

## TABLE OF CONTENTS

STATEMENT .....	iii
ACKNOWLEDGEMENT .....	iv
MEANING OF SYMBOLS AND ABBREVIATIONS .....	vi
ABSTRACT .....	vii
TABLE OF CONTENTS .....	viii
LIST OF FIGURES .....	x
LIST OF TABLES .....	xi
CHAPTER 1 INTRODUCTION .....	1
1.1. Background.....	1
1.2. Problem Formulation.....	3
1.3. Scope of Problem .....	4
1.4. The originality of the research.....	4
1.5. Research Objectives .....	7
1.6. Benefits of Research.....	7
CHAPTER 2 LITERATURE REVIEW AND BASIC THEORIES .....	8
2.1. Literature Reviews.....	8
2.2. Basic Theories .....	10
2.2.1. Feature Engineering.....	11
2.2.1.1. Bag of Words .....	11
2.2.1.2. Term Frequency – Inverse Document Frequency (TF-IDF).....	11
2.2.2. Classification models.....	13
2.2.2.1. Support Vector Machines (SVMs).....	13
2.2.2.2. Naïve Bayes .....	17
2.2.2.3. Logistic Regression.....	19
2.2.3. Resampling methods .....	21
2.2.3.1. Over-sampling techniques.....	21
2.2.3.1.1. Random Over-sampling .....	21
2.2.3.1.2. Synthetic Minority Over-sampling TEchnique (SMOTE) .....	21
2.2.3.1.3. Adaptive Synthetic (ADASYN) .....	22
2.2.3.2. Under-sampling techniques.....	23
2.2.3.2.1. Random Under-sampling.....	23
2.2.4. Regularization.....	23

2.2.5. K-fold Cross-validation .....	24
CHAPTER 3 METHODOLOGY .....	25
3.1. Tools and Materials of Research .....	25
3.1.1. Research Tools .....	25
3.1.2. Research Materials .....	26
3.2. The Course of Study .....	27
3.2.1. Literature Review Study .....	28
3.2.2. Data Acquisition .....	28
3.2.3. Data Preprocessing .....	28
3.2.4. Data Resampling .....	28
3.2.5. Feature Extraction .....	29
3.2.6. Classification .....	34
3.2.7. Evaluation .....	35
CHAPTER 4 RESULT AND DISCUSSIONS .....	37
4.1. Result of Data Resampling .....	37
4.1.1. Result of Class Distribution in Oversampling Methods .....	37
4.1.2. Result of Class Distribution in Undersampling Methods .....	38
4.2. Result of Classification Models .....	40
4.2.1. Performance of Support Vector Machines (SVM) .....	40
4.2.2. Performance of Logistic Regression (LR) .....	41
4.2.3. Performance of Naïve Bayes .....	42
4.2.4. Result Comparison and Discussions .....	43
4.3. Strength and Weaknesses of Proposed Approaches .....	44
4.3.1. Strengths .....	44
4.3.2. Weaknesses .....	44
CHAPTER 5 CONCLUSION AND SUGGESTIONS .....	45
5.1. Conclusion .....	45
5.2. Suggestion .....	45
REFERENCES .....	46