

گزارش کامل کلاس بندی فایل :

KNN classification report :

	precision	recall	f1-score	support
804	0.81	0.85	0.78	1
577	0.71	0.67	0.76	0
accuracy		0.77		1381
macro avg	0.77	0.76	0.76	1381
weighted avg	0.77	0.77	0.77	1381

SVM linear classification report :

	precision	recall	f1-score	support
804	0.53	0.38	0.91	1
577	0.67	0.95	0.52	0
accuracy		0.62		1381
macro avg	0.72	0.66	0.60	1381
weighted avg	0.75	0.62	0.59	1381

SVM non_linear classification report :

	precision	recall	f1-score	support
804	0.90	0.92	0.89	1
577	0.86	0.84	0.88	0
accuracy		0.88		1381
macro avg	0.88	0.88	0.88	1381
weighted avg	0.88	0.88	0.88	1381

LogisticRegression classification report :

	precision	recall	f1-score	support
804	0.92	0.95	0.89	1
577	0.88	0.83	0.93	0

accuracy		0.90		1381
macro avg	0.91	0.89	0.90	1381
weighted avg	0.90	0.90	0.90	1381

```
*****
*****
*****
*****
```

NaiveBayes classification report :

	precision	recall	f1-score	support
804	0.87	0.81	0.95	1
577	0.85	0.94	0.78	0

accuracy		0.86		1381
macro avg	0.86	0.87	0.86	1381
weighted avg	0.88	0.86	0.86	1381

```
*****
*****
*****
*****
```

DecisionTree classification report :

	precision	recall	f1-score	support
--	-----------	--------	----------	---------

804	0.93	0.95	0.91	1
-----	------	------	------	---

577	0.90	0.87	0.93	0
-----	------	------	------	---

accuracy		0.92		1381
----------	--	------	--	------

macro avg	0.92	0.91	0.91	1381
-----------	------	------	------	------

weighted avg	0.92	0.92	0.91	1381
--------------	------	------	------	------

AdaBoost classification report :

	precision	recall	f1-score	support
--	-----------	--------	----------	---------

804	0.94	0.97	0.92	1
-----	------	------	------	---

577	0.92	0.89	0.95	0
-----	------	------	------	---

accuracy		0.93		1381
----------	--	------	--	------

macro avg	0.94	0.93	0.93	1381
-----------	------	------	------	------

weighted avg	0.93	0.93	0.93	1381
--------------	------	------	------	------

RandomForrest classification report :

	precision	recall	f1-score	support
804	0.95	0.97	0.94	1
577	0.93	0.91	0.95	0
accuracy		0.94		1381
macro avg	0.94	0.94	0.94	1381
weighted avg	0.94	0.94	0.94	1381

For this SPECIFIC dataset :

train times :

{ 'NaiveBayes': 0.005340576171875,	'KNN': 0.005997657775878906,
'DecisionTree': 0.054894208908081055,	'SVM linear': 0.26819419860839844,
'SVM non_linear': 0.58949589729,	'LogisticRegression': 1.255558967,
'AdaBoost': 1.507398843,	'RandomForrest': 1.74252605 }

test times :

{'LogisticRegression': 0.0025055408, 'NaiveBayes': 0.004504203,
'DecisionTree': 0.00514984130859375, 'SVM linear': 0.03130459785461426,
'RandomForrest': 0.048981666564941406, 'AdaBoost': 0.07058262825012207,
'KNN': 0.12496495246887207, 'SVM non_linear': 0.222419500}

accuracies :

{'RandomForrest': 0.941346850108617, 'AdaBoost': 0.9333816075307748,
'DecisionTree': 0.9152787834902245, 'LogisticRegression': 0.90296886314 ,
'SVM non_linear': 0.88269370021, 'NaiveBayes': 0.8616944243301955,
'KNN': 0.7733526430123099, 'SVM linear': 0.6169442433019551}

مقایسه زمان تمرین :

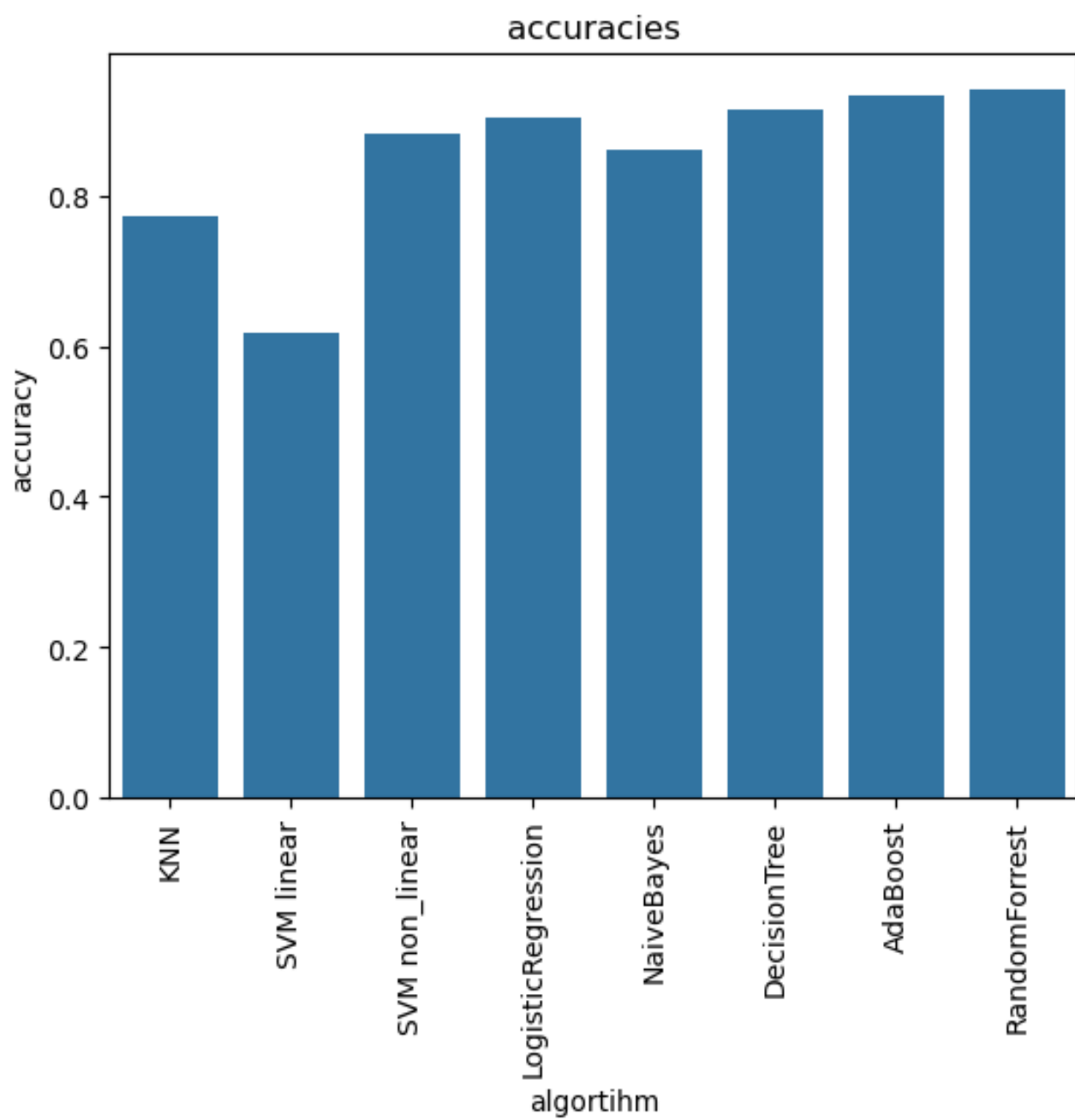
1. KNN
2. Naïve Bayes
3. Decision tree
4. Svm linear
5. Svm non linear
6. Logistic regression
7. Adaboost
8. Radnom Forrest

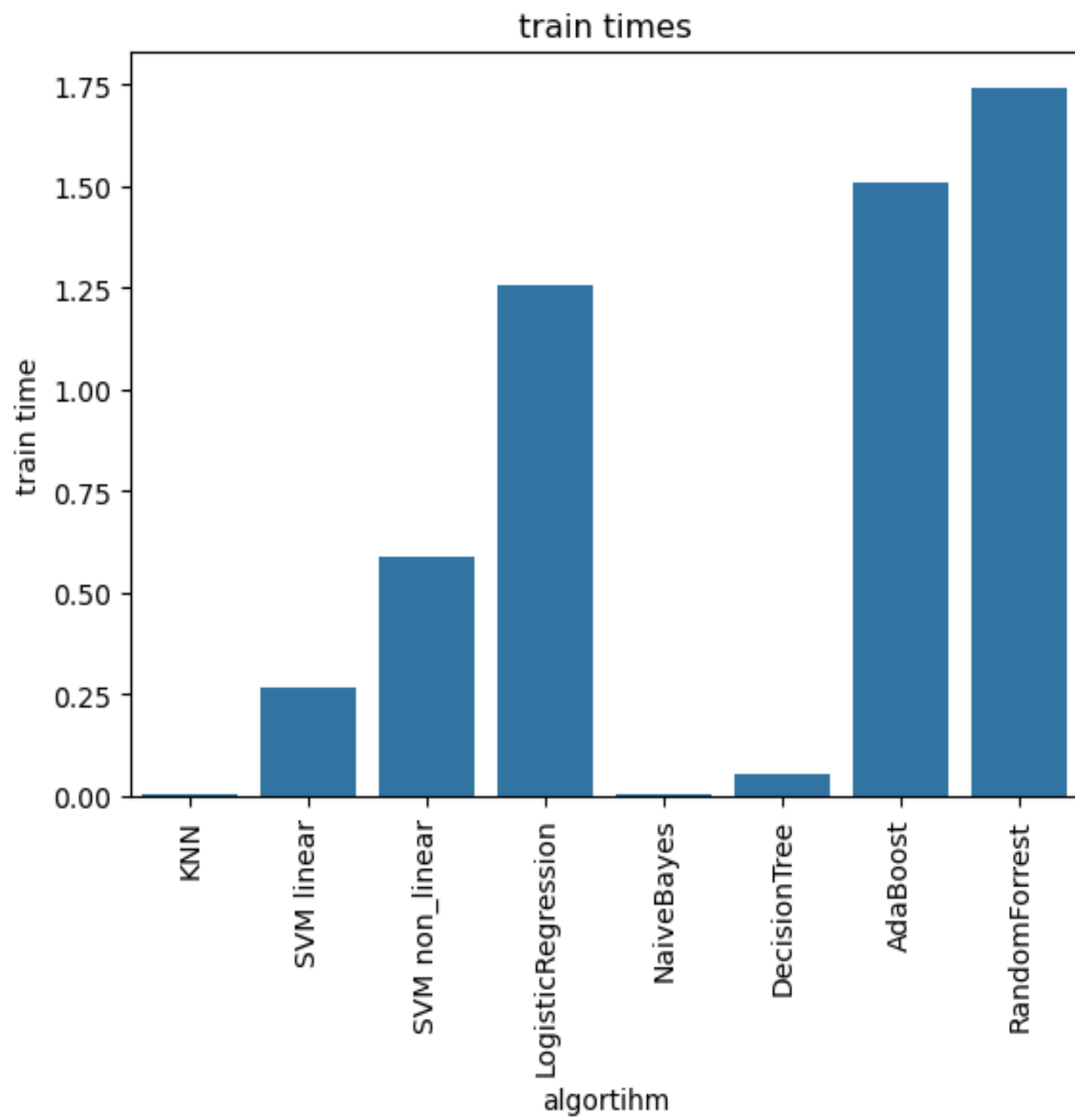
مقایسه زمان تست :

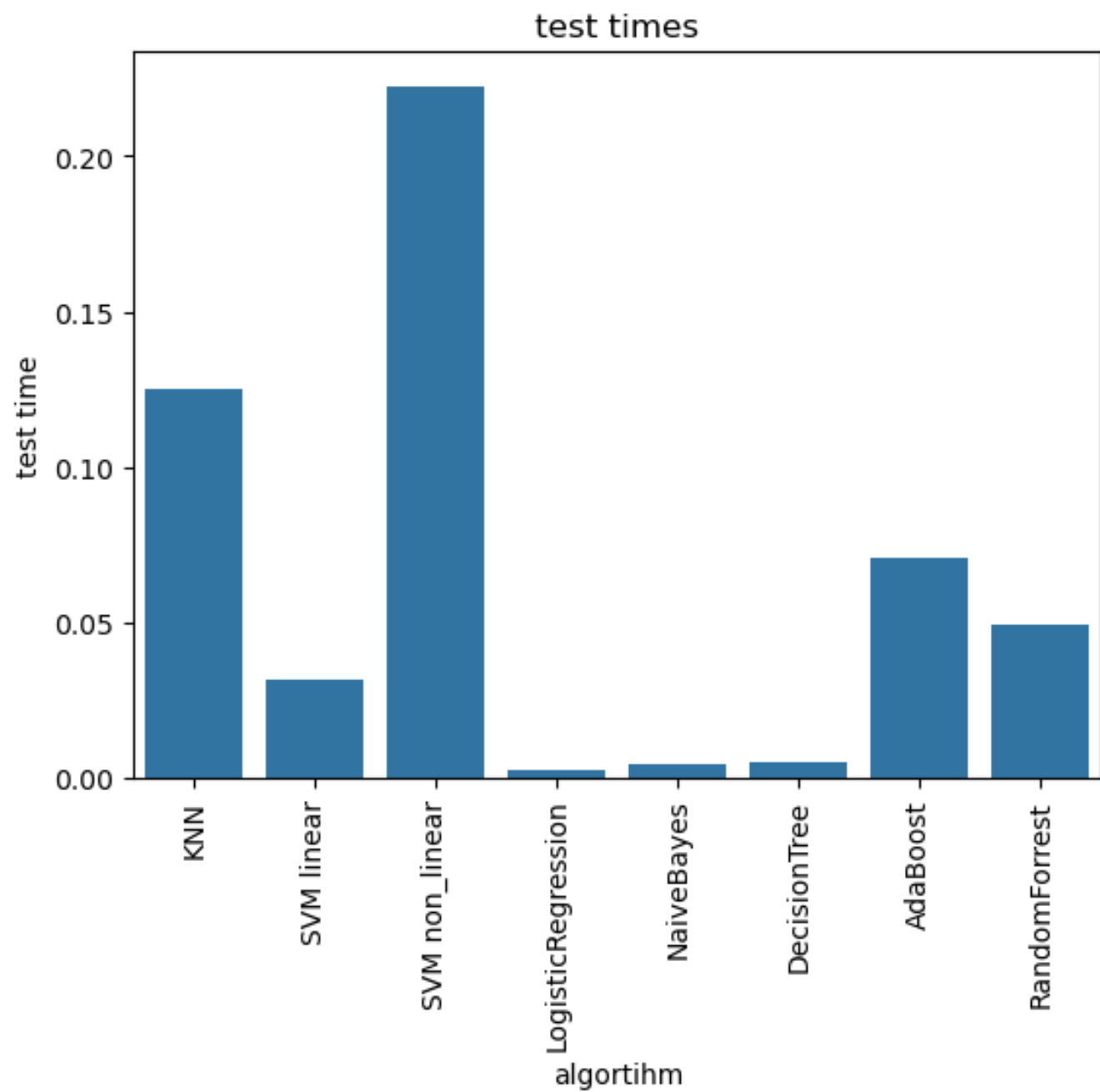
1. Logistic regression
2. Naïve bayse
3. Decision Tree
4. SVM linear
5. Random Forrest
6. Adaboost
7. Knn
8. Svm non_linear

مقایسه دقت :

1. Random Forrest
2. Adaboost
3. Decision tree
4. Logistic Regression
5. Svm non linear
6. Naïve bayse
7. KNN
8. Svm linear







- الگوریتم همسایه های نزدیک و ماشین پشتیبان خطی بدترین انتخاب است، هر د. دارای دقت پایین . زمان بالا
- جنگل رندوم با وجود زمان زیاد ، دقت بسیار بالایی دارد
- درخت تصمیم بهترین انتخاب است، زمان تمرین و تست پایین و دقت بالا دارد
- پایین ترین دقت متعلق به ماشین خطی پشتیبان هست حدود 66 درصد
- تمامی پارامتر ها در فایل نوتبوک به صورت بهینه انتخاب شده اند
- بالاترین دقت رندوم فارست هست حدود 95 درصد
- حذف کلمات غیر تاثیر گذار باعث افزایش دقت تا 95 درصد شد ، نه تنها حذف نکردن باعث کاهش بیشترین دقت تا 88 درصد میشد ، بلکه در برخی الگوریتم ها مانند رگرسیون خطی باعث OVERFITTING نیز میشد ، پس حذف کلمات مهم هست و تاثیر گذار