

**ENSEMBLE METHOD OF TWEET, URL AND OTHER FEATURES
CLASSIFIERS IN TWITTER BOT DETECTION**



UNIVERSITI TEKNIKAL MALAYSIA MELAKA

BORANG PENGESAHAN STATUS LAPORAN

JUDUL: [ENSEMBLE METHOD OF TWEET, URL AND OTHER FEATURES CLASSIFIERS IN TWITTER BOT DETECTION]

SESI PENGAJIAN: [2022 / 2023]

Saya: AHMAD SYAHMI FARHAN BIN SABRI

mengaku membenarkan tesis Projek Sarjana Muda ini disimpan di Perpustakaan Universiti Teknikal Malaysia Melaka dengan syarat-syarat kegunaan seperti berikut:

1. Tesis dan projek adalah hakmilik Universiti Teknikal Malaysia Melaka.
2. Perpustakaan Fakulti Teknologi Maklumat dan Komunikasi dibenarkan membuat salinan untuk tujuan pengajian sahaja.
3. Perpustakaan Fakulti Teknologi Maklumat dan Komunikasi dibenarkan membuat salinan tesis ini sebagai bahan pertukaran antara institusi pengajian tinggi.
4. * Sila tandakan (✓)

SULIT

(Mengandungi maklumat yang berdarjah keselamatan atau kepentingan Malaysia seperti yang termaktub di dalam AKTA RAHSIA RASMI 1972)

TERHAD

(Mengandungi maklumat TERHAD yang telah ditentukan oleh organisasi / badan di mana penyelidikan dijalankan)

TIDAK TERHAD

(TANDATANGAN PELAJAR)

Alamat: SU 14 JALAN SENTOSA,
TAMAN SENTOSA SUNGEI BARU
ULU, 78300 MASJID TANAH,
MELAKA

Tarikh: 21/9/2023

(TANDATANGAN PENYELIA)

TS. NOR AZMAN BIN MAT ARIFF

Nama Penyelia

Tarikh: 22/9/2023

ENSEMBLE METHOD OF TWEET, URL AND OTHER FEATURES
CLASSIFIERS IN TWITTER BOT DETECTION

AHMAD SYAHMI FARHAN BIN SABRI



This report is submitted in partial fulfillment of the requirements for the
Bachelor of [Computer Science (Computer Security)] with Honours.

UNIVERSITI TEKNIKAL MALAYSIA MELAKA

FACULTY OF INFORMATION AND COMMUNICATION TECHNOLOGY
UNIVERSITI TEKNIKAL MALAYSIA MELAKA

2023

DECLARATION

I hereby declare that this project report entitled
**[ENSEMBLE METHOD OF TWEET, URL AND OTHER FEATURES
CLASSIFIERS IN TWITTER BOT DETECTION]**

is written by me and is my own effort and that no part has been plagiarized
without citations.

STUDENT :



([AHMAD SYAHMI FARHAN BIN SABRI])

Date : 21/9/2023



UNIVERSITI TEKNIKAL MALAYSIA MELAKA

I hereby declare that I have read this project report and found
this project report is sufficient in term of the scope and quality for the award of
Bachelor of [Computer Science (Computer Security)] with Honours.

SUPERVISOR :



([TS. NOR AZMAN BIN MAT ARIFF])

Date : 22/9/2023

DEDICATION

To my beloved parents, I like to extend my thanks to my father and mother who always support me despite of many obstacles in life. I would like to express my gratitude and appreciation to them because they have given me a lot of encouragement and keep praying for successful future. This work hard is for my parents, and I want to make them happy.

To my fellow friends, thank you for being there for me throughout the entire bachelor program and their cooperation while conducting the research. I would like to express my gratitude to my supervisor for encouraging and believing in me to complete this research.



ACKNOWLEDGEMENTS

All praise is due to the All-Mighty Allah SWT, who has granted me the faith, strength, and abilities to comprehend, study, and complete this research. Peace and prayers be upon our most beloved Prophet Muhammad S.A.W., the most beautiful soul whose sayings, actions, and stories have profoundly inspired me to believe that there are no limits to what I can accomplish when fully committed to achieving something with Allah on my side.

I also revere the assistance and direction of my supervisor, Mr. Nor Azman bin Mat Ariff, for his guidance, encouragement, and patience in preparing this research. I am fortunate to have had such wonderful, loving, and supportive parents, who provided me with a home education while I was growing up and paid for my education until I graduated. They have **been** my pillar of fortitude, and to this day, they want to be the first to congratulate me on even the smallest of my accomplishments. To my colleagues who assisted in the arduous process of gathering data and preparing this research. May Allah bless you all for your perseverance and altruism. I would also like to thank my beloved parents who have supported and motivated me throughout my project.

UNIVERSITI TEKNIKAL MALAYSIA MELAKA

ABSTRACT

Twitter has experienced a remarkable rise in popularity and influence. However, twitter's popularity and open nature make it a desirable target for bots known as Twitter bots. This research proposes an Ensemble method for detecting Twitter bots. In the initial phase, four models are developed from Twitter features: a model that extracts tweet features using words, a model that extracts tweet features using n-gram, a model that extracts URL features using n-gram, and one that extracts additional features. Information Gain feature selection is applied and evaluated for multiple threshold values for all models to achieve the most accurate representation. The model with the threshold value that has the highest accuracy on the training set is chosen as the input for the ensemble method. The final prediction is derived by combining the probability outputs of these four models. This Ensemble method strategy aims to improve the classifier's overall performance by capitalizing on the strengths of the four fundamental models. To evaluate the effectiveness of the proposed method, extensive experiments are conducted on the Cresci-2017 dataset. The test results show that the proposed method using four models provides a satisfactory mean accuracy of 97.50%, which surpasses the accuracy of the models together, which are 70.50% for the tweets word model, 85.00% for the tweets n-gram model, 93.50% for the URL n-gram model and 82.00% for other features model. This demonstrates the efficacy of the ensemble approach and the importance of incorporating diverse features to achieve outstanding Twitter bot detection accuracy.

ABSTRAK

Twitter telah mengalami peningkatan yang luar biasa dalam populariti dan pengaruh. Walau bagaimanapun, populariti dan sifat terbuka Twitter menjadikannya sasaran yang wajar untuk bot yang dikenali sebagai bot Twitter. Penyelidikan ini mencadangkan Kaedah Ensemble untuk mengesan bot Twitter. Pada fasa awal, empat model dibangunkan daripada ciri Twitter: model yang mengekstrak ciri tweet menggunakan perkataan, model yang mengekstrak ciri tweet menggunakan n-gram, model yang mengekstrak ciri URL menggunakan n-gram, dan model yang mengekstrak ciri tambahan. Pemilihan filter *Information Gain* digunakan dan dinilai untuk nilai ambang berbilang bagi kedua-dua model untuk mencapai perwakilan yang paling tepat. Model dengan nilai ambang yang mempunyai ketepatan tertinggi pada set latihan dipilih sebagai input untuk kaedah *Ensemble*. Ramalan akhir diperoleh dengan menggabungkan output kebarangkalian bagi empat model ini. Strategi kaedah *Ensemble* ini bertujuan untuk meningkatkan prestasi keseluruhan pengelasan dengan memanfaatkan kekuatan empat model asas. Untuk menilai keberkesanan kaedah yang dicadangkan, eksperimen yang meluas dijalankan pada dataset awam Cresci-2017. Keputusan ujian menunjukkan bahawa kaedah yang dicadangkan hanya menggunakan empat ciri memberikan ketepatan min yang memuaskan iaitu 97.50%, yang mengatasi ketepatan model bersama, iaitu 70.50% untuk model perkataan *tweets*, 85.00% untuk model *tweets* n-gram, 93.50% untuk model URL n-gram dan 82.00% untuk model ciri lain. Ini menunjukkan keberkesanan pendekatan *ensemble* dan kepentingan menggabungkan ciri yang pelbagai untuk mencapai ketepatan pengesanan bot Twitter yang cemerlang.

TABLE OF CONTENTS

	PAGE
DECLARATION	II
DEDICATION	III
ACKNOWLEDGEMENTS	IV
ABSTRACT	V
ABSTRAK	VI
TABLE OF CONTENTS	VII
LIST OF TABLES	XVI
LIST OF FIGURES	XVIII
LIST OF ABBREVIATIONS	XXI
LIST OF ATTACHMENTS	XXII
CHAPTER 1: INTRODUCTION	1
1.1 Introduction.....	1
1.2 Problem Statement (PS).....	2
1.3 Project Question (PQ).....	2

1.4	Project Objective (PO).....	3
1.5	Project Scope	3
1.6	Project Contribution (PC)	4
1.7	Report Organization.....	4
1.7.1	Introduction.....	5
1.7.2	Literature Review	5
1.7.3	Project Methodology	6
1.7.4	Design	6
1.7.5	Implementation	6
1.7.6	Testing and Analysis.....	6
1.7.7	Project Conclusion.....	6
1.8	Conclusion	7
CHAPTER 2: LITERATURE REVIEW.....		8
2.1	Introduction.....	8
2.2	ISA Definition.....	10
2.2.1	Denial-of-Service Attack (DoS)	10
2.2.2	Phishing Attack.....	11
2.2.3	Spam Attack.....	11
2.2.4	Malware	12
2.2.5	Virus	12
2.2.6	Botnet Attack	12
2.3	Classification of ISA.....	13
2.3.1	Active Attack	13
2.3.2	Passive Attack.....	13
2.4	Twitter.....	14

2.4.1	Twitter Definition	14
2.4.2	Bots	14
2.4.3	Twitter Bots	15
2.4.4	Twitter Bot Features	18
2.4.4.1	Content.....	18
2.4.4.2	Sentiment	18
2.4.4.3	Account Information.....	19
2.4.4.4	Account Usage.....	19
2.4.4.5	Social Network	19
2.5	Machine learning	20
2.5.1	Types of Machine Learning.....	20
2.5.1.1	Supervised Learning	21
2.5.1.2	Semi-supervised Learning	22
2.5.1.3	Unsupervised Learning.....	22
2.5.2	Text Classification	22
2.5.3	Dataset	23
2.5.3.1	Kaggle.....	23
2.5.3.2	GitHub	24
2.5.3.3	Twitter API	24
2.5.4	Data Preprocessing	24
2.5.4.1	Definition.....	24
2.5.4.2	Preprocessing Type.....	24
2.5.5	Feature Extraction.....	26

2.5.5.1	Bag-Of-Words (BOW)	26
2.5.5.2	N-Gram	26
2.5.5.3	Lexical Features	27
2.5.5.4	TF-IDF	27
2.5.6	Feature Selection	28
2.5.6.1	Filter Methods.....	29
2.5.6.2	Wrapper Methods	30
2.5.6.3	Embedded Methods	31
2.5.6.4	Document Frequency.....	31
2.5.6.5	Information Gain	31
2.5.6.6	Gini Index	31
2.5.6.7	Chi-Square	32
2.5.6.8	Best Term.....	32
2.5.6.9	Ambiguity Measure	33
2.5.6.10	Analysis of Feature Selection	33
2.5.7	Classification	34
2.5.7.1	Definition.....	34
2.5.7.2	Classification Type	34
2.5.7.2.1	Discriminative.....	34
2.5.7.2.2	Generative.....	35
2.5.7.3	Analysis of Classifiers	37
2.6	Critical Review	39

2.6.1	Previous Research on Twitter Bot Detection.....	39
2.6.1.1	Research Paper 1.....	39
2.6.1.2	Research Paper 2.....	39
2.6.1.3	Research Paper 3.....	40
2.6.1.4	Research Paper 4.....	41
2.6.1.5	Research Paper 5.....	42
2.6.1.6	Research Paper 6.....	42
2.6.1.7	Research Paper 7.....	43
2.6.1.8	Twitter Bot Detection Literature Analysis Error! Bookmark not defined.	
2.7	Conclusion	46
CHAPTER 3: PROJECT METHODOLOGY		47
3.1	Introduction.....	47
3.2	Methodology	47
3.2.1	Previous Research.....	48
3.2.2	Information Gathering	48
3.2.3	Define Scope.....	48
3.2.4	Design and Implementation.....	48
3.2.5	Testing and Model Evaluation.....	49
3.2.6	Documentation.....	49
3.3	Project Schedule and Milestones	49
3.3.1	Project Flowchart.....	49
3.3.2	Project Milestones	50
3.4	Requirement Analysis	52

3.4.1	Software Requirements.....	52
3.4.2	Hardware Requirements	53
3.5	Conclusion	54
CHAPTER 4: DESIGN		55
4.1	Introduction.....	55
4.2	Problem Analysis	55
4.3	Project Design.....	56
4.3.1	Dataset	58
4.3.2	Data Preprocessing	60
4.3.2.1	Removal of Special Characters.....	60
4.3.2.2	Removal of Stop-words	60
4.3.2.3	Stemming	61
4.3.2.4	One-Hot Encoding	61
4.3.3	Feature Extraction.....	61
4.3.4	Generate Feature Vector.....	61
4.3.5	Feature Selection	62
4.3.6	Split Train and Test Set	63
4.3.7	Normalization	63
4.3.8	Classification	63
4.3.9	Ensemble Method	66
4.4	Conclusion	66
CHAPTER 5: IMPLEMENTATION.....		68
5.1	Introduction.....	68
5.2	Environment Setup.....	68

5.3	Process Module	69
5.3.1	Collection of Dataset	69
5.3.2	Data Preprocessing	69
5.3.2.1	Sorted	70
5.3.2.2	Content	73
5.3.2.3	Remove Special Character	74
5.3.2.4	Convert to Lowercase	75
5.3.2.5	Remove Stop Words	77
5.3.2.6	Stemming	77
5.3.2.7	Random Subsampling	78
5.3.2.8	One-Hot Encoding	79
5.3.3	Feature Extraction	80
5.3.3.1	Tweets Data	80
5.3.3.2	URLs Data	82
5.3.4	Feature Vector	84
5.3.4.1	Tweets Data	84
5.3.4.2	URL Data	86
5.3.4.3	Other Features Data	88
5.3.5	Data Training and Testing	88
5.3.5.1	Splitting Data	89
5.3.5.2	Normalization	90
5.3.5.3	Feature Selection	91
5.4	Classification	91

5.4.1	Model 1 (Tweets word)	91
5.4.1.1	Attributes	91
5.4.1.2	10 Runs	92
5.4.1.3	Accuracy Table	93
5.4.2	Model 2 (Tweets N-grams).....	94
5.4.2.1	Attributes	94
5.4.2.2	10 Runs	95
5.4.2.3	Accuracy Table	96
5.4.3	Model 3 (URLs N-grams).....	99
5.4.3.1	Attributes	99
5.4.3.2	10 Runs	100
5.4.3.3	Accuracy Table.....	101
5.4.4	Model 4 (Other Features).....	104
5.4.4.1	Attributes	104
5.4.4.2	10 Runs	105
5.4.4.3	Accuracy Table	106
5.5	Confusion Matrix	106
5.6	Ensemble.....	108
CHAPTER 6: DISCUSSION		111
6.1	Introduction.....	111
6.2	Discussion of Project	111
6.3	Discussion on The Newly Proposed Method.....	112
6.4	Conclusion	114

CHAPTER 7: PROJECT CONCLUSION	116
7.1 Introduction.....	116
7.2 Project Summary.....	116
7.3 Project Constraint.....	117
7.4 Project Contribution.....	117
7.5 Project Limitation	118
7.6 Future Work.....	118
7.7 Conclusion	119
REFERENCES.....	120



LIST OF TABLES

	PAGE
Table 1.1: Problem Statement.....	2
Table 1.2: Project Question.....	2
Table 1.3: Project Objective.....	3
Table 1.4: Project Contribution.....	4
Table 2.1: Comparative analysis of feature selection methods.....	33
Table 2.2: Summary of machine learning algorithms.....	37
Table 2.3: Twitter Bot Detection Literature.....	44
Table 3.1: Project Milestones.....	51
Table 3.3: Software Requirements.....	52
Table 4.1: Description of data attributes.....	58
Table 4.2: Total dataset based on class categories.....	59
Table 5.1: Data extraction of tweet based on feature extraction.....	81
Table 5.2: Data extraction of URL based on feature extraction.....	83
Table 5.3: Tweet feature descriptor based on the data extracted.....	84
Table 5.4: Tweet feature vector.....	85
Table 5.5: URL feature descriptor.....	86
Table 5.6: URL feature vector.....	87
Table 5.7: Sorted Number of Instances of Each Dataset.....	89
Table 5.8: Training and Testing Size.....	89
Table 5.9: Total Instances attributes based on th3 threshold value.....	92
Table 5.10: Total instances attributes based on the threshold value.....	94
Table 5.11: Tweets n-gram total instances attributes based on the threshold value.	95
Table 5.12: Train and Test Accuracy of Tweets using 1-gram method.....	97

Table 5.13: Train and Test Accuracy of Tweets using 2-gram method.....	98
Table 5.14: Train and Test Accuracy of Tweets using 3-gram method.....	99
Table 5.15: URL n-gram total instances attributes based on the threshold value	100
Table 5.16: Train and Test Accuracy of URL using 1-gram method.....	102
Table 5.17: Train and Test Accuracy of URL using 2-gram method.....	103
Table 5.18: Train and Test Accuracy of URL using 3-gram method.....	104
Table 5.19: Other features total instances attributes based on the threshold value.	105
Table 5.20: Train and Test Accuracy of other features.....	106
Table 5.21: Result of each model based on the highest training accuracy.	109
Table 5.22: Test and Product Accuracy based on 10 run model.	110
Table 5.23: The ensemble method of accuracy.....	110
Table 6.1: The Benchmark.....	113
Table 6.2: Final Result.....	114



LIST OF FIGURES

	PAGE
Figure 1.1: Report Organization	5
Figure 2.2.1: Taxonomy of Twitter bot	Error! Bookmark not defined.
Figure 2.2: 2017-2021 Bad bot v good bot v human traffic 2017-2021 Graph (Imperva Bad Bot Report, 2022.)	15
Figure 2.3: Example of former of Twitter CEO tweets about spam account, Source:.....	17
Figure 2.4: Machine learning phase	20
Figure 2.5: Types of machine learning	21
Figure 2.6: Categories of machine learning algorithms according to training data nature	21
Figure 2.7: Filter methods	29
Figure 2.8: Wrapper methods	30
Figure 2.9: Information Gain formula	31
Figure 2.10: Gini index formula	32
Figure 3.1: Project Framework	47
Figure 3.2: Project Flowchart	50
Figure 4.1: Design of the project.....	57
Figure 4.2: Example of the Twitter dataset.	60
Figure 4.3: Data Train and Test 80/20	63
Figure 4.4: SVM Graph, Source: SVM classifier, 2013.....	64
Figure 4.5: Hyperplane Placement, Source: SVM classifier,2013	65
Figure 4.6: Real Distribution Data Example, Source: Support Vector Machine (SVM) algorithm, 2017	65
Figure 4.7: Non-Linear Graph, Source: The SVM classifier, 2013.....	66

Figure 5.1: Process Module	69
Figure 5.2: Data preprocessing process	70
Figure 5.3: Combined file command	70
Figure 5.4: Tweets text dataset	71
Figure 5.5: Twitter URL.....	71
Figure 5.6: Twitter User ID instances	72
Figure 5.7: Other relevant features instances.....	73
Figure 5.8: The dataset with headers description	73
Figure 5.9: The dataset after remove headers	74
Figure 5.10: remove non-ASCII Character	74
Figure 5.11: Remove special character	75
Figure 5.12: dataset after remove special character	75
Figure 5.13: Method to change uppercase to lowercase	76
Figure 5.14: Tweets text change to lowercase.....	76
Figure 5.15: Tweets text change to lowercase.....	77
Figure 5.16: Tweets before stemming.....	78
Figure 5.17: Tweets after stemming.	78
Figure 5.18: Random Sample Number of Instances	79
Figure 5.19: Feature vector module	84
Figure 5.20: Other Features Feature Vector	88
Figure 5.21: Train and Test Set of Data.....	90
Figure 5.22: The accuracy of Model 1 in 10 runs.....	93
Figure 5.23: The accuracy of tweets using 1-gram extraction method in 10 runs	96
Figure 5.24: The accuracy of tweets using 2-gram extraction method in 10 runs	96
Figure 5.25: The accuracy of tweets using 3-gram extraction method in 10 runs	96
Figure 5.26: The accuracy of URL using 1-gram extraction method	101
Figure 5.27: The accuracy of URL using 2-gram extraction method	101
Figure 5.28: The accuracy of URL using 3-gram extraction method	101
Figure 5.29: The accuracy of other features in 10 runs.....	105
Figure 5.30: Confusion matrix for model 1	107
Figure 5.31: Confusion matrix for model 2	107

Figure 5.32: Confusion matrix for model 3	107
Figure 5.33: Confusion matrix for model 4	108
Figure 5.34: Graph of the accuracy of each model and ensemble models.....	110
Figure 6.1: Recall of the architecture of ensemble method.....	113



LIST OF ABBREVIATIONS

FYP - **Final Year Project**

BOW - **Bag-of-Words**

CS - **Chi-Square**

DoS - **Denial-of-Service**

SVM - **Support Vector Machine**

IG - **Information Gain**

TP - **True Positive**

TN - **True Negative**

FP - **False Positive**

FN - **False Negative**

URL - **Uniform Resource Locator**

LIST OF ATTACHMENTS

		PAGE
Attachment 1	Gantt Chart	127
Attachment 2	Run Scripts	128

