Theme B

2.1 Experimental Setup:

I. **Features:** We have considered three feature set for our experimental study. First setup we considered only Local measures, second setup we considered only global measure and finally we considered Local and global measure together.

II. **Classifier:** In order to have better understanding of the system we have considered very wide range of classifier. The detail and parameter setting of classifiers is presented in Table1:

Classifier	Parameter's setting
Decision Tree	Id3 Estimator
SVM	Linear, RBF, Polynomial Kernel
Naïve Bayes	
Adaboost	50 estimators with learning rate 1
Bagging	20 estimators, 50 percent samples, full set of features.

III. **Performance Measure:** we have used average accuracy, F1 Measure, Precision, Recall and AUC score and ROC curve as performance metrics.

2.2 Performance Evaluation:

2.2.1 Performance valuation of different classifier:

Classifier	Local	Global	Local+Global
Naive Bayes	0.92754	**0.93784**	0.93269
DT	**0.98**	**0.98**	**0.98**
SVM_LIN	**0.95665**	0.94334	0.949995
SCM_Poly	**0.95122**	0.92608	0.93865
SVM_RBF	**0.95678**	0.94344	0.95011
Adaboost	**0.989231099**	**0.989231099**	**0.989231099**
Ada_DT	**0.986666666**	0.918	0.952333333
Ada_SVM	0.69988	**0.9112**	0.80554
Ada_Naive	0.58656	**0.80131**	0.693935
Bag_DT	**0.989637366**	0.9377	0.963668683
Bag_SVM	**0.95664**	0.94458	0.95061
Bag_Naive	0.927646	**0.93776**	0.932703

Table 2: Accuracy comparison of different classifier for 3 feature set setups for restaurant data sets.

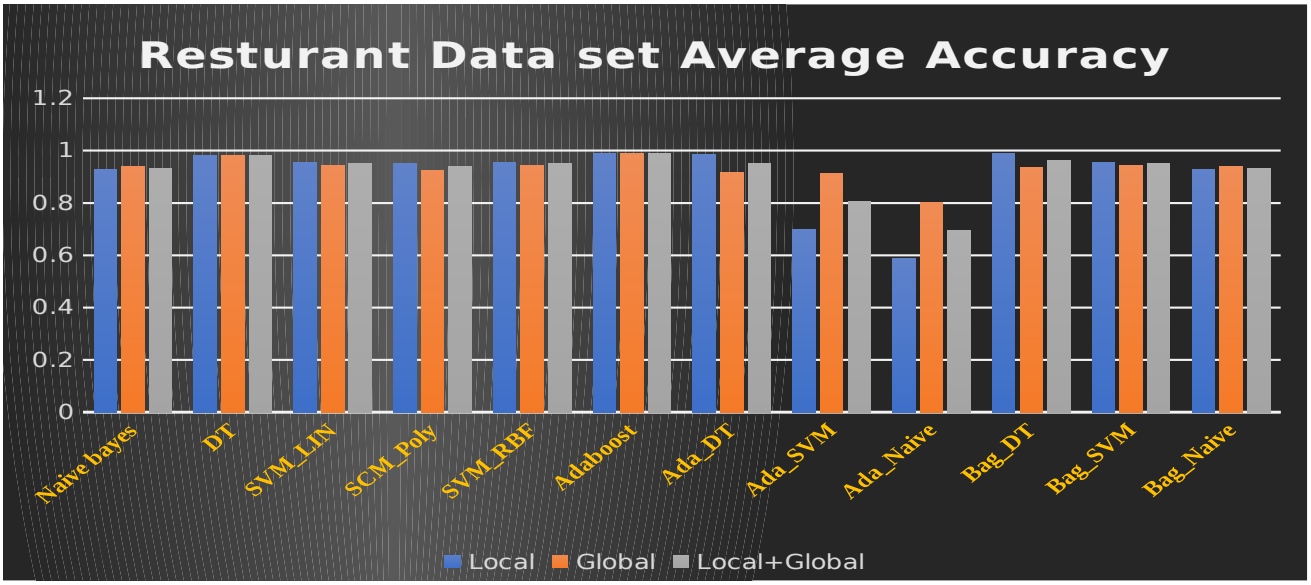


Figure: 1 Graphical comparison of different classifier for 3 feature set setups for restaurant data sets

Observations: It can be observed that some classifier performed well when we have only local measure as Feature sets and some classifier performed when we have only global measure as feature sets and some performed well when we have local as well as global measure as feature sets.

Classifier	**Local**	**Global**	**Local+Global**
Naive Bayes	0.92536	0.93668	**0.937054**
DT	0.855432902	0.91778	**0.97962**
SVM_LIN	0.95928	0.9436	**0.97688**
SCM_Poly	0.95132	0.93	**0.95132**
SVM_RBF	0.955988	0.94404	**0.97876**
Adaboost	0.89645946	0.94484	**0.975932**
Ada_DT	0.890631564	0.91794	**0.9793**
Ada_SVM	0.73022	0.89504	**0.95**
Ada_Naive	0.69536	**0.79138**	0.70778
Bag_DT	0.901357192	0.9371	**0.9684754**
Bag_SVM	0.955852	0.947354	**0.98414**

Bag_Naive	0.92536	0.94752	**0.95712**

Table 3: F1 score comparison of different classifier for 3 feature set setups for restaurant data sets.

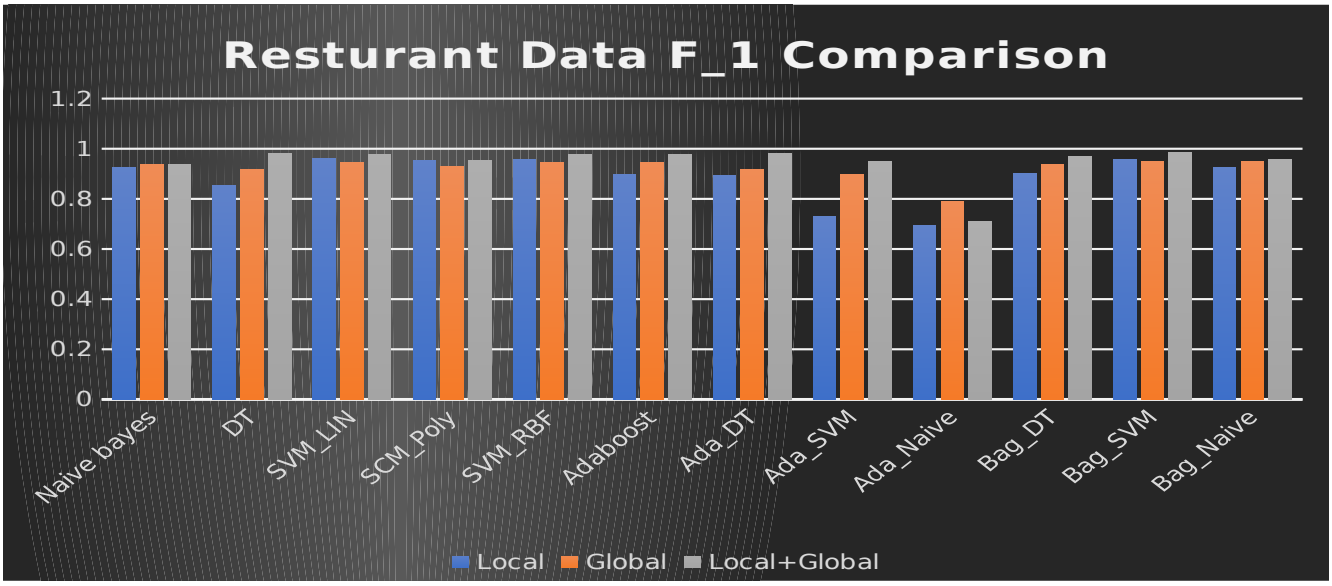


Figure 2: F1 score Graphical comparison of different classifier for 3 feature set setups for restaurant data sets

Observation: All the classifer give best performance when Local and Global feature is used together except the case of Adaboost with naïve bayes as weak learner.

Classifer	**Local**	**Global**	**Local+Global**
Naive bayes	0.95414	0.9405	**0.95734**
DT	0.855432902	0.9175	**0.9793**
SVM_LIN	0.973906	0.93946	**0.97944**
SCM_Poly	0.948482	0.88294	**0.948482**
SVM_RBF	0.97434	0.93804	**0.9808**
Adaboost	0.949491956	0.93378	**0.97718**
Ada_DT	0.931368082	0.9157	**0.97952**
Ada_SVM	0.71184	**0.973**	0.85488
Ada_Naive	0.66558	**0.91448**	0.67994
Bag_DT	0.951352966	0.93254	**0.98522**
Bag_SVM	0.96656	0.94674	**0.98556**
Bag_Naive	0.954056	0.95416	**0.957632**

Table: Precision score comparison of different classifier for 3 feature set setups for restaurant data sets.

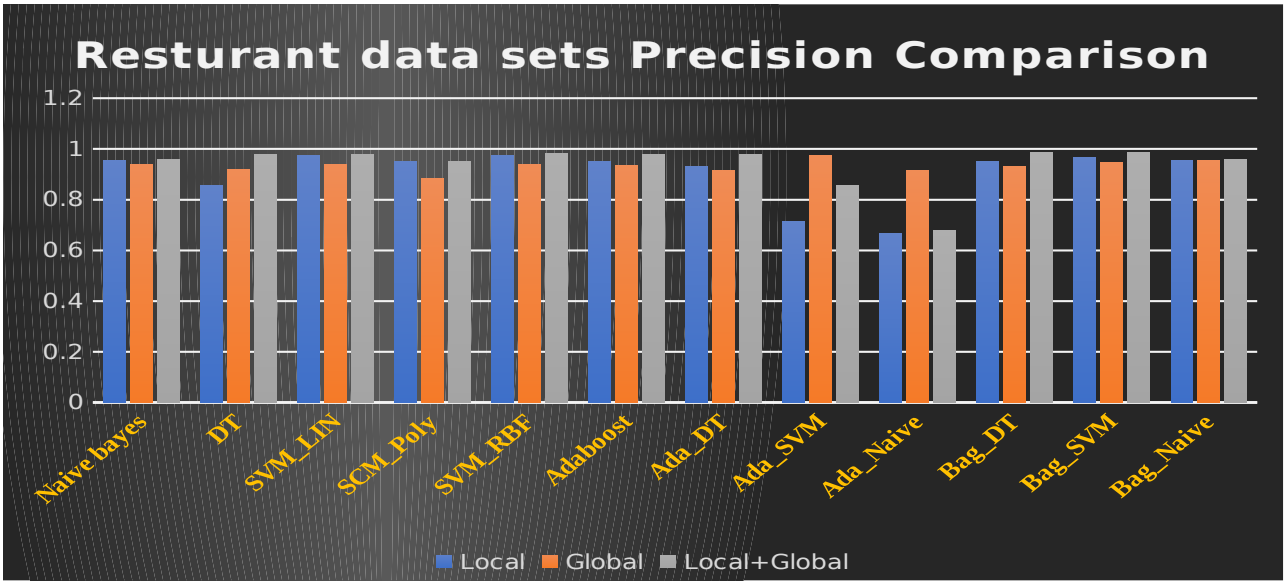


Figure 3: Precision score Graphical comparison of different classifier for 3 feature set setups for restaurant data sets

Observation: All the classifier performed well when we have local measure as well as global measure as feature sets.

Classifier	Local	Global	Local+Global
Naive bayes	0.898287	0.93404	**0.95676**
DT	0.856259686	0.91858	**0.97946**
SVM_LIN	0.93706	0.94788	**0.9518**
SCM_Poly	0.95322	**0.98238**	0.96964
SVM_RBF	0.938818	0.95466	**0.977**
Adaboost	0.849051812	0.95432	**0.97462**
Ada_DT	0.85844812	0.918612	**0.98026**
Ada_SVM	0.97854	0.84418	**0.98934**
Ada_Naive	0.91472	0.91472	**0.93872**
Bag_DT	0.854644414	0.94266	**0.98388**
Bag_SVM	0.9385	0.94258	**0.96838**
Bag_Naive	0.89856	0.920206	**0.96228**

Table: Recall score comparison of different classifier for 3 feature set setups for restaurant data sets

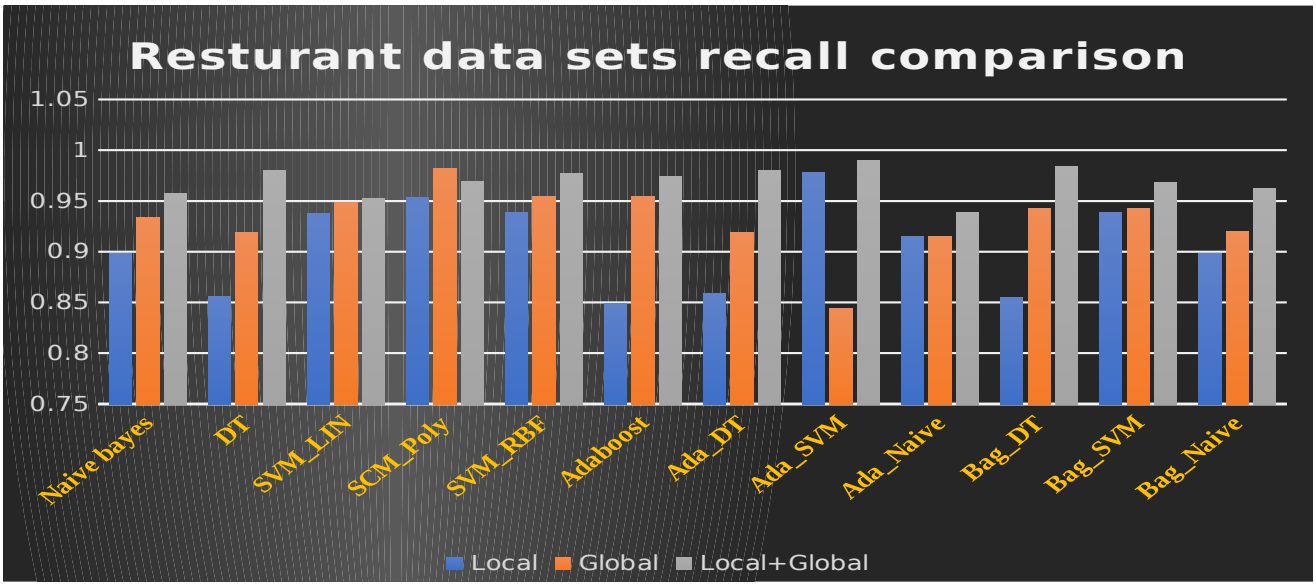


Figure 4: Recall score Graphical comparison of different classifier for 3 feature set setups for restaurant data sets

Blog Data sets

Classifier	Local	Global	Local+Global
Naive bayes	0.78788	0.901	**0.9414**
DT	**0.98**	**0.98**	**0.98**
SVM_LIN	0.8904	0.99248	**1**
SCM_Poly	0.8281	0.93722	**0.944**
SVM_RBF	0.90436	0.99398	**0.99698**
Adaboost	**0.9892311**	0.92216	0.989231099
Ada_DT	0.986666666	0.91326	**0.99848**
Ada_SVM	0.747878	0.78292	**1**
Ada_Naive	0.39838	0.7892	**0.91808**
Bag_DT	0.989637366	0.92982	**0.99706**
Bag_SVM	0.87868	0.90262	**0.99698**
Bag_Naive	0.7921	0.90854	**0.9491**

Table: Accuracy score comparison of different classifier for 3 feature set setups for Blog data sets

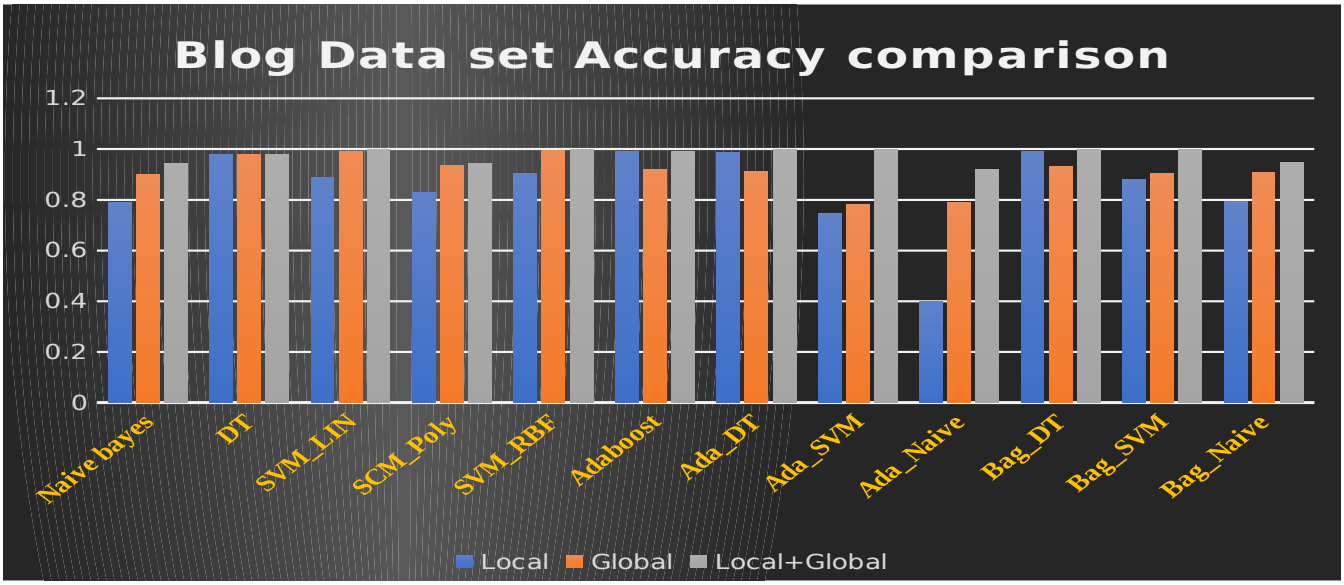


Figure 5: Accuracy score Graphical comparison of different classifier for 3 feature set setups for Blog data sets

Observation: Decision tree give same performance irrespective of feature sets. Adaboost performed slightly better when we have only local measure as feature sets. All other classifier performed better when we have Local measure as well as global measure as feature sets.

Classifier	**Local**	**Global**	**Local+Global**
Naive bayes	0.7362	0.8933	**0.94086**
DT	0.855432902	0.99694	**0.9971**
SVM_LIN	0.8845	0.99384	**1**
SCM_Poly	0.837104	0.94738	**0.95654**
SVM_RBF	0.90266	0.99084	**0.99708**
Adaboost	0.89645946	0.92156	**0.99854**
Ada_DT	0.890631564	0.9153	**0.99854**
Ada_SVM	0.7405	0.67684	**0.99854**
Ada_Naive	0.46144	0.82328	**0.89546**
Bag_DT	0.901357192	0.93262	**0.99854**
Bag_SVM	0.86656	0.894566	**0.99704**
Bag_Naive	0.735732	0.9014	**0.95613**

Table: F1 score comparison of different classifier for 3 feature set setups for Blog data sets

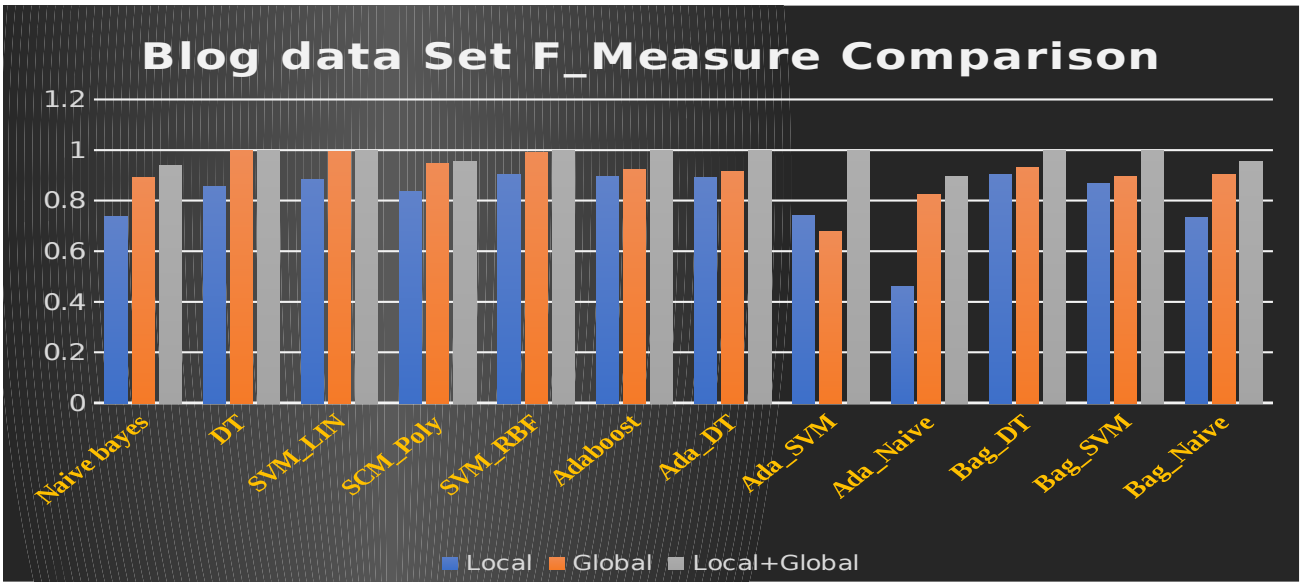


Figure 6: F1 score Graphical comparison of different classifier for 3 feature set setups for Blog data sets

Observations: All the classifier performed well when have Local as well as Global feature.

Classifier	**Local**	**Global**	**Local+Global**
Naive bayes	**0.96982**	**0.96852**	0.94086
DT	0.855432902	**0.997**	**0.9971**
SVM_LIN	0.93692	**1**	**1**
SCM_Poly	0.79596	0.91366	**0.95654**
SVM_RBF	0.91868	**1**	0.99708
Adaboost	0.949491956	0.9165	**0.99854**
Ada_DT	0.931368082	0.90624	**0.99854**
Ada_SVM	0.77428	0.93596	**0.99854**
Ada_Naive	0.48792	0.79386	**0.89546**
Bag_DT	0.951352966	0.93024	**0.99854**
Bag_SVM	0.938752	0.9265	**0.99704**
Bag_Naive	**0.96972**	0.9548	0.95613

Table: Precision score comparison of different classifier for 3 feature set setups for Blog data sets

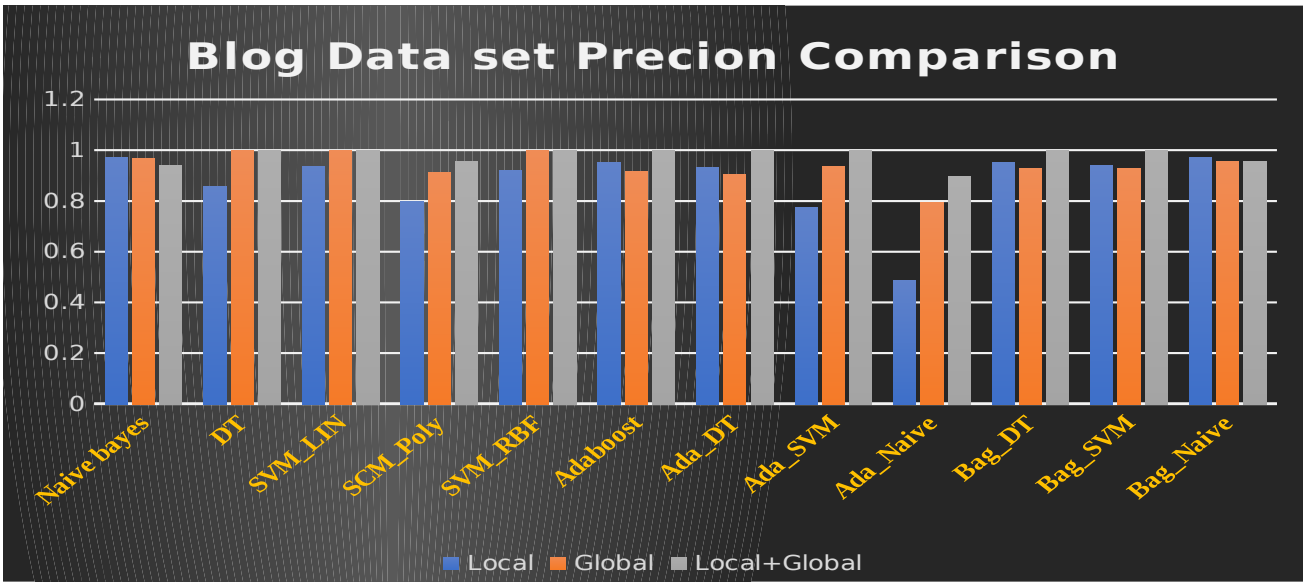


Figure 7: Precision score Graphical comparison of different classifier for 3 feature set setups for Blog data sets

Classifier	Local	Global	Local+Global
Naive bayes	**0.96982**	**0.96852**	0.94086
DT	0.855432902	**0.997**	**0.9971**
SVM_LIN	0.93692	**1**	**1**
SCM_Poly	0.79596	0.91366	**0.95654**
SVM_RBF	0.91868	**1**	0.99708
Adaboost	0.949491956	0.9165	**0.99854**
Ada_DT	0.931368082	0.90624	**0.99854**
Ada_SVM	0.77428	0.93596	**0.99854**
Ada_Naive	0.48792	0.79386	**0.89546**
Bag_DT	0.951352966	0.93024	**0.99854**
Bag_SVM	0.938752	0.9265	**0.99704**
Bag_Naive	**0.96972**	0.9548	0.95613

Table: Recall score comparison of different classifier for 3 feature set setups for Blog data sets.

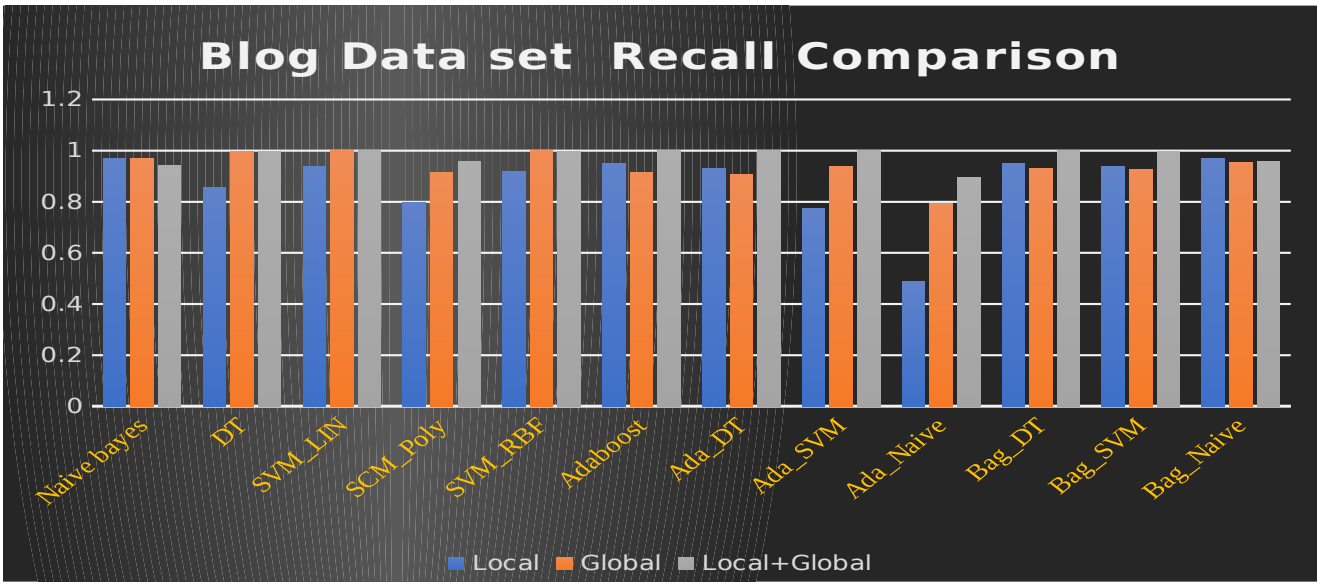


Figure 8: Recall score Graphical comparison of different classifier for 3 feature set setups for Blog data sets.

2.2.2 Performance evaluation of different dimensionality reductions:

PCA for restaurant data sets.

			Resturant Data sets			Local	Measure	Feature
Dimesnions	Naive baye	DT	SVM	Ada_SVM	Ada_Naive	Bag_SVM	Bag_Naive	
1	0.9	0.873199	0.92	0.77	0.63	0.91	0.9	
2	0.88	0.866400386	0.93	0.94	0.5	0.91	0.89	
3	0.87	0.866005158	0.93	0.95	0.68	0.95	0.87	
4	0.9	0.865901854	0.95	0.95	0.69	0.96	0.9	

Observations: From above figure it can be observed that different classifier gives its best performance in term of accuracy for different dimension.

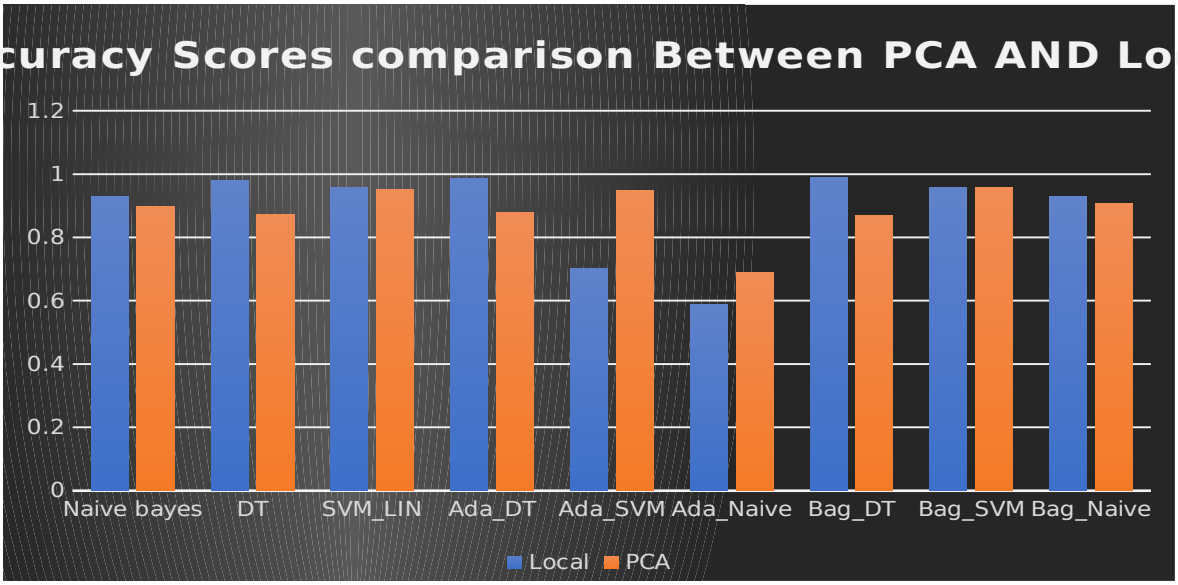


Figure 9: Accuracy score comparison between performance of classifer on reduced dimension and performance of classifer on actual dimenson.

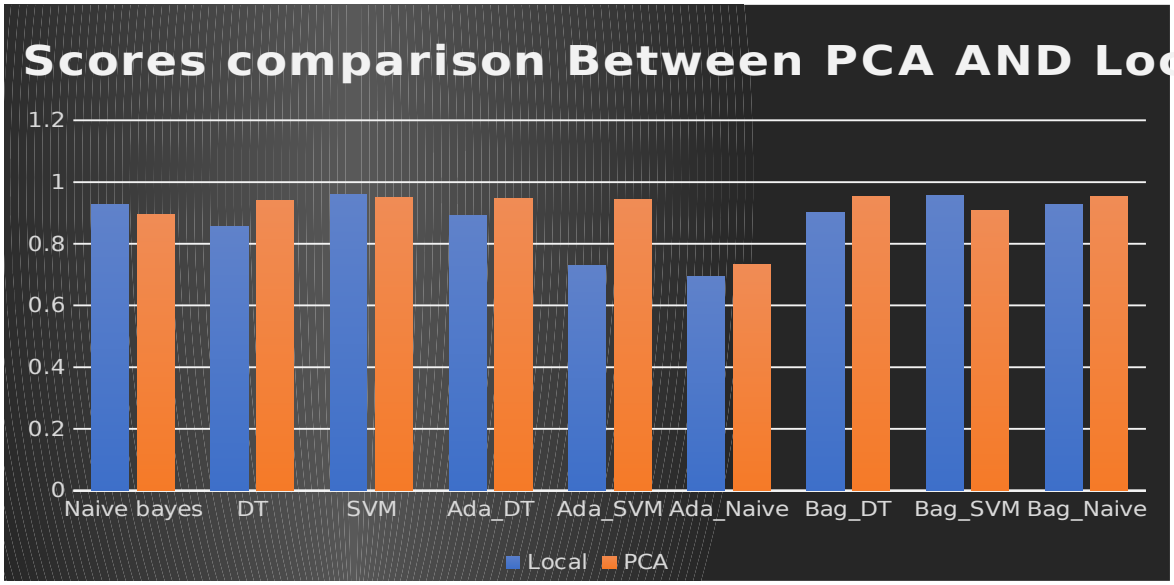


Figure 10: F1 score comparison between performance of classifer on reduced dimension and performance of classifer on actual dimenson.

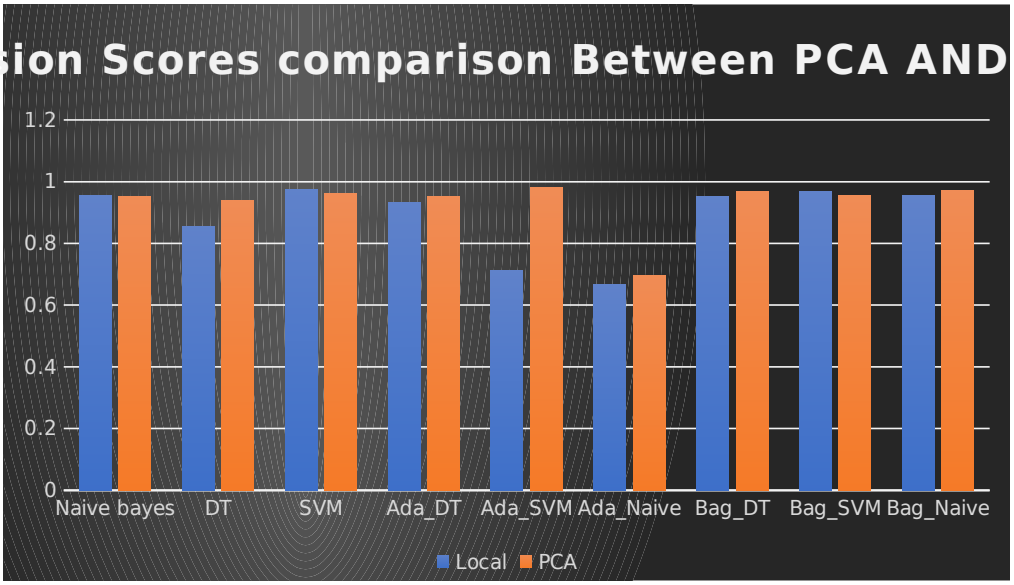


Figure 11: Precision score comparison between performance of classifier on reduced dimension and performance of classifier on actual dimenson.

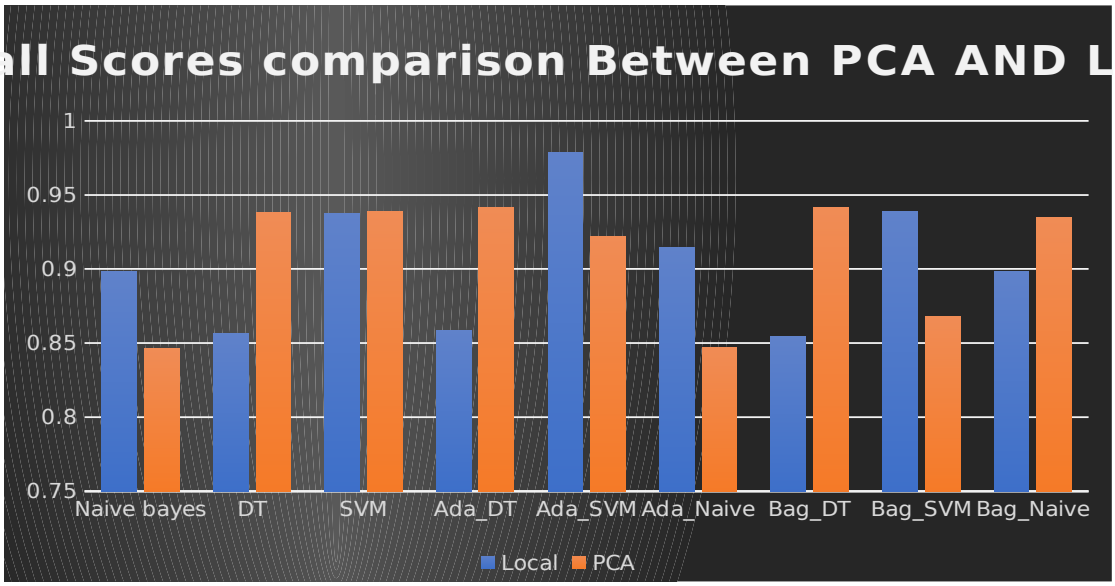


Figure 12: Recall score comparison between performance of classifier on reduced dimension and performance of classifier on actual dimenson

Blog data sets

			Blog	Data sets		Local	Measure	Feature
Dimesnions	Naive baye	DT	SVM	Ada_SVM	Ada_Naive	Bag_Naive	Bag_SVM	
1	0.9	0.892056844	0.92	0.77	0.63	0.9	0.91	
2	0.88	0.921628062	0.93	0.94	0.5	0.89	0.91	
3	0.87	0.937539018	0.93	0.95	0.68	0.877	0.95	
4	0.9	0.937761	0.95	0.95	0.69	0.9	0.96	

Observations: From above figure it can be observed that different classifier gives its best performance in term of accuracy for different dimension.

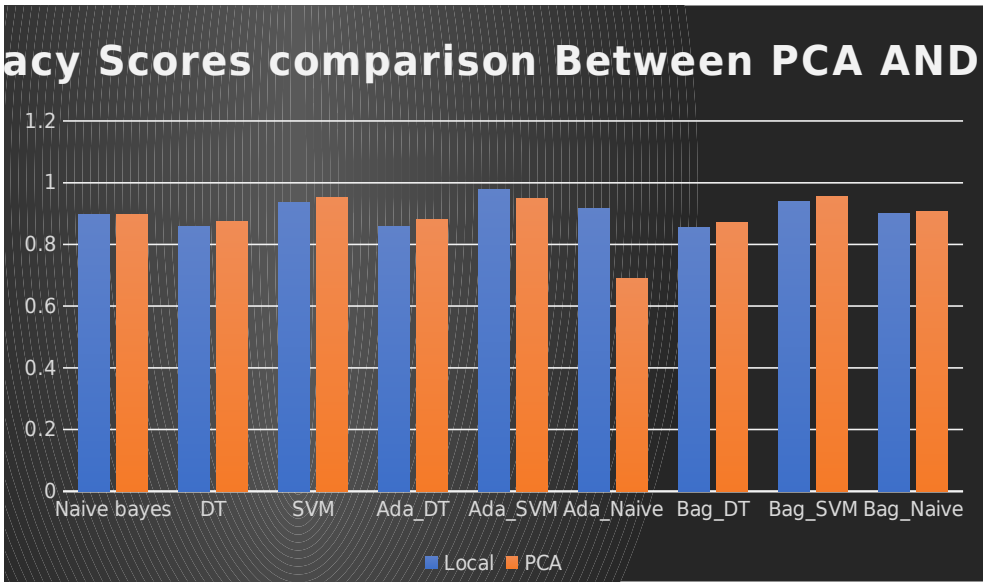


Figure 13: accuracy score comparison between performance of classifier on reduced dimension and performance of classifier on actual dimenson.

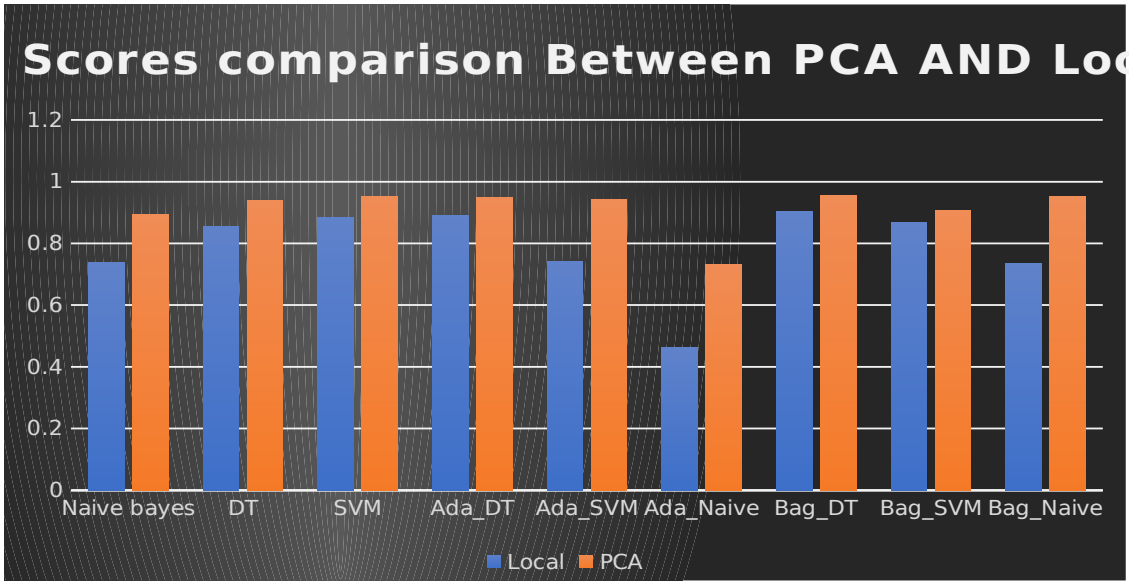


Figure 14: F1 score comparison between performance of classifier on reduced dimension and performance of classifier on actual dimenson.

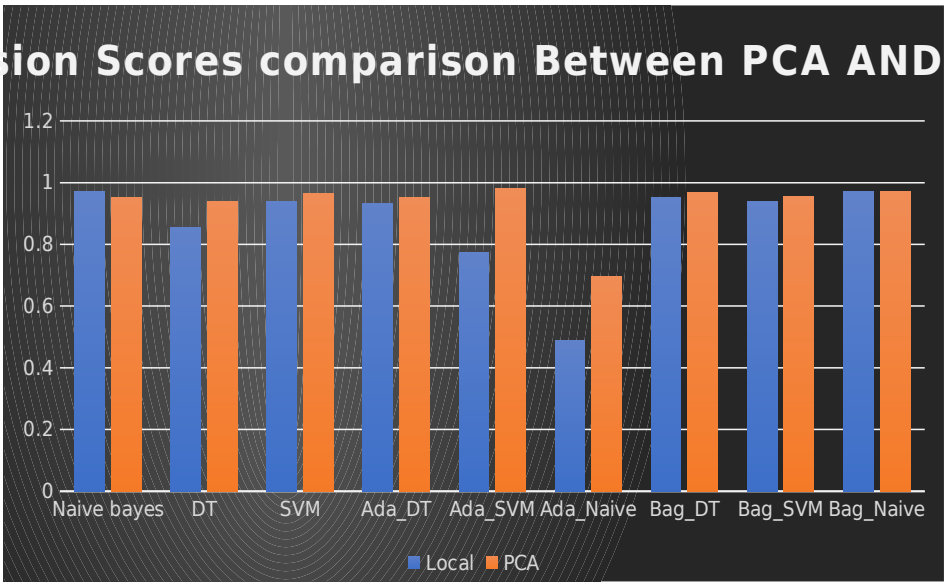


Figure 15: Precision comparison between performance of classifier on reduced dimension and performance of classifier on actual dimenson.

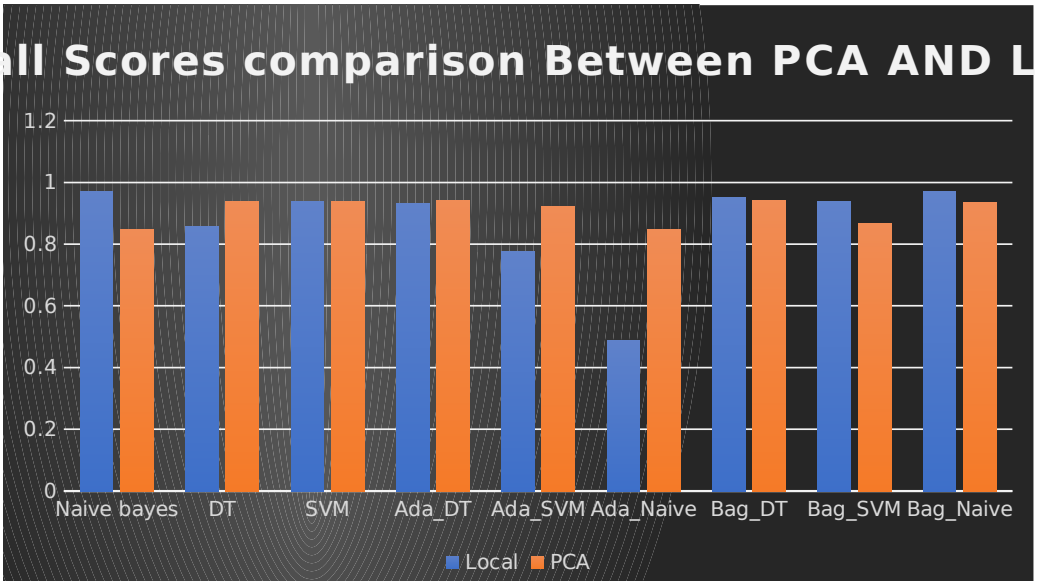


Figure 16: Recall comparison between performance of classifier on reduced dimension and performance of classifier on actual dimenson.

Restaurant Data sets Local +Global Features

			Resturant Data sets			Local+Global Measure Feature			
Dimesnions	Naive bayes	DT	SVM	Ada_DT	Ada_SVM	Ada_Naive	Bag_DT	Bag_SVM	Bag_Naive
1	*0.92*	0.89	**0.93**	**0.89**	0.79	**0.54**	**0.91**	**0.92**	**0.92**
2	0.92	0.92	**0.94**	**0.91**	**0.93**	**0.63**	**0.94**	**0.93**	**0.92**
3	0.92	**0.96**	**0.97**	**0.92**	0.93	**0.74**	**0.97**	0.94	0.92
4	0.9	0.96	0.97	**0.94**	0.93	0.67	0.97	**0.97**	**0.85**

Observations: From above figure it can be observed that different classifier gives its best performance in term of accuracy for different dimension

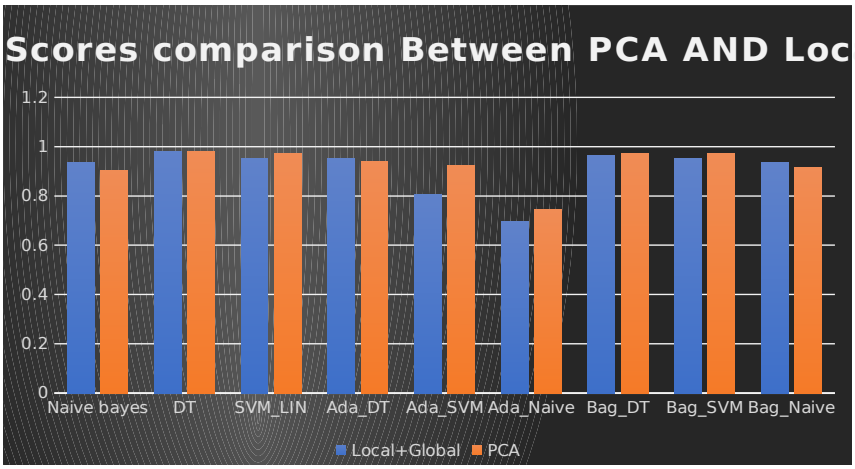


Figure 17: Recall comparison between performance of classifier on reduced dimension and performance of classifier on actual dimension.

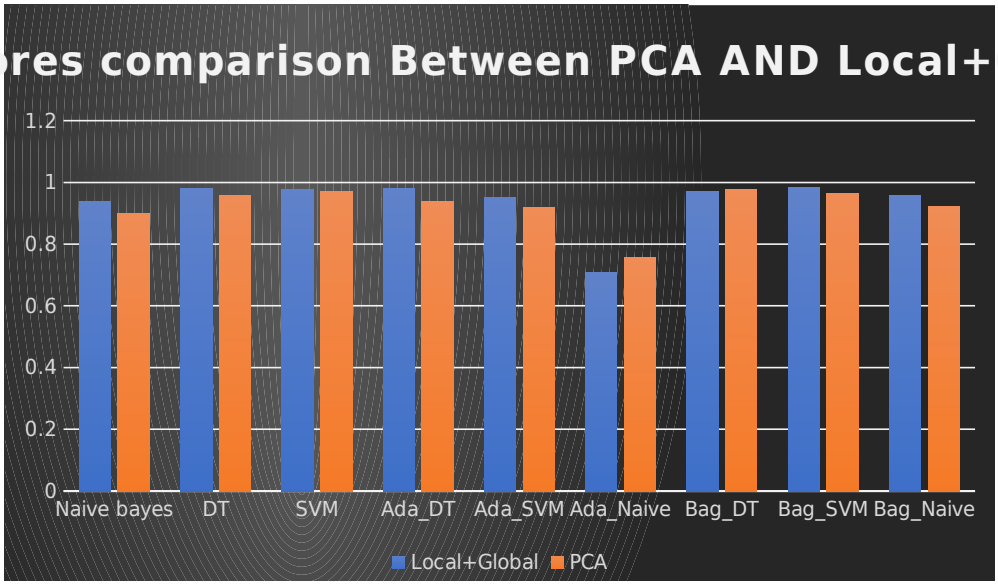


Figure 18: F1 Score comparison between performance of classifier on reduced dimension and performance of classifier on actual dimension.

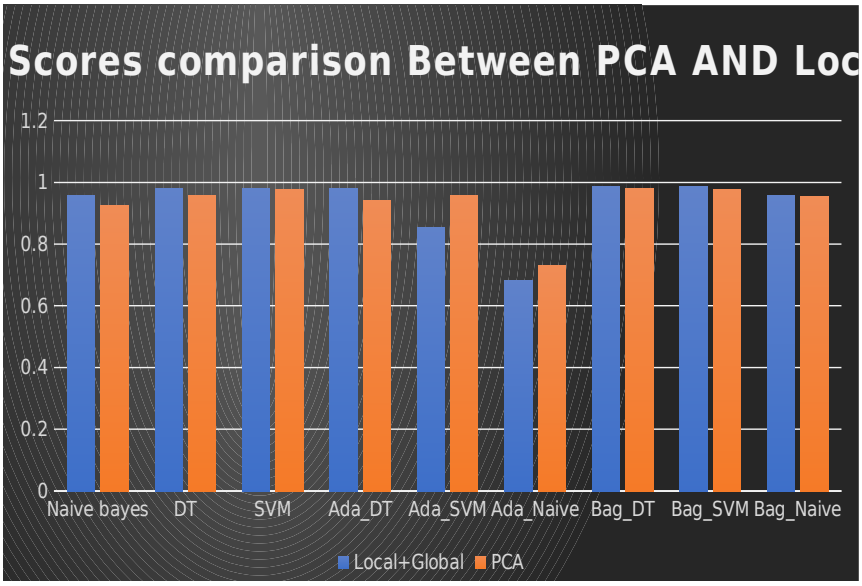


Figure 19: precision Score comparison between performance of classifier on reduced dimension and performance of classifier on actual dimension.

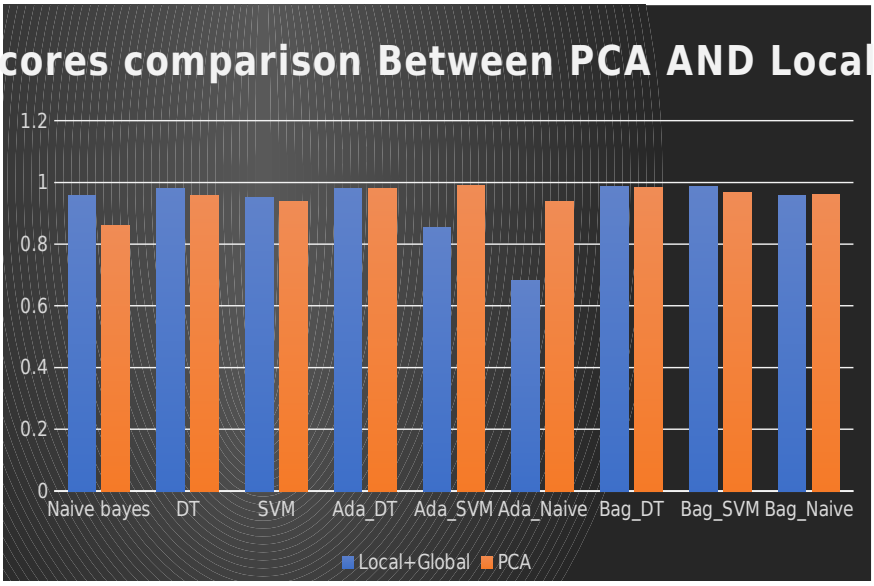


Figure 20: Recall Score comparison between performance of classifier on reduced dimension and performance of classifier on actual dimension.

Blog Data Sets Local and Global Features

			Blog	Data sets		Local +Global	Measure	Feature	
Dimesnions	Naive bayes	DT	SVM	Ada_DT	Ada_SVM	Ada_Naive	Bag_DT	Bag_SVM	Bag_Naive
1	0.83	0.85	0.87	0.86	0.79	0.78	0.86	0.87	0.84
2	0.94	0.96	0.96	0.96	0.97	0.76	0.96	0.96	0.94
3	0.94	0.98	0.97	0.97	0.97	0.86	0.97	0.97	0.95
4	0.93	0.99	0.98	0.99	100	0.88	0.99	1	0.93

Observations: From above figure it can be observed that different classifier gives its best performance in term of accuracy for different dimension

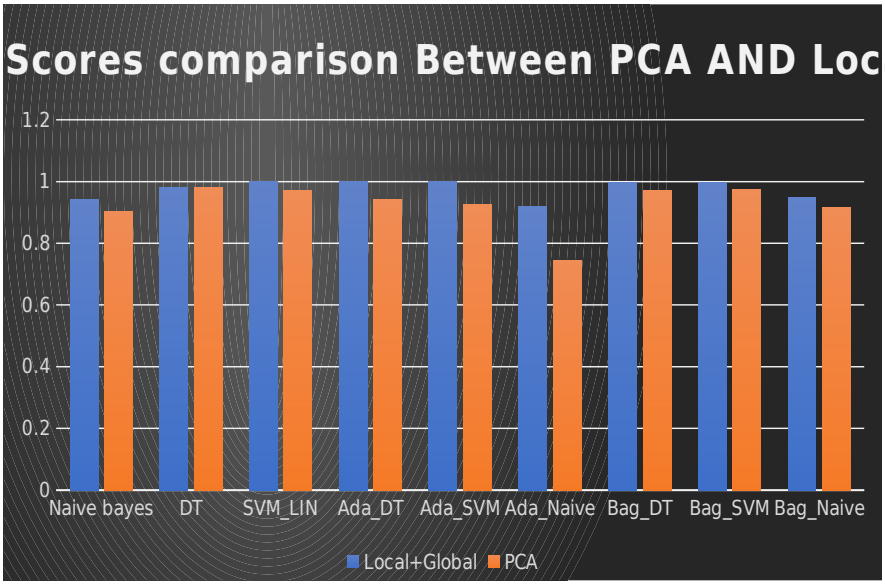


Figure 21: Accuracy Score comparison between performance of classifier on reduced dimension and performance of classifier on actual dimension

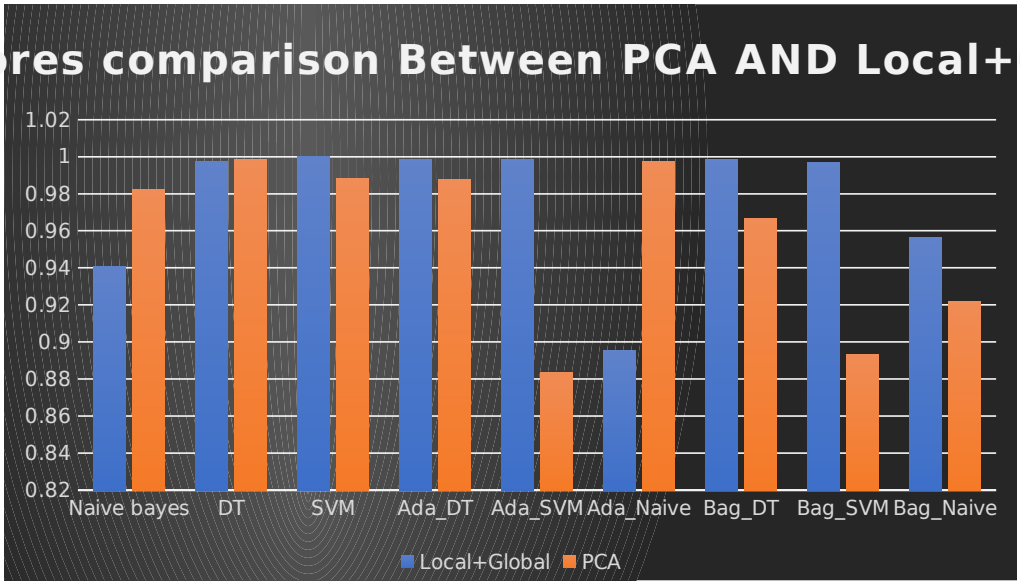


Figure 22: F1 Score comparison between performance of classifier on reduced dimension and performance of classifier on actual dimension

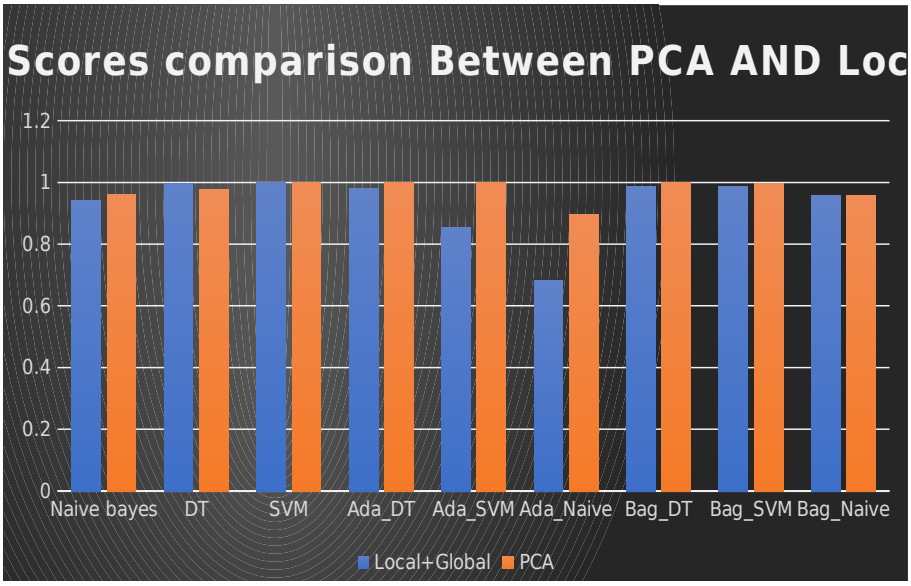


Figure 23: Precision Score comparison between performance of classifier on reduced dimension and performance of classifier on actual dimension

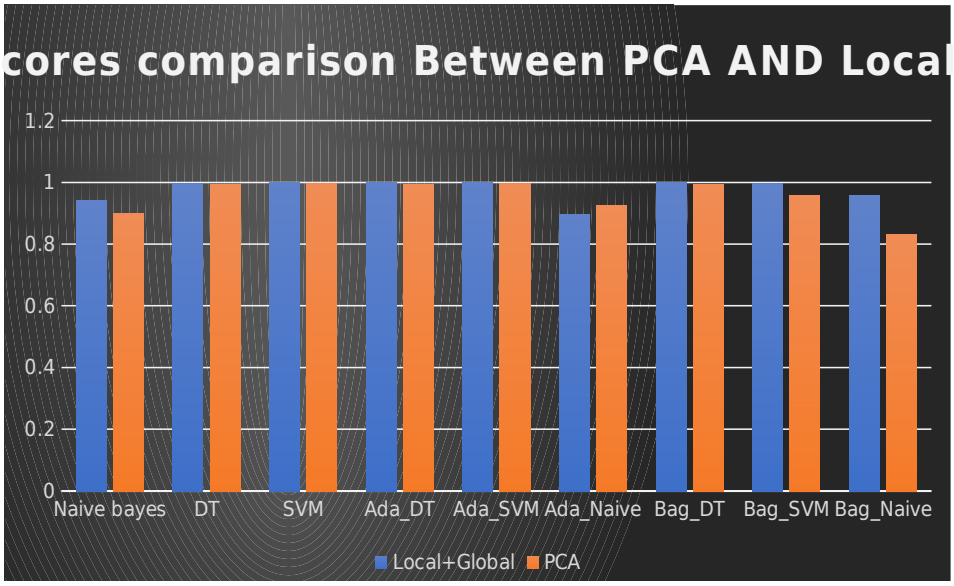


Figure 24: Recall Score comparison between performance of classifier on reduced dimension and performance of classifier on actual dimension

SVD Analysis

Restaurant Data sets:

			Resturant Data sets			Local Measure Feature		
Dimesnions	Naive baye	DT	SVM	Ada_DT	Ada_SVM	Ada_Naive	Bag_SVM	Bag_Naive
1	0.873199	0.92	0.91	0.77	0.8731986	0.632	0.91	0.845
2	0.866400386	0.93	0.95	0.94	0.76	0.589	0.91	0.856
3	0.866005158	0.93	0.95	0.95	0.79	0.645	0.95	0.9559
4	0.865901854	0.96	0.95	0.95	0.56	0.69	0.96	0.9452

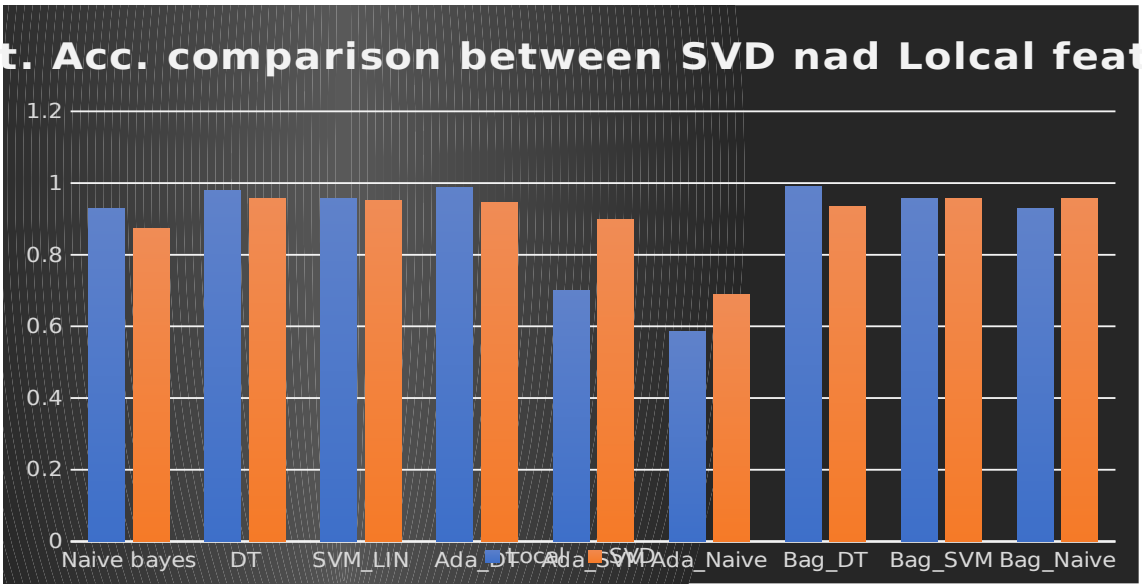


Figure 25: Accuracy Score comparison between performance of classifier on reduced dimension and performance of classifier on actual dimension

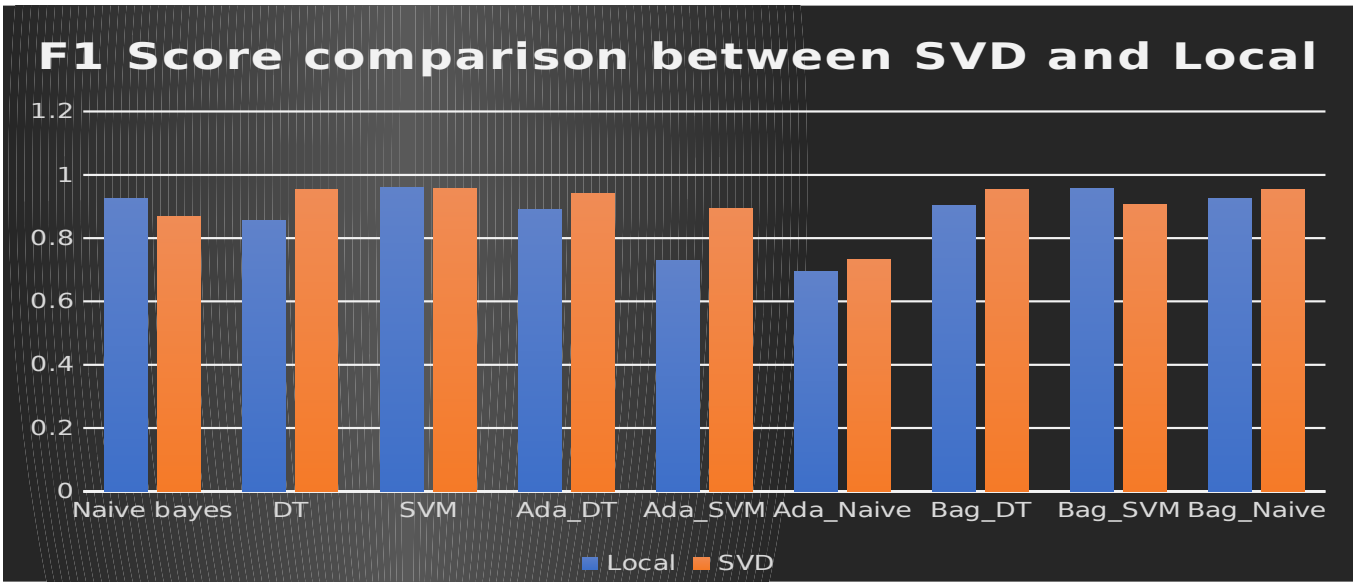


Figure 26: F1 Score comparison between performance of classifier on reduced dimension and performance of classifier on actual dimension

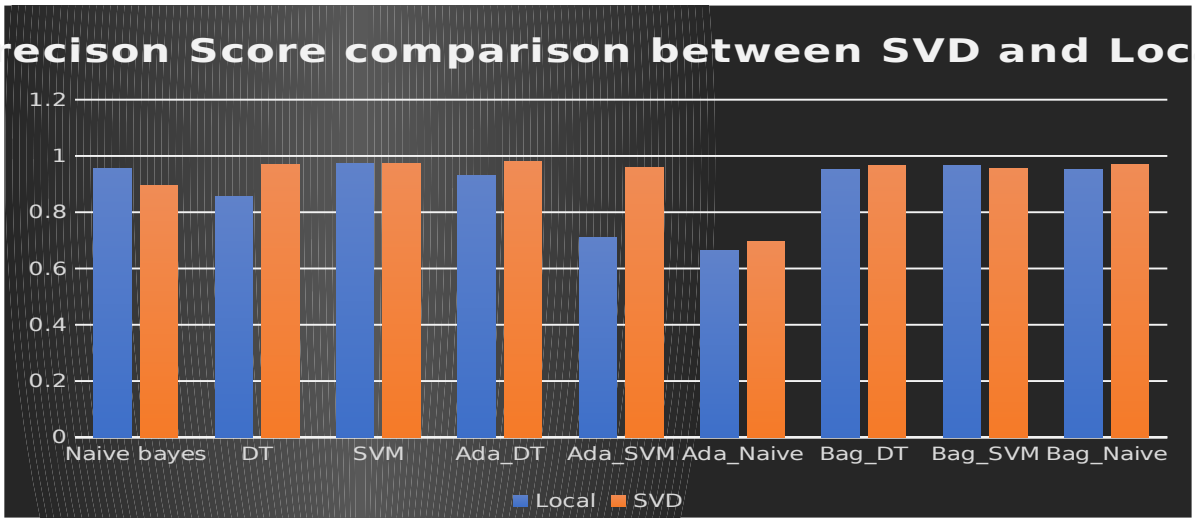


Figure 27: Accuracy Score comparison between performance of classifier on reduced dimension and performance of classifier on actual dimension

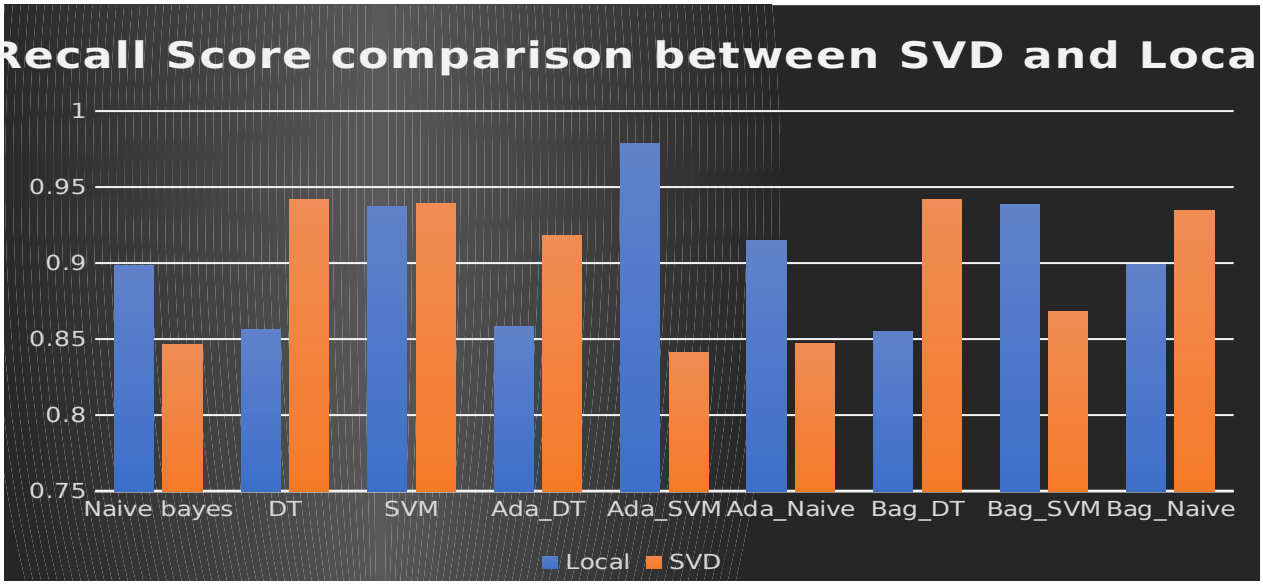


Figure 28: recall Score comparison between performance of classifier on reduced dimension and performance of classifier on actual dimension

Restaurant Data sets Local and Global feature:

			Resturant Data sets			local+Globa Measure Feature			
Dimesnions	Naive bayes	DT	SVM	Ada_DT	Ada_SVM	Ada_Naive	Bag_DT	Bag_SVM	Bag_Naive
1	0.6325	0.73	**0.93**	0.759	0.8	0.872	0.945	**0.92**	0.79
2	0.562	0.91	**0.94**	0.864	0.96	0.856	0.892	**0.93**	**0.93**
3	**0.69**	**0.94**	**0.97**	**0.915**	0.97	0.879	0.96	0.97	0.93
4	**0.69**	0.93	0.97	0.95	**0.99**	**0.88**	**0.9702**	0.97	0.93

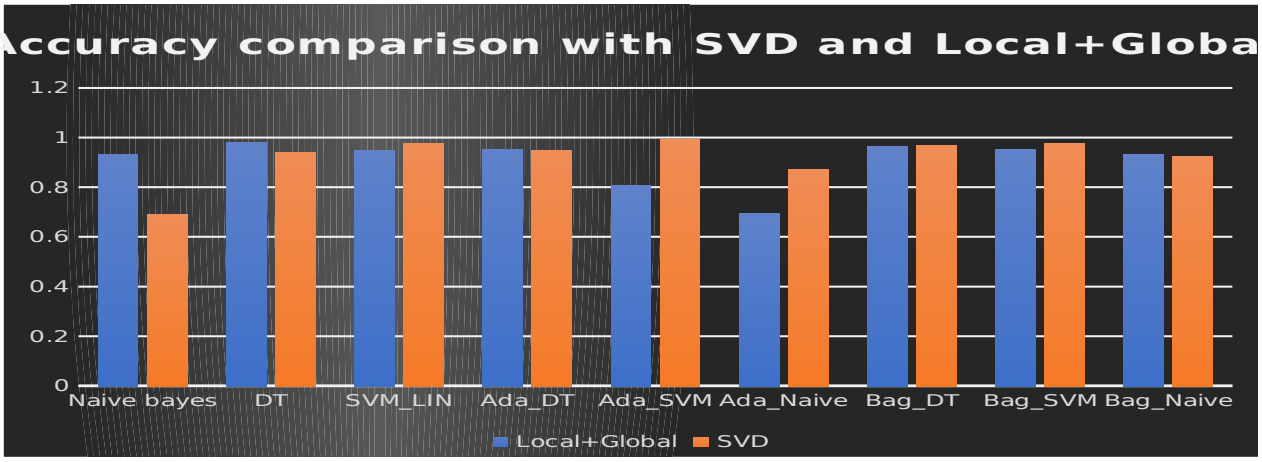


Figure 29: Accuracy Score comparison between performance of classifier on reduced dimension and performance of classifier on actual dimension

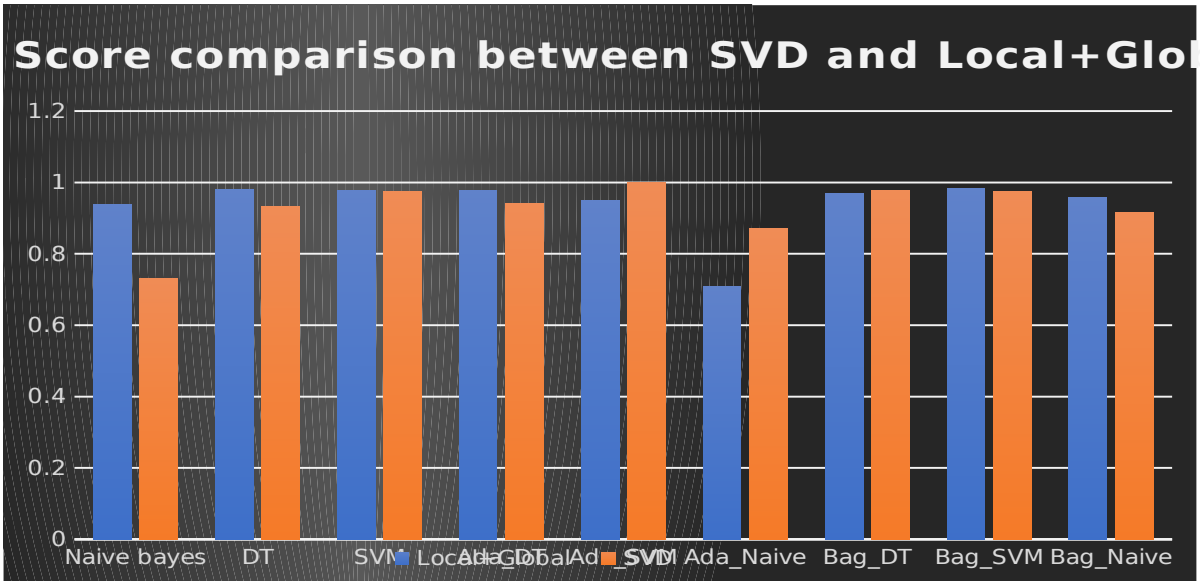


Figure 30: F1 Score comparison between performance of classifier on reduced dimension and performance of classifier on actual dimension

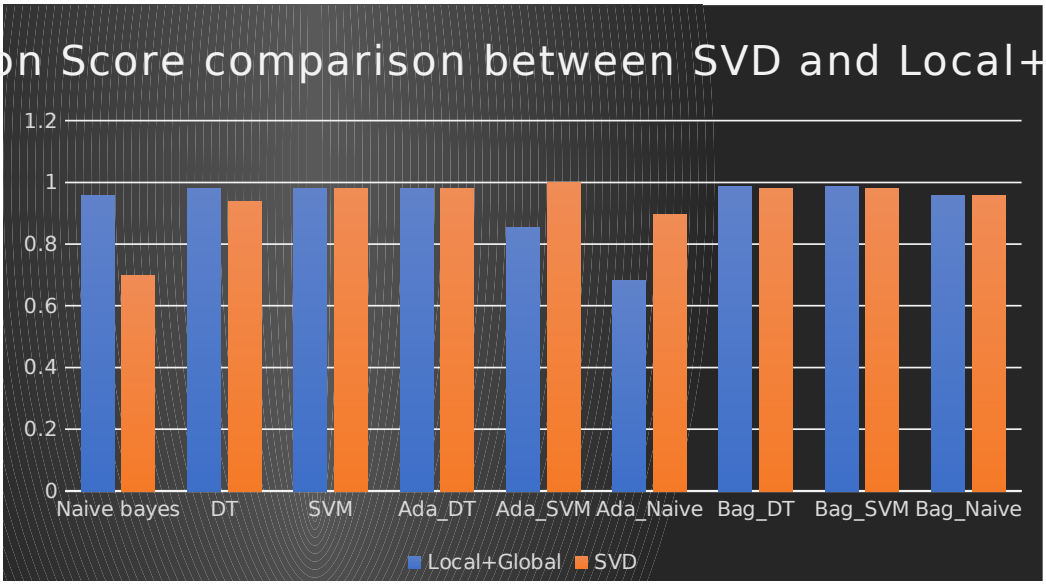


Figure 31: Precision Score comparison between performance of classifier on reduced dimension and performance of classifier on actual dimension

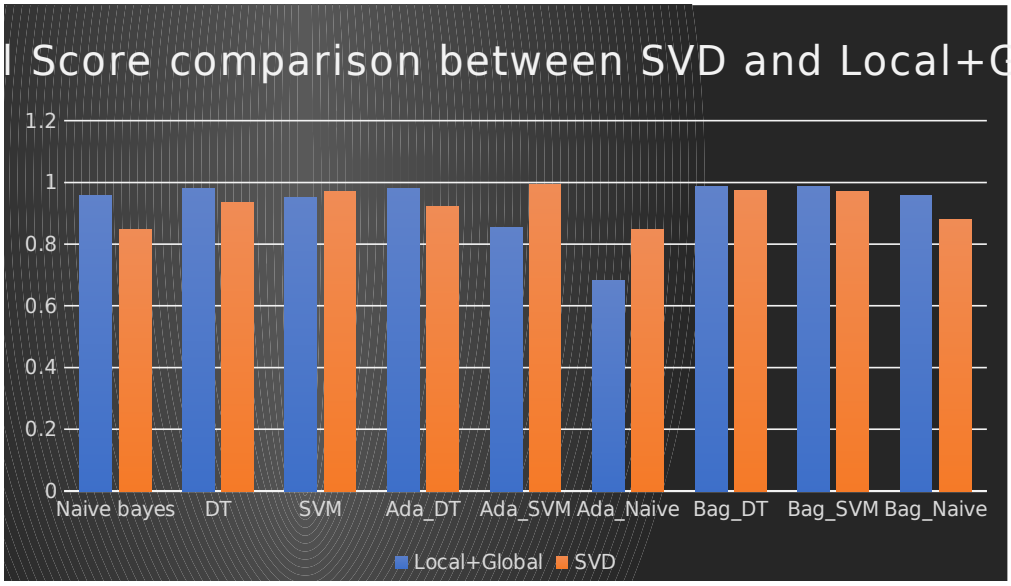


Figure 32: F1 Score comparison between performance of classifier on reduced dimension and performance of classifier on actual dimension

Blog data sets Local Feature:

			Data sets			Local +Gl Measure	Feature	
Dimesnions	Naive baye	DT	Ada_SVM		Ada_SVM	Ada_Naive	Bag_DT	Bag_SVM
1	0.892056844	0.9	0.77		0.9	0.856	0.92	0.635
2	0.921628062	0.89	0.94		0.88	0.9145	0.98	0.845
3	0.937539018	0.87	0.95		0.87	0.92	0.8962	0.9845
4	0.937761	0.9	0.95		0.9	0.94117	0.8476	0.5

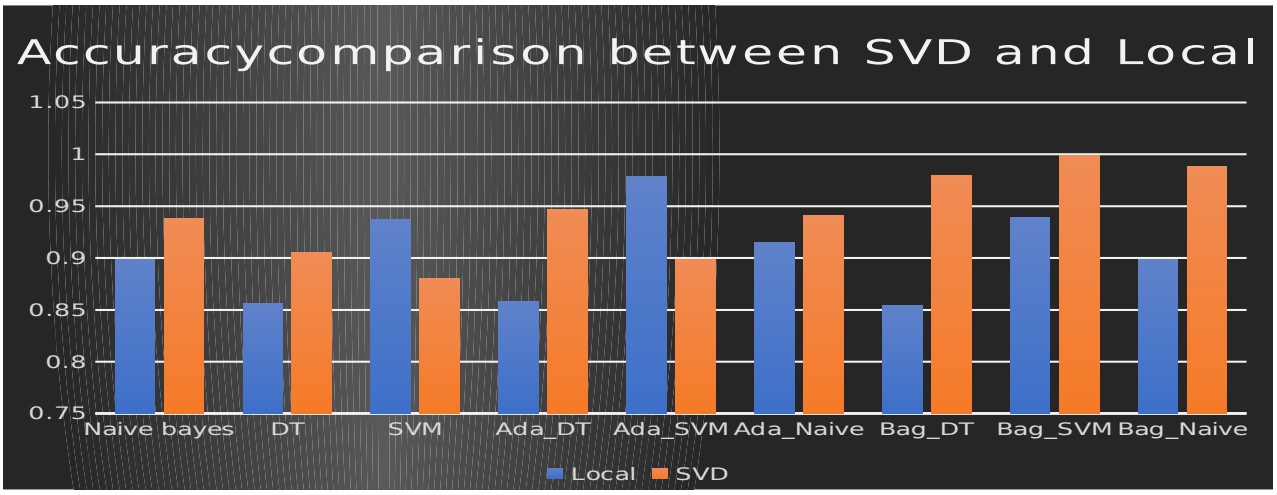


Figure 33: Accuracy Score comparison between performance of classifier on reduced dimension and performance of classifier on actual dimension

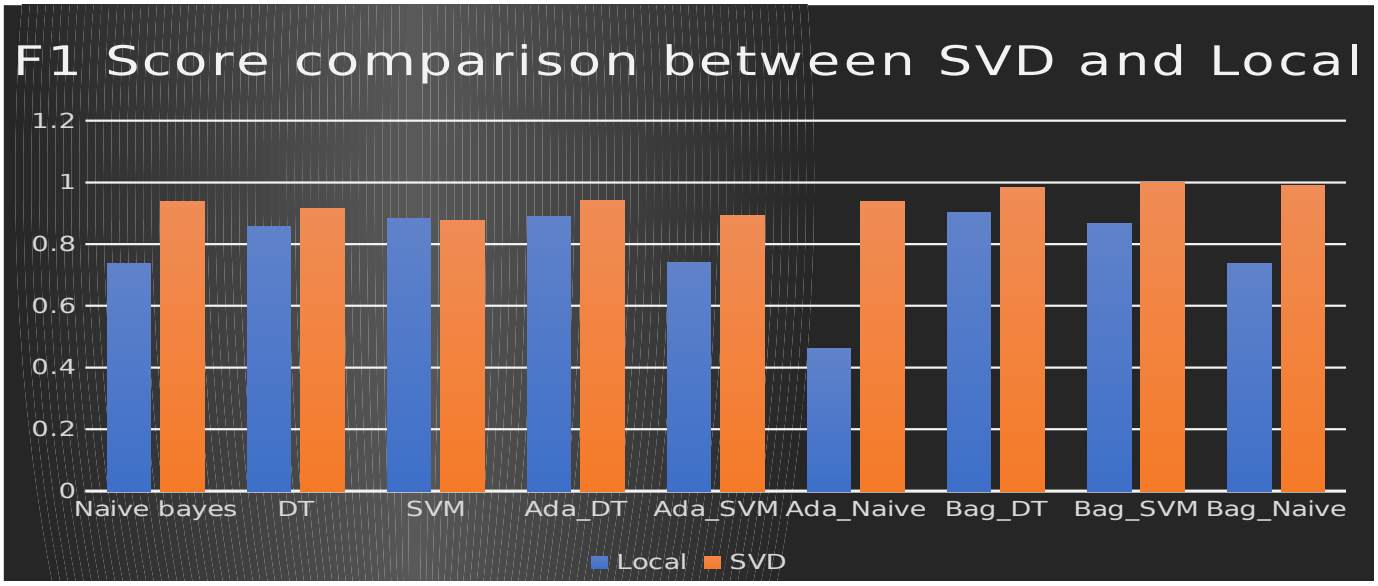


Figure 34: F1 Score comparison between performance of classifier on reduced dimension and performance of classifier on actual dimension

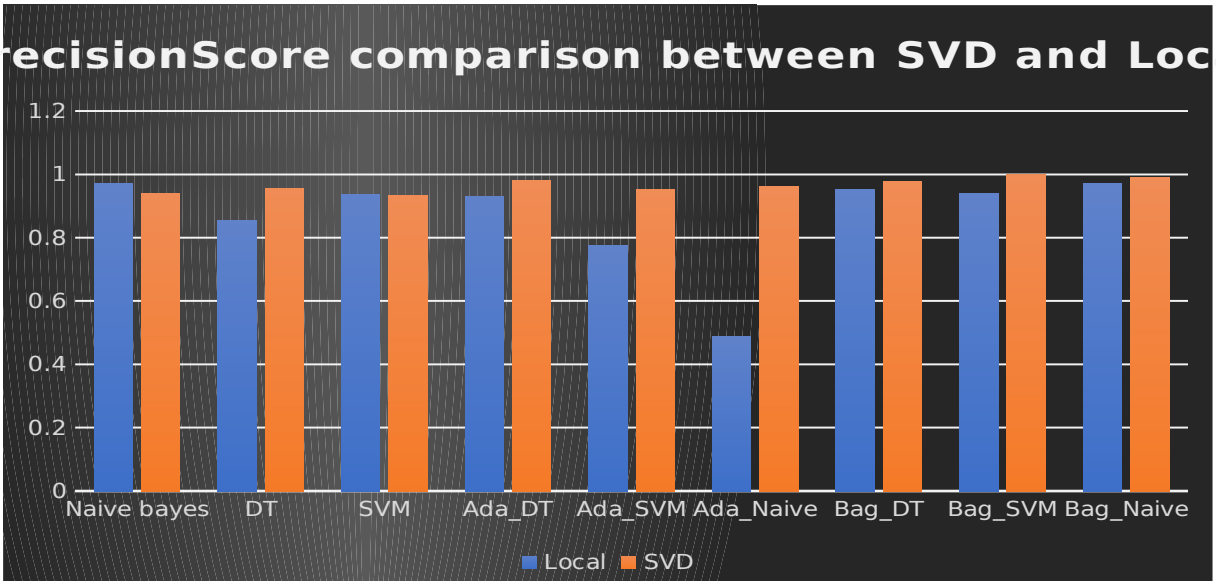


Figure 35: Precision Score comparison between performance of classifier on reduced dimension and performance of classifier on actual dimension

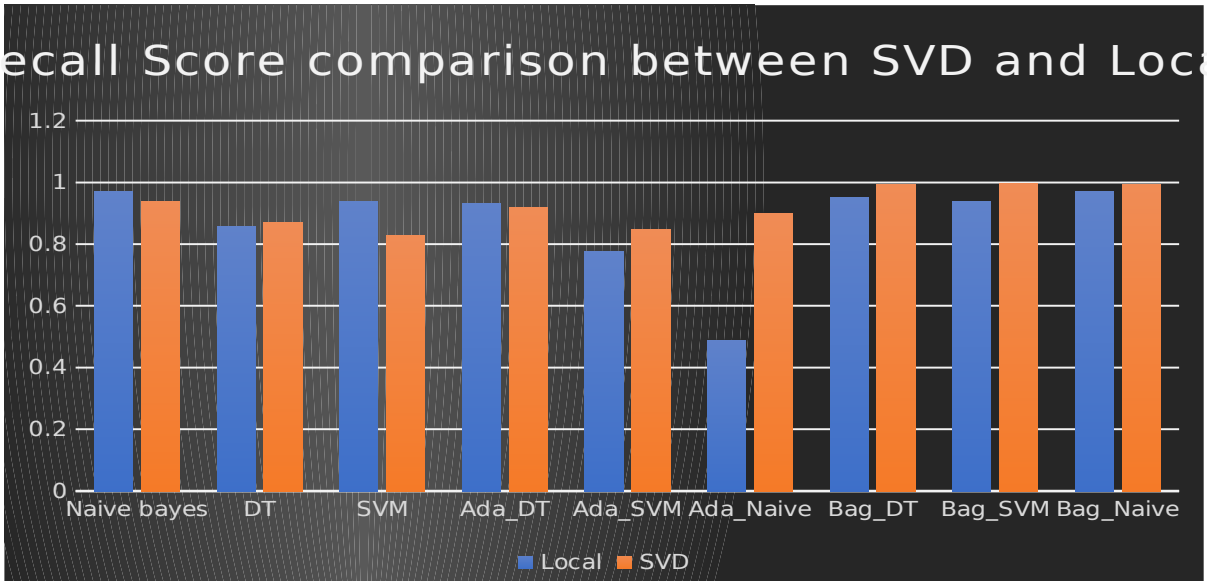


Figure 36: Precision Score comparison between performance of classifier on reduced dimension and performance of classifier on actual dimension

Blog Data Sets Local and Global Feature:

			Blog Data sets			Local +Glol Measure Feature			
Dimesnions	Naive bayes	DT	SVM	Ada_DT	Ada_SVM	Ada_Naive	Bag_DT	Bag_SVM	Bag_Naive
1	0.78	0.86	0.87	0.8695	0.8	0.68	0.854	0.87	0.5423
2	0.79	0.96	0.96	0.8956	0.96	0.66	0.912	0.96	0.586
3	0.85	0.97	0.97	0.9256	0.97	0.658	0.9355	0.97	0.6482
4	0.98	0.99	0.99	0.95	1	0.69	0.9103	1	0.359

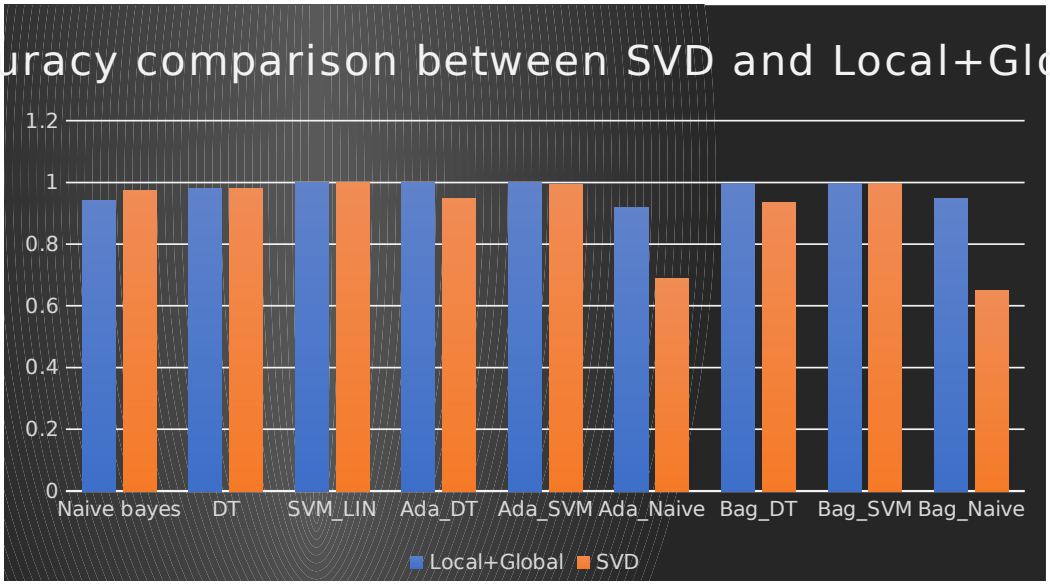


Figure 37: Accuracy Score comparison between performance of classifier on reduced dimension and performance of classifier on actual dimension

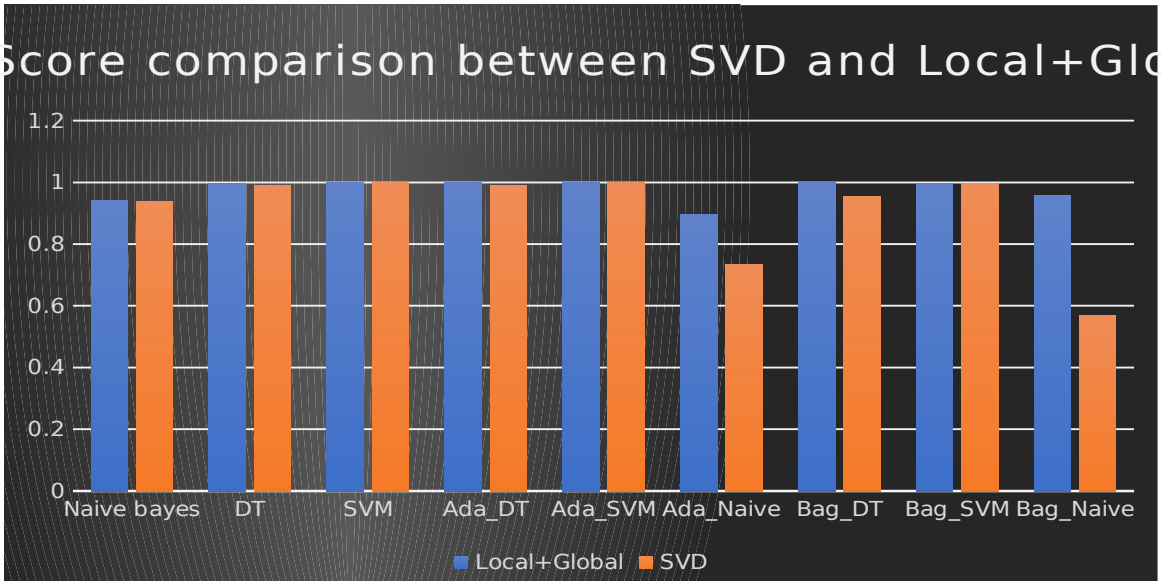


Figure 38: F1 Score comparison between performance of classifier on reduced dimension and performance of classifier on actual dimension

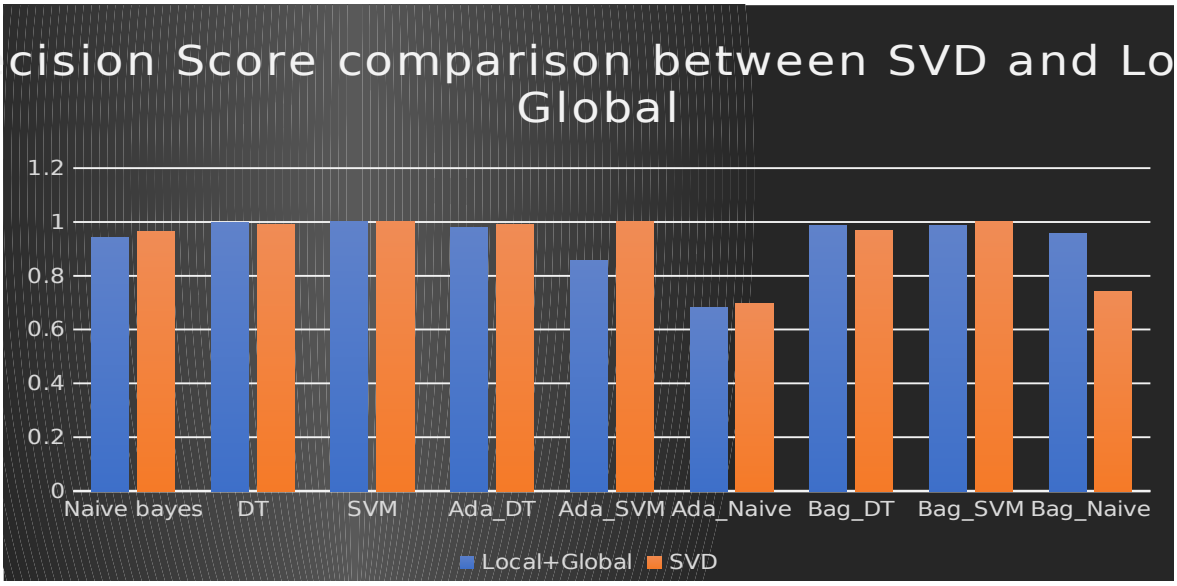


Figure 39: Precision Score comparison between performance of classifier on reduced dimension and performance of classifier on actual dimension

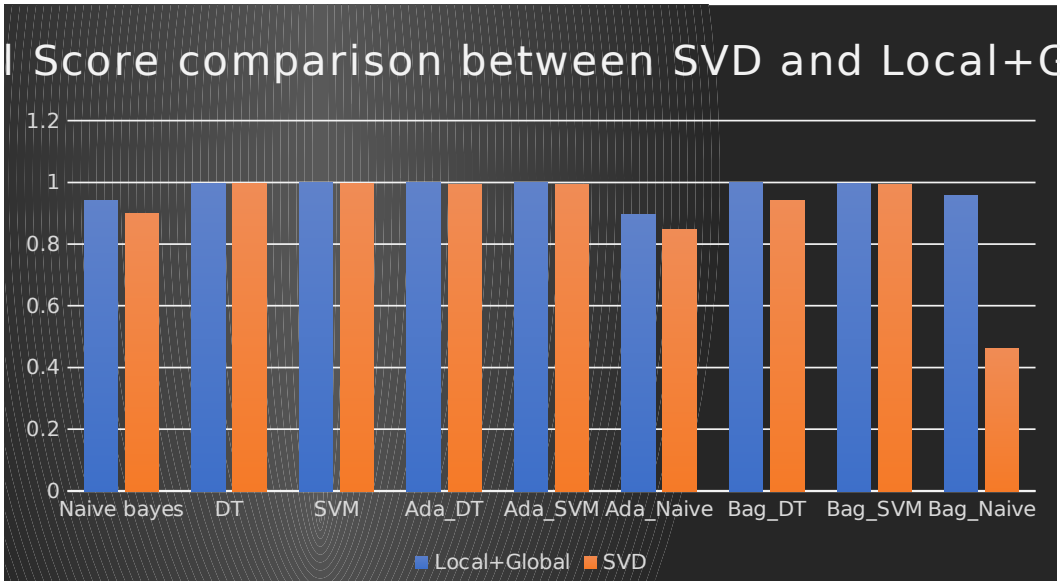


Figure 40: Precision Score comparison between performance of classifier on reduced dimension and performance of classifier on actual dimension

Deep Walk:

Restaurant Data sets:

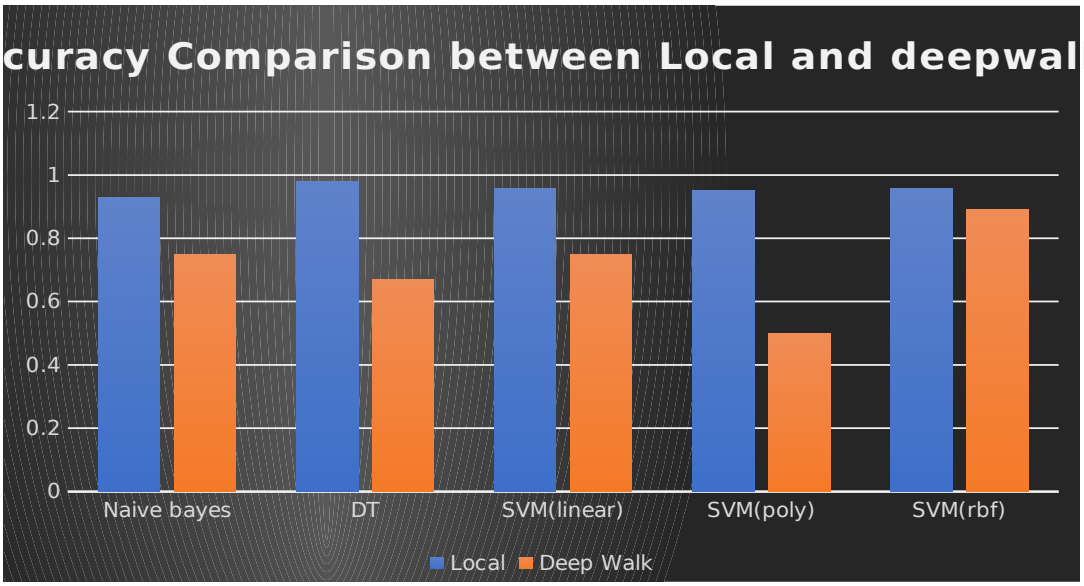


Figure 41: Accuracy comparison between performance of classifier on Deep walk and performance of classifier on actual dimension

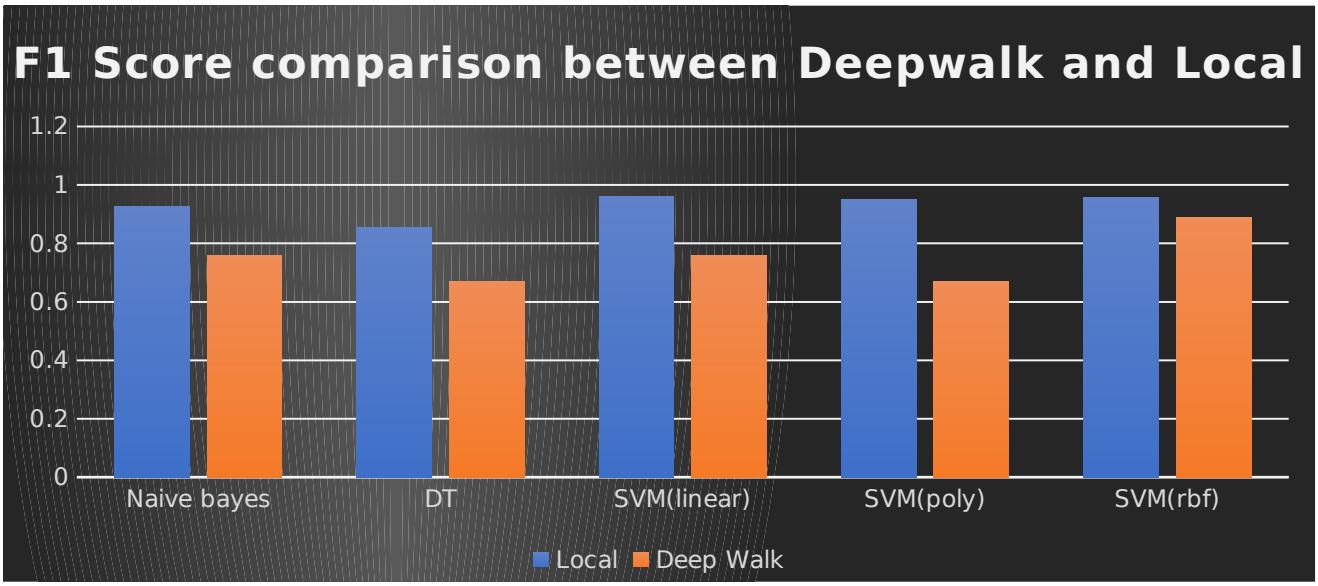


Figure 42: F1 comparison between performance of classifier on Deep walk and performance of classifier on actual dimension

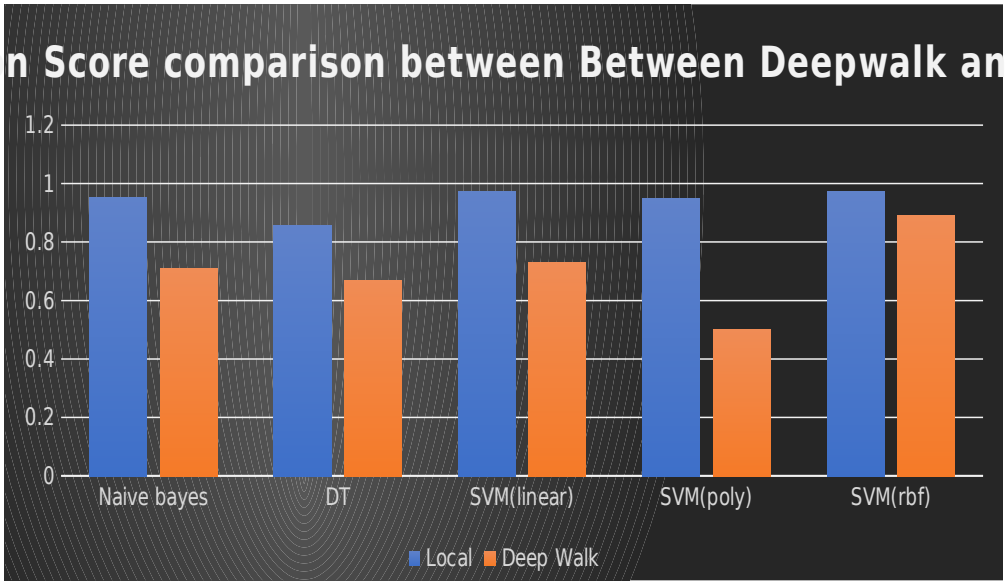


Figure 43: Precision score comparison between performance of classifier on Deep walk and performance of classifier on actual dimension

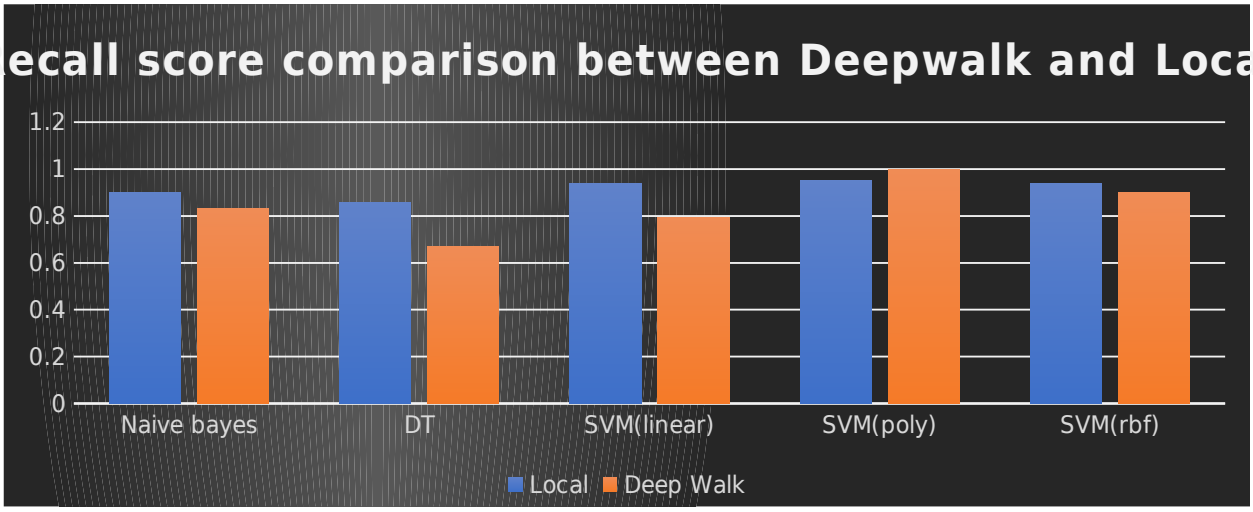


Figure 44: Recall score comparison between performance of classifier on Deep walk and performance of classifier on actual dimension.

Blog Data sets.

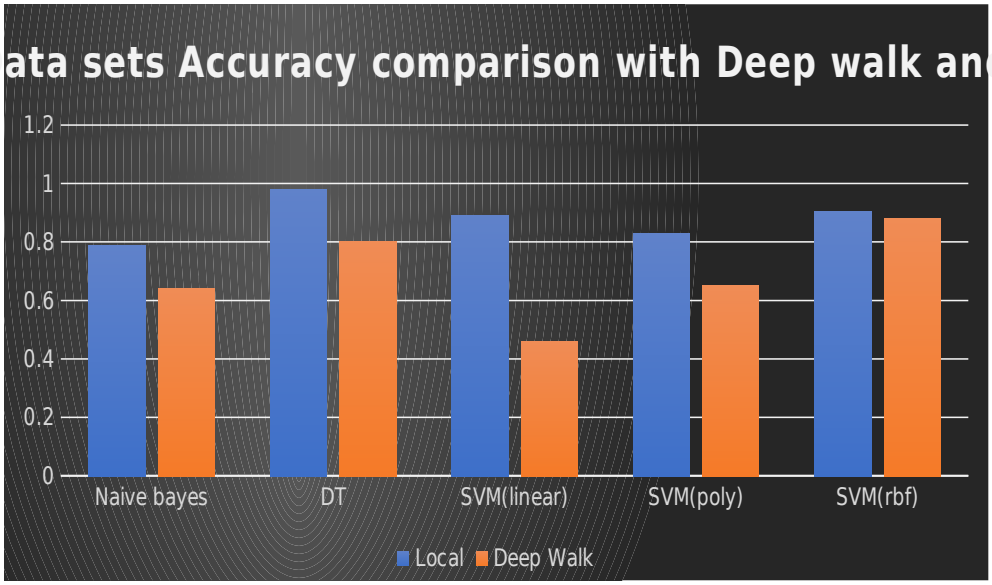


Figure 45: accuracy score comparison between performance of classifier on Deep walk and performance of classifier on actual dimension

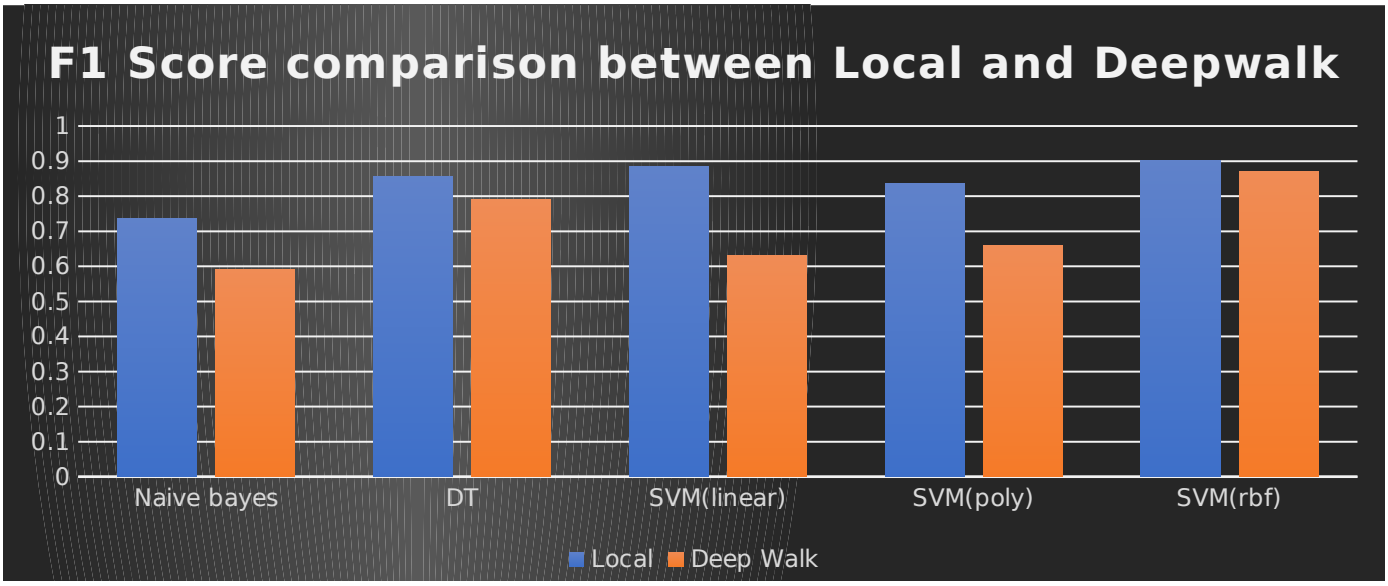


Figure 46: F1 score comparison between performance of classifier on Deep walk and performance of classifier on actual dimension

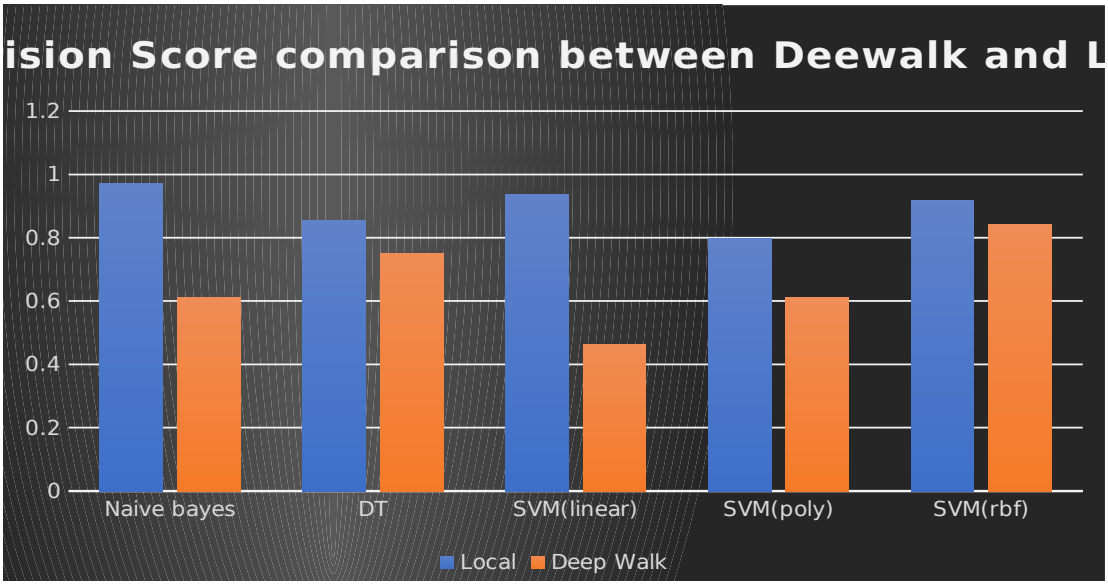


Figure 47: Precision score comparison between performance of classifier on Deep walk and performance of classifier on actual dimension

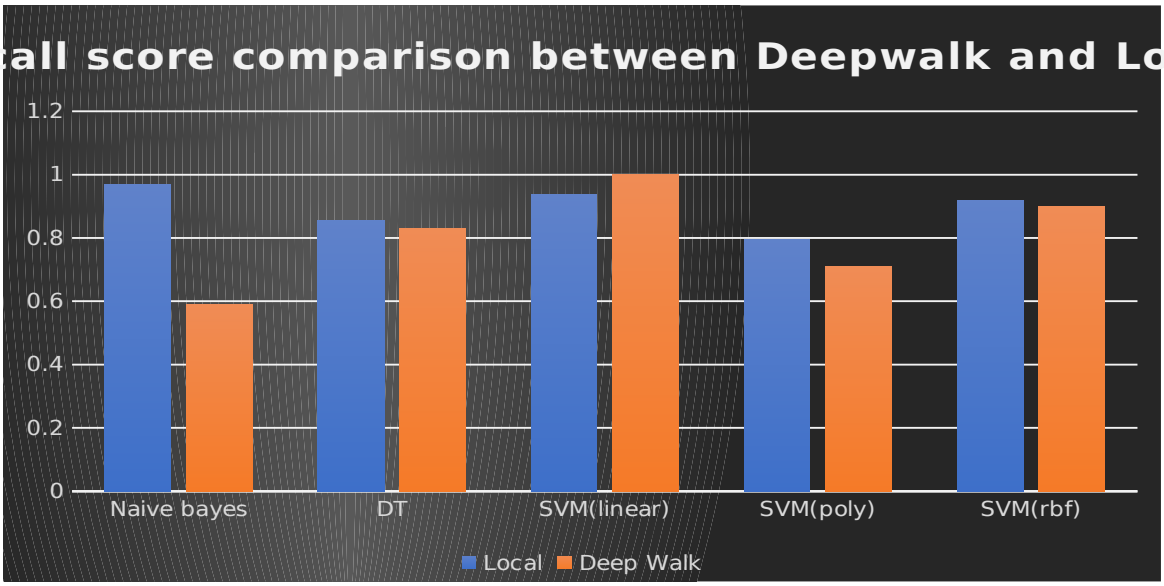


Figure 48: Recall score comparison between performance of classifier on Deep walk and performance of classifier on actual dimension

ROC Curves

Restaurant Dataset (Local measures)

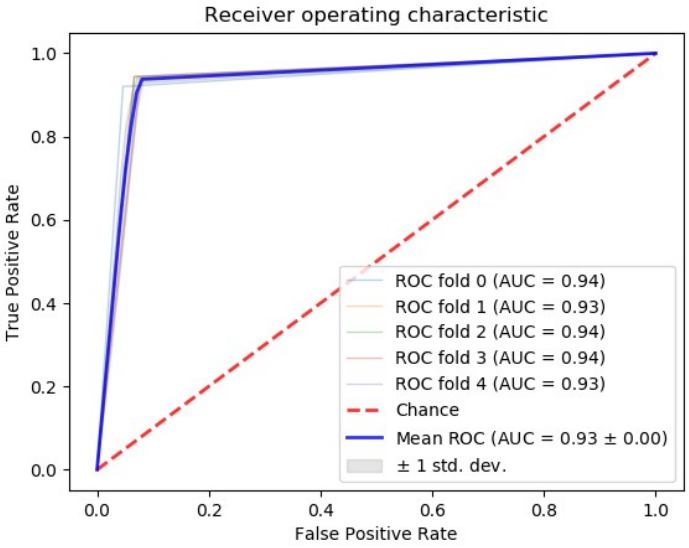


Fig 49: Decision Tree

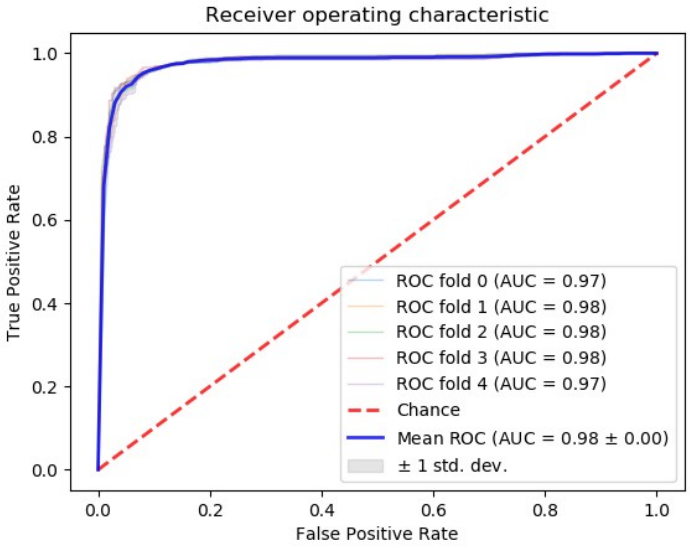


Fig 50: Naive Bayes

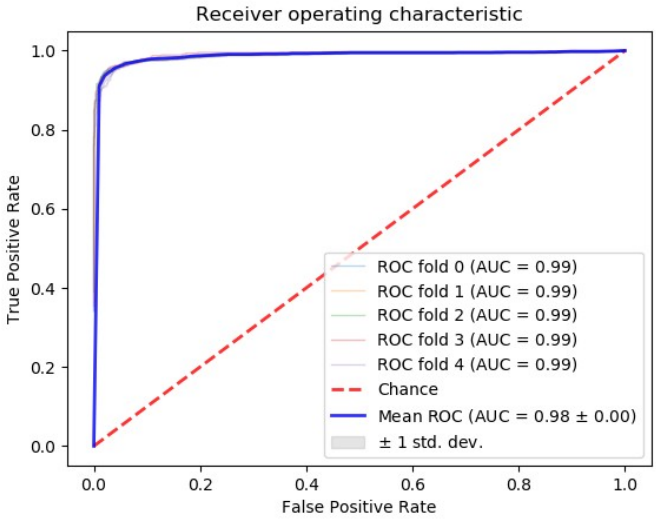


Fig 51: SVM(linear)

Observation: SVM and Naive Bayes are giving better classification with AUC score=0.98

Restaurant Dataset (Global measures)

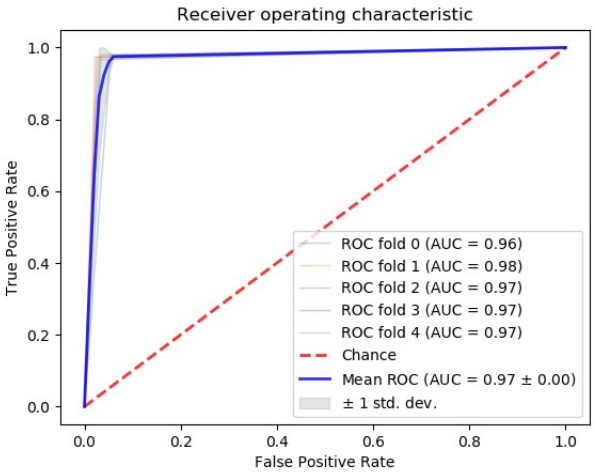


Fig 52: Decision Tree

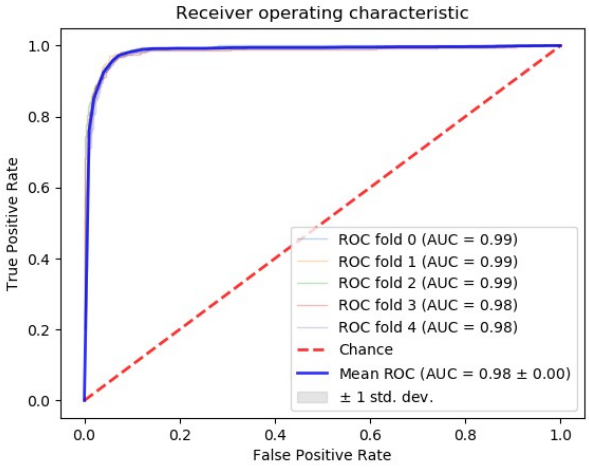


Fig 53: Naive Bayes

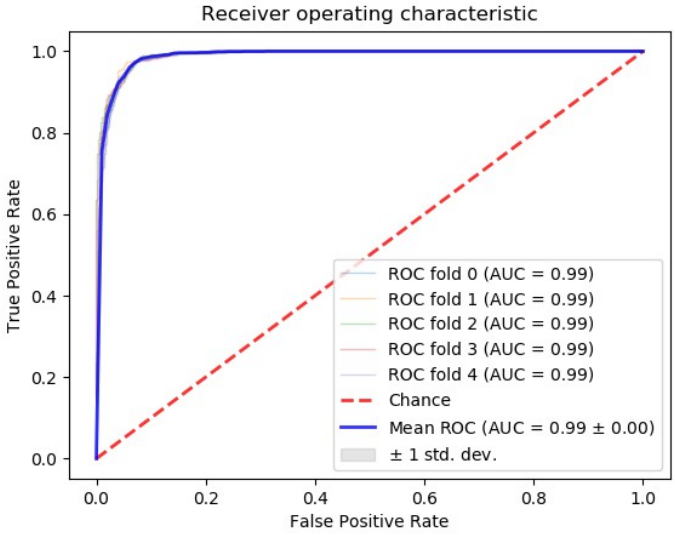


Fig 54: SVM(linear)

Observation: SVM is giving better classification with AUC score=0.99

Restaurant Dataset (local+Global measures)

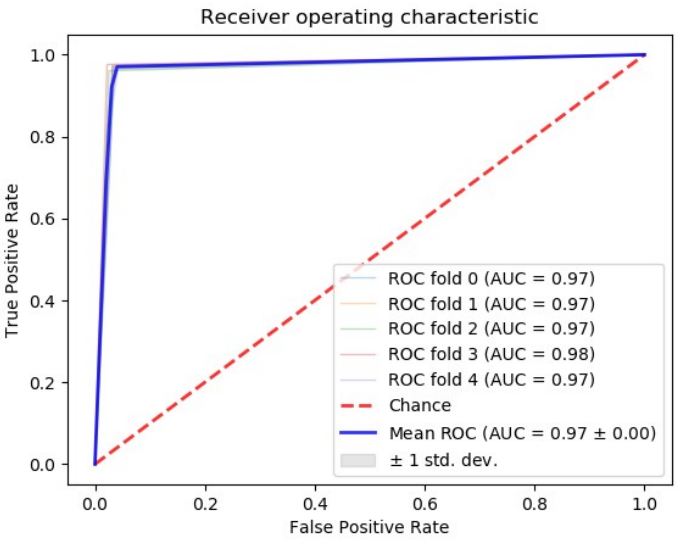


Fig 55: Decision Tree

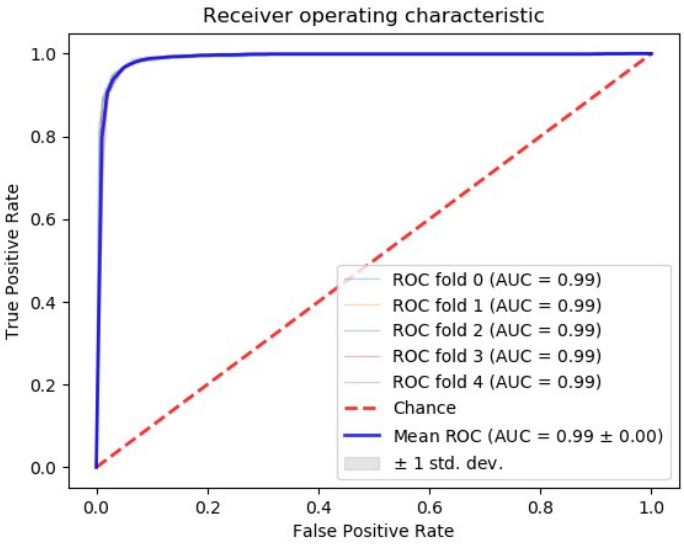


Fig 56: Naive Bayes

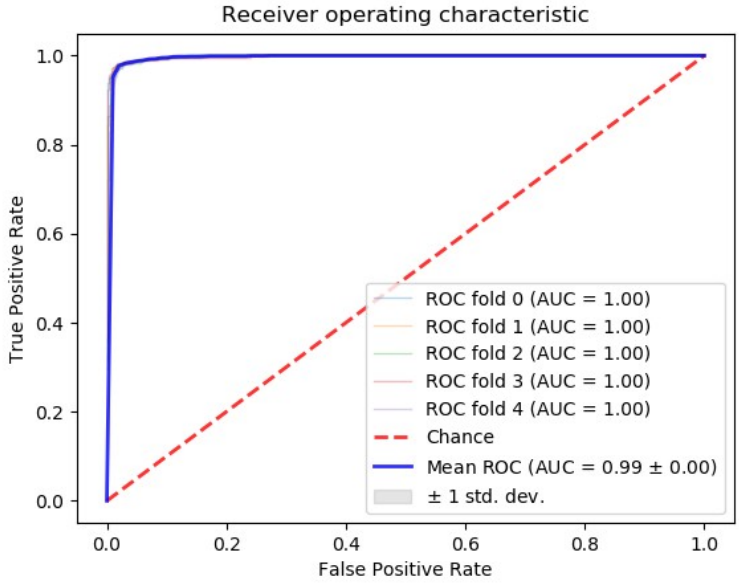


Fig 57: SVM(linear)

Observation: SVM and Naive Bayes are giving better classification with AUC score=0.99

Blog Dataset (Local measures)

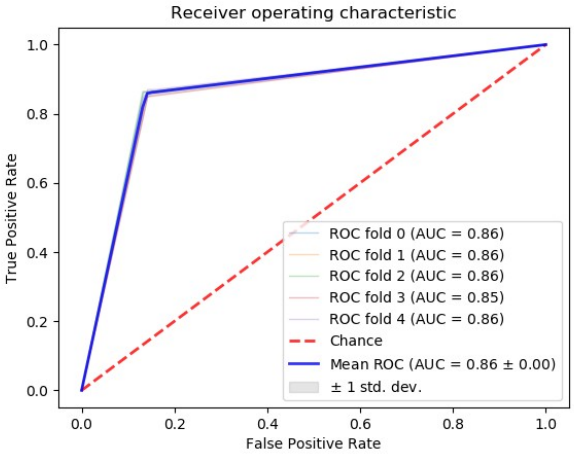


Fig 58: Decision Tree

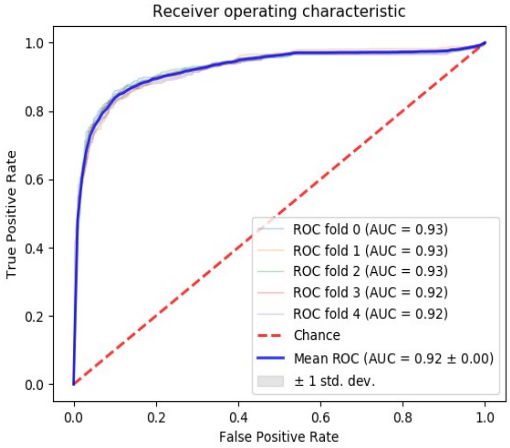


Fig 59: Naive Bayes

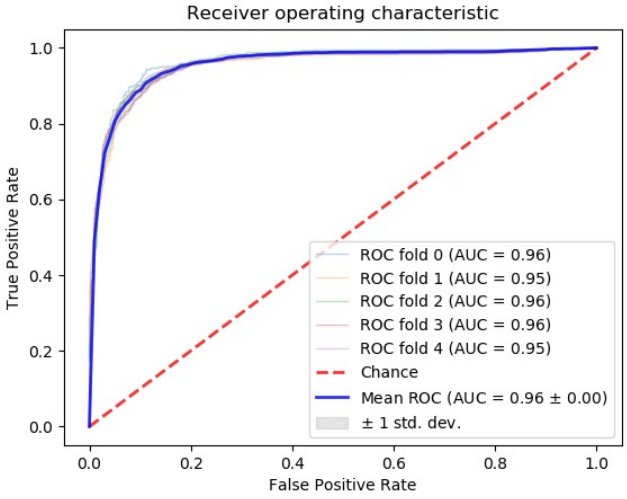


Fig 60: SVM(linear)

Observation: SVM is giving better classification with AUC score=0.96

Blog Dataset (Global measures)

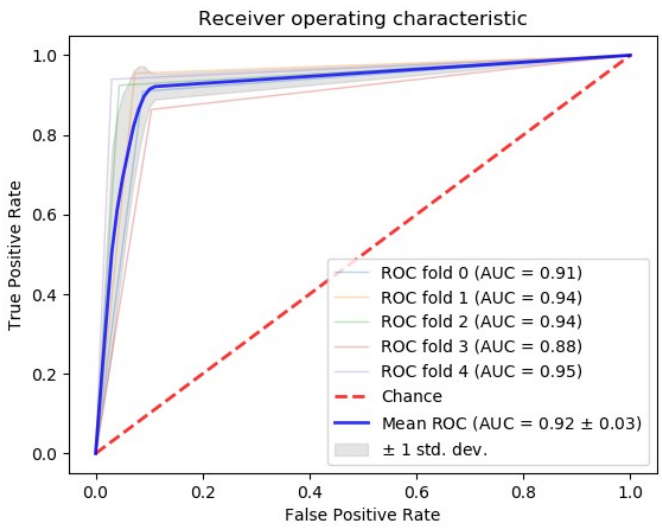


Fig 61: Decision Tree

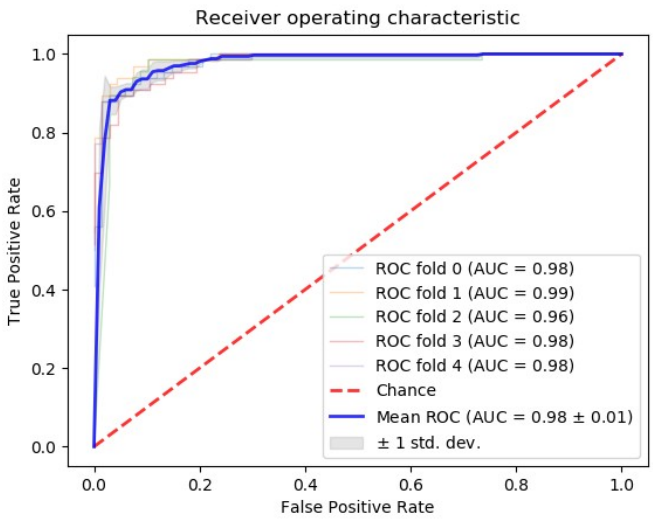


Fig 62: Naive Bayes

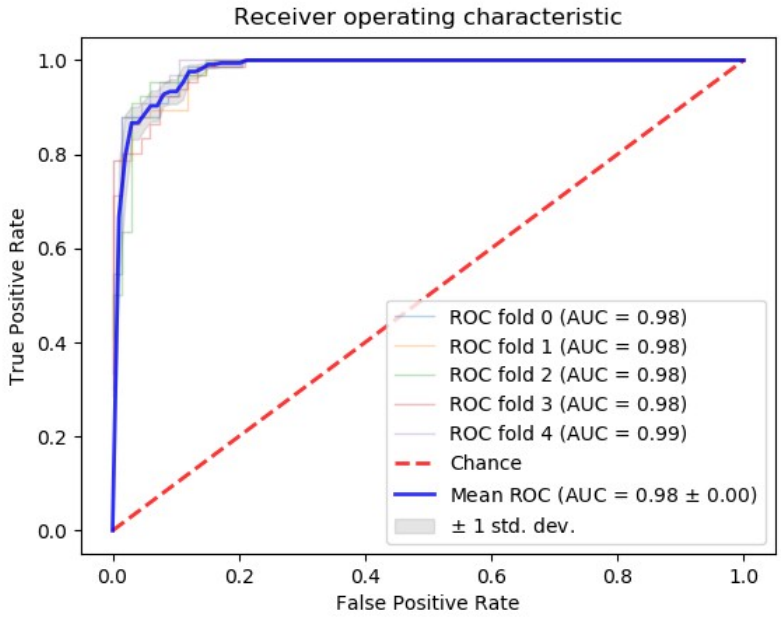


Fig 63: SVM(linear)

Observation: SVM and Naive Bayes are both giving better classification with same AUC score=0.98

Blog Dataset (Local+Global measures)

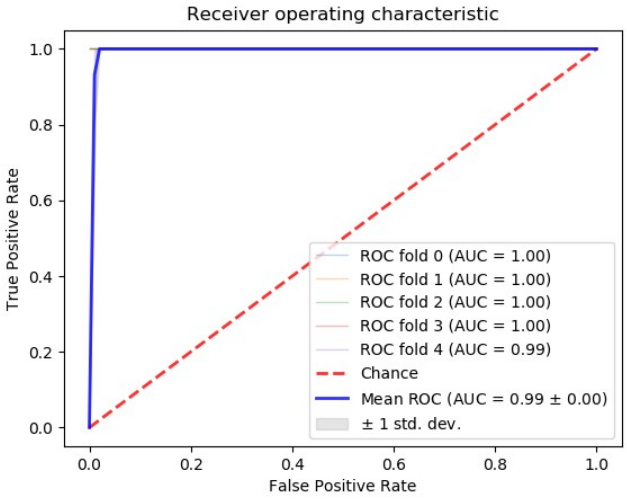


Fig 64: Decision Tree

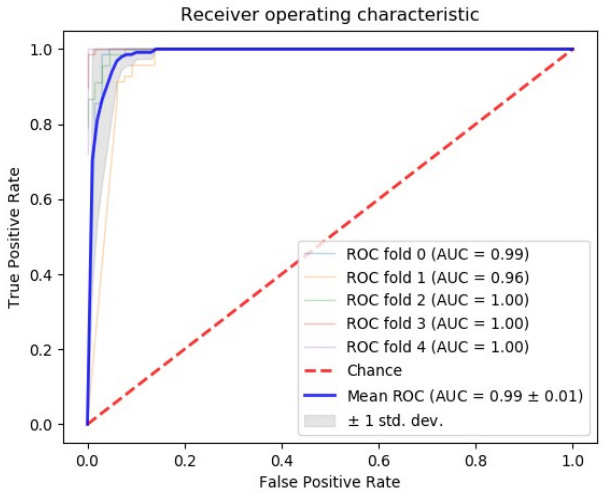


Fig 65: Naive Bayes

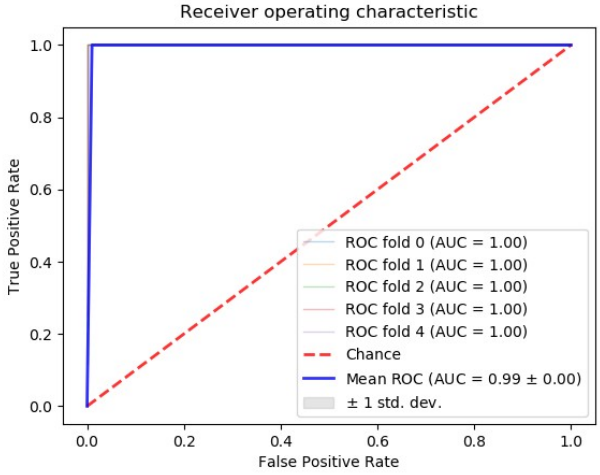
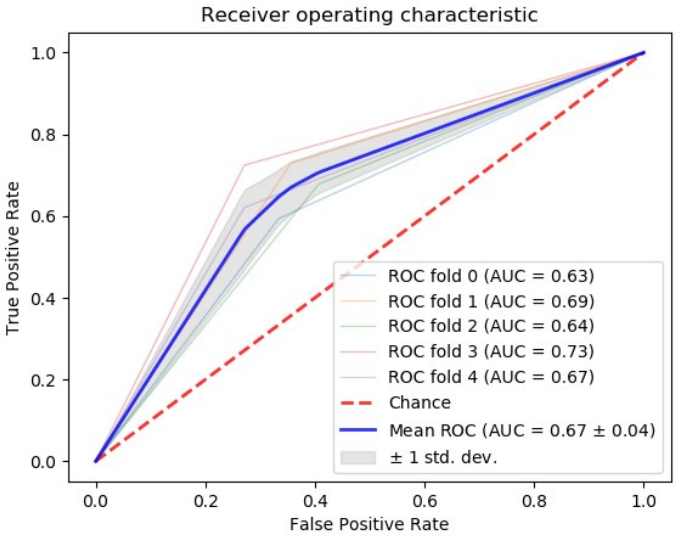
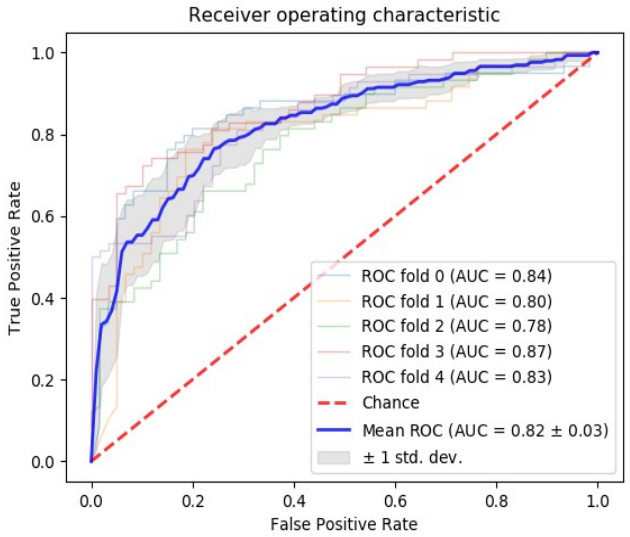


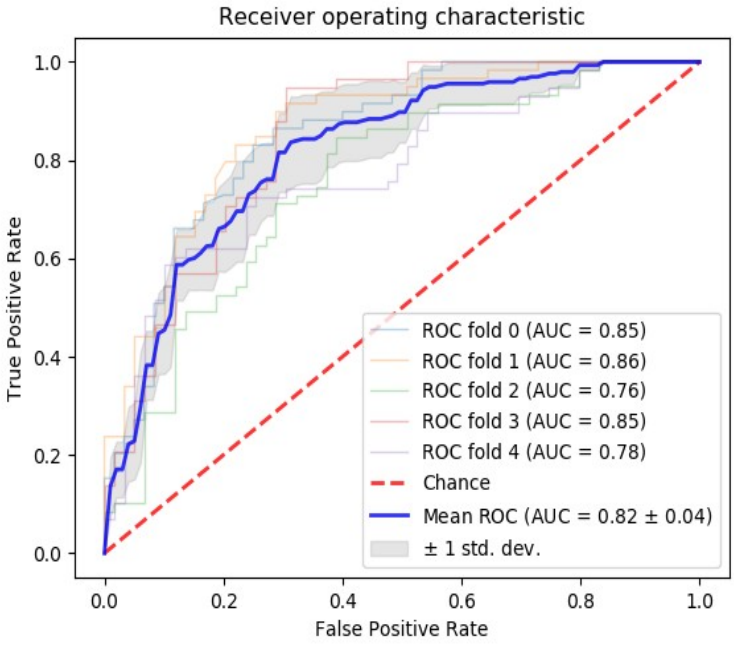
Fig 66: SVM(linear)

Observation: Here SVM, decision tree and Naive Bayes are giving the same AUC score=0.99

Deep Walk(Restaurant Dataset)

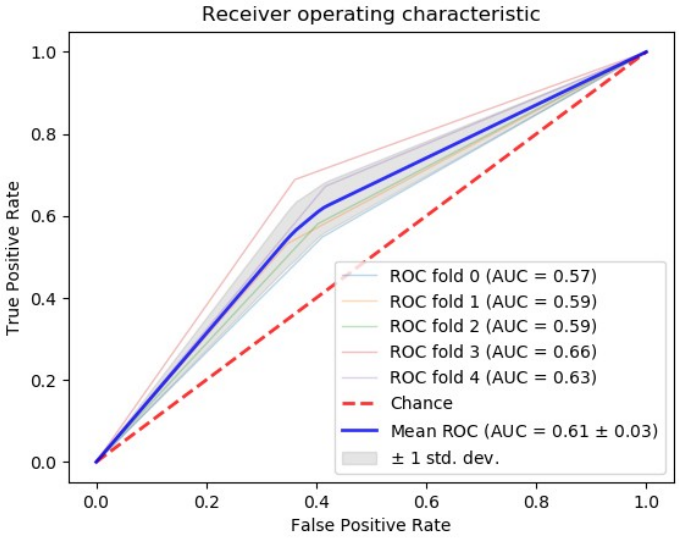


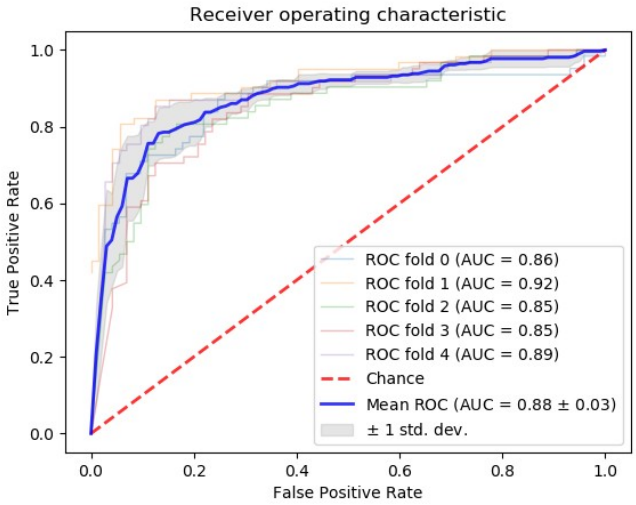


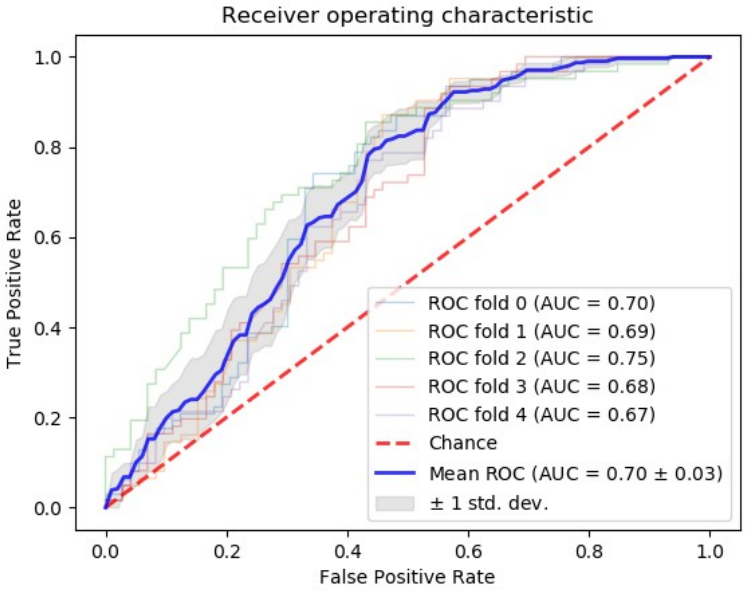


Observation: SVM and Naive Bayes are both giving better classification with same AUC score=0.82

Deep Walk(Blog Dataset)







Observation: Naive Bayes are both giving better classification with same AUC score=0.88

Dataset generated by Topological measure is giving better classification than node embedding by Deep walk .