CHAPTER 5

THE IMPLEMENTATION, RESULT AND ANALYSIS

5.1 Introduction

This Chapter consists of two (2) main parts; description of the experimental setup for testing and the analysis of the results and findings of the testing. The whole experiment is arranged in a structured manner to determine five (5) different outcomes. Every experiment is measured for its performance with evaluation metrics such as true positive and accuracy value.

5.2 Experimental Setup: Validating The Algorithm Via RiCCA

The initial test is conducted to identify and verify the initial version of the proposed algorithm. Hence, in order to verify the findings in the preliminary testing, a larger size of the dataset is required. This experiment measures the intensity or degree of severity for a text spam message consists of five (5) series of simulation, as shown below:

Table 5.1: Series Of Experiments

Experiment	Description and Justification
1	To identify the most effective term weighting schemes, risk scale, signal weighting, and anomaly threshold value that are adequately suitable for DCA and dDCA. The weight is used to define the interrelationship between three (3) different input signals. A sensitivity analysis of a range of weights is performed to ensure that the derived weights are suitable for the signal processing (Greensmith, 2007).
2	To identify which immune classifier is better, either DCA or dDCA by applying the best result from Experiment 1.
3	To study the effect of an imbalanced number of spam and ham messages in initial population and also to identify the suitable proportion of spam and ham messages in retrieving the risk level with optimum accuracy rate. This experiment is executed using the best specifications identified from Experiment 1.
4	To study if there is any significant impact of antigen multiplication, both with the deployment of DCA and dDCA. Characteristic of weighting schemes across various sizes of antigen corpus and its frequency may behave differently. The experiment is executed using the best specifications identified from Experiment 1.
5	To evaluate the ability of non-immune classifiers in spam filtering and severity assessment. This experiment also intended to recognize its differences with immune classifier.

5.2.1 Dataset Deployment

As aforementioned in Chapter 3, a set of data downloaded from UCI Machine Learning Repository is deployed. For simulation using RiCCA, in addition to UCI dataset, there are 1,012 self-collected spam messages. These messages are combined with the UCI and the partition of the dataset is as follows:

Table 5.2: Source Of Dataset

Dataset ID	Source	Ham Messages	Spam Messages
A	UCI Machine Learning	4,827	747
B	Self-collected spam messages	0	1,012
C	Combined Dataset A and B	4,827	1,759
	Total	**6,586**	

All the dataset are merged as Dataset C and to ease the identification of these messages, every message is labelled as the following:

i. H#.txt – refer to text ham message, with number (#) from Dataset A
ii. S#.txt – refer to text spam message, with number (#) from Dataset A
iii. ES#.txt - refer to text spam message, with number (#) from Dataset B

For self-collected spam messages, a mobile application named SMS Backup & Restore version 9.74.1 is used to extract the messages from an Android phone. The data extracted is amounted 1,012 spam messages which dated in between 15th January 2015 until 30th June 2017, and labelled as Dataset B. This self-collected dataset also has been shared with intention that other researchers could make use of it.

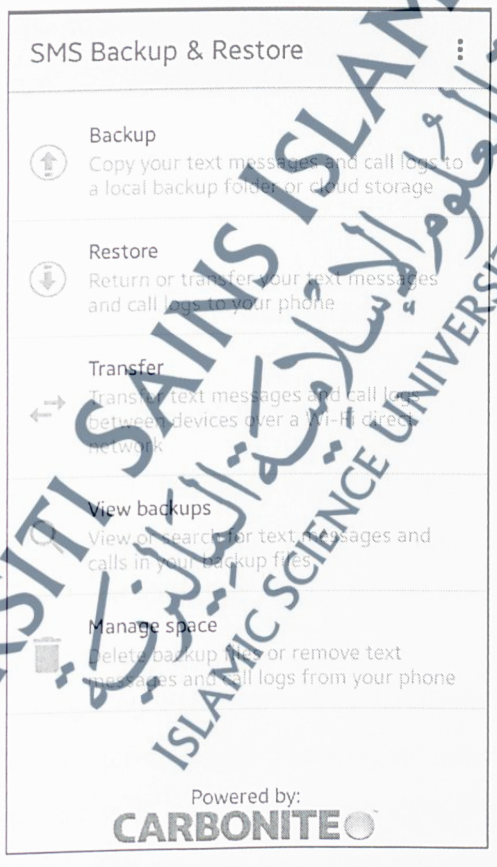


Figure 5.1: The Screenshot Of SMS Backup & Restore Mobile Application

In the following experiment, Dataset C has been deployed to create the initial population. Subsequently, 600 messages have been randomly selected to be tested in the testing phase.

5.3 Series Of Experiments

A detailed of experiment process flow is depicted in the following Figure 5.2.

No.	Process Flow	Software
1.	• Pre-processing (tokenization); and • Calculate weight for every tokenized words using Term Weighting Scheme	RapidMiner (version 5.3.015)
2.	Every SMS message is stored in a different file, one file .txt consist of one SMS message only.	Notepad (version 6.1)
3*.	Calculate the risk of SMS messages using DCA and dDCA.	**RiCCA prototype**
4.	SMS messages are labelled according to its measured risk.	

*Process flow number 3 is executed automatically using the developed prototype, RiCCA.

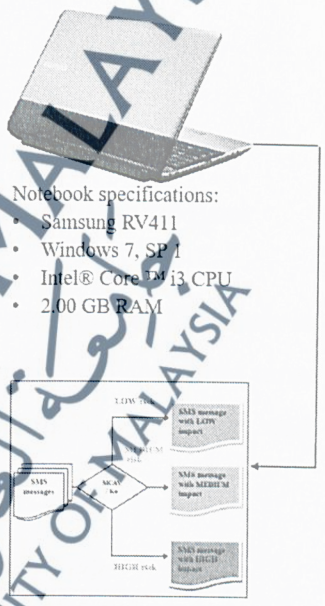


Figure 5.2: The Process Flow Of Simulated Experiment Using RiCCA

5.3.1 Experiment 1 – Influential Parameters With Effective Values

In this experiment, the most effective term weighting schemes, risk scale, signal weighting, and anomaly threshold value that adequately suitable for DCA is required to be identified. These factors may influence the results of the assessment.

i. Term Weighting Schemes – TF, IG Ratio and CHI^2 are used to calculate the term weight. The weights indicate the input signals of antigen and it is calculated using data mining tool, RapidMiner version 5.3. The weights were then retrieved and stored in RiCCA prototype for further processes.

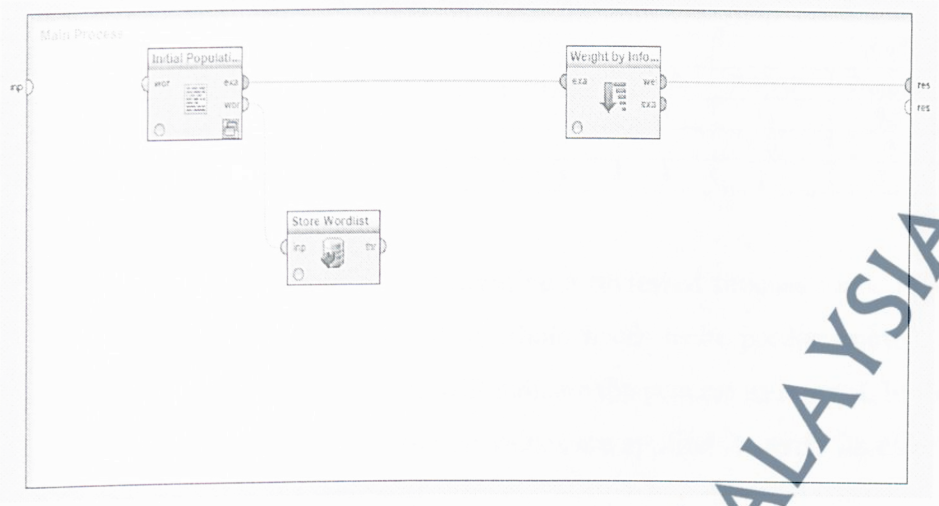


Figure 5.3: The Screenshot Of IG Ratio Process Using RapidMiner

ii. Risk Scale – Three (3) options of risk scale are applied. The scale is used to differentiate the levels of input signals (calculated from term weighting schemes) and output signals (calculated from DCA and dDCA).

Table 5.3: Three (3) Different Ranges Of Risk Scale

S1	S2	S3	Risk level	
			Input signals	**Output signals**
1.00 - 0.70	1.00 - 0.80	1.00 - 0.90	PAMPs	High
0.69 - 0.40	0.79 - 0.50	0.89 - 0.30	Danger	Medium
0.39 - 0.00	0.49 - 0.00	0.29 - 0.00	Safe	Low

As explained in Chapter 4, for dDCA, PAMPs signal is considered as Danger input signals. While for Low output signals, the measurement using dDCA must return a negative value.

iii. Signal Weighting – Three (3) options of signal weighting are applied. This is based on Figure 2.5 in Chapter 2 which applied to DCA only.

Table 5.4: Three (3) Different Transforming Weights

Signals	WM1			WM2			WM3		
	PAMPs	Danger	Safe	PAMPs	Danger	Safe	PAMPs	Danger	Safe
CSM	1	0.5	1.5	2	0	2	1	0.5	1
smDC	0	0	1	1	0	1	0	0	1
mDC	1	0.5	1.5	2	3	3	1.5	0.5	1.5

iv. Anomaly threshold – To determine a reviewed process tends to be anomalous, thus a value of threshold needs to be predetermined. A value below this threshold will indicate the process as normal. In this experiment, three (3) different values are applied to verify its effects on assessing the risk level.

- Initial population – based on a number of messages in Dataset C, which the number of spam is divided with the total messages. In this case, anomaly threshold is derived from the initial population proportion and set as 0.2671;
- Depending on the minimum value of mature class from the risk scale. For instance, anomaly threshold for S1 is 0.4000, S2 is 0.5000 and S3 is 0.3000;
- The third option for the value of anomaly threshold is based on the equal division of two (2) categories (mature and semi-mature) and the value is set as 0.5000; and
- For dDCA, there is one additional value considered as anomaly threshold. As described in Chapter 2, output signal that return a negative value (below 0) from the severity calculation is tagged as normal. Hence, as depicted in Figure 5.14, value 0 is suggested to be considered in this experiment as the threshold value which this value is the minimum value for mature class, retrieved from the risk scale for output signal.

v. The effect of pre-processing and without pre-processing. As explained in Chapter and 3, there are different opinions on pre-processing effects in text mining.

5.3.2 Experiment 2 – DCA Versus dDCA

From Experiment 1, the results of assessing risk for spam messages are compared between DCA and dDCA. The performance is compared in terms of accuracy, true positive detection, false positive and false negative rate. Other than that, the accurateness of concentration value also comparatively analyzed.

5.3.3 Experiment 3 – Various Proportions Of Initial Population

By applying the best value for required parameters that are earlier identified in Experiment 1, an imbalanced number of ham and spam messages for the initial population is used for the third experiment. The number of ham and spam messages that are varies in between 0 to 100% from the collected data is tested to identify the right proportion of initial population messages in order to result in optimum accuracy value. This experiment is also executed as conducted in other publication such as in Mahmoud et al., (2014). These papers suggested that different proportion of ham and spam messages in sampling the initial population would produce a different accuracy rate for spam filtering task.

The test is repeated 11 times with the following specifications, as tabulated in Table 5.5. Dataset C is deployed in this experiment and the best parameters identified in Experiment 1 are applied.

Table 5.5: The Proportion Of Spam And Ham Messages For Initial Population Sampling

Test	Spam (%)	Ham (%)	No. of spam	No. of ham	Total messages as initial population	Dataset C		
						Labelled Spam (ES and S)		Labelled Ham (H)
1	0	100	0	4827	4827	0	0	All
2	10	90	176	4344	4520	ES1-ES101	S1-S75	H1-H4344
3	20	80	352	3862	4214	ES1-ES202	S1-S150	H1-H3862
4	30	70	528	3379	3907	ES1-ES304	S1-S224	H1-H3379
5	40	60	704	2896	3600	ES1-ES405	S1-S299	H1-H2896
6	50	50	880	2414	3294	ES1-ES506	S1-S374	H1-H2414
7	60	40	1055	1931	2986	ES1-ES607	S1-S448	H1-H1931
8	70	30	1231	1448	2679	ES1-ES708	S1-S523	H1-H1448
9	80	20	1407	965	2372	ES1-ES809	S1-S598	H1-H965
10	90	10	1583	483	2066	ES1-ES911	S1-S672	H1-H483
11	100	0	1759	0	1759	All	All	0

5.3.4 Experiment 4 – Antigen Multiplication

Applying the best result from Experiment 1 and 2 for parameters that are identified as influence factors (such as term weighting schemes, risk scale, signal weighting, and anomaly threshold value), a test for antigen multiplication is executed.

Most researchers claimed that the antigen multiplication may overcome the problem of antigen deficiency or signal decay (Greensmith, 2007). Furthermore, this test might assist in optimizing the accuracy rate of DCA and dDCA using identified best parameters value. The experiment is conducted 10 times where every simulation is tested with 10 times of multiplication of antigen. For example, 10 times of antigen multiplication means the initial population is developed with 17, 590 spam messages and 4,827 ham messages and 70 times of antigen multiplication means the initial population is developed with 123, 130 spam messages and no changes for a number of ham messages. The antigen multiplication testing is repeated until it has reached 100 times of multiplication. The number of messages used in this antigen multiplication test for the initial population is simplified as tabulated in Table 5.6.

The initial population with antigen multiplication is using Dataset C and all testing using the same 600 messages that randomly selected earlier.

Table 5.6: The Number Of Messages Deployed To Construct The Initial Population Database Library According To Times Of Antigen Multiplication

Antigen multiplication	No. of spam messages	No. of ham messages	Total messages as initial population
10	17,590	4,827	22,417
20	35,180	4,827	40,007
30	52,770	4,827	57,597
40	70,360	4,827	75,187
50	87,950	4,827	92,777
60	105,540	4,827	110,367
70	123,130	4,827	127,957
80	140,720	4,827	145,547
90	158,310	4,827	163,137
100	175,900	4,827	180,727

5.3.5 Experiment 5 – Non-immune Classifiers

For a competitive reason, an experiment using non-immune classifiers was also conducted. In this experiment, a prominent classifier used in spam filtering Naïve Bayesian (NB) and Support Vector Machine (SVM) are applied to verify their detection ability. However, these classifiers cannot be compared in terms of risk analysis since the tools are only appropriate and meant for spam filtering only.

By deploying the same set of data, RapidMiner is used to conduct the experiment. Dataset C is deployed in training phase and the same 600 messages that were used to test DCA and dDCA, is used in testing phase for SVM and NB.

Figure 5.4: The Screenshot Of Spam Detection Process For SVM And NB Using RapidMiner

5.4 Performance Measurement

The model developed in this study needs to be evaluated to verify its performance. The performance of this model is identified to ensure that the reliability of the proposed model for this specific problem. The evaluation metrics are helpful in analysing the model performance and also to avoid any misleading predictions. Hence, choosing the right evaluation metrics for an experiment in order to obtain maximum

accuracy and to derive relevant information is important for validating any proposed approach (Abdulhamid et al., 2017).

Table 5.7: Confusion Matrix

		True Class	
		Mature	**Semi-mature**
Hypothesized Class	**Mature**	True Positive (TP)	False Positive (FP)
	Semi-mature	False Negative (FN)	True Negative (TN)

Based on Table 5.7, the possible metrics that are practical to be used for this performance measurement (Japkowicz & Shah (2011); Abdulhamid et al., (2017)) include:

i. True Positive (TP): spam messages are identified precisely as malicious (high and medium risk);

ii. True Negative (TN): spam messages are identified precisely as benign (low risk);

iii. False Positive (FP): benign messages are incorrectly predicted as malicious (high and medium risk); and

iv. False Negative (FN): malicious spam messages are incorrectly predicted as benign (low risk)

v. Accuracy (Acc): the total number of spam messages identified precisely as malicious (high and medium risk) and benign (low risk)

$$Acc = \frac{TP+TN}{TP+TN+FP+FN} \quad (5.1)$$

Classification effectiveness is usually measured in terms of precision and recall, which are described as follows:

vi. Precision, P: refers to the closeness of two (2) or more measurements to each other and it is the extended edition of accuracy. The higher the precision rate, the fewer messages likely is judged as malicious.

$$P = \frac{TP}{TP+FP} \quad (5.2)$$

vii. Recall, R: Also known as sensitivity. The higher the recall rate, the less error is denied. Sensitivity measure the accurateness of the model to detect malicious spam message as a malicious class.

$$R = \frac{TP}{TP+FN} \quad (5.3)$$

viii. F-measure: this score is a valuable and efficient metric for unbalanced data

$$F - measure = 2\ x(\frac{Precision\ x\ Recall}{Precision+Recall}) \quad (5.4)$$

ix. Matthews Correlation Coefficient (MCC): used in machine learning evaluation as a determinant of the value of binary classifications. It computes and returns a real value within the range [-1, +1]. A coefficient of

- +1 signifies a perfect prediction;
- 0 signifies a normal arbitrary prediction; and
- -1 signifies an inverse prediction.

$$MCC = \frac{(TPxTN)-(FPxFN)}{\sqrt{(TP+FP)x(TP+FN)x(TN+FP)x(TN+FN)}} \quad (5.5)$$

For TP and TN, the higher the value the better the performance is, while for FP and FN, the lower the value the better it is. Other than that, specificity is an additional metric that can be considered for the model evaluation.

x. Specificity measure the ability of the model to detect a benign message as a normal class

$$Specificity = \frac{TN}{TN+FP} \quad (5.6)$$

5.4.1 End-Users' Judgment As Baseline Comparison

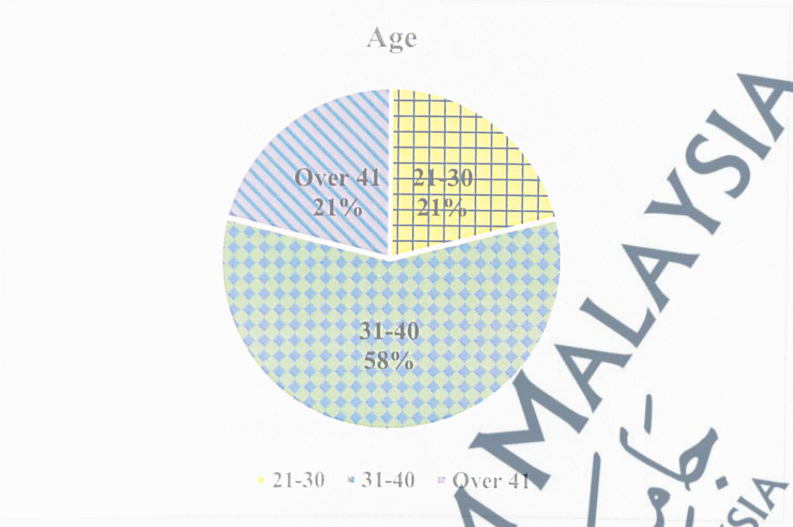
In order to compare the result measured from RiCCA, a list of the deployed dataset with its true label of risk level must be established. To develop the list, a questionnaire of some samples from the dataset is distributed to nineteen (19) participants. This survey is performed with the purpose of getting the idea of how spam messages can be rated from its risk or impact perspective. Participants are chosen from these aspects of:

i. Age (above 18 years old);

ii. Own a smartphone;

iii. Have experienced receiving SMS spam; and

iv. Have some idea about impact of SMS spam.



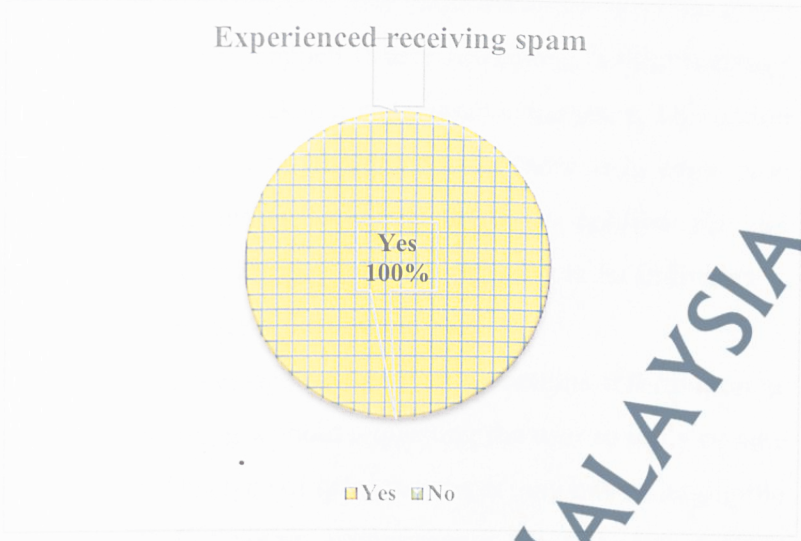


Figure 5.5: The Demographic Details For Participants

This questionnaire is consisting of 20 samples SMS messages for participants to select its risk level according to the provided impact definition. In this study, the risk is differentiated into three (3) levels of impact as the following:

i. High Risk: the spam message may cause severe impact if the user is responds accordingly as requested such as clicking on a given URL link. As elaborated in Jain & Gupta (2018), this action literally could activate phishing activity or SMiShing that will lead to scam or fraud action by the perpetrators. These may cause users to lose money and lead to identity theft. More than one spammy terms also may contribute in high risk spam.

ii. Medium Risk: the spam message may cause serious impact but lesser than high risk. This may happen if user responding to the message such as return call or text back to given unknown numbers. This action may impose some unnecessary/hidden cost to users or in worst case scenario; social-engineering technique may be applied for the perpetrator stealing confidential information such as an individual's bank card number and identification number.

iii. Low Risk: the spam message basically only contains information or notification of something without requesting the user to reply or take any further action. This type of spam messages may have a negligible risk or none at all. For certain circumstances, the very low risk for spam messages may validate as non-spam messages.

A copy of the distributed questionnaire is attached in Appendix E. From the survey, most participants (95%) are agreed with the definition of risk for three (3) different levels as the following that is tagged as a true label, while only one (1) participant suggested that the risk should be differentiated into two (2) classes only; high and low risk. Overall, participants are understanding the potential risk that could occur caused by spam messages.

5.5 Results And Analysis

5.5.1 Experiment 1– Influential Parameters With Effective Values

For this first experiment, DCA is employed as the classifier.

5.5.1.1 TF With Pre-processing

Firstly, using DCA classifier, the test is executed using TF as the weighting scheme, with pre-processing and combined with the following parameters; risk scale (S1, S2, and S3), weight matrix (WM1, WM2, and WM3) and four (4) different values of anomaly threshold (0.2671, 0.4000, 0.5000 and 0.3000).

Table 5.8: The Performance Measurement Results Using DCA With TF And Pre-processing

	Anomaly threshold, t_m	TP	TN	FP	FN	Accuracy	Recall
S1WM1	0.2671	0.3550	0.5417	0.0833	0.0200	0.8967	0.9467
	0.4000	0.3550	0.5517	0.0733	0.0200	**0.9067**	**0.9467**
	0.5000	0.3467	0.5533	0.0750	0.0250	**0.9000**	**0.9327**
S1WM2	0.2671	0.3333	0.5333	0.0917	0.0417	0.8666	0.8888
	0.4000	0.3333	0.5450	0.0800	0.0417	0.8783	0.8888
	0.5000	0.3255	0.5459	0.0785	0.0501	0.8714	0.8666
S1WM3	0.2671	0.3567	0.5400	0.0850	0.0183	0.8967	0.9512
	0.4000	0.3567	0.5500	0.0750	0.0183	**0.9067**	**0.9512**
	0.5000	0.3483	0.5500	0.0750	0.0267	0.8983	0.9288
S2WM1	0.2671	0.3233	0.5383	0.0867	0.0517	0.8616	0.8621
	0.5000	0.3233	0.5683	0.0567	0.0517	0.8916	0.8621
	0.5000	0.3233	0.5683	0.0567	0.0517	0.8916	0.8621
S2WM2	0.2671	0.3200	0.5383	0.0867	0.0550	0.8583	0.8533
	0.5000	0.3200	0.5417	0.0833	0.0550	0.8617	0.8533
	0.5000	0.3200	0.5417	0.0833	0.0550	0.8617	0.8533
S2WM3	0.2671	0.3267	0.5433	0.0817	0.0483	0.8700	0.8712
	0.5000	0.3267	0.5683	0.0567	0.0483	0.8950	0.8712
	0.5000	0.3267	0.5683	0.0567	0.0483	0.8950	0.8712
S3WM1	0.2671	0.3650	0.4067	0.2183	**0.0100**	0.7717	**0.9733**
	0.3000	0.3650	0.4167	0.2083	**0.0100**	0.7817	**0.9733**
	0.5000	0.3533	0.4167	0.2083	0.0217	0.7700	0.9421
S3WM2	0.2671	0.3750	0.3917	0.2333	**0.0000**	0.7667	**1.0000**
	0.3000	0.3750	0.3917	0.2333	**0.0000**	0.7667	**1.0000**
	0.5000	0.3633	0.3917	0.2333	0.0117	0.7550	0.9688
S3WM3	0.2671	0.3650	0.4067	0.2183	**0.0100**	0.7717	**0.9733**
	0.3000	0.3650	0.4167	0.2083	**0.0100**	0.7817	**0.9733**
	0.5000	0.3533	0.4167	0.2083	0.0217	0.7700	0.9421

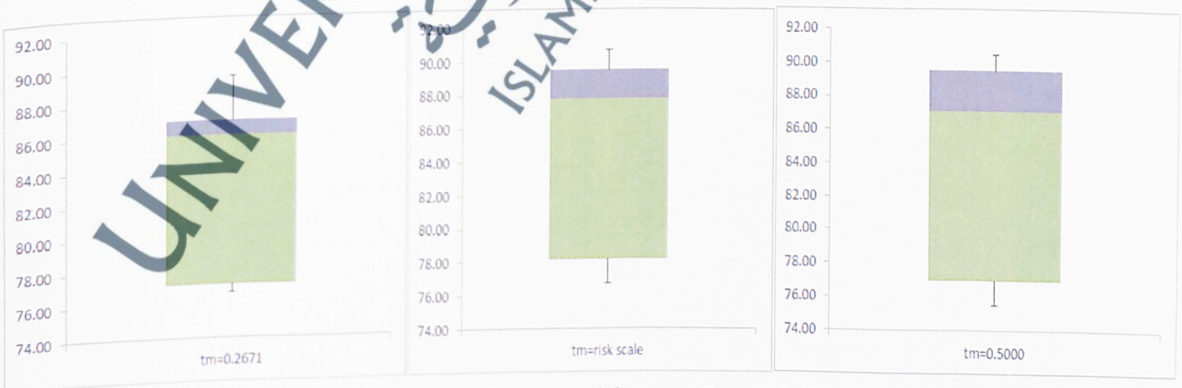


Figure 5.6: Various Anomaly Thresholds, t_m Tested For DCA Using TF, With Pre-processing

From Table 5.8, it is shown that for DCA using TF with pre-processing:

i. Combination of S1WM1 and S1WM3 have resulted in as the highest accuracy with the same rate value 90.67%. However, considering the second parameter of metric, S1WM3 is better than S1WM1 because the value of recall is higher (0.9512) compared to S1WM1 (0.9467);
ii. For all options of risk scale, S and signal weight, WM, by considering the value from risk scale as anomaly threshold have resulted in as the higher accuracy rate compared to the other two (2) options of anomaly threshold; and
iii. The third option of the risk scale, S3 has resulted in as a very low rate of false negative (almost 0%). However, this risk scale did not give a high accuracy detection rate.

5.5.1.2 TF Without Pre-processing

The test using classifier DCA is further executed using the same weighting scheme, TF but without pre-processing. All the options available of risk scales, and signal weights and anomaly threshold are employed together with the DCA.

Table 5.9: The Performance Measurement Results Using DCA With TF And Without Pre-processing

	Anomaly threshold, t_m	TP	TN	FP	FN	Accuracy	Recall
S1WM1	0.2671	0.3417	0.4067	0.2183	0.0333	0.7484	0.9112
	0.4000	0.3417	0.4233	0.2017	0.0333	**0.7650**	**0.9112**
	0.5000	0.3233	0.4233	0.2017	0.0517	0.7466	0.8621
S1WM2	0.2671	0.3533	0.2667	0.3583	0.0217	0.6200	0.9421
	0.4000	0.3533	0.2667	0.3583	0.0217	0.6200	0.9421
	0.5000	0.3333	0.2667	0.3583	0.0417	0.6000	0.8888
S1WM3	0.2671	0.3400	0.4067	0.2183	0.0350	0.7467	0.9067
	0.4000	0.3400	0.4200	0.2050	0.0350	**0.7600**	**0.9067**
	0.5000	0.3233	0.4200	0.2050	0.0517	0.7433	0.8621
S2WM1	0.2671	0.2283	0.4883	0.1367	0.1467	0.7166	0.6088
	0.5000	0.2283	0.5283	0.0967	0.1467	0.7566	0.6088
	0.5000	0.2283	0.5283	0.0967	0.1467	0.7566	0.6088
S2WM2	0.2671	0.2367	0.4100	0.2150	0.1383	0.6467	0.6312
	0.5000	0.2367	0.4100	0.2150	0.1383	0.6467	0.6312
	0.5000	0.2367	0.4100	0.2150	0.1383	0.6467	0.6312
S2WM3	0.2671	0.2283	0.4883	0.1367	0.1467	0.7166	0.6088
	0.5000	0.2283	0.5300	0.0950	0.1467	0.7583	0.6088
	0.5000	0.2283	0.5300	0.0950	0.1467	0.7583	0.6088
S3WM1	0.2671	0.3450	0.2617	0.3633	**0.0300**	0.6067	0.9200
	0.3000	0.3450	0.2733	0.3517	**0.0300**	0.6183	0.9200
	0.5000	0.3417	0.2733	0.3517	0.0333	0.6150	0.9112
S3WM2	0.2671	0.3700	0.1300	0.4950	**0.0050**	0.5000	0.9867
	0.3000	0.3700	0.1300	0.4950	**0.0050**	0.5000	0.9867
	0.5000	0.3667	0.1300	0.4950	0.0083	0.4967	0.9779
S3WM3	0.2671	0.3450	0.2617	0.3633	**0.0300**	0.6067	0.9200
	0.3000	0.3450	0.2733	0.3517	**0.0300**	0.6183	0.9200
	0.5000	0.3417	0.2733	0.3517	0.0333	0.6150	0.9112

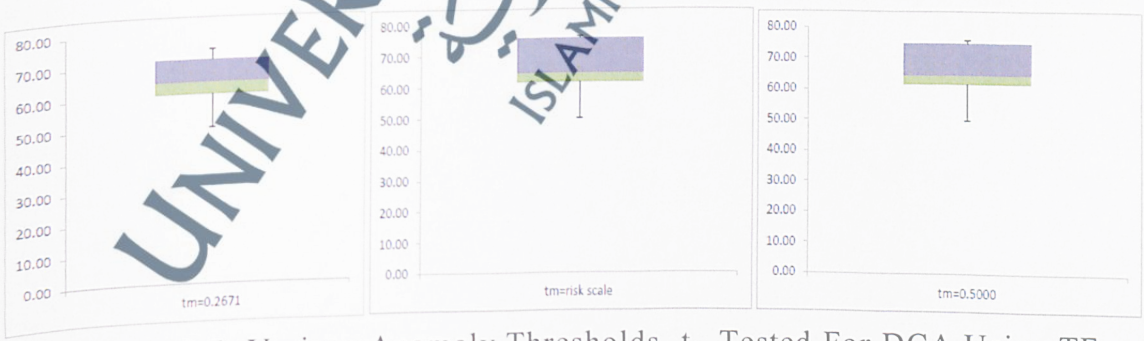


Figure 5.7: Various Anomaly Thresholds, t_m Tested For DCA Using TF, Without Pre-processing

From Table 5.9, it is shown that for DCA using TF without pre-processing:

i. Combination of S1WM1 has resulted in as the highest accuracy among all options with the rate value 76.50%;
ii. The accuracy rate value is not high (not more than 80%) using TF without pre-processing;
iii. The value for false positive rate is quite high using TF without pre-processing;
iv. For all options of risk scale, S and signal weight, WM, by considering the value from risk scale as anomaly threshold still resulted in a higher accuracy rate compared to the other two (2) options of anomaly threshold; and
v. The third option of the risk scale, S3 has resulted in a very low rate of false negative (almost 0%). However, this risk scale still did not give a high accuracy detection rate using TF without pre-processing.

5.5.1.3 IG Ratio And CHI^2 With Pre-processing

Experiencing that the process without pre-processing has resulted a low accuracy rate (below 80%) and high false positive rate (more than 10%), the experiment is further executed with other two (2) options of term weighting schemes (IG Ratio and CHI^2) with pre-processing phase only.

Table 5.10: The Performance Measurement Results Using DCA With IG Ratio And With Pre-processing

	Anomaly threshold, t_m	TP	TN	FP	FN	Accuracy	Recall
S1WM1	0.2671	0.2400	0.5883	0.0367	0.1350	**0.8283**	**0.6400**
	0.4000	0.2400	0.5883	0.0367	0.1350	**0.8283**	**0.6400**
	0.5000	0.2067	0.5883	0.0367	0.1683	0.7950	0.5512
S1WM2	0.2671	0.2883	0.5800	0.0450	0.0867	**0.8683**	**0.7688**
	0.4000	0.2883	0.5800	0.0450	0.0867	**0.8683**	**0.7688**
	0.5000	0.2417	0.5800	0.0450	0.1333	0.8217	0.6445
S1WM3	0.2671	0.2400	0.5883	0.0367	0.1350	**0.8283**	**0.6400**
	0.4000	0.2400	0.5883	0.0367	0.1350	**0.8283**	**0.6400**
	0.5000	0.2067	0.5883	0.0367	0.1683	0.7950	0.5512
S2WM1	0.2671	0.0983	0.6000	0.0250	0.2767	0.6983	0.2621
	0.5000	0.0983	0.6000	0.0250	0.2767	0.6983	0.2621
	0.5000	0.0983	0.6000	0.0250	0.2767	0.6983	0.2621
S2WM2	0.2671	0.1583	0.5950	0.0300	0.2167	0.7533	0.4221
	0.5000	0.1583	0.5950	0.0300	0.2167	0.7533	0.4221
	0.5000	0.1583	0.5950	0.0300	0.2167	0.7533	0.4221
S2WM3	0.2671	0.0967	0.6000	0.0250	0.2783	0.6967	0.2579
	0.5000	0.0967	0.6000	0.0250	0.2783	0.6967	0.2579
	0.5000	0.0967	0.6000	0.0250	0.2783	0.6967	0.2579
S3WM1	0.2671	0.2467	0.5683	0.0567	0.1283	0.8150	0.6579
	0.3000	0.2467	0.5683	0.0567	0.1283	0.8150	0.6579
	0.5000	0.2200	0.5683	0.0567	0.1550	0.7883	0.5867
S3WM2	0.2671	0.3483	0.5567	0.0683	0.0267	0.9050	0.9288
	0.3000	0.3483	0.5567	0.0683	0.0267	0.9050	0.9288
	0.5000	0.2967	0.5567	0.0683	0.0783	0.8534	0.7912
S3WM3	0.2671	0.2467	0.5683	0.0567	0.1283	0.8150	0.6579
	0.3000	0.2467	0.5683	0.0567	0.1283	0.8150	0.6579
	0.5000	0.2133	0.5683	0.0567	0.1617	0.7816	0.5688

Figure 5.8: Various Anomaly Thresholds, t_m Tested For DCA Using IG Ratio, With Pre-processing

From Table 5.10, it is shown that for DCA using IG Ratio with pre-processing:

i. Combination of S1WM2 has resulted in as the highest accuracy among all options with the rate value 86.83%. Although the accuracy rate is higher than S1WM1 and S1WM3, its accuracy value is not high as using TF as the term weighting scheme;
ii. The value for false negative rate is high; and
iii. For all options of the risk scale, S and signal weight, WM, by considering the value from risk scale as anomaly threshold still result in a higher accuracy rate compared to the other two (2) options of anomaly threshold.

Table 5.11: The Performance Measurement Results Using DCA With CHI^2 And With Pre-processing

	Anomaly threshold, t_m	TP	TN	FP	FN	Accuracy	Recall
S1WM1	0.2671	0.0000	0.6183	0.0067	0.3750	0.6183	0.0000
	0.4000	0.0000	0.6183	0.0067	0.3750	0.6183	0.0000
	0.5000	0.0000	0.6183	0.0067	0.3750	0.6183	0.0000
S1WM2	0.2671	0.0000	0.6183	0.0067	0.3750	0.6183	0.0000
	0.4000	0.0000	0.6183	0.0067	0.3750	0.6183	0.0000
	0.5000	0.0000	0.6183	0.0067	0.3750	0.6183	0.0000
S1WM3	0.2671	0.0000	0.6183	0.0067	0.3750	0.6183	0.0000
	0.4000	0.0000	0.6183	0.0067	0.3750	0.6183	0.0000
	0.5000	0.0000	0.6183	0.0067	0.3750	0.6183	0.0000
S2WM1	0.2671	0.0000	0.6183	0.0067	0.3750	0.6183	0.0000
	0.5000	0.0000	0.6183	0.0067	0.3750	0.6183	0.0000
	0.5000	0.0000	0.6183	0.0067	0.3750	0.6183	0.0000
S2WM2	0.2671	0.0000	0.6183	0.0067	0.3750	0.6183	0.0000
	0.5000	0.0000	0.6183	0.0067	0.3750	0.6183	0.0000
	0.5000	0.0000	0.6183	0.0067	0.3750	0.6183	0.0000
S2WM3	0.2671	0.0000	0.6183	0.0067	0.3750	0.6183	0.0000
	0.5000	0.0000	0.6183	0.0067	0.3750	0.6183	0.0000
	0.5000	0.0000	0.6183	0.0067	0.3750	0.6183	0.0000
S3WM1	0.2671	0.0083	0.6100	0.0150	0.3667	0.6183	0.0221
	0.3000	0.0083	0.6100	0.0150	0.3667	0.6183	0.0221
	0.5000	0.0033	0.6100	0.0150	0.3717	0.6133	0.0088
S3WM2	0.2671	0.0317	0.6100	0.0150	0.3433	0.6417	0.0845
	0.3000	0.0317	0.6100	0.0150	0.3433	0.6417	0.0845
	0.5000	0.0117	0.6100	0.0150	0.3633	0.6217	0.0312
S3WM3	0.2671	0.0083	0.6100	0.0150	0.3667	0.6183	0.0221
	0.3000	0.0083	0.6100	0.0150	0.3667	0.6183	0.0221
	0.5000	0.0033	0.6100	0.0150	0.3717	0.6133	0.0088

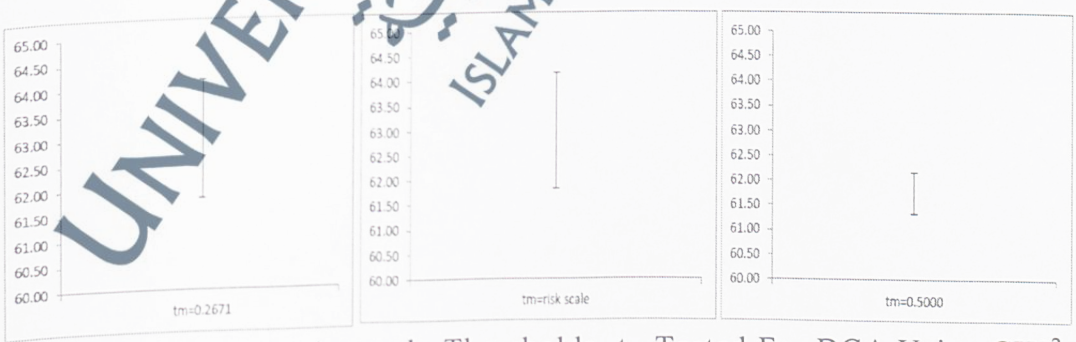


Figure 5.9: Various Anomaly Thresholds, t_m Tested For DCA Using CHI^2, With Pre-processing

From Table 5.11, it is shown that for DCA using CHI^2 with pre-processing:

i. All combination of risk scale, S and signal weight, WM resulted in a low accuracy rate;
ii. The value for false negative rate is high (more than 30%), meaning this scheme is not able to detect effectively the message with mature category (high and medium risk); and
iii. The high value of true negative indicated that this scheme is able to detect message with semi-mature category (low risk).

5.5.2 Experiment 2 – DCA Versus dDCA

For this second experiment, dDCA is employed as the classifier with all options of term weighting scheme, risk scale and various anomaly thresholds, T_k value. This process is executed with pre-processing only, whereby the process without pre-processing has been proved as not efficient in Chapter 3 and Experiment 1 of this chapter.

Table 5.12: The Performance Measurement Results Using dDCA With All Term Weighting Schemes

Weighting Scheme	Risk Scale	Anomaly threshold, T_k	TP	TN	FP	FN	Accuracy	Recall
TF	S1	0.2671	0.3533	0.5733	0.0517	0.0217	**0.9266**	0.9421
		0.4000	0.2850	0.5733	0.0517	0.0900	0.8583	0.7600
		0.5000	0.2400	0.5733	0.0517	0.1350	0.8133	0.6400
		0.0000	0.3683	0.5733	0.0517	0.0067	**0.9416**	**0.9821**
	S2	0.2671	0.3117	0.5767	0.0483	0.0633	0.8884	0.8312
		0.5000	0.2083	0.5767	0.0483	0.1667	0.7850	0.5555
		0.5000	0.2083	0.5767	0.0483	0.1667	0.7850	0.5555
		0.0000	0.3533	0.5783	0.0467	0.0217	**0.9316**	0.9421
	S3	0.2671	0.3583	0.4533	0.1717	0.0167	0.8116	0.9555
		0.3000	0.3467	0.4533	0.1717	0.0283	0.8000	0.9245
		0.5000	0.2800	0.4533	0.1717	0.0950	0.7333	0.7467
		0.0000	0.3733	0.4533	0.1717	0.0017	0.8266	0.9955
IG Ratio	S1	0.2671	0.1717	0.5967	0.0283	0.2033	0.7684	0.4579
		0.4000	0.0850	0.5967	0.0283	0.2900	0.6817	0.2267
		0.5000	0.0417	0.5967	0.0283	0.3333	0.6384	0.1112
		0.0000	0.3000	0.5967	0.0283	0.0750	0.8967	0.8000
	S2	0.2671	0.0783	0.6033	0.0217	0.2967	0.6816	0.2088
		0.5000	0.0150	0.6033	0.0217	0.3600	0.6183	0.0400
		0.5000	0.0150	0.6033	0.0217	0.3600	0.6183	0.0400
		0.0000	0.2000	0.6033	0.0217	0.1750	0.8033	0.5333
	S3	0.2671	0.2233	0.5783	0.0467	0.1517	0.8016	0.5955
		0.3000	0.1967	0.5783	0.0467	0.1783	0.7750	0.5245
		0.5000	0.0617	0.5783	0.0467	0.3133	0.6400	0.1645
		0.0000	0.3467	0.5783	0.0467	0.0283	**0.9250**	0.9245
CHI^2	S1	0.2671	0.0000	0.6183	0.0067	0.3750	0.6183	0.0000
		0.4000	0.0000	0.6183	0.0067	0.3750	0.6183	0.0000
		0.5000	0.0000	0.6183	0.0067	0.3750	0.6183	0.0000
		0.0000	0.0000	0.6183	0.0067	0.3750	0.6183	0.0000
	S2	0.2671	0.0000	0.6183	0.0067	0.3750	0.6183	0.0000
		0.5000	0.0000	0.6183	0.0067	0.3750	0.6183	0.0000
		0.5000	0.0000	0.6183	0.0067	0.3750	0.6183	0.0000
		0.0000	0.0000	0.6183	0.0067	0.3750	0.6183	0.0000
	S3	0.2671	0.0033	0.6100	0.0150	0.3717	0.6133	0.0088
		0.3000	0.0033	0.6100	0.0150	0.3717	0.6133	0.0088
		0.5000	0.0000	0.6100	0.0150	0.3750	0.6100	0.0000
		0.0000	0.0400	0.6100	0.0150	0.3350	0.6500	0.1067

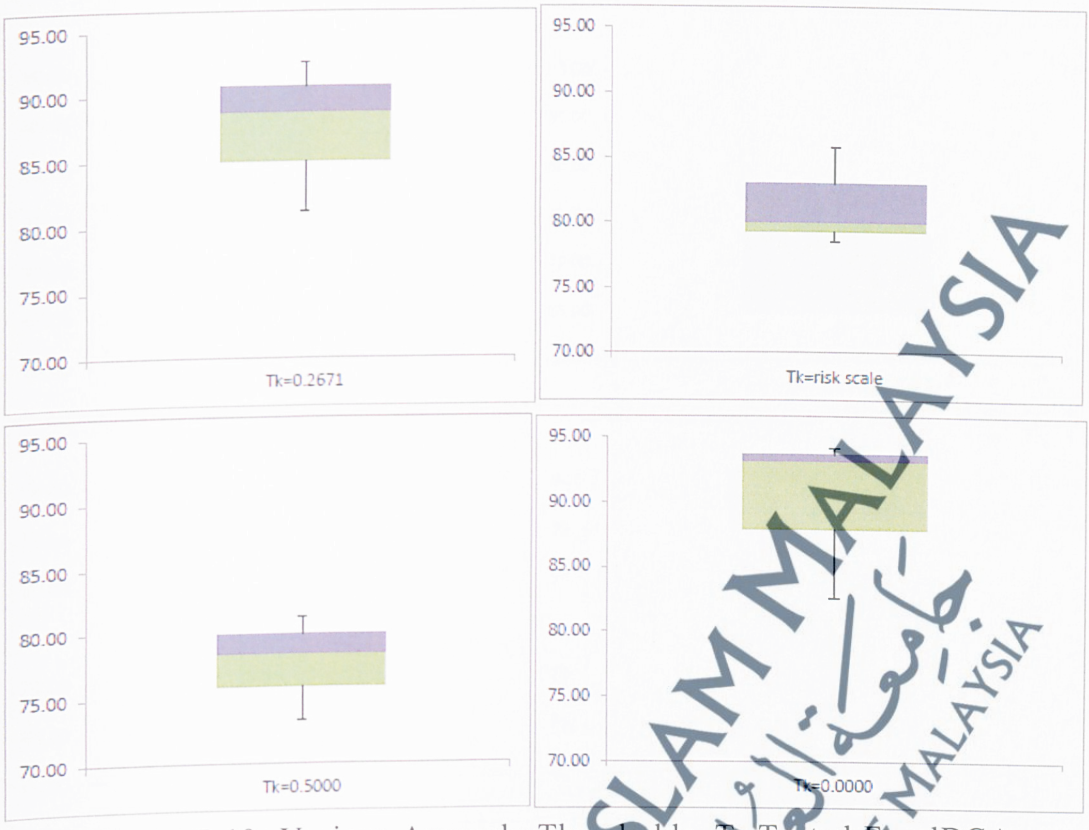


Figure 5.10: Various Anomaly Thresholds, T_k Tested For dDCA Using TF, With Pre-processing

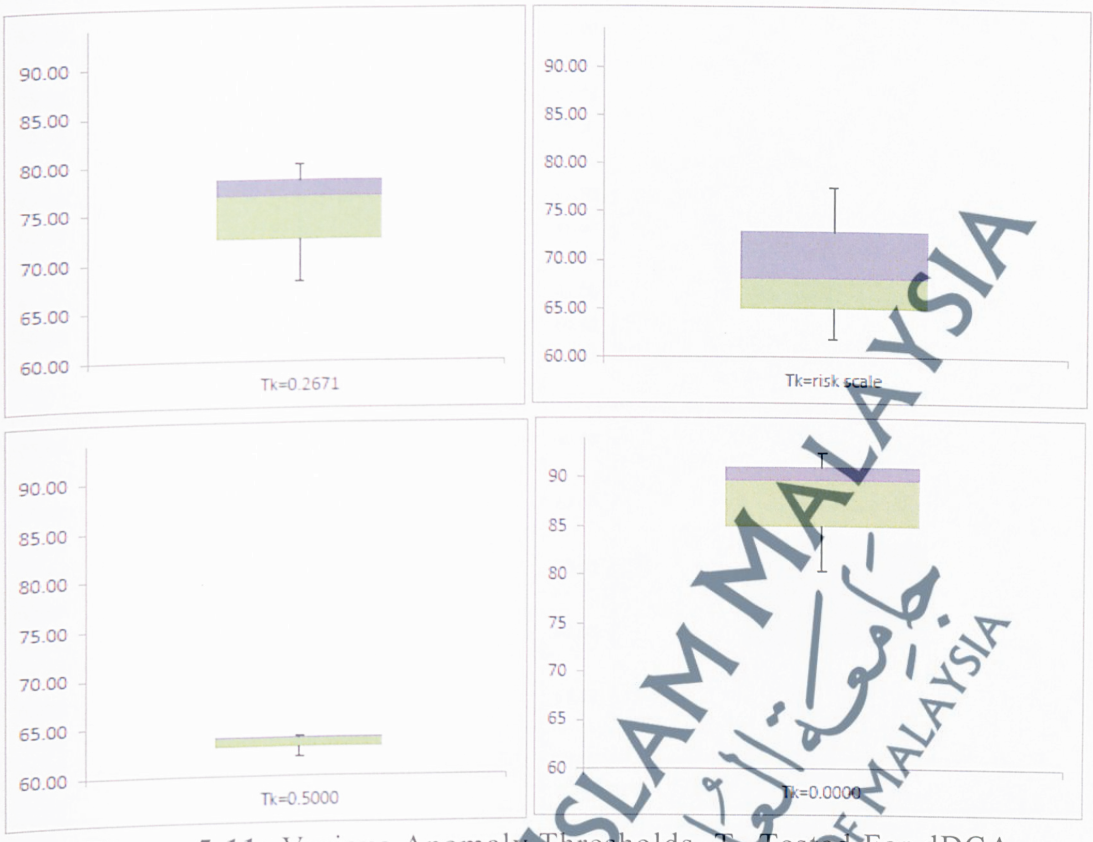


Figure 5.11: Various Anomaly Thresholds, T_k Tested For dDCA Using IG Ratio, With Pre-processing

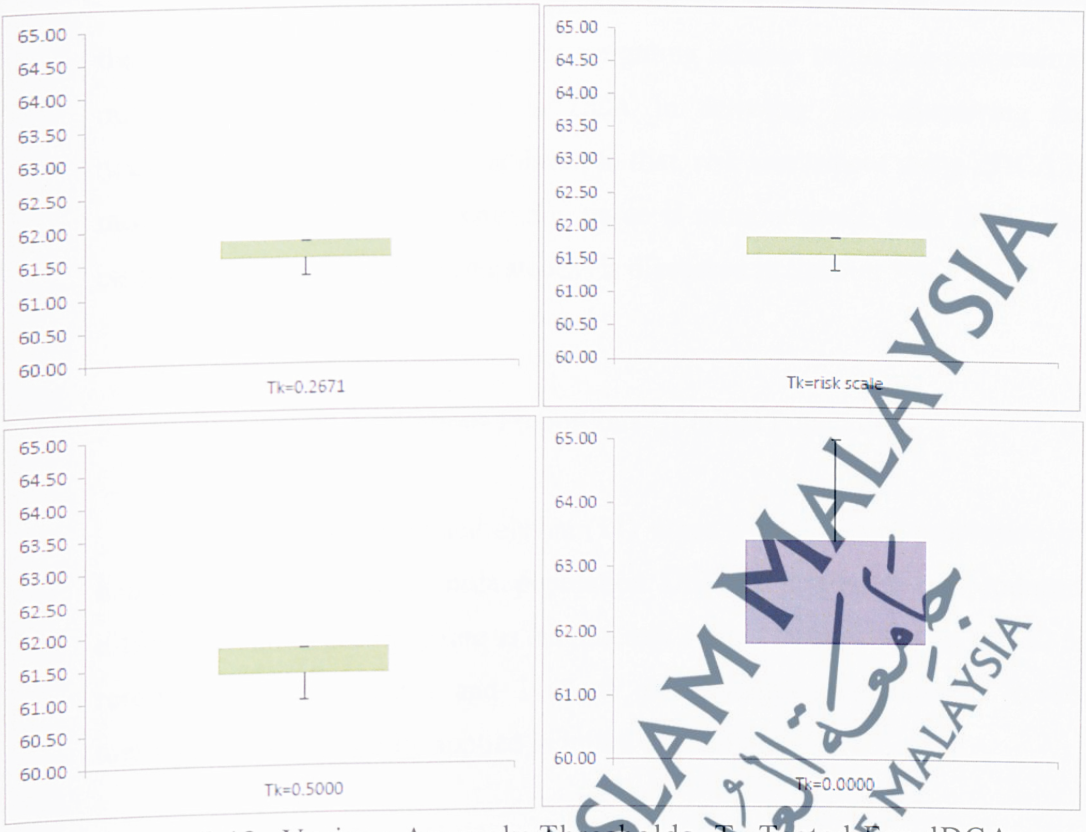


Figure 5.12: Various Anomaly Thresholds, T_k Tested For dDCA Using CHI^2, With Pre-processing

From Table 5.12, it is shown that for dDCA:

i. Combination of TF with S1 has resulted in as the highest accuracy among all others with the rate value of 94.16% with T_k=0 and 92.66% with T_k=0.2671. TF with S2 risk scale and T_k=0 also resulted in high accuracy of 93.16%;

ii. For all weighting schemes, by considering the value of 0 as the anomaly threshold has resulted in a higher accuracy rate compared from the other three (3) options of anomaly threshold; and

iii. IG Ratio with risk scale S3 and anomaly threshold as 0 has resulted in a competitive high accuracy rate (92.50%) compared to TF. However, CHI^2 scheme did not result in a high accuracy rate (below 80%).

After executing these two (2) experiments, it is shown from the result that classifier dDCA with TF as the weighting scheme (with pre-processing) outperformed the performance of DCA in detecting and measuring the maliciousness of messages. In addition to that, risk assessment using dDCA is more accurate, and the concentration value is more accurate than DCA. The issue concerning risk level accurateness is discussed in Section 5.6.3.

5.5.3 Experiment 3 – Various Proportions Of Initial Population

The testing is repeated eleven (11) times with different proportion of ham and spam messages in initial population. Different proportion has produced different result of accuracy rate as tabulated in Table 5.5. In this table, test 12 is referring to Experiment 1 and 2 which all messages in Dataset C (6,586 messages) have been fully applied in initial population.

Table 5.13: The Performance Of DCA And dDCA With Different Proportions Of Initial Population For Messages

Test	TP	TN	FP	FN	Accuracy	Recall
DCA with TF pre-processing, S1WM3 and t_m=0.4000						
1	0.0000	0.6150	0.0100	0.3750	0.6150	0.0000
2	0.2117	0.5850	0.0400	0.1633	0.7967	0.5645
3	0.3000	0.5467	0.0783	0.0750	**0.8467**	0.8000
4	0.3283	0.5067	0.1183	0.0467	0.8350	0.8755
5	0.3467	0.4867	0.1383	0.0283	0.8334	0.9245
6	0.3583	0.4483	0.1767	0.0167	0.8066	0.9555
7	0.3700	0.3967	0.2283	0.0050	0.7667	0.9867
8	0.3717	0.3517	0.2733	0.0033	0.7234	0.9912
9	**0.3750**	0.2767	0.3483	0.0000	0.6517	**1.0000**
10	**0.3750**	0.1400	0.4850	0.0000	0.5150	**1.0000**
11	**0.3750**	**0.0000**	0.6250	0.0000	0.3750	**1.0000**
12	0.3567	0.5500	0.0750	0.0183	**0.9067**	0.9512
dDCA with TF pre-processing, S1and T_k=0.0000						
1	0.0000	0.0000	0.6250	0.3750	0.0000	0.0000
2	0.2800	0.3817	0.2433	0.0950	0.6617	0.7467
3	0.3367	0.4067	0.2183	0.0383	0.7434	0.8979
4	0.3600	0.4150	0.2100	0.0150	0.7750	0.9600
5	0.3600	0.4333	0.1917	0.0150	**0.7933**	0.9600
6	0.3700	0.4200	0.2050	0.0050	0.7900	0.9867
7	0.3733	0.3717	0.2533	0.0017	0.7450	0.9955
8	0.3733	0.3367	0.2883	0.0017	0.7100	0.9955
9	0.3733	0.2717	0.3533	0.0017	0.6450	0.9955
10	**0.3750**	0.1367	0.4883	0.0000	0.5117	**1.0000**
11	**0.3750**	0.0000	0.6250	0.0000	0.3750	**1.0000**
12	0.3683	0.5733	0.0517	0.0067	**0.9416**	0.9821

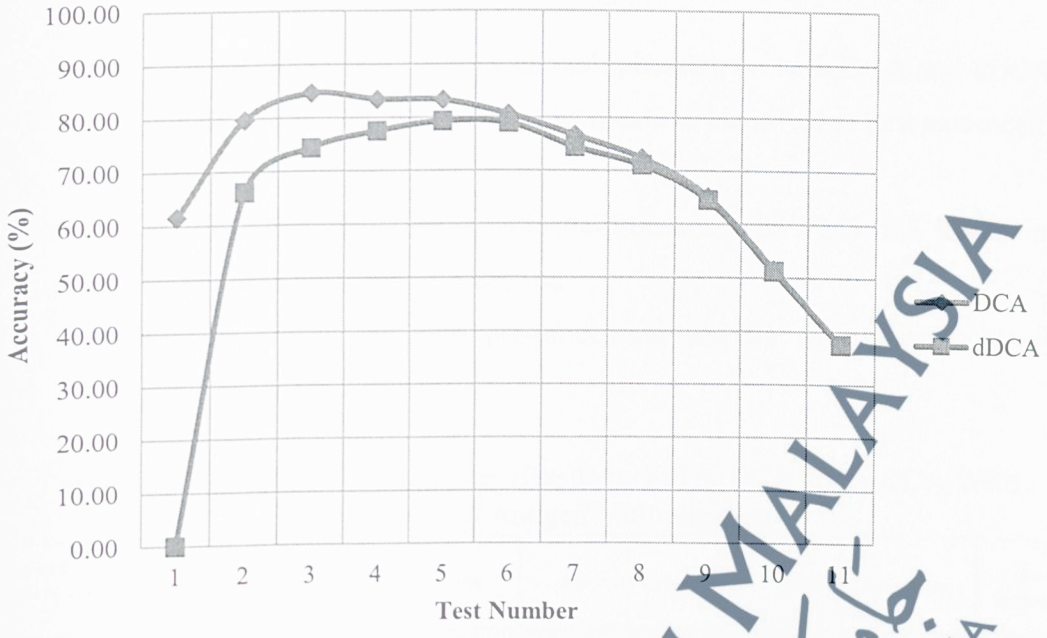


Figure 5.13: DCA And dDCA Performance With Various Proportions Rate Of Messages In Initial Population

From Table 5.13 and Figure 5.13, it is shown that:

i. DCA result in for highest accuracy rate of 84.67% in test 3, which its initial population consists of 20% spam and 80% ham messages;

ii. In test 9, 10 and 11, DCA able to identify and tag all the spam messages, with minimum of 80% spam messages in its initial population. This is shown with value 1.000 for recall and 0.0000 for false negative rate.

iii. While for dDCA, it has result in highest accuracy rate at 79.33% in test 5 with 40% spam and 60% ham messages for its initial population;

iv. dDCA able to identify and tag all the spam messages in test 10 and 11 with minimum of 90% spam messages in its initial population; and

v. Low population of ham has resulted in high rate of false positive, while low population of spam has caused low rate in classifying the malicious messages precisely.

5.5.4 Experiment 4 – Antigen Multiplication

To verify the effect of antigen multiplication on both DCA and dDCA, a test is conducted with the following specification, identified as best parameters from Experiment 1 and 2:

i. for DCA, TF scheme with pre-processing, S1WM3 and 0.4000 as anomaly threshold value; and

ii. for dDCA, TF with pre-processing and S1 with 0 as anomaly threshold value.

Table 5.14: Results Of Performance Measurement For DCA And dDCA With Various Times Of Antigen Multiplication

Antigen Multiplication (%)	TP	TN	FP	FN	Acc	Precision	Recall	Specificity	F-measure	MCC
DCA										
10	**0.3750**	0.1367	0.4883	**0.0000**	0.5117	0.4344	**1.0000**	0.2187	0.6057	0.0513
20	**0.3740**	0.0618	0.5643	**0.0000**	0.4358	0.3986	**1.0000**	0.0987	0.5700	0.0231
30	**0.3750**	0.0417	0.5833	**0.0000**	0.4167	0.3913	**1.0000**	0.0667	0.5625	0.0156
40	**0.3750**	0.0200	0.6050	**0.0000**	0.3950	0.3827	**1.0000**	0.0320	0.5535	0.0075
50	**0.3750**	0.0200	0.6050	**0.0000**	0.3950	0.3827	**1.0000**	0.0320	0.5535	0.0075
60	**0.3750**	0.0200	0.6050	**0.0000**	0.3950	0.3827	**1.0000**	0.0320	0.5535	0.0075
70	**0.3750**	0.0200	0.6050	**0.0000**	0.3950	0.3827	**1.0000**	0.0320	0.5535	0.0075
80	**0.3750**	0.0117	0.6133	**0.0000**	0.3867	0.3794	**1.0000**	0.0187	0.5501	0.0044
90	**0.3750**	0.0117	0.6133	**0.0000**	0.3867	0.3794	**1.0000**	0.0187	0.5501	0.0044
100	**0.3750**	0.0100	0.6150	**0.0000**	0.3850	0.3788	**1.0000**	0.0160	0.5495	0.0038
dDCA										
10	**0.3750**	0.1300	0.4950	**0.0000**	0.5050	0.4310	**1.0000**	0.2080	0.6024	0.0488
20	**0.3750**	0.0483	0.5767	**0.0000**	0.4233	0.3940	**1.0000**	0.0773	0.5653	0.0181
30	**0.3750**	0.0317	0.5933	**0.0000**	0.4067	0.3873	**1.0000**	0.0507	0.5583	0.0119
40	**0.3750**	0.0100	0.6150	**0.0000**	0.3850	0.3788	**1.0000**	0.0160	0.5495	0.0038
50	**0.3750**	0.0100	0.6150	**0.0000**	0.3850	0.3788	**1.0000**	0.0160	0.5495	0.0038
60	**0.3750**	0.0100	0.6150	**0.0000**	0.3850	0.3788	**1.0000**	0.0160	0.5495	0.0038
70	**0.3750**	0.0100	0.6150	**0.0000**	0.3850	0.3788	**1.0000**	0.0160	0.5495	0.0038
80	**0.3750**	0.0000	0.6250	**0.0000**	0.3750	0.3750	**1.0000**	0.0000	0.5455	0.0000
90	**0.3750**	0.0000	0.6250	**0.0000**	0.3750	0.3750	**1.0000**	0.0000	0.5455	0.0000
100	**0.3750**	0.0000	0.6250	**0.0000**	0.3750	0.3750	**1.0000**	0.0000	0.5455	0.0000

Surprisingly, the result from this testing showed that the accuracy rate has been tremendously decreased with the application of antigen multiplication, as tabulated in Table 5.14. It is shown that multiplication of antigen has caused the signals to be excessively amplified. Hence, this caused a high false positive value which caused ham messages or low-risk spam messages detected as

malicious. This is due to the value of potential spam or risk term weight is increasing when the multiplication is higher.

Overall, this method is inappropriate and does not assist in optimizing the result. It is only effective to detect malicious message but the concentration value might be amplified inappropriately, such as a message with medium risk is detected as too high-risk message. For example, ES1010.txt message supposedly calculated as low risk, but by applying the antigen multiplication, the assessment becomes falsely classified as medium or high risk.

Table 5.15: Results For One Sample Message, ES1010.txt With Antigen Multiplication Using DCA And dDCA

ES1010.txt	RM0.00 Airbnb security code: 2359. Use this to finish verification.	
Antigen multiplication (%)	**DCA**	**dDCA**
10	0.0000 – Low risk	-0.2564 – Low risk
20	0.0000 – Low risk	-0.4546 – Low risk
30	0.0000 – Low risk	-0.6122 – Low risk
40	1.0000 – High risk	-0.7408 – Low risk
50	1.0000 – High risk	0.4237 – Medium risk
60	1.0000 – High risk	0.4688 – Medium risk
70	1.0000 – High risk	0.5072 – Medium risk
80	1.0000 – High risk	0.5405 – Medium risk
90	1.0000 – High risk	0.5696 – Medium risk
100	1.0000 – High risk	0.5952 – Medium risk

For example, as tabulated in Table 5.15, the message is truly classified up to 30% of antigen multiplication for DCA and 40% for dDCA. This message then has been falsely classified when the multiplication gets higher, both in DCA and dDCA. This is affected by the term weight that is used for calculating the output signal has been excessively amplified and affects the measurement result in the inverse direction. In this case, the weight for term '*finish*' has been increased when the antigen is multiplied incrementally.

Table 5.16: The Weight Values For Term 'finish'

Multiply (%)	**0**	**10**	**20**	**30**	**40**	**50**	**60**	**70**	**80**	**90**	**100**
finish	0.0145	0.1282	0.2273	0.3061	0.3704	0.4237	0.4688	0.5072	0.5405	0.5696	0.5952

With referring to Table 5.16, it is shown that the weight value for a term is increased when the antigen is multiplied incrementally. This data is using TF scheme with pre-processing.

Other than that, antigen multiplication caused the initial population or the database library become to grow excessively and that consequently required larger storage and even takes a longer time to process the risk assessment.

5.5.5 Experiment 5 – Non-immune Classifiers

The testing is also furthered using non-immune classifier. SVM and NB that are well-known as efficient spam filtering classifier are employed in this experiment. Dataset C is deployed in training and the randomly selected 600 messages are deployed in the testing phase.

Table 5.17: The Performance Measurement Results Using Non-immune Classifiers

Classifier	TP	TN	FP	FN	Accuracy	Recall
With pre-processing						
SVM	0.3617	0.5067	**0.0000**	0.1317	0.8683	0.7331
NB	0.4933	0.4017	0.1050	**0.0000**	0.8950	**1.0000**
Without pre-processing						
SVM	0.3667	0.5067	**0.0000**	0.1267	0.8733	0.7432
NB	0.4933	0.4600	0.0467	**0.0000**	0.9533	**1.0000**

From Table 5.17, it is shown that for non-immune classifiers:

i. SVM performed well in detecting ham messages with 0% of false positive value, while NB able to detect spam messages effectively with 0% of false negative;

ii. SVM and NB performed well with both pre-processing and without pre-processing to distinguished spam messages; and

iii. SVM and NB have a high rate of accuracy for filtering the true label of spam and ham messages without pre-processing which NB has a better accuracy value compared to SVM.

5.6 Discussion

5.6.1 Text Mining

Potential factors that influence the accuracy of spam detection and risk assessment have been explored in series of experiments. The requirements of Danger Theory to identify the antigens and its associated signals weight has become a mandatory element in this study. With assistance of text mining process, majority of required elements in Danger Theory has been fulfilled. These are discussed as the following:

5.6.1.1 Term Weighting Schemes

The processes in text mining have been validated as a proper method to derive the contents of messages as antigen and the terms with weight value as antigen's signal. This is executed by applying term weighting scheme to calculate the weight for every tokenized word. TF, IG Ratio, and CHI^2 are identified as a suitable scheme as these have been applied successfully for spam filtering in previous work. However, in this study, it is demonstrated that not all three (3) schemes are reliable to be applied in risk assessment. TF is revealed as the most appropriate scheme while CHI^2 recognized as an inappropriate scheme for this task. The weight value derived from CHI^2 is generally low even for the spammy or risky term. Unable to produce the right value has caused this scheme failed to detect malicious messages.

5.6.1.2 Pre-processing

Several studies investigating the effect of pre-processing have been carried out by Almeida & Hidalgo, (2012) and Almeida et al., (2012). There is a significant difference between the two (2) conditions, whereby the text mining with pre-processing results a higher accuracy rate compared to without pre-processing that caused high false rate due to noise data. By applying pre-processing, noise or uninformative terms

can be removed, which reduce the computational complexity and enhance effectiveness of the classifier.

5.6.1.3 Dictionary List For Stop Word And Root Word

On top of that, an updated list for stop word and root word also highly influenced the calculated output for risk level. The unnecessary term may behave as noise and stopword list supposedly ensure only the required term is calculated for the output signal. Unwanted or unnecessary terms do influenced the measurement of risk value.

5.6.2 Weight Sensitivity

Weight is one of the important elements in DCA and dDCA. These algorithms are very sensitive to weight deviation even for a slight change. In this study, the parameter such as term weights, risk scale, signal weights (for DCA) and anomaly threshold values are identified as factors that determine the calculated output signals. Additionally, the multiplication of antigen also had shown the effect which its application in initial population caused the signal to be excessively amplified. This has caused the algorithm as being unable to detect non-spam and low-risk spam messages, and hence the performances of the classifiers are not optimized with the antigen multiplication. The findings of this experiment suggest that weight value has the ability in spam detection and measurement of severity.

In addition to that, the amount of messages and its proportion for different categories to develop the initial population as the database library is also influences the accuracy rate. Experiment 3 has demonstrated that it is required to prepare sufficient amount of data in order for the classifier to function efficiently.

5.6.3 DCA Versus dDCA

The findings of these experiments suggest that although DCA and dDCA are reliable to measure the severity level of messages, however, dDCA has resulted in a more accurate risk level via its concentration value. For example,

as tabulated in Table 5.18, using the best specification for term weighting scheme, risk scale, signal weights and anomaly threshold value that has been identified in Experiment 1 and 2, these three (3) messages are differently categorized for their risk level, even though they are classified in the right class (mature) as defined in Section 5.4.1. dDCA demonstrated as the classifier that is able to calculate risk in a finer manner (finer grained classification) which is considered as better and more optimized than DCA.

Table 5.18: The Differentiation In Terms Of Levelling Risk Using DCA And dDCA

Message ID	Content	DCA	dDCA
S722.txt	IMPORTANT MESSAGE. This is a final contact attempt. You have important messages waiting out our customer claims dept. Expires 13/4/04. Call 08717507382 NOW!	0.7500 – High risk	0.5544 – Medium risk
ES757.txt	RM0.00 OTP: Request for One Time Password. OTP : 337354. Did NOT request? Call : 1300886688. 18 Dec 15:26:16. TQ	1.000 – High risk	0.5486 – Medium risk
S278.txt	Reply to win Â£100 weekly! Where will the 2006 FIFA World Cup be held? Send STOP to 87239 to end service	0.8571 – High risk	0.5560 – Medium risk

Besides the accurateness issue, dDCA has fewer parameters to be considered prior the risk measurement. This certainly would make the calculation easier and less error would occur.

Other than that, both classifiers are able to execute the risk calculation for messages that contain terms existed in both ham and spam messages. These terms are some of the example words that exist in both classes of messages.

"click, com, send, gift, account, urgent, collect, voucher, info, promo, chat, net, win, login, log, visit, call, download, free, text, cash, loan, offer, stop, reward, sign"

For example, using the best specifications identified in Experiment 1, a message with ID H111.txt is examined for its severity level. Message H111.txt has the following content;

'Dear, will *call* Tmorrow.pls accomodate.'

DCA measured this message as low risk with concentration value is 0.2500; while dDCA measured the same message also as low risk with concentration value -0.1418.

Besides that, dDCA has a wider magnitude for output signals. Greensmith & Aickelin (2008), Chelly & Elouedi (2015) and Musselle (2010) studied that the anomaly metric in dDCA or k value generates real-valued anomaly scores and may assist in the polarization of normal and anomalous processes. In Greensmith & Aickelin (2008), the metric k is tested, and it is shown to be more sensitive to the minor fluctuations in the resulting output of the cells and provides a more accurate overview of the classification of the various antigen types.

As shown in Figure 5.14, the value that distinguishes the semi-mature and mature category is the threshold of between benign and malicious situation (indicated as dotted line).

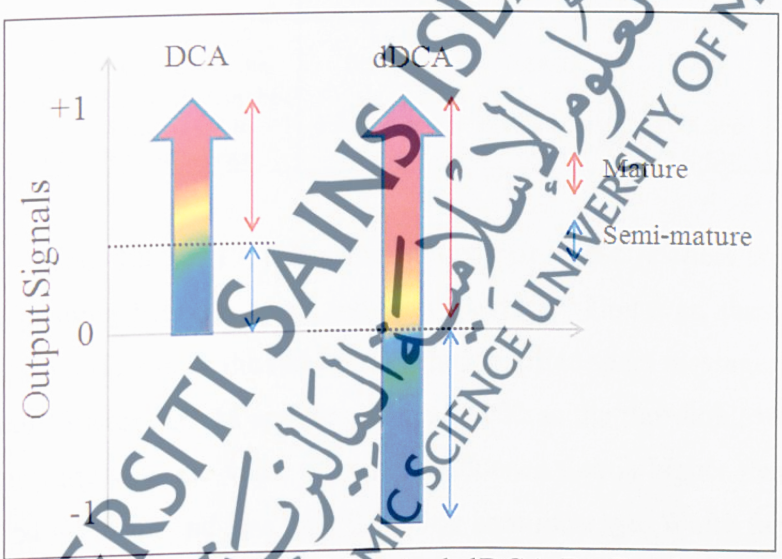


Figure 5.14: Levelling Risk In DCA And dDCA By Retrieving The Value Of Output Signals From The Risk Scale

5.6.4 Non-immune Classifiers

As demonstrated in Experiment 5 that the non-immune classifiers SVM and NB are performing well in spam filtering. However, these two (2) classifiers are less contributing in assessment for its potential risk. Using the same example

as in Table 5.18, the result of testing using these classifiers are simplified as tabulated in Table 5.19.

Table 5.19: The Distinguished Of Spam Messages Using Non-immune Classifiers; SVM And NB

Message ID	Content	SVM		NB	
		With pre-processing	Without pre-processing	With pre-processing	Without pre-processing
		Confidence value for spam			
S722.txt	IMPORTANT MESSAGE. This is a final contact attempt. You have important messages waiting out our customer claims dept. Expires 13/4/04. Call 08717507382 NOW!	0.6693 Predicted as spam	0.6840 Predicted as spam	1.000 Predicted as spam	1.000 Predicted as spam
ES757.txt	RM0.00 OTP: Request for One Time Password. OTP : 337354. Did NOT request? Call : 1300886688. 18 Dec 15:26:16. TQ	0.2145 Predicted as ham	0.2537 Predicted as ham	1.000 Predicted as spam	1.000 Predicted as spam
S278.txt	Reply to win A£100 weekly! Where will the 2006 FIFA World Cup be held? Send STOP to 87239 to end service	0.6700 Predicted as spam	0.6840 Predicted as spam	1.000 Predicted as spam	1.000 Predicted as spam

From Table 5.19, it is demonstrated that both classifiers are feasible in distinguishing between spam and ham messages. However, these classifiers unable to differentiate the risk level of the identified spam message. In this case of spam filtering, SVM applying value 0.5000 as the threshold, which means the message with the value of spam confidence that is higher than 0.5000 is labelled as spam, and less than 0.5000 is ham message. While for NB, if the confidence value for spam of the message is 1.000, then it is tagged as spam; and if the confidence value for spam is 0, then the message is tagged as ham. In addition to that, the message with ID ES757.txt is falsely classified as ham by SVM, both with and without pre-processing.

5.7 Summary

From the conducted experiments series, as overall it is shown that DCA and dDCA from Danger Theory are capable to perform as an algorithm in assessing the risk level of text spam messages. The first experiment validated that the same term returned a various value calculated from different term weighting scheme. This may differentiate the level of input signals to be processed by DCA and dDCA. It is also shown that DCA and dDCA are very sensitive to weights distribution for input signals. Hence, parameters that involve numbers as weights such as term weighting scheme, risk scale, signal weight and chosen value for anomaly threshold highly influences the classifier to perform.

In the second experiment, although dDCA outperformed the performance of DCA, the performance of DCA is still reliable with more than 80% of accuracy rate. DCA is able to perform well if all the involved parameters have the right weight value. These influence factors indicate that the detection of anomalous processes is sensitive to changes in the values of the weights. The weight sensitivity also validated via the third experiment of antigen multiplication. Applying multiplication of antigen in the initial population caused the signal to be amplified excessively. It has been demonstrated that a high number of spam messages in initial population results in low accuracy to detect valid messages.

Due to less parameter that is required to be considered in dDCA and better rate of accuracy, has resulted in dDCA being a better algorithm compared to DCA in risk assessing. Other than that, the accurateness of risk measurement using dDCA is better and the concentration value is more accurate compared to DCA.

In addition, the updated lists for term databases such as stop word list and root word list are critically influenced and ended up becoming a challenge in this task. The outcome is highly dependent on these lists.

In the last experiment, despite non-immune classifier (SVM and NB) acting as efficient classifiers for spam detection, they are however functioning without the concentration value to depict the maliciousness. Hence, the severity level of spam is unable to be identified using these classifiers and the probability of risk or potential impact loss is hardly seen by users.