

## FOREWORD

Praises be unto God for only because of His guidance the author is able to finish this thesis.

The thesis entitled "STUDENT GRADE PREDICTION USING DATA MINING TECHNIQUES" is meant to fulfill a part of academic requirements to obtain Sarjana Matematika Strata Satu Universitas Pelita Harapan, Tangerang.

There are a lot of people who have given help and support so that finishing this thesis is possible, and in this chance, the author wants to express her gratitude to every person who has helped her in the completion of this thesis.

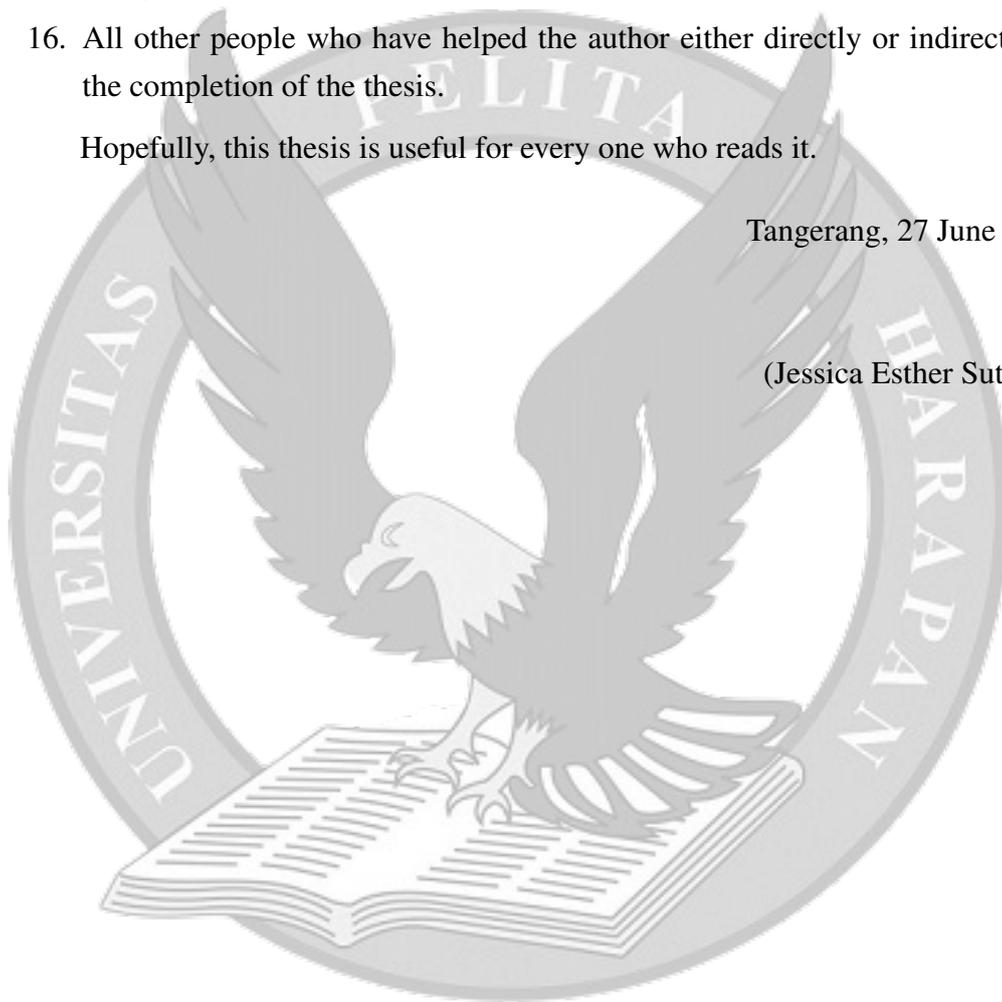
1. Mr. Eric Jobilong, Ph.D., as Dean of Faculty of Science and Technology.
2. Mrs. Sunie Rahardja, M.S.CE, as Acting Vice Dean of Faculty of Science and Technology.
3. Mr Laurence, M.T, as Finance & Student Affairs director of Faculty of Science and Technology.
4. Mr. Kie Van Ivanky Saputra, Ph.D., the Head of Mathematics Department of Universitas Pelita Harapan as well as the author's thesis advisor, who has given many insights which have helped the author to develop as a student.
5. Mrs. Lina Cahyadi, M.Si., the thesis advisor, who has given a lot of suggestions and helped, motivated, and encouraged the author countless times in order to complete the thesis.
6. Lecturers and staffs in Mathematics who have shared their knowledge to the author during the four years of study.
7. Father, mother and brother who have prayed and given a lot of support.
8. Nam Woo Joon, who has supported and accompanied the author throughout many pleasant and difficult times.
9. Felicia Sofian and Viola Citra Dewi who have stayed by author's side for the last five years, being author's best friends, strength, and role models.
10. Licia Kurniawan and Melisa Tiffany, who have spent many years together supporting and accompanying author throughout the whole process of university and thesis writing.
11. Regina Clara Widjaya, who has been author's good friend in good and difficult times, having fun, doing thesis together and supporting each other.
12. Kim Chae Eun, who even when she can't be with author physically, she gave encouragement and prayer for the author.

13. Rheza Dwiputra Tamin, Yordan Saputra, Ryan Natagawa, and Abe Vallerian, who have been a source of happiness, experiences, knowledge, and motivation for the author during university years.
14. Mathematics 2014 classmates, with whom the author has spent time studying and having great memories together.
15. The author's friends and classmates, who have spent pleasant times with the author.
16. All other people who have helped the author either directly or indirectly in the completion of the thesis.

Hopefully, this thesis is useful for every one who reads it.

Tangerang, 27 June 2019

(Jessica Esther Sutedja)



## TABLE OF CONTENTS

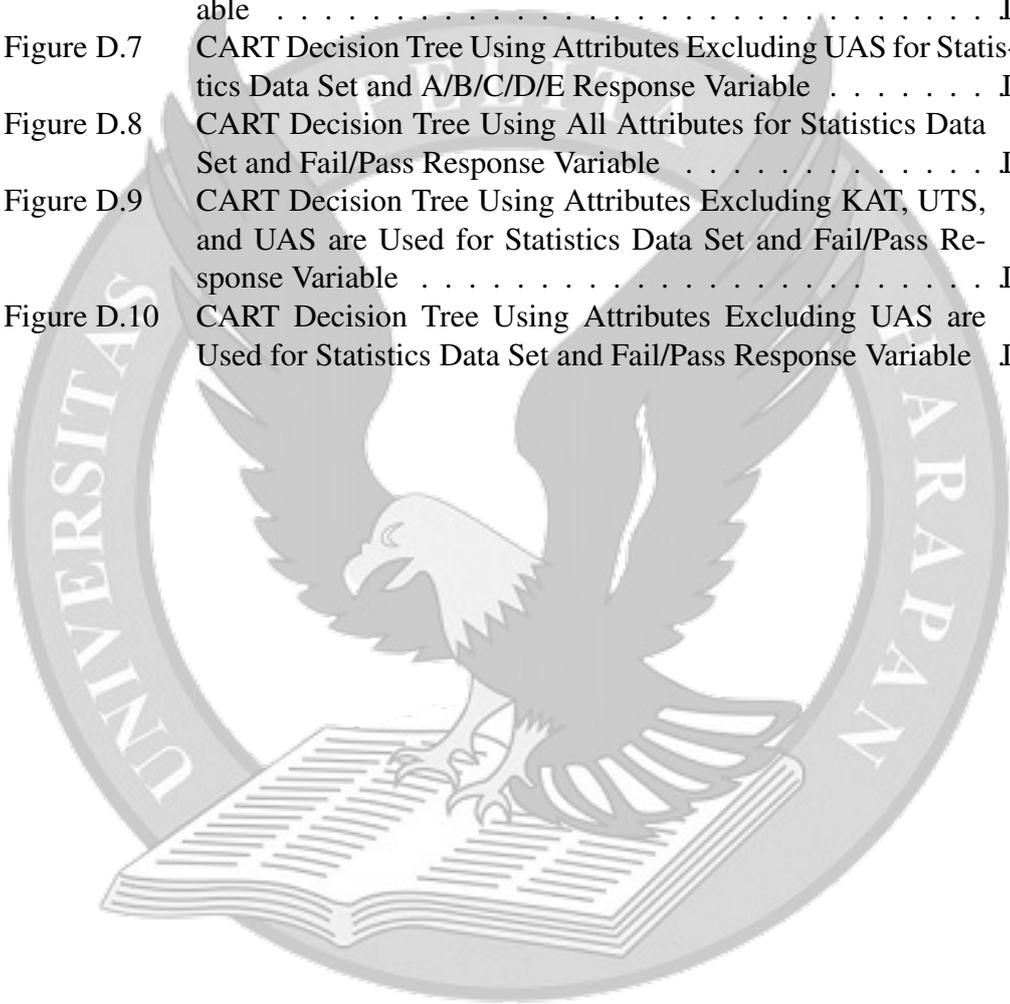
	page
TITLE PAGE	
<i>PERNYATAAN KEASLIAN KARYA SKRIPSI</i>	
<i>PERSETUJUAN DOSEN PEMBIMBING</i>	
<i>PERSETUJUAN TIM PENGUJI SKRIPSI</i>	
ABSTRAK	v
ABSTRACT	vi
FOREWORD	vii
TABLE OF CONTENTS	ix
LIST OF FIGURES	xii
LIST OF TABLES	xxiv
CHAPTER 1 INTRODUCTION	1
1.1 Background . . . . .	1
1.2 Problem Statement . . . . .	3
1.3 Objectives . . . . .	3
1.4 Restrictions and Assumptions . . . . .	4
1.5 Benefits . . . . .	5
1.5.1 Theoretical Benefits . . . . .	5
1.5.2 Practical Benefits . . . . .	5
1.6 Thesis Structure . . . . .	5
CHAPTER 2 THEORETICAL FRAMEWORK	7
2.1 Statistical Learning . . . . .	7
2.1.1 Estimating $f$ . . . . .	7
2.1.2 The Trade-Off Between Prediction Accuracy and Model Interpretability . . . . .	8
2.2 Measuring Performance in Classification Models . . . . .	9
2.2.1 Training Error and Test Error Definitions . . . . .	9
2.2.2 Bias-Variance Trade-Off . . . . .	10
2.2.3 Confusion Matrices . . . . .	11
2.2.4 The Kappa Statistics . . . . .	12
2.2.5 Sensitivity and specificity . . . . .	13
2.2.6 Precision and recall . . . . .	14
2.3 Naïve Bayes . . . . .	14
2.3.1 Laplace Estimator . . . . .	16
2.4 Decision Tree . . . . .	16
2.4.1 Gini Index . . . . .	17
2.4.2 Cross Entropy . . . . .	17
2.4.3 Boosting . . . . .	18
2.5 Support Vector Machine (SVM) . . . . .	20
2.5.1 Hyperplane . . . . .	20

2.5.2	Linearly Separatable . . . . .	21
2.5.3	Non-Linearly Separatable . . . . .	22
2.5.4	Kernels for Non-Linear Spaces . . . . .	23
2.5.5	One-versus-One Training . . . . .	25
2.6	Resampling Methods . . . . .	25
2.6.1	Cross Validation . . . . .	26
2.6.1.1	The Validation Set Approach . . . . .	27
2.6.2	k-Fold Cross-Validation . . . . .	28
2.7	Literature Review . . . . .	28
<b>CHAPTER III METHODOLOGY</b>		30
3.1	Knowledge Discovery in Databases and Data Mining . . . . .	30
3.2	Data Collection and Data Understanding . . . . .	32
3.2.1	Data Visualization . . . . .	32
3.3	Data Pre-processing and Implementation of Data Mining Techniques	34
3.3.1	Dealing with Missing Values . . . . .	34
3.3.2	Naïve Bayes Algorithm Application . . . . .	35
3.3.3	Decision Tree Algorithm Application . . . . .	35
3.3.4	Support Vector Machine Algorithm Application . . . . .	35
<b>CHAPTER IV EXPERIMENT AND RESULTS</b>		36
4.1	Data Exploration . . . . .	36
4.2	Naïve Bayes . . . . .	36
4.2.1	Calculus Data Set . . . . .	38
4.2.2	Statistics Data Set . . . . .	44
4.3	Decision Tree . . . . .	45
4.3.1	Calculus Data Set . . . . .	46
4.3.2	Statistics Data Set . . . . .	54
4.4	Support Vector Machine . . . . .	57
4.4.1	Calculus Data Set . . . . .	57
4.4.2	Statistics Data Set . . . . .	59
4.5	Analysis . . . . .	60
4.5.1	Calculus Data Set . . . . .	61
4.5.2	Statistics Data Set . . . . .	65
<b>CHAPTER V CLOSING</b>		71
5.1	Conclusions . . . . .	71
5.2	Future Works . . . . .	71
<b>REFERENCES</b>		74
<b>APPENDIX</b>		A-1

## LIST OF FIGURES

		page
Figure 1.1	Knowledge Discovery in Databases . . . . .	1
Figure 1.2	Proportions of Mathematics Students' Basic Calculus Final Grade from Academic Year 2013 to 2017 . . . . .	2
Figure 2.1	Confusion Matrices for 2 Classes and 3 Classes . . . . .	11
Figure 2.2	Hyperplanes in 2 Dimensions and 3 Dimensions . . . . .	20
Figure 2.3	Support Vectors and Maximum Margin . . . . .	21
Figure 2.4	Two Points Falling on the Wrong Side of Line . . . . .	23
Figure 2.5	Application of Kernel Trick . . . . .	24
Figure 2.6	One-versus-One Training: Triangles versus Hexagons . . . . .	26
Figure 2.7	One-versus-One Training: Hexagons versus Diamonds . . . . .	26
Figure 2.8	One-versus-One Training: Diamonds versus Triangles . . . . .	26
Figure 2.9	Validation Set Approach . . . . .	27
Figure 2.10	A Schematic Display of 5-fold CV . . . . .	28
Figure 4.1	CART Decision Tree Using All Attributes for Calculus Data Set and A/B/C/D/E Response Variable . . . . .	48
Figure 4.2	CART Decision Tree Using All Attributes for Calculus Data Set and Fail/Pass Response Variable . . . . .	52
Figure B.1	Histogram of Majors Students Enrolled in . . . . .	B-1
Figure B.2	Histogram of Number of Students in Each Batch . . . . .	B-1
Figure B.3	Histogram of Travel Time to University . . . . .	B-2
Figure B.4	Histogram of Independent Study Time per Week . . . . .	B-2
Figure B.5	Histogram of Time Used to Do Non-Academic Related Activ- ities per Week . . . . .	B-3
Figure B.6	Histogram of Confidence in Mathematical Abilities . . . . .	B-3
Figure B.7	Histogram of Level of Interest in Mathematics . . . . .	B-4
Figure B.8	Histogram of Thoughts on Importance of Mathematics . . . . .	B-4
Figure B.9	Histogram of Level of Class Satisfaction . . . . .	B-5
Figure B.10	Histogram of Assignments and Quizzes Scores . . . . .	B-5
Figure B.11	Histogram of Mid-term Score . . . . .	B-6
Figure B.12	Histogram of Final Exam Score . . . . .	B-6
Figure B.13	Histogram of Final Grades (A/B/C/D/E) . . . . .	B-7
Figure B.14	Histogram of Passing / Failing the Course . . . . .	B-7
Figure D.1	CART Decision Tree Using Attributes Excluding KAT, UTS, and UAS for Calculus Data Set and A/B/C/D/E Response Vari- able . . . . .	D-2
Figure D.2	CART Decision Tree Using Attributes Excluding UAS for Cal- culus Data Set and A/B/C/D/E Response Variable . . . . .	D-6

Figure D.3	CART Decision Tree Using All Attributes for Calculus Dataset and Fail/Pass Response Variable . . . . .	D-9
Figure D.4	CART Decision Tree Using Attributes Excluding UAS for Calculus Data Set and Fail/Pass Response Variable . . . . .	D-12
Figure D.5	CART Decision Tree Using All Attributes for Statistics Dataset and A/B/C/D/E Response Variable . . . . .	D-15
Figure D.6	CART Decision Tree Using Attributes Excluding KAT, UTS, and UAS for Statistics Data Set and A/B/C/D/E Response Variable . . . . .	D-19
Figure D.7	CART Decision Tree Using Attributes Excluding UAS for Statistics Data Set and A/B/C/D/E Response Variable . . . . .	D-23
Figure D.8	CART Decision Tree Using All Attributes for Statistics Data Set and Fail/Pass Response Variable . . . . .	D-26
Figure D.9	CART Decision Tree Using Attributes Excluding KAT, UTS, and UAS are Used for Statistics Data Set and Fail/Pass Response Variable . . . . .	D-28
Figure D.10	CART Decision Tree Using Attributes Excluding UAS are Used for Statistics Data Set and Fail/Pass Response Variable . . . . .	D-30



## LIST OF TABLES

		page
Table 2.1	Model $k \times k$ Contingency Table to Assess Agreement about $k$ Categories by Two Different Raters . . . . .	13
Table 3.1	Student Related Variables . . . . .	33
Table 3.2	Output Variable with 5 Classes . . . . .	33
Table 3.3	Output Variable with 2 Classes . . . . .	34
Table 4.1	Number of Students in Each Academic Year . . . . .	36
Table 4.2	Student Related Variables for Calculus Data Set . . . . .	37
Table 4.3	Student Related Variables for Statistics Data Set . . . . .	38
Table 4.4	Confusion Matrix of Naïve Bayes Model Using All Attributes with No Laplace Estimator for Calculus Data Set and A/B/C/D/E as Response Variable . . . . .	39
Table 4.5	Confusion Matrix of Naïve Bayes Model Using All Attributes with Laplace Estimator Set to 1 for Calculus Data Set and A/B/C/D/E as Response Variable . . . . .	39
Table 4.6	Naïve Bayes Models Using All Attributes Output Comparison for Calculus Data Set and A/B/C/D/E as Response Variable . . . . .	40
Table 4.7	Interesting Bayes Probabilities from the Model Using All Attributes and No Laplace Smoothing for Calculus Data Set and A/B/C/D/E as Response Variable . . . . .	41
Table 4.8	Interesting Bayes Probabilities from the Model Using All Attributes and Laplace Smoothing Set to 1 for Calculus Data Set and A/B/C/D/E as Response Variable . . . . .	42
Table 4.9	Confusion Matrix of Naïve Bayes Model Using All Attributes with No Laplace Estimator for Calculus Data Set and Fail/Pass as Response Variable . . . . .	42
Table 4.10	Confusion Matrix of Naïve Bayes Model Using All Attributes with Laplace Estimator = 1 for Calculus Data Set and Fail/Pass as Response Variable . . . . .	43
Table 4.11	Naïve Bayes Models Using All Attributes Output Comparison for Calculus Data Set and Fail/Pass as Response Variable . . . . .	43
Table 4.12	Interesting Bayes Probabilities from the Model Using All Attributes with No Laplace Smoothing for Calculus Data set and Fail/Pass as Response Variable . . . . .	43
Table 4.13	Interesting Bayes Probabilities from the Model Using All Attributes with Laplace Smoothing set to 1 for Calculus Data Set and Fail/Pass as Response Variable . . . . .	43
Table 4.14	Attributes Usage for C5.0 When All Attributes are Used for Calculus Data Set and A/B/C/D/E Response Variable . . . . .	46

Table 4.15	Attributes Usage for Boosted C5.0 When All Attributes are Used for Calculus Data Set and A/B/C/D/E Response Variable	47
Table 4.16	Variable Importance for CART When All Attributes are Used for Calculus Data Set and A/B/C/D/E Response Variable . . . .	47
Table 4.17	Confusion Matrix of C5.0 Decision Tree Using All Attributes for Calculus Dataset and A/B/C/D/E Response Variable . . . .	48
Table 4.18	Confusion Matrix of Boosted C5.0 Decision Tree Using All Attributes for Calculus Data Set and A/B/C/D/E Response Variable . . . . .	48
Table 4.19	Confusion Matrix of CART Decision Tree Using All Attributes for Calculus Data Set and A/B/C/D/E Response Variable . . . .	49
Table 4.20	Decision Tree Models Using All Attributes Output Comparison for Calculus Data Set and A/B/C/D/E Response Variable .	49
Table 4.21	Attributes Usage for C5.0 when All Attributes are Used for Calculus Data Set and Fail/Pass Response Variable . . . . .	51
Table 4.22	Attributes Usage for Boosted C5.0 when All Attributes are Used for Calculus Data Set and Fail/Pass Response Variable .	51
Table 4.23	Variable Importance for CART when All Attributes are Used for Calculus Data Set and Fail/Pass Response Variable . . . . .	51
Table 4.24	Confusion Matrix of C5.0 Decision Tree Using All Attributes for Calculus Data Set and Fail/Pass Response Variable . . . . .	52
Table 4.25	Confusion Matrix of Boosted C5.0 Decision Tree Using All Attributes for Calculus Data Set and Fail/Pass Response Variable	52
Table 4.26	Confusion Matrix of CART Decision Tree Using All Attributes for Calculus Data Set and Fail/Pass Response Variable . . . . .	53
Table 4.27	Decision Tree Models Using All Attributes Output Comparison for Calculus Data Set and Fail/Pass Response Variable . .	53
Table 4.28	Confusion Matrix of SVM Using All Attributes for Calculus Data Set and A/B/C/D/E Response Variable . . . . .	57
Table 4.29	SVM Using All Attributes Output for Calculus Data Set and A/B/C/D/E Response Variable . . . . .	58
Table 4.30	Confusion Matrix of SVM Using All Attributes for Calculus Data Set and Fail/Pass Response Variable . . . . .	58
Table 4.31	SVM Using All Attributes Output for Calculus Data Set and Fail/Pass Response Variable . . . . .	59
Table 4.32	Comparison of Models Constructed Using All Attributes for Calculus Data Set with A/B/C/D/E as Output Variable . . . . .	61
Table 4.33	Comparison of Models Constructed Using Attributes Excluding KAT, UTS, and UAS for Calculus Data Set with A/B/C/D/E as Output Variable . . . . .	62
Table 4.34	Comparison of Models Constructed Using Attributes Excluding UAS for Calculus Data Set with A/B/C/D/E as Output Variable . . . . .	62

Table 4.35	Comparison of Models Constructed Using All Attributes for Calculus Data Set with Pass/Fail as Output Variable . . . . .	63
Table 4.36	Comparison of Models Constructed Using Attributes Excluding KAT, UTS, and UAS for Calculus Data Set with Pass/Fail as Output Variable . . . . .	64
Table 4.37	Comparison of Models Constructed Using Attributes Excluding UAS for Calculus Data Set with Pass/Fail as Output Variable . . . . .	64
Table 4.38	Comparison of Models Constructed Using All Attributes for Statistics Data Set with A/B/C/D/E as Output Variable . . . . .	66
Table 4.39	Comparison of Models Constructed Using Attributes Excluding KAT, UTS, and UAS for Statistics Data Set with A/B/C/D/E as Output Variable . . . . .	67
Table 4.40	Comparison of Models Constructed Using Attributes Excluding KAT, UTS, and UAS for Statistics Data Set with A/B/C/D/E as Output Variable . . . . .	67
Table 4.41	Comparison of Models Constructed Using All Attributes for Statistics Data Set with Pass/Fail as Output Variable . . . . .	68
Table 4.42	Comparison of Models Constructed Using Attributes Excluding KAT, UTS, and UAS for Statistics Data Set with Pass/Fail as Output Variable . . . . .	68
Table 4.43	Comparison of Models Constructed Using Attributes Excluding UAS for Statistics Data Set with Pass/Fail as Output Variable . . . . .	69
Table C.1	Confusion Matrix of Naïve Bayes Model Using Attributes Excluding KAT, UTS, and UAS with No Laplace Estimator for Calculus Data Set and A/B/C/D/E as Response Variable . . . . .	C-1
Table C.2	Confusion Matrix of Naïve Bayes Model Using Attributes Excluding KAT, UTS, and UAS with Laplace Estimator = 1 for Calculus Data Set and A/B/C/D/E as Response Variable . . . . .	C-1
Table C.3	Naïve Bayes Models Using Attributes Excluding KAT, UTS, and UAS Output Comparison for Calculus Data Set and A/B/C/D/E as Response Variable . . . . .	C-2
Table C.4	Interesting Bayes Probabilities from the Model Using Attributes Excluding KAT, UTS, and UAS with No Laplace Smoothing for Calculus Data Set and A/B/C/D/E as Response Variable . . . . .	C-2
Table C.5	Interesting Bayes Probabilities from the Model Using Attributes Excluding KAT, UTS, and UAS with Laplace Smoothing Set to 1 for Calculus Data Set and A/B/C/D/E as Response Variable . . . . .	C-3
Table C.6	Confusion Matrix of Naïve Bayes Model Using Attributes Excluding UAS with No Laplace Estimator for Calculus Data Set and A/B/C/D/E as Response Variable . . . . .	C-3
Table C.7	Confusion Matrix of Naïve Bayes Model Using Attributes Excluding UAS with Laplace Estimator = 1 for Calculus Data Set and A/B/C/D/E as Response Variable . . . . .	C-3

Table C.8	Naïve Bayes Models Using Attributes Excluding UAS Output Comparison for Calculus Data Set and A/B/C/D/E as Response Variable . . . . .	C-4
Table C.9	Interesting Bayes Probabilities from the Model Using Attributes Excluding UAS with No Laplace Smoothing for Calculus Dataset and A/B/C/D/E as Response Variable . . . . .	C-5
Table C.10	Interesting Bayes Probabilities from the Model Using Attributes Excluding UAS with Laplace Smoothing Set to 1 for Calculus Data Set and A/B/C/D/E as Response Variable . . . . .	C-6
Table C.11	Confusion Matrix of Naïve Bayes Model Using Attributes Excluding KAT, UTS, and UAS with No Laplace Estimator for Calculus Data Set and Fail/Pass as Response Variables . . . . .	C-6
Table C.12	Confusion Matrix of Naïve Bayes Model Using Attributes Excluding KAT, UTS, and UAS with Laplace Estimator = 1 for Calculus Data Set and Fail/Pass as Response Variables . . . . .	C-6
Table C.13	Naïve Bayes Models Using Attributes Excluding KAT, UTS, and UAS Output Comparison for Calculus Data Set and Fail/Pass as Response Variables for Calculus Dataset and Fail/Pass as Response Variables . . . . .	C-7
Table C.14	Confusion Matrix of Naïve Bayes Model Using Attributes Excluding UAS with No Laplace Estimator for Calculus Data Set and Fail/Pass as Response Variables . . . . .	C-7
Table C.15	Confusion Matrix of Naïve Bayes Model Using Attributes Excluding UAS with Laplace Estimator = 1 for Calculus Data Set and Fail/Pass as Response Variables . . . . .	C-7
Table C.16	Naïve Bayes Models Using Attributes Excluding UAS Output Comparison for Calculus Data Set and Fail/Pass as Response Variables . . . . .	C-7
Table C.17	Interesting Bayes Probabilities from the Model Using Attributes Excluding UAS with no Laplace Smoothing for Calculus Data Set and Fail/Pass as Response Variables . . . . .	C-8
Table C.18	Interesting Bayes Probabilities from the Model Using Attributes Excluding UAS with Laplace Smoothing set to 1 for Calculus Data Set and Fail/Pass as Response Variables . . . . .	C-8
Table C.19	Confusion Matrix of Naïve Bayes Model Using All Attributes with No Laplace Estimator for Statistics Data Set and A/B/C/D/E as Response Variable . . . . .	C-8
Table C.20	Confusion Matrix of Naïve Bayes Model Using All Attributes with Laplace Estimator = 1 for Statistics Data Set and A/B/C/D/E as Response Variable . . . . .	C-8
Table C.21	Naïve Bayes Models Using All Attributes Output Comparison for Statistics Data Set and A/B/C/D/E as Response Variable . . . . .	C-9

Table C.22	Interesting Bayes Probabilities from the Model Using All Attributes and No Laplace Smoothing for Statistics Data Set and A/B/C/D/E as Response Variable . . . . .	C-10
Table C.23	Interesting Bayes Probabilities from the Model Using All Attributes and Laplace Smoothing set to 1 for Statistics Data Set and A/B/C/D/E as Response Variable . . . . .	C-11
Table C.24	Confusion Matrix of Naïve Bayes Model Using Attributes Excluding KAT, UTS, and UAS with No Laplace Estimator for Statistics Data Set and A/B/C/D/E as Response Variable . . . . .	C-11
Table C.25	Confusion Matrix of Naïve Bayes Model Using Attributes Excluding KAT, UTS, and UAS with Laplace Estimator = 1 for Statistics Data Set and A/B/C/D/E as Response Variable . . . . .	C-12
Table C.26	Naïve Bayes Models Using Attributes Excluding KAT, UTS, and UAS Output Comparison for Statistics Data Set and A/B/C/D/E as Response Variable . . . . .	C-12
Table C.27	Interesting Bayes Probabilities from the Model Using Attributes Excluding KAT, UTS, and UAS with No Laplace Smoothing for Statistics Data Set and A/B/C/D/E as Response Variable . . . . .	C-13
Table C.28	Interesting Bayes Probabilities from the Model Using Attributes Excluding KAT, UTS, and UAS with Laplace Smoothing set to 1 for Statistics Data Set and A/B/C/D/E as Response Variable . . . . .	C-13
Table C.29	Confusion Matrix of Naïve Bayes Model Using Attributes Excluding UAS with No Laplace Estimator for Statistics Data Set and A/B/C/D/E as Response Variable . . . . .	C-14
Table C.30	Confusion Matrix of Naïve Bayes Model Using Attributes Excluding UAS with Laplace Estimator = 1 for Statistics Data Set and A/B/C/D/E as Response Variable . . . . .	C-14
Table C.31	Naïve Bayes Models Using Attributes Excluding UAS Output Comparison for Statistics Data Set and A/B/C/D/E as Response Variable . . . . .	C-15
Table C.32	Interesting Bayes Probabilities from the Model Using Attributes Excluding UAS with No Laplace Smoothing for Statistics Data set and A/B/C/D/E as Response Variable . . . . .	C-16
Table C.33	Interesting Bayes Probabilities from the Model Using Attributes Excluding UAS with Laplace Smoothing Set to 1 for Statistics Data Set and A/B/C/D/E as Response Variable . . . . .	C-17
Table C.34	Confusion Matrix of Naïve Bayes Model Using All Attributes with No Laplace Estimator for Statistics Data Set and Fail/Pass as Response Variable . . . . .	C-17
Table C.35	Confusion Matrix of Naïve Bayes Model Using All Attributes with Laplace Estimator = 1 for Statistics Data Set and Fail/Pass as Response Variable . . . . .	C-17
Table C.36	Naïve Bayes Models Using All Attributes Output Comparison for Statistics Data Set and Fail/Pass as Response Variable . . . . .	C-18

Table C.37	Interesting Bayes Probabilities from the Model Using All Attributes with No Laplace Smoothing for Statistics Data Set and Fail/Pass as Response Variable . . . . .	C-18
Table C.38	Interesting Bayes Probabilities from the Model Using All Attributes with Laplace Smoothing Set to 1 for Statistics Data Set and Fail/Pass as Response Variable . . . . .	C-19
Table C.39	Confusion Matrix of Naïve Bayes Model Using Attributes Excluding KAT, UTS, and UAS with No Laplace Estimator for Statistics Data Set and Fail/Pass as Response Variable . . . . .	C-19
Table C.40	Confusion Matrix of Naïve Bayes Model Using Attributes Excluding KAT, UTS, and UAS with Laplace Estimator = 1 for Statistics Data Set and Fail/Pass as Response Variable . . . . .	C-19
Table C.41	Naïve Bayes Models Using Attributes Excluding KAT, UTS, and UAS Output Comparison for Statistics Data Set and Fail/Pass as Response Variable . . . . .	C-20
Table C.42	Interesting Bayes Probabilities from the Model Using Attributes Excluding KAT, UTS and UAS with No Laplace Smoothing for Statistics Data Set and Fail/Pass as Response Variable . . . . .	C-20
Table C.43	Interesting Bayes Probabilities from the Model Using Attributes Excluding KAT, UTS and UAS with Laplace Smoothing Set to 1 for Statistics Data Set and Fail/Pass as Response Variable . . . . .	C-20
Table C.44	Confusion Matrix of Naïve Bayes Model Using Attributes Excluding UAS with No Laplace Estimator for Statistics Data Set and Fail/Pass as Response Variable . . . . .	C-21
Table C.45	Confusion Matrix of Naïve Bayes Model Using Attributes Excluding UAS with Laplace Estimator = 1 for Statistics Data Set and Fail/Pass as Response Variable . . . . .	C-21
Table C.46	Naïve Bayes Models Using Attributes Excluding UAS Output Comparison for Statistics Data Set and Fail/Pass as Response Variable . . . . .	C-21
Table C.47	Interesting Bayes Probabilities from the Model Using Attributes Excluding UAS with No Laplace Smoothing for Statistics Data Set and Fail/Pass as Response Variables . . . . .	C-22
Table C.48	Interesting Bayes Probabilities from the Model Using Attributes Excluding UAS with Laplace Smoothing set to 1 for Statistics Data Set and Fail/Pass as Response Variable . . . . .	C-22
Table D.1	Attributes Usage for C5.0 When Attributes Excluding KAT, UTS, and UAS are Used for Calculus Data Set and A/B/C/D/E Response Variable . . . . .	D-1
Table D.2	Attributes Usage for Boosted C5.0 When Attributes Excluding KAT, UTS, and UAS are Used for Calculus Data Set and A/B/C/D/E Response Variable . . . . .	D-1

Table D.3	Variable Importance for CART When Attributes Excluding KAT, UTS, and UAS are Used for Calculus Data Set and A/B/C/D/E Response Variable . . . . .	D-1
Table D.4	Confusion Matrix of C5.0 Decision Tree Using Attributes Excluding KAT, UTS and UAS for Calculus Data Set and A/B/C/D/E Response Variable . . . . .	D-2
Table D.5	Confusion Matrix of Boosted C5.0 Decision Tree Using Attributes Excluding KAT, UTS and UAS for Calculus Data Set and A/B/C/D/E Response Variable . . . . .	D-3
Table D.6	Confusion Matrix of CART Decision Tree Using Attributes Excluding KAT, UTS, and UAS for Calculus Data Set and A/B/C/D/E Response Variable . . . . .	D-3
Table D.7	Decision Tree Models Using Attributes Excluding KAT, UTS, and UAS Output Comparison for Calculus Data Set and A/B/C/D/E Response Variable . . . . .	D-4
Table D.8	Attributes Usage for C5.0 when Attributes Excluding UAS are Used for Calculus Data Set and A/B/C/D/E Response Variable	D-4
Table D.9	Attributes Usage for Boosted C5.0 when Attributes Excluding UAS are Used for Calculus Data Set and A/B/C/D/E Response Variable . . . . .	D-5
Table D.10	Variable Importance for CART When Attributes Excluding UAS are Used for Calculus Data Set and A/B/C/D/E Response Variable . . . . .	D-5
Table D.11	Confusion Matrix of C5.0 Decision Tree Using Attributes Excluding UAS for Calculus Data Set and A/B/C/D/E Response Variable . . . . .	D-5
Table D.12	Confusion Matrix of Boosted C5.0 Decision Tree Using Attributes Excluding UAS for Calculus Data Set and A/B/C/D/E Response Variable . . . . .	D-6
Table D.13	Confusion Matrix of CART Decision Tree Using Attributes Excluding UAS for Calculus Data Set and A/B/C/D/E Response Variable . . . . .	D-6
Table D.14	Decision Tree Models Using Attributes Excluding UAS Output Comparison for Calculus Data Set and A/B/C/D/E Response Variable . . . . .	D-7
Table D.15	Attributes Usage for C5.0 When Attributes Excluding KAT, UTS, and UAS are Used for Calculus Data Set and Fail/Pass Response Variable . . . . .	D-7
Table D.16	Attributes Usage for Boosted C5.0 When Attributes Excluding KAT, UTS, and UAS are Used for Calculus Data Set and Fail/Pass Response Variable . . . . .	D-8
Table D.17	Variable Importance for CART When Attributes Excluding KAT, UTS, and UAS are Used for Calculus Dataset and Fail/Pass Response Variable . . . . .	D-8

Table D.18	Confusion Matrix of C5.0 Decision Tree Using Attributes Excluding KAT, UTS, and UAS for Calculus Dataset and Fail/Pass Response Variable . . . . .	D-9
Table D.19	Confusion Matrix of Boosted C5.0 Decision Tree Using Attributes Excluding KAT, UTS, and UAS for Calculus Data Set and Fail/Pass Response Variable . . . . .	D-9
Table D.20	Confusion Matrix of CART Decision Tree Using Attributes Excluding KAT, UTS, and UAS for Calculus Data Set and Fail/Pass Response Variable . . . . .	D-10
Table D.21	Decision Tree Models Using Attributes Excluding KAT, UTS, and UAS Output Comparison for Calculus Data Set and Fail/Pass Response Variable . . . . .	D-10
Table D.22	Attributes Usage for C5.0 when Attributes Excluding UAS are Used for Calculus Data Set and Fail/Pass Response Variable . . . . .	D-10
Table D.23	Attributes Usage for Boosted C5.0 when Attributes Excluding UAS are Used for Calculus Data Set and Fail/Pass Response Variable . . . . .	D-11
Table D.24	Variable Importance for CART when Attributes Excluding UAS are Used for Calculus Data Set and Fail/Pass Response Variable . . . . .	D-11
Table D.25	Confusion Matrix of C5.0 Decision Tree Using Attributes Excluding UAS for Calculus Data Set and Fail/Pass Response Variable . . . . .	D-12
Table D.26	Confusion Matrix of Boosted C5.0 Decision Tree Using Attributes Excluding UAS for Calculus Data Set and Fail/Pass Response Variable . . . . .	D-12
Table D.27	Confusion Matrix of CART Decision Tree Using Attributes Excluding UAS for Calculus Data Set and Fail/Pass Response Variable . . . . .	D-13
Table D.28	Decision Tree Models Using Attributes Excluding UAS Output Comparison for Calculus Data Set and Fail/Pass Response Variable . . . . .	D-13
Table D.29	Attributes Usage for C5.0 When All Attributes are Used for Statistics Data Set and A/B/C/D/E Response Variable . . . . .	D-13
Table D.30	Attributes Usage for Boosted C5.0 When All Attributes are Used for Statistics Data Set and A/B/C/D/E Response Variable . . . . .	D-14
Table D.31	Variable Importance for CART When All Attributes are Used for Statistics Data Set and A/B/C/D/E Response Variable . . . . .	D-14
Table D.32	Confusion Matrix of C5.0 Decision Tree Using All Attributes for Statistics Data Set and A/B/C/D/E Response Variable . . . . .	D-15
Table D.33	Confusion Matrix of Boosted C5.0 Decision Tree Using All Attributes for Statistics Data Set and A/B/C/D/E Response Variable . . . . .	D-16
Table D.34	Confusion Matrix of CART Decision Tree Using All Attributes for Statistics Data Set and A/B/C/D/E Response Variable . . . . .	D-16

Table D.35	Decision Tree Models Using All Attributes Output Comparison for Statistics Data Set and A/B/C/D/E Response Variable	D-17
Table D.36	Attributes Usage for C5.0 When Attributes Excluding KAT, UTS, and UAS are Used for Statistics Data Set and A/B/C/D/E Response Variable	D-17
Table D.37	Attributes Usage for Boosted C5.0 When Attributes Excluding KAT, UTS, and UAS are Used for Statistics Data Set and A/B/C/D/E Response Variable	D-18
Table D.38	Variable Importance for CART When Attributes Excluding KAT, UTS, and UAS are Used for Statistics Data Set and A/B/C/D/E Response Variable	D-18
Table D.39	Confusion Matrix of C5.0 Decision Tree Using Attributes Excluding KAT, UTS and UAS for Statistics Data Set and A/B/C/D/E Response Variable	D-19
Table D.40	Confusion Matrix of Boosted C5.0 Decision Tree Using Attributes Excluding KAT, UTS and UAS for Statistics Data Set and A/B/C/D/E Response Variable	D-20
Table D.41	Confusion Matrix of CART Decision Tree Using Attributes Excluding KAT, UTS, and UAS for Statistics Data Set and A/B/C/D/E Response Variable	D-20
Table D.42	Decision Tree Models Using Attributes Excluding KAT, UTS, and UAS Output Comparison for Statistics Data Set and A/B/C/D/E Response Variable	D-21
Table D.43	Attributes Usage for C5.0 When Attributes Excluding UAS are Used for Statistics Data Set and A/B/C/D/E Response Variable	D-21
Table D.44	Attributes Usage for Boosted C5.0 When Attributes Excluding UAS are Used for Statistics Data Set and A/B/C/D/E Response Variable	D-22
Table D.45	Variable Importance for CART When Attributes Excluding UAS are Used for Statistics Data Set and A/B/C/D/E Response Variable	D-22
Table D.46	Confusion Matrix of C5.0 Decision Tree Using Attributes Excluding UAS for Statistics Data Set and A/B/C/D/E Response Variable	D-23
Table D.47	Confusion Matrix of Boosted C5.0 Decision Tree Using Attributes Excluding UAS for Statistics Data Set and A/B/C/D/E Response Variable	D-24
Table D.48	Confusion Matrix of CART Decision Tree Using Attributes Excluding UAS for Statistics Data Set and A/B/C/D/E Response Variable	D-24
Table D.49	Decision Tree Models Using Attributes Excluding UAS Output Comparison for Statistics Data Set and A/B/C/D/E Response Variable	D-25

Table D.50	Attributes Usage for C5.0 When All Attributes are Used for Statistics Data Set and Fail/Pass Response Variable . . . . .	D-25
Table D.51	Attributes Usage for Boosted C5.0 When All Attributes are Used for Statistics Data Set and Fail/Pass Response Variable . . . . .	D-25
Table D.52	Variable Importance for CART When All Attributes are Used for Statistics Data Set and Fail/Pass Response Variable . . . . .	D-26
Table D.53	Confusion Matrix of C5.0 Decision Tree Using All Attributes for Statistics Data Set and Fail/Pass Response Variable . . . . .	D-26
Table D.54	Confusion Matrix of Boosted C5.0 Decision Tree Using All Attributes for Statistics Data Set and Fail/Pass Response Variable . . . . .	D-27
Table D.55	Confusion Matrix of CART Decision Tree Using All Attributes for Statistics Data Set and Fail/Pass Response Variable . . . . .	D-27
Table D.56	Decision Tree Models Using All Attributes Output Comparison for Statistics Data Set and Fail/Pass Response Variable . . . . .	D-27
Table D.57	Variable Importance for CART When Attributes Excluding KAT, UTS, and UAS are Used for Statistics Data Set and Fail/Pass Response Variable . . . . .	D-27
Table D.58	Confusion Matrix of C5.0 Decision Tree Using Attributes Excluding KAT, UTS, and UAS are Used for Statistics Data Set and Fail/Pass Response Variable . . . . .	D-28
Table D.59	Confusion Matrix of CART Decision Tree Using Attributes Excluding KAT, UTS, and UAS are Used for Statistics Data Set and Fail/Pass Response Variable . . . . .	D-28
Table D.60	Decision Tree Models Using Attributes Excluding KAT, UTS, and UAS Output Comparison for Statistics Data Set and Fail/Pass Response Variable . . . . .	D-29
Table D.61	Attributes Usage for C5.0 when Attributes Excluding UAS are Used for Statistics Data Set and Fail/Pass Response Variable . . . . .	D-29
Table D.62	Attributes Usage for Boosted C5.0 when Attributes Excluding UAS for Statistics Data Set and Fail/Pass Response Variable . . . . .	D-29
Table D.63	Variable Importance for CART when Attributes Excluding UAS are Used for Statistics Data Set and Fail/Pass Response Variable . . . . .	D-29
Table D.64	Confusion Matrix of C5.0 Decision Tree Using Attributes Excluding UAS are Used for Statistics Data Set and Fail/Pass Response Variable . . . . .	D-30
Table D.65	Confusion Matrix of Boosted C5.0 Decision Tree Using Attributes Excluding UAS are Used for Statistics Data Set and Fail/Pass Response Variable . . . . .	D-30
Table D.66	Confusion Matrix of CART Decision Tree Using Attributes Excluding UAS are Used for Statistics Data Set and Fail/Pass Response Variable . . . . .	D-31
Table D.67	Decision Tree Models Using Attributes Excluding UAS Output Comparison for Statistics Data Set and Fail/Pass Response Variable . . . . .	D-31

Table E.1	Confusion Matrix of SVM Using Attributes Excluding KAT, UTS, and UAS for Calculus Data Set and A/B/C/D/E Response Variable . . . . .	E-1
Table E.2	SVM Using Attributes Excluding KAT, UTS and UAS Output for Calculus Data Set and A/B/C/D/E Response Variable . . . . .	E-1
Table E.3	Confusion Matrix of SVM Using Attributes Excluding UAS for Calculus Data Set and A/B/C/D/E Response Variable . . . . .	E-2
Table E.4	SVM Using Attributes Excluding UAS Output for Calculus Data Set and A/B/C/D/E Response Variable . . . . .	E-2
Table E.5	Confusion Matrix of SVM Using Attributes Excluding KAT, UTS, and UAS for Calculus Data Set and Fail/Pass Response Variable . . . . .	E-3
Table E.6	SVM Using Attributes Excluding KAT, UTS and UAS Output for Calculus Data Set and Fail/Pass Response Variable . . . . .	E-3
Table E.7	Confusion Matrix of SVM Using Attributes Excluding UAS for Calculus Data Set and Fail/Pass Response Variable . . . . .	E-3
Table E.8	SVM Using Attributes Excluding UAS Output for Calculus Data Set and Fail/Pass Response Variable . . . . .	E-3
Table E.9	Confusion Matrix of SVM Using All Attributes for Calculus Statistics and A/B/C/D/E Response Variable . . . . .	E-4
Table E.10	SVM Using All Attributes Output for Statistics Data Set and A/B/C/D/E Response Variable . . . . .	E-4
Table E.11	Confusion Matrix of SVM Using Attributes Excluding KAT, UTS, and UAS for Statistics Data Set and A/B/C/D/E Response Variable . . . . .	E-5
Table E.12	SVM Using Attributes Excluding KAT, UTS and UAS Output for Statistics Data Set and A/B/C/D/E Response Variable . . . . .	E-5
Table E.13	Confusion Matrix of SVM Using Attributes Excluding UAS for Statistics Data Set and A/B/C/D/E Response Variable . . . . .	E-6
Table E.14	SVM Using Attributes Excluding UAS Output for Statistics Data Set and A/B/C/D/E Response Variable . . . . .	E-6
Table E.15	Confusion Matrix of SVM Using All Attributes for Statistics Data Set and Fail/Pass Response Variable . . . . .	E-7
Table E.16	SVM Using All Attributes Output for Statistics Data Set and Fail/Pass Response Variable . . . . .	E-7
Table E.17	Confusion Matrix of SVM Using Attributes Excluding KAT, UTS, and UAS for Statistics Data Set and Fail/Pass Response Variable . . . . .	E-7
Table E.18	SVM Using Attributes Excluding KAT, UTS and UAS Output for Statistics Data Set and Fail/Pass Response Variable . . . . .	E-7
Table E.19	Confusion Matrix of SVM Using Attributes Excluding UAS for Statistics Data Set and Fail/Pass Response Variable . . . . .	E-8
Table E.20	SVM Using Attributes Excluding UAS Output for Statistics Data Set and Fail/Pass Response Variable . . . . .	E-8

Tabel

