 

NLP BASED EMOTION RECOGNITION USING TEXT & AUDIO

**A MINI PROJECT REPORT**

***Submitted by***

|  |  |
| --- | --- |
| **ASWIN N** | **(710720104011)** |
| **JANANI KEERTHANA U** | **(710720104034)** |
| **JERISH PRINCETON R** | **(710720104037)** |
| **SIVA PRAKASH** | **(710720104307)** |

***In partial fulfillment for the award of the degree of***

**BACHELOR OF TECHNOLOGY IN**

**COMPUTER SCIENCE AND ENGINEERING**

**DR.N.G.P INSTITUTE OF TECHNOLOGY**

**COIMBATORE 641048**

**MAY 2023**

ANNA UNIVERSITY: CHENNAI 600 025

# 

**BONAFIDE CERTIFICATE**

Certified that this Mini project report **“NLP BASED EMOTION RECOGNITION USING TEXT & AUDIO”** is the work of ***ASWIN N(710720104011), JANANI KEERTHANA(710720104034), JERISH PRINCETON(710720104037),***

***SIVA PRAKAS( 710720104307),*** who carried out the mini project work under my supervision.

**SIGNATURE**

|  |
| --- |
| **HEAD OF THE DEPARTMENT** |
| **Dr. D.PALANIKKUMAR M.E.,Ph.D** |
| HEAD OF THE DEPARTMENT |
| Department of Computer Science and Engineering |
| Dr.N.G.P. Institute of Technology |
| Coimbtore-641048 |

**SIGNATURE**

|  |
| --- |
| **SUPERVISOR** |
| **MR.SELVARAJ ME** |
| Assistant Professor |
| Department of Computer Science and Engineering |
| Dr.N.G.P. Institute of Technology |
| Coimbatore-641048 |

**SIGNATURE SIGNATURE**

The mini project report submitted for the Viva-Voce held on …………………

**INTERNAL EXAMINER EXTERNAL EXAMINER**

# ACKNOWLEDGEMENT

Words act as gateway to express tokens of acknowledgement. First of all, we would like to thank the supreme power, the **Almighty God**, who has given us the strength and courage to complete our work successfully.

We express our profound gratitude and deep sense of thanks to **Dr. Nalla G. Palaniswami MD., AB (USA),** Chairman of Kovai Medical Center& Hospital, for providing us with the necessary facilities to complete our project work effectively.

Our heartfelt gratitude to **Dr. Thavamani D Palaniswami MD., AB (USA)**,Secretary of Dr.

N. G. P. Institute of Technology, for her generous attitude and constant motivation, which had been one of the sole reasons to complete our project.

We are sincerely grateful to **Dr. S.U. Prabha M.E., Ph.D.**, Principal, who has always been a source of inspiration, well-wisher and a pillar of support for all the students in our institution by rendering full motivation whenever required.

We are highly indebted to **Dr. D. Palanikkumar M.E., Ph.D.**, Head of Department, Department of Computer Science and Engineering, for her dedication, keen interest, meticulous scrutiny and overwhelming attitude to help her students that have helped us to a very great extent to accomplish this task.

We wish to thank our Project Coordinator **Mrs.SUJITHA,** Professor, Department of Computer Science and Engineering, for his excellent assistance, aspiring guidance, regular feedback and invaluably constructive ideas.

Our sincere gratitude to our Supervisor **MR.SELVARAJ M.E.,** Assistant Professor, Department of Computer Science and Engineering for willingly share his precious time by giving us useful comments, remarks, timely suggestion with kindness, enthusiasm and dynamism have enabled us to complete our thesis. Finally, we owe huge thanks to our **Parents**, all the **Faculty members** and our **classmates**, whose love and insights have so deeply enriched our work.

# ABSTRACT

Emotion analysis and recognition has become an interesting topic of research among the computer vision research community. The database consists of audio and video sequences of actors displaying three different intensities of expressions of 23 different emotions along with facial feature tracking, skeletal tracking and the corresponding physiological data. The advancements in neural networks and the on- demand need for accurate and real-time Speech Emotion Recognition (SER) in human– computer interactions make it mandatory to compare available methods and databases in SER to achieve feasible solutions and a firmer understanding of this open-ended problem.

A Novel deep learning model architecture is presented that further improves the state-of-the-art in multimodal emotion recognition with speech and text on the IEMOCAP & RAVDESS datasets. Results from models trained on individual corpora show that combining speech and text improves performance only on corpora where the text of utterances varies across different emotions, while it reduced performance on corpora with fixed text expressed in different emotions, where the speech-only models performed better. Further, cross-corpus investigations are presented to understand the robustness to changing acoustic and textual content. Results show that models perform significantly better in matched conditions in particular single corpus models perform better than multi-corpus models, with the latter showing a tendency to be more robust to acoustic variations, while performance still depends on characteristics of both training corpora and test corpus.

iv

**LIST OF FIGURES**

|  |  |  |
| --- | --- | --- |
| **FIGURE**  **NO** | **NAME OF THE FIGURES** | **PAGE**  **NO** |
| 3.1 | Facial Emotion Recognition – AffectNet | 4 |
| 3.2 | Acoustic Emotion Recognition – SPEAR | 5 |
| 5.1 | Model Architecture | 7 |
| 7.1.1 | Text Model Training & Validation Loss | 20 |
| 7.1.2 | Text Model Training & Validation Accuracy | 20 |
| 7.2.1 | Audio Model Training & Validation Loss | 21 |
| 7.2.2 | Audio Model Training & Validation Accuracy | 21 |

v

# LIST OF ABBREVATIONS

|  |  |  |
| --- | --- | --- |
| **S. NO** | **ABBREVATION** | **EXPANSION** |
| 1 | SER | Speech Emotion Recognition |
| 2 | IEMOCAP | Interactive Emotional Dyadic Motion Capture |
| 3 | RAVDESS | Ryerson Audio-Visual Database of Emotional Speech and Song |
| 4 | CNN | Convolution Neural Network |
| 5 | RNN | Recurrent Neural Network |
| 6 | ASR | Automatic Speech Recognition |
| 7 | NLP | Natural Language Processing |
| 8 | DNN | Deep Neural Networks |
| 9 | MFCC | Mel-Frequency Cepstral Coefficients |
| 10 | SPEAR | Speech Emotion Analysis and Recognition |
| 11 | FC Layer | Fully Connected Layer |
| 12 | LSTM | Long Short-Term Memory |

**TABLE OF CONTENTS**

|  |  |  |
| --- | --- | --- |
| **CHAPTER NO** | **TITLE** | **PAGE NO** |
|  | **ABSTRACT** | **iv** |
|  | **LIST OF FIGURES** | **v** |
|  | **LIST OF ABBREVIATION** | **vi** |
| **1** | **INTRODUCTION** | **1** |
|  | 1.1 Scope of the project | 2 |
|  | 1.2 Objective of the project | 2 |
|  | 1.3 Review Summary | 2 |
| **2** | **LITERATURE SURVEY** | **3** |
|  | 2.1 Introduction | 3 |
| **3** | **EXISTING SYSTEM** | **4** |
|  | 3.1 AffectNet | 4 |
|  | 3.2 SPEAR | 5 |
| **4** | **PROPOSED SYSTEM** | **6** |
| **5** | **MODEL ARCHITECTURE** | **7** |
| **6** | **SYSTEM SPECIFICATION** | **9** |
|  | 6.1 Software & Library Requirements | 9 |
|  | 6.2 Hardware Requirement | 13 |
|  | 6.3 Installation Procedure | 13 |
| **7** | **IMPLEMENTATION AND RESULTS** | **20** |
|  | 7.1 Text Only Recognition | 20 |
|  | 7.2 Audio Only Recognition | 21 |
| **8** | **CONCLUSION & FUTURE SCOPE** | **22** |
|  | **APPENDIX – I** | **23** |
|  | **APPENDIX – II** | **33** |