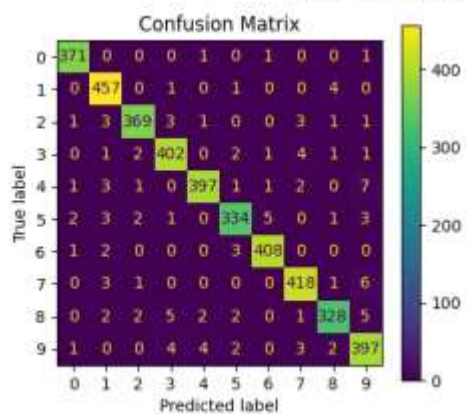
mnist SVC with Linear Kernel C=0.1



Classification report:

	precision	recall	f1-score	support
0	0.95	0.98	0.97	374
1	0.95	0.97	0.96	463
2	0.89	0.91	0.90	382
3	0.86	0.87	0.87	414
4	0.91	0.94	0.92	413
5	0.89	0.87	0.88	351
6	0.94	0.97	0.96	414
7	0.92	0.92	0.92	429
8	0.92	0.82	0.86	347
9	0.90	0.86	0.88	413
accuracy			0.91	4000
macro avg	0.91	0.91	0.91	4000
weighted avg	0.91	0.91	0.91	4000

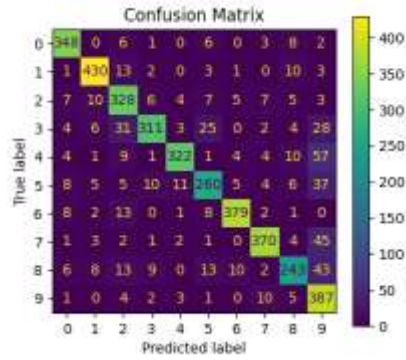
mnist SVC with Polynomial Kernel C=0.1



Classification report:

	precision	recall	f1-score	support
0	0.98	0.99	0.99	374
1	0.96	0.99	0.98	463
2	0.98	0.97	0.97	382
3	0.97	0.97	0.97	414
4	0.98	0.96	0.97	413
5	0.97	0.95	0.96	351
6	0.98	0.99	0.98	414
7	0.97	0.97	0.97	429
8	0.97	0.95	0.96	347
9	0.94	0.96	0.95	413
accuracy			0.97	4000
macro avg	0.97	0.97	0.97	4000
weighted avg	0.97	0.97	0.97	4000

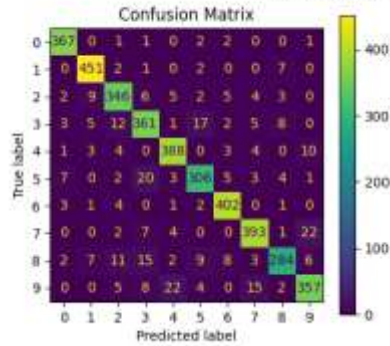
mnist LinearSVC C=0.5



Classification report:

	precision	recall	f1-score	support
0	0.90	0.93	0.91	374
1	0.92	0.93	0.93	463
2	0.77	0.86	0.81	382
3	0.91	0.75	0.82	414
4	0.93	0.78	0.85	413
5	0.80	0.74	0.77	351
6	0.94	0.92	0.93	414
7	0.92	0.86	0.89	429
8	0.82	0.70	0.76	347
9	0.64	0.94	0.76	413
accuracy			0.84	4000
macro avg	0.85	0.84	0.84	4000
weighted avg	0.86	0.84	0.85	4000

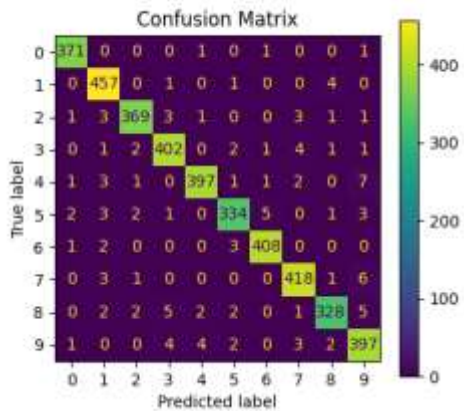
mnist SVC with Linear Kernel C=0.5



Classification report:

	precision	recall	f1-score	support
0	0.95	0.98	0.97	374
1	0.95	0.97	0.96	463
2	0.89	0.91	0.90	382
3	0.86	0.87	0.87	414
4	0.91	0.94	0.92	413
5	0.89	0.87	0.88	351
6	0.94	0.97	0.96	414
7	0.92	0.92	0.92	429
8	0.92	0.82	0.86	347
9	0.90	0.86	0.88	413
accuracy			0.91	4000
macro avg	0.91	0.91	0.91	4000
weighted avg	0.91	0.91	0.91	4000

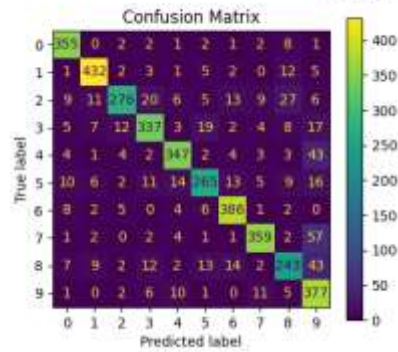
mnist SVC with Polynomial Kernel C=0.5



Classification report:

	precision	recall	f1-score	support
0	0.98	0.99	0.99	374
1	0.96	0.99	0.98	463
2	0.98	0.97	0.97	382
3	0.97	0.97	0.97	414
4	0.98	0.96	0.97	413
5	0.97	0.95	0.96	351
6	0.98	0.99	0.98	414
7	0.97	0.97	0.97	429
8	0.97	0.95	0.96	347
9	0.94	0.96	0.95	413
accuracy			0.97	4000
macro avg	0.97	0.97	0.97	4000
weighted avg	0.97	0.97	0.97	4000

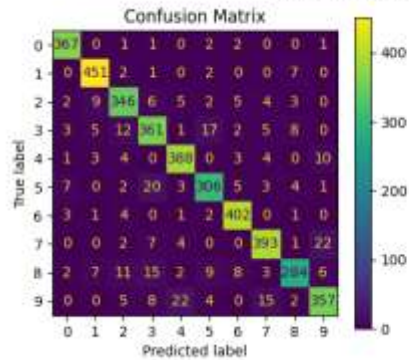
mnist LinearSVC C=1.0



Classification report:

	precision	recall	f1-score	support
0	0.89	0.95	0.92	374
1	0.92	0.93	0.93	463
2	0.90	0.72	0.80	382
3	0.85	0.81	0.83	414
4	0.89	0.84	0.86	413
5	0.83	0.75	0.79	351
6	0.89	0.93	0.91	414
7	0.91	0.84	0.87	429
8	0.76	0.70	0.73	347
9	0.67	0.91	0.77	413
accuracy			0.84	4000
macro avg	0.85	0.84	0.84	4000
weighted avg	0.85	0.84	0.84	4000

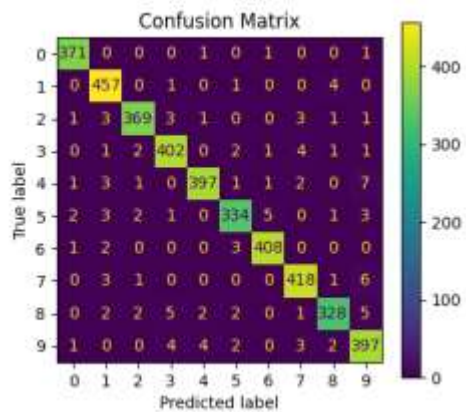
mnist SVC with Linear Kernel C=1.0



Classification report:

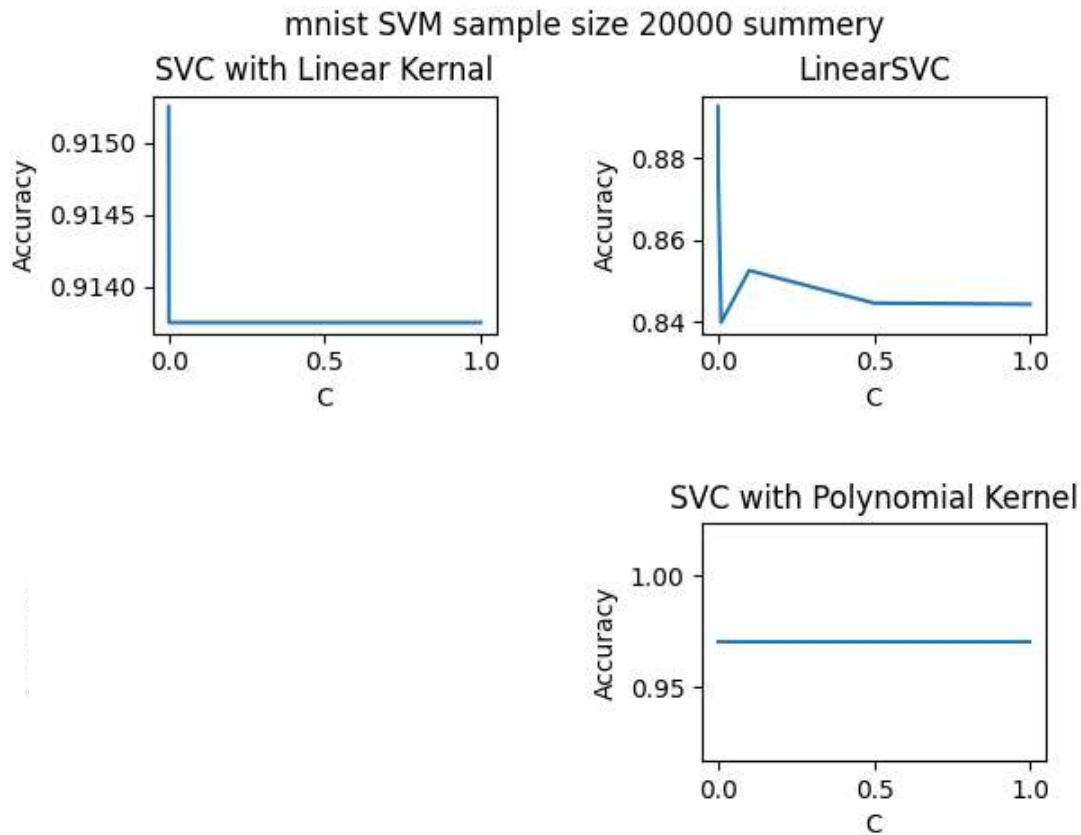
	precision	recall	f1-score	support
0	0.95	0.98	0.97	374
1	0.95	0.97	0.96	463
2	0.89	0.91	0.90	382
3	0.86	0.87	0.87	414
4	0.91	0.94	0.92	413
5	0.89	0.87	0.88	351
6	0.94	0.97	0.96	414
7	0.92	0.92	0.92	429
8	0.92	0.82	0.86	347
9	0.90	0.86	0.88	413
accuracy			0.91	4000
macro avg	0.91	0.91	0.91	4000
weighted avg	0.91	0.91	0.91	4000

mnist SVC with Polynomial Kernel C=1.0



Classification report:

	precision	recall	f1-score	support
0	0.98	0.99	0.99	374
1	0.96	0.99	0.98	463
2	0.98	0.97	0.97	382
3	0.97	0.97	0.97	414
4	0.98	0.96	0.97	413
5	0.97	0.95	0.96	351
6	0.98	0.99	0.98	414
7	0.97	0.97	0.97	429
8	0.97	0.95	0.96	347
9	0.94	0.96	0.95	413
accuracy			0.97	4000
macro avg	0.97	0.97	0.97	4000
weighted avg	0.97	0.97	0.97	4000



מן הנתונים הללו ניתן לראות כי התוצאות הטובות ביותר מגיעות משימוש באלגוריתם SVC עם גרעין פולינומיאלי (דרגה 3). יתר על כן, ניתן להגיד כי ערכי ה-C המשתנים אינם משפיעים רבות על רמת הדיוק של האלגוריתמים שנבדקו (Poly SVC לא מושפע בכלל).

למרות שרמת הדיוק של האלגוריתם המדובר טובה יותר משל אלגוריתם kNN, אני הייתי משתמש ב-kNN בשל זמן הריצה הארוך של האלגוריתמים המוזכרים בחלק זה (כ-50 דקות עבור 20,000 דוגמיות של MNIST).