|  |  |  |  |
| --- | --- | --- | --- |
| **TABLE OF CONTENTS** | | | |
| **CHAPTER NO.** |  | **TITLE** | **PAGE NO.** |
|  | **ABSTRACT** | | **iii** |
|  | **LIST OF FIGURES** | | **viii** |
|  | **LIST OF TABLES** | | **x** |
|  | **LIST OF ABBREVIATION** | | **xi** |
| **1** | **INTRODUCTION** | | **1** |
|  | 1.1 | CYBERBULLYING | 1 |
|  |  | 1.1.1 Different Types Of Cyberbullying | 2 |
|  |  | 1.1.2 Main Effects Of Cyberbullying | 4 |
|  | 1.2 | TWITTER | 5 |
|  | 1.3 | CYBERBULLYING IN TWITTER | 5 |
|  | 1.4 | SENTIMENT ANALYSIS | 7 |
|  | 1.5 | TEXT CLASSIFICATION | 9 |
|  |  | ALGORITHMS FOR SENTIMENT |  |
|  |  | ANALYSIS |  |
|  | 1.6 | SCOPE | 12 |
|  | 1.7 | CHALLENGES | 12 |
| **2** | **LITERATURE REVIEW** | | **13** |
|  | 2.1 | REVIEW OF RESEARCH PAPERS | 13 |
|  | 2.2 | SUMMARY OF EXISTING PAPERS | 20 |
| **3** | **PROPOSED WORK** | | **22** |
|  | 3.1 | TWITTER DATASET | 22 |
|  | 3.2 | ARCHITECTURE DIAGRAM | 23 |
|  |  | 3.2.1 Preprocessing Tweets | 24 |

|  |  |  |  |
| --- | --- | --- | --- |
|  |  | 3.2.2 Feature Extraction | 25 |
|  |  | 3.2.2.1 Term Frequency-Inverse | 25 |
|  |  | Document Frequency |  |
|  |  | 3.2.2.2 Tokenizers | 26 |
|  |  | 3.2.3 Supervised Text Classification | 27 |
|  |  | Models |  |
|  |  | 3.2.4 Evaluation Metrics | 30 |
|  |  | 3.2.5 Classifying Tweets | 32 |
| **4** | **SYSTEM REQUIREMENTS** | | **33** |
|  | 4.1 | HARDWARE REQUIREMENTS | 33 |
|  | 4.2 | SOFTWARE REQUIREMENTS | 33 |
|  |  | 4.2.1 Visual Studio Code | 33 |
|  |  | 4.2.2 Tensorflow | 34 |
|  |  | 4.2.3 Transformers | 34 |
|  |  | 4.2.4 Angular JS | 35 |
|  |  | 4.2.5 Selenium | 35 |
| **5** | **IMPLEMENTATION MODULES** | | **36** |
|  | 5.1 | WEB SCRAPING | 36 |
|  | 5.2 | TWEETS COLLECTION AND | 38 |
|  |  | PREPROCESSING |  |
|  | 5.3 | FEATURE EXTRACTION | 39 |
|  | 5.4 | REAL-TIME TWEETS | 40 |
|  |  | CLASSIFICATION |  |
|  | 5.5 | REPORT GENERATION | 40 |

|  |  |  |  |
| --- | --- | --- | --- |
| **6** | **RESULTS AND DISCUSSIONS** | | **41** |
|  | 6.1 | PERFORMANCE METRICS | 41 |
|  | 6.2 | VISUALIZATION | 42 |
|  | 6.3 | ANALYZING BEST CLASSIFICATION | 45 |
|  |  | MODEL |  |
| **7** | **CONCLUSION AND FUTURE WORK** | | **48** |
|  | 7.1 | CONCLUSION | 48 |
|  | 7.2 | FUTURE WORK | 48 |
|  | **REFERENCES** | | 49 |

|  |  |  |
| --- | --- | --- |
| **LIST OF FIGURES** | | |
| **FIGURE NO.** | **TITLE** | **PAGE NO.** |
| 1.1 | Percentage of People Cyberbullied Across Social | **2** |
|  | Media Platforms |  |
| 1.2 | Examples of Non Bullying Tweet | **6** |
| 1.3 | Examples of Bullying Tweets | **7** |
| 3.1 | Corona NLP Twitter Dataset | **22** |
| 3.2 | Country-wise Tweets Count | **23** |
| 3.3 | Architecture Diagram of Cyberbullying Detection | **24** |
|  | in Tweets |  |
| 5.1 | A Tweet Posted by Elon Musk | **37** |
| 5.2 | XPath Extraction From Tweets For Scraping | **37** |
| 5.3 | Code Snippet For Scraping Tweets | **37** |
| 5.4 | Collection of Real-Time Tweets after Webscraping | **38** |
| 6.1 | Accuracy Metric Between Text Classification | **43** |
|  | Models |  |
| 6.2 | Precision Metric Between Text Classification | **43** |
|  | Models |  |
| 6.3 | Recall Metric Between Text Classification Models | **44** |
| 6.4 | F1 Score Metric Between Text Classification | **45** |
|  | Models |  |
| 6.5 | BERT Model Summary | **46** |
| 6.6 | Classification of Real-Time Tweets by BERT | **46** |
|  | Model |  |
|  |  |  |
|  |  |  |
| 6.7 | CSV File of the Generated Report | **47** |
| 6.8 | Report Generated in Frontend | **47** |

|  |  |  |
| --- | --- | --- |
| **LIST OF TABLES** | | |
| **TABLE NO.** | **TITLE** | **PAGE NO.** |
| 3.1 | TF-IDF Calculation | **26** |
| 3.2 | Confusion Matrix | **30** |
| 6.1 | Performance Metrics Result for Bullying Class | **41** |
| 6.2 | Performance Metrics Result for Non Bullying | **45** |
|  | Class |  |

**LIST OF ABBREVIATIONS**

API Application Programming Interface

BERT Bidirectional Encoder Representations from Transformers

Bi-GRU Bidirectional Gated Recurrent Units

Bi-LSTM Bidirectional Long Short Term Memory Networks

CNN Convolution Neural Network

GRU Gated Recurrent Units

IDF Inverse Document Frequency

LR Logistic Regression

LSTM Long Short Term Memory Networks

NB Naive Bayes

NLP Natural Language Processing

NN Neural Network

RNN Recurrent Neural Networks

SA Sentiment Analysis

SPC Spanish Cyberbullying Prevention System

SVM Support Vector Machine

TF Term Frequency

TF-IDF Term Frequency - Inverse Document Frequency