

MITS5509 Intelligent Systems for Analytics

Assignment 3



NOTE: This Document is used in conjunction with MITS5509

Objective(s)

This assessment item relates to the unit learning outcomes as in the unit descriptor. This assessment is designed to improve student collaborative skills in a team environment and to give students experience in constructing a range of documents as deliverables form different stages of the Intelligent Systems for Analytics

INSTRUCTIONS

Assignment 3:- Group Assignment (30%) and submission at week 12

In this assignment students will work in small groups to develop components of the Documents discussed in lectures. Student groups should be formed by Session four. Each group needs to complete the group participation form attached to the end of this document. Assignments will not be graded unless the student has signed a group participation form.

Carefully read the following two questions and provide the appropriate answer.

Question 1. The bankruptcy-prediction problem can be viewed as a problem of classification. The data set you will be using for this problem includes two ratios that have been computed from the financial statements of real-world firms. These two ratios have been used in studies involving bankruptcy prediction. The first sample (training set) includes 68 data value on firms that went bankrupt and firms that didn't. This will be your training sample. The second sample (testing set) of 68 firms also consists of some bankrupt firms and some non bankrupt firms. Your goal is to use different classifiers to build a training model, by randomly selecting the 40 data points (20 points from category 1 and 20 points from category 0), and then test its performance on the testing model by randomly selecting 40 data points from the testing set. (Try to analyze the new cases yourself manually before you run the neural network and see how well you do). Both Data Sets are provided below:

Students have to use the following classifiers. The selection of the classifiers depend upon the members of the group. E.g. If the group has four members then they will use the four classifiers from the following six classifiers.

- 1. Neural networks
- 2. Support vector machines
- 3. Nearest neighbor algorithms
- 4. Decision trees
- 5. Naive Bayes
- 6. Any other classifier



The following tables show the training sample and test data you should use for this exercise.

	Training set			
Firm	WC	DC	Category	
1	3338.61	0.56555	1	
2	3801.72	0.570567	1	
3	2818.817	0.572058		
4	1250.953	0.568258	1	
5	2444.406	0.553276		
6	937.917	0.561066	1	
7	1600.792	0.534662	1	
8	3128.813	0.564714	1	
9	2486.803	0.564239	1	
10	4220.996	0.58465	1	
11	2585.41	0.572457	1	
12	3512.085	0.550878	1	
13	4170.333	0.569516	1	
14	938.879	0.545574	1	
15	1437.695	0.529922	1	
16	627.985	0.51941	1	
17	4430.049	0.567547	1	
18	989.568	0.534501	1	
19	3275.474	0.555306	1	
20	1500.437	0.565886	1	
21	848.989	0.548603	1	
22	1386.494	0.56229	1	
23	1554.257	0.562346		
24	2228.338	0.565556	1	
25	2568.391	0.54973 1		
26	1720.128	0.568458 1		
27	4106.106	0.57767 1		
28	3500.883	0.557197 1		
29	1217.846	0.525333	1	
30	3544.406	0.568735	1	
31	2082.873	0.557527	1	
32	709.01	0.541673	1	
33	2523.939	0.55366	1	
34	2781.307	0.569188	1	
35	309.577	0.557668	0	
36	363.79	0.561751	0	
37	341.399	0.550717	0	
38	363.616	0.568882	0	



	39	323.673	0.554499	0
	40	323.353	0.558233	0
	41	350.371	0.566447	0
	42	240.602	0.5656	0
	43	220.057	0.544182	0
	44	287.837	0.522119	0
	45	274.6	0.551492	0
	46	278.494	0.550846	0
	47	234.267	0.554828	0
	48	284.923	0.533586	0
	49	190.62	0.54899	0
	50	327.76	0.538896	0
	51	211.94	0.551569	0
	52	373.571	0.549753	0
	53	219.891	0.546936	0
	54	193.489	0.56059	0
	55	204.333	0.550777	0
	56	205.657	0.550677	0
	57	362.361	0.551315	0
	58	285.562	0.578965	0
	59	352.649	0.541763	0
	60	400.44	0.557809	0
	61	307.301	0.578949	0
	62	240.314	0.548355	0
	63	322.995	0.569978	0
	64	408.197	0.574972	0
	65	209.027	0.554203	0
	66	198.979	0.559771	0
	67	340.418	0.57343	0
_	68	320.154	0.560661	0



Testing set				
Firm	WC	DC		
1	4204.066	0.578231		
2	1411.733	0.560415		
3	4197.206	0.565368		
4	1121.866	0.540554		
5	820.683	0.566067		
6	1349.887	0.524683		
7	3128.736	0.547596		
8	2551.433	0.57368		
9	809.115	0.552148		
10	2866.623	0.559484		
11	1193.951	0.515996		
12	2014.445	0.564598		
13	4400.268	0.578645		
14	266.396	0.550131		
15	243.554	0.559966		
16	172.184	0.566274		
17	362.479	0.553563		
18	249.981	0.55274		
19	327.877	0.565451		
20	286.696	0.572919		
21	182.762	0.56313		
22	338.347	0.546618		
23	302.57	0.551846		
24	1781.718	0.564307		
25	3711.358	0.570857		
26	2030.189	0.564332		
27	845.019	0.550468		
28	1925.183	0.574114		
29	1549.089	0.538726		
30	1953.371	0.577015		
31	932.5	0.564721		
32	924.554	0.554162		
33	2386.011	0.545268		
34	2112.875	0.560262		
35	3568.877	0.561775		
36	4104.984	0.570978		
37	367.325	0.533232		
38	347.513	0.552354		
39	330.226	0.549799		



40	178.106	0.574406
41	378.899	0.531441
42	257.212	0.565379
43	333.088	0.54545
44	182.324	0.569686
45	238.099	0.563344
46	329.643	0.558005
47	294.644	0.556574
48	1058.649	0.54729
49	956.021	0.546774
50	2089.824	0.572031
51	2198.033	0.558597
52	4538.527	0.560383
53	3137.934	0.544445
54	2002.459	0.58141
55	2136.376	0.562953
56	281.666	0.553904
57	308.086	0.553646
58	317.079	0.560538
59	245.139	0.567829
60	354.662	0.548939
61	292.256	0.557991
62	306.79	0.57065
63	222.396	0.547811
64	367.628	0.53711
65	342.115	0.562531
66	353.326	0.548094
67	336.39	0.539131
68	298.008	0.562856

From the above data set, the group has to prepare a report which include the following:

- 1. List the values (40 values) in the Table used for Training set
- 2. List the values (40 values) in the Table used for Testing set
- 3. The output results of each classifier for the testing set in Table form
- 4. Snapshot or Screenshot of each of the steps

Note: Students can use any open source free data mining software such as Statistica Data Miner, Weka, RapidMiner, KNIME and MATLAB etc.



Question 2. Create a DASHBOARD. For creating a dashboard, the group can use the above database or any other database. The group have to prepare a report which include the following:

- 1. List of the values in the Table used for creating the dashboard
- 2. A Snapshot or Screenshot of each of the steps

The above list of documents is not necessarily in any order. The chronological order we cover these topics in lectures is not meant to dictate the order in which you collate these into one coherent document for your assignment.

Your report must include a Title Page with the title of the Assignment and the name and ID numbers of all group members. A contents page showing page numbers and titles of all major sections of the report. All Figures included must have captions and Figure numbers and be referenced within the document. Captions for figures placed below the figure, captions for tables placed above the table. Include a footer with the page number. Your report should use 1.5 spacing with a 12 point Times New Roman font. Include references where appropriate. Citation of sources (if using any) is mandatory and must be in the Harvard style.

Only one submission is to be made per group. The group should select a member to submit the assignment by the due date and time. All members of the group will receive the same grade unless special arrangement is made due to group conflicts. Any conflict should be resolved by the group, but failing that, please contact your lecture who will then resolve any issues which may involve specific assignment of work tasks, or removal of group members.

What to Submit

All submissions are to be submitted through turn-it-in. Drop-boxes linked to turn-it-in will be set up in the Unit of Study Moodle account. Assignments not submitted through these drop-boxes will not be considered.

Submissions must be made by the due date and time (which will be in the session detailed above) and determined by your Unit coordinator. Submissions made after the due date and time will be penalized at the rate of 10% per day (including weekend days).

The turn-it-in similarity score will be used in determining the level if any of plagiarism. Turn-it-in will check conference web-sites, Journal articles, the Web and your own class member submissions for plagiarism. You can see your turn-it-in similarity score when you submit your assignment to the appropriate drop-box. If this is a concern you will have a chance to change your assignment and re-submit. However, re-submission is only allowed prior to the submission due date and time. After the due date and time have elapsed you cannot make re-submissions



and you will have to live with the similarity score as there will be no chance for changing. Thus, plan early and submit early to take advantage of this feature. You can make multiple submissions, but please remember we only see the last submission, and the date and time you submitted will be taken from that submission

Please Note: All work is due by the due date and time. Late submissions will be penalized at the rate of 10% per day including weekends.



Group Participation Form

This form is to be completed by the group and returned to your tutor/lecturer as soon as possible.

We, the undersigned, agree to contribute individually and as a team to complete the Group Assignment for MITS5509 Intelligent Systems for Analytics in the time specified. (It should be noted that failure to participate in a group may result in a fail for the assignment component of the subject.)

Group membership:

	Surname	First name	Student ID	Date	Signature
1.				/	
2.				/	
3.				/	
4.				/	

^{*} All members in the team will receive the same mark for an assignment, unless there are extenuating circumstances whereby an individual's mark has to be altered by the tutor/lecturer, or if the peer group assessment warrants it.

^{**} Team members should contact their tutor/lecturer immediately if problems arise within the team that may cause completion of an assignment to be severely delayed, or the quality of the submission to be substantially lowered.

^{***} No additions or deletions of Team Members from this form allowed unless agreed to by your Instructor