

DATA ANALYTICS WITH R, EXCEL AND TABLAEU

ASSIGNMENT model evaluation assessment 15.1

ANSWERS

By ASHISH S SHANBHAG

ashishshanbhag108@gmail.com

5. Problem Statement

1. Use the below given data set DataSet

a. Predict the no of comments in next H hrs

Ans

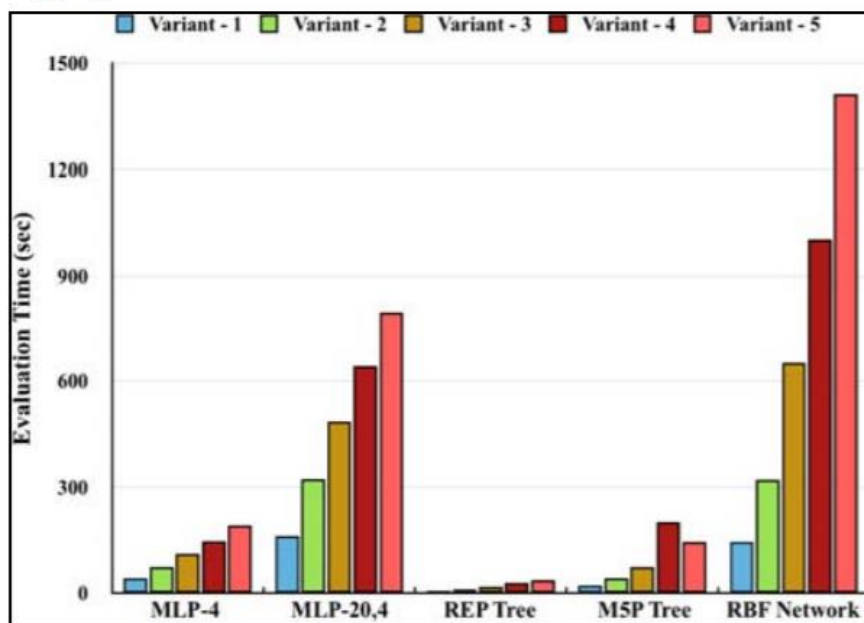


Figure 12. Evaluation time for comment volume prediction is presented in this graph.

In case of REP tree the evaluation time is decreased from 10.844 Sec to 6.034 Sec for variant - 1 and from 46.871 Sec to 37.627 Sec for variant - 5. For MSP tree it again decreased from 34.440 Sec to 21.616 Sec for variant - 1 and from 518.638 Sec to 146.027 Sec for variant - 5. This shows that the evaluation time is directly proportional to the training dataset

b. Use regression technique

For analysis, we crawled the Facebook pages for raw data, pre-processed it, and made a temporal split of the data to prepare the training and testing set. Then, this training set is used to train the regressor and performance of regressor is then estimated using testing data (whose target value is hidden) using some evaluation metrics.

c. Report the training accuracy and test accuracy

Ans Training Accuracy

Model		Variant – 1	Variant – 2	Variant – 3	Variant – 4	Variant – 5
MLP – 4	Hits@10	5.300 ± 1.487	4.700 ± 1.676	4.700 ± 1.792	5.200 ± 1.720	4.800 ± 1.470
	AUC@10	0.704 ± 0.168	0.729 ± 0.176	0.727 ± 0.160	0.809 ± 0.182	0.780 ± 0.168
	Time Taken	42.769 Sec	74.114 Sec	111.313 Sec	148.069 Sec	196.082 Sec
MLP – 20,4	Hits@10	5.400 ± 1.428	5.500 ± 1.628	4.900 ± 1.700	5.800 ± 1.077	5.600 ± 1.356
	AUC@10	0.797 ± 0.141	0.817 ± 0.151	0.755 ± 0.164	0.822 ± 0.145	0.787 ± 0.140
	Time Taken	163.152 Sec	323.098 Sec	484.597 Sec	642.405 Sec	798.430 Sec
REP Tree	Hits@10	6.400 ± 1.200	6.300 ± 1.269	6.200 ± 1.249	5.600 ± 1.020	6.500 ± 1.204
	AUC@10	0.878 ± 0.085	0.823 ± 0.092	0.846 ± 0.095	0.766 ± 0.097	0.834 ± 0.079
	Time Taken	6.034 Sec	10.757 Sec	18.397 Sec	28.802 Sec	37.627 Sec
MSP Tree	Hits@10	6.000 ± 0.775	6.400 ± 0.917	6.400 ± 0.917	5.900 ± 0.943	6.400 ± 1.114
	AUC@10	0.814 ± 0.140	0.808 ± 0.141	0.734 ± 0.152	0.747 ± 0.152	0.774 ± 0.144
	Time Taken	21.616 Sec	41.800 Sec	74.506 Sec	202.465 Sec	146.027 Sec
RBF Network (90 Clusters)	Hits@10	4.600 ± 0.800	4.100 ± 1.375	4.700 ± 0.781	4.300 ± 1.100	4.100 ± 1.044
	AUC@10	0.846 ± 0.290	0.714 ± 0.323	0.890 ± 0.175	0.934 ± 0.073	0.935 ± 0.058
	Time Taken	145.172 Sec	322.602 Sec	651.349 Sec	1004.657 Sec	1413.476 Sec

Testing Accuracy

Model		Variant – 1	Variant – 2	Variant – 3	Variant – 4	Variant – 5
MLP – 4	Hits@10	5.300 ± 1.269	6.000 ± 0.894	6.300 ± 0.781	6.200 ± 1.249	5.900 ± 1.375
	AUC@10	0.681 ± 0.134	0.781 ± 0.227	0.833 ± 0.183	0.836 ± 0.204	0.761 ± 0.194
	Time Taken	360.626 Sec	723.550 Sec	1067.561 Sec	1574.212 Sec	1782.111 Sec
MLP – 20,4	Hits@10	5.600 ± 1.428	6.100 ± 0.943	6.400 ± 0.917	6.400 ± 1.281	6.200 ± 0.980
	AUC@10	0.735 ± 0.147	0.816 ± 0.198	0.811 ± 0.197	0.828 ± 0.186	0.792 ± 0.177
	Time Taken	1576.265 Sec	3138.339 Sec	5065.178 Sec	6462.098 Sec	7773.225 Sec
REP Tree	Hits@10	6.800 ± 1.536	6.000 ± 1.000	6.300 ± 1.418	5.800 ± 1.249	6.300 ± 1.187
	AUC@10	0.746 ± 0.122	0.727 ± 0.113	0.761 ± 0.121	0.634 ± 0.128	0.722 ± 0.098
	Time Taken	52.722 Sec	137.485 Sec	229.443 Sec	302.766 Sec	412.973 Sec
MSP Tree	Hits@10	6.200 ± 1.077	6.400 ± 1.356	6.500 ± 1.025	6.900 ± 0.831	6.000 ± 1.183
	AUC@10	0.781 ± 0.148	0.714 ± 0.164	0.754 ± 0.181	0.610 ± 0.165	0.812 ± 0.156
	Time Taken	318.466 Sec	655.642 Sec	1142.914 Sec	1411.762 Sec	1922.949 Sec
RBF Network (90 Clusters)	Hits@10	5.000 ± 1.095	4.600 ± 1.020	4.100 ± 0.831	4.100 ± 1.221	4.400 ± 1.200
	AUC@10	0.964 ± 0.047	0.939 ± 0.076	0.949 ± 0.064	0.945 ± 0.064	0.945 ± 0.063
	Time Taken	2445.239 Sec	4619.697 Sec	10171.881 Sec	11996.384 Sec	16344.734 Sec