

## TABLE OF CONTENTS

|                                                |           |
|------------------------------------------------|-----------|
| <b>TABLE OF CONTENTS .....</b>                 | <b>2</b>  |
| <b>LIST OF TABLES .....</b>                    | <b>5</b>  |
| <b>LIST OF FIGURES .....</b>                   | <b>6</b>  |
| <b>ABSTRAK .....</b>                           | <b>8</b>  |
| <b>ABSTRACT .....</b>                          | <b>9</b>  |
| <b>CHAPTER I INTRODUCTION .....</b>            | <b>10</b> |
| 1.1 BACKGROUND .....                           | 10        |
| 1.2 RESEARCH PROBLEM.....                      | 12        |
| 1.3 RESEARCH SCOPE .....                       | 13        |
| 1.4 RESEARCH OBJECTIVE.....                    | 13        |
| 1.5 RESEARCH BENEFITS.....                     | 13        |
| <b>CHAPTER II LITERATURE REVIEW.....</b>       | <b>15</b> |
| <b>CHAPTER III THEORITICAL BASIS .....</b>     | <b>18</b> |
| 3.1 CROSS-SITE SCRIPTING (XSS).....            | 18        |
| 3.2 TYPES OF XSS ATTACKS .....                 | 19        |
| 3.2.1 <i>Reflected XSS Attacks</i> .....       | 19        |
| 3.2.2 <i>Stored XSS Attacks</i> .....          | 20        |
| 3.2.3 <i>DOM-based XSS Attacks</i> .....       | 20        |
| 3.3 NATURAL LANGUAGE PROCESSING (NLP).....     | 20        |
| 3.4 DATA PREPROCESSING .....                   | 22        |
| 3.5 CODEBERT .....                             | 23        |
| 3.5.1 <i>CodeBERT Algorithm</i> .....          | 27        |
| 3.6 TRANSFER LEARNING .....                    | 30        |
| 3.6.1 <i>Transfer Learning Algorithm</i> ..... | 31        |
| 3.7 EVALUATION METRICS.....                    | 33        |
| 3.7.1 <i>Accuracy</i> .....                    | 33        |
| 3.7.2 <i>Precision</i> .....                   | 34        |
| 3.7.3 <i>Recall</i> .....                      | 35        |
| 3.7.4 <i>F1-Score</i> .....                    | 36        |
| <b>CHAPTER IV RESEARCH METHODOLOGY .....</b>   | <b>37</b> |

|                                                                       |            |
|-----------------------------------------------------------------------|------------|
| 4.1 RESEARCH GENERAL DESCRIPTION .....                                | 37         |
| 4.2 DATASETS .....                                                    | 38         |
| 4.3 SETUPS .....                                                      | 45         |
| 4.4 DATA PREPROCESSING .....                                          | 45         |
| 4.4.1 Data Splitting .....                                            | 51         |
| 4.5 MODEL TRAINING .....                                              | 53         |
| 4.6 ANALYSIS OF RESULTS .....                                         | 55         |
| 4.6.1 Hyperparameter Tuning .....                                     | 56         |
| <b>CHAPTER V IMPLEMENTATION .....</b>                                 | <b>60</b>  |
| 5.1 DATA PREPROCESSING AND DATASET BUILDING FOR PHP DATASETS .....    | 60         |
| 5.1.1 Extracting PHP Code Blocks .....                                | 60         |
| 5.1.2 Dataset Diversification .....                                   | 61         |
| 5.1.3 Code Standardization .....                                      | 63         |
| 5.1.4 Feature Engineering .....                                       | 64         |
| 5.1.5 Data Encoding .....                                             | 66         |
| 5.1.6 Feature Normalization .....                                     | 67         |
| 5.2 DATA PREPROCESSING AND DATASET BUILDING FOR NODEJS DATASETS ..... | 68         |
| 5.2.1 Simplifying NodeJS Code .....                                   | 68         |
| 5.2.2 Dataset Diversification .....                                   | 69         |
| 5.2.3 Code Standardization .....                                      | 71         |
| 5.2.4 Feature Engineering .....                                       | 72         |
| 5.2.5 Data Encoding and Normalization .....                           | 76         |
| 5.2.6 Feature Normalization .....                                     | 77         |
| 5.3 MODEL BUILDING AND TRAINING .....                                 | 78         |
| 5.3.1 Model Architecture .....                                        | 78         |
| 5.3.2 Training Process .....                                          | 80         |
| <b>CHAPTER VI RESULT AND DISCUSSION .....</b>                         | <b>92</b>  |
| 6.1 TRAINING RESULTS .....                                            | 92         |
| 6.2 TRAINING RESULTS OF CODEBERT MODEL .....                          | 106        |
| 6.3 COMPARISON OF MODEL RESULTS WITH CODEBERT .....                   | 113        |
| 6.4 COMPARISON OF MODEL RESULTS WITH EXISTING LITERATURE .....        | 118        |
| 6.5 COMPARATIVE ANALYSIS WITH STATIC ANALYSIS TOOLS .....             | 121        |
| <b>CHAPTER VII CONCLUSION AND FUTURE WORKS .....</b>                  | <b>127</b> |

|                                       |            |
|---------------------------------------|------------|
| 7.1 CONCLUSION .....                  | 127        |
| 7.2 LIMITATIONS AND FUTURE WORK ..... | 129        |
| 7.2.1 <i>Limitations</i> .....        | 129        |
| 7.2.2 <i>Future Work</i> .....        | 130        |
| <b>REFERENCES.....</b>                | <b>132</b> |