

CONTENTS

COVER	i
APPROVAL PAGE	ii
PLAGIARISM STATEMENT	iii
ACKNOWLEDGMENT	iv
EPIGRAPH PAGE	v
PREFACE	vi
CONTENTS	viii
LIST OF TABLES	xii
LIST OF FIGURES	xiii
NOMENCLATURE	xv
ABSTRACT	xvii
I INTRODUCTION	1
1.1. Research Background	1
1.2. Research Objective and Advantage	3
1.3. Literature Review	3
1.4. Research Methodology	12
1.5. Document Structure	12
1.5.1. Chapter I Introduction	12
1.5.2. Chapter II Basic Theory	12
1.5.3. Chapter III The Implementation of Bootstrap Aggregating Classifier for Auto Insurance Claims Fraud Detection	13
1.5.4. Chapter IV Study Case of The Implementation of Bootstrap Aggregating Classifier for Auto Insurance Claims Fraud Detection	13
1.5.5. Chapter V Conclusion and Recommendation	13
II BASIC THEORY	14
2.1. Fraud Definition	14
2.1.1. Typology of Insurance Fraud	14
2.2. Types of Fraud	16
2.2.1. Credit Card Fraud	16
2.2.2. Cryptocurrency Fraud	17
2.2.3. Financial Statement Fraud	17
2.2.4. Insurance Fraud	18

2.2.5.	Mortgage Fraud	18
2.2.6.	Money Laundering	18
2.2.7.	Securities and Commodities Fraud	19
2.3.	The Categories of Auto Insurance Claim Fraud	19
2.3.1.	Automobile Property	19
2.3.2.	Automobile Collisions	20
2.4.	Probability	20
2.4.1.	Properties of Probability	21
2.4.2.	Conditional Probability	22
2.5.	Data Mining	23
2.6.	Machine Learning	24
2.6.1.	The Categories of Machine Learning	25
2.7.	Classification	26
2.8.	Logistic Regression	27
2.8.1.	Logistic Regression Algorithm	27
2.8.2.	Estimating Logistic Regression Parameters Using Maximum Likelihood Estimation	30
2.8.3.	Logistic Function and Odds Ratio	33
2.8.4.	Inference and Interpretation for Logistic Regression	33
2.8.5.	The Advantages and Disadvantages of Logistic Regression	36
2.9.	Support Vector Machines	36
2.9.1.	Support Vector Machines Algorithm	37
2.9.2.	Kernel Function	41
2.9.3.	Kernel Trick for Dual Problem	42
2.9.4.	The Advantages and Disadvantages of Support Vector Machines	43
2.9.5.	Platt Scaling	44
2.10.	Naïve Bayes	46
2.10.1.	Naïve Bayes Algorithm	46
2.10.2.	Probability Model for Naïve Bayes	46
2.10.3.	Laplace Smoothing	48
2.10.4.	Maximum Likelihood Estimator in Naïve Bayes Model	49
2.10.5.	The Advantages and Disadvantages of Naïve Bayes	50
2.11.	Algorithm Selection Based on Data Characteristics	51
2.11.1.	Logistic Regression	51
2.11.2.	Support Vector Machines	52

2.11.3. Naïve Bayes	52
2.12. Performance Evaluation	54
2.13. Machine Learning Interpretation	56
III THE IMPLEMENTATION OF BOOTSTRAP AGGREGATING CLASSIFIER FOR AUTO INSURANCE CLAIMS FRAUD DETECTION	58
3.1. Auto Insurance Fraud Detection	58
3.2. Ensemble Method	58
3.2.1. Types of Ensemble Algorithms	59
3.2.2. The Ensemble Paradigms	61
3.3. Bootstrap Aggregating (Bagging)	62
3.3.1. Bagging Algorithm	62
3.3.2. The Advantage and Disadvantages of Bootstrap Aggregating	65
3.4. Bagging Algorithm for Classification	66
3.5. Research Flowchart	67
3.5.1. Exploratory Data Analysis	67
3.5.2. Data Preprocessing	69
3.5.3. Data Splitting and Balancing	71
3.5.4. Modelling	73
3.5.5. Model Evaluation	74
IV STUDY CASE OF THE IMPLEMENTATION OF BOOTSTRAP AGGREGATING CLASSIFIER FOR AUTO INSURANCE CLAIMS FRAUD DETECTION	75
4.1. Data Description	75
4.2. Exploratory Data Analysis	76
4.3. Feature Engineering	83
4.3.1. Feature Creation	83
4.3.2. Feature Encoding	84
4.3.3. Imputation and Binning	86
4.3.4. Feature Selection	88
4.3.5. Feature Scaling	92
4.3.6. Data Splitting and Balancing	92
4.4. Fraud Detection with Bootstrap Aggregating Classifier	93
4.4.1. Bootstrap Aggregating-Logistic Regression	93
4.4.2. Bootstrap Aggregating-Support Vector Machines	96
4.4.3. Bootstrap Aggregating-Naïve Bayes	100
4.4.4. Empirical Analysis of Model Performance and Its Interpretation	103

4.4.5. Risk Profile of Policyholder with the Highest Fraud Claim	
Probability	113
V CONCLUSIONS	116
5.1. CONCLUSIONS	116
5.2. RECOMMENDATIONS	117
REFERENCES	119
A DATA	124
B PROGRAM	128
C BOOTSTRAP AGGREGATING-SUPPORT VECTOR MACHINE MODEL PREDICTIONS	179
D BOOTSTRAP AGGREGATING-LOGISTIC REGRESSION MODEL COEFFICIENT PARAMETER	180
E BOOTSTRAP AGGREGATING-NAIVE BAYES MODEL CONDITIONAL FEATURE PROBABILITIES OF THE CLASS	181
F BOOTSTRAP AGGREGATING-NAIVE BAYES MODEL PRIOR PROBABILITY OF THE CLASS	183