**PLEASE MENTION TITLE OF THE THESIS**

*A Project Report*

*Submitted in partial fulfillment of the requirements for the award of the Degree of*

**BACHELOR OF TECHNOLOGY**

**IN**

# ELECTRONICS & COMMUNICATION ENGINEERING

BY

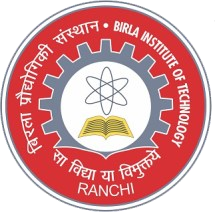
**NAME OF THE STUDENT**

**(ROLL NO.)**

Under the supervision of

**(Guide Name)**

**(Designation)**



DEPT. OF ELECTRONICS AND COMMUNICATION ENGINEERING

BIRLA INSTITUTE OF TECHNOLOGY

MESRA-835215, RANCHI

2025

### APPROVAL OF THE GUIDE

Recommended that the Project entitled **<Title of the report>** submitted by **<Roll No. and Name of the Student>** is approved by me for submission. This should be accepted as fulfilling the partial requirements for tthe award of Degree of Bachelor of Technology in **<Branch Name>**. To the best of my knowledge, the content of this report does not form a basis for the award of any previous degree to anyone else.

**Date: <Name of the mentor>**

**<Designation of the mentor>**

**Department of Electronics & Communication Engineering**

**Birla Institute of Technology, Mesra**

### DECLARATION CERTIFICATE

I certify that,

1. The work contained in the report is original and has been done by myself under the general supervision of my supervisor.
2. The work has not been submitted to any other Institute for any other degree or diploma.
3. I have followed the guidelines provided by the Institute in writing the report.
4. I have conformed to the norms and guidelines given in the Ethical Code of Conduct of the Institute.
5. Whenever I have used materials (data, theoretical analysis, and text) from other sources, I have given due credit to them by citing them in the text of the report and giving their details in the references.
6. Whenever I have quoted written materials from other sources, I have put them under quotation marks and given due credit to the sources by citing them and giving required details in the references.

**Date: <Name of the Student>**

**<Roll No. of the Student>**

**Department of Electronics & Communication Engineering**

**Birla Institute of Technology, Mesra**

# CERTIFICATE OF APPROVAL

This is to certify that the work embodied in this thesis entitled **“\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_”**, is carried out by **Name of the Student (Roll Number)** has been approved for the degree of Bachelor of Technology of Birla Institute of Technology, Mesra, Ranchi.

Date:

Place:

**Internal Examiner External Examiner**

**(Chairman)**

**Head of Department**

#### ABSTRACT

Abstract is a short summary of a research / industry project. It should include the statement of the problem you are trying to solve, purpose of the research / work, the methods used to find the solution, the results, and the implications of your findings. Word limit of Abstract for this report is 250. Use Times New Roman as Font Face, Size of the Font is 12, Line and the paragraph spacing is 1.5 lines. Text distribution should be even within the margins i.e., use justified text. Do not use citation in Abstract section of the report.

#### ACKNOWLEDGEMENT

This section of a project report is used to express your gratitude to the persons or institutions who significantly contributed during the project or research and helped you to accomplish your project. In this section you may mention your Guide, Head of the Department, Teachers of the department, Parents, Classmates, Lab staff, the department, institute etc. The maximum word limit of this section is 300.

Date: (Signature)

Name of the Student

Roll No. of the Student

#### CONTENTS

[ABSTRACT i](#_TOC_250003)

[ACKNOWLEDGEMENT ii](#_TOC_250002)

[LIST OF FIGURES iii](#_TOC_250001)

[LIST OF TABLES iv](#_TOC_250000)

CHAPTER 1 <NAME OF THE CHAPTER> 1

<SECTION NO. (e.g., 1.1)> 1

<SUBSECTION NO. (e.g., 1.1.4)> 3

CHAPTER 2 <NAME OF THE CHAPTER> 10

**The maximum number of Chapters should be Five.**

1. **Introduction: Overview of the project**
2. **Literature Review: Background of the project**
3. **Methodology: Description of the methods or algorithms applied.**
4. **Experimental Results and Discussion: Discussion and analysis of results.**
5. **Conclusion: Summary of the work accomplished and future scopes.**

**APPENDIX A ACCEPTED / COMMUNICATED PAPERS 34**

**REFERENCES 35**

#### LIST OF FIGURES

|  |  |  |
| --- | --- | --- |
| Figure 1.1 | Caption or Label of the Figure | 5 |
| Figure 2.1 | Caption or Label of the Figure | 8 |

#### LIST OF TABLES

|  |  |  |
| --- | --- | --- |
| Table 2.1 | Label of the Table No. 2.1 | 10 |
| Table 3.2 | Label of the Table No. 3.2 | 15 |

***CHAPTER 1***

# TEXT CLUSTERING

### 1.1 INTRODUCTION

Text Mining is one of the recent research domains in the field of Machine Learning. It is the process of transforming unstructured textual content into structured data for simple analysis. Text Mining is one of the recent research domains in the field of Machine Learning [1]. It is the process of transforming unstructured textual content into structured data for simple analysis.

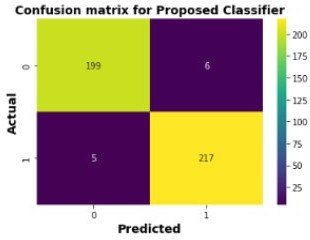


Figure 1.1 Confusion Matrix

Text Mining is one of the recent research domains in the field of Machine Learning [1]. It is the process of transforming unstructured textual content into structured data for simple analysis.

Table 2: Performance Comparison of ML Algorithms

|  |  |  |  |
| --- | --- | --- | --- |
| Classifiers | Precision | Recall | Accuracy |
| KNN | 71.42 | 66.34 | 90.23 |
| LR | 79.42 | 62.34 | 91.23 |
| SVM | 80.30 | 72.40 | 92.93 |
| XGB | 91.42 | 77.23 | 93.12 |

Text Mining is one of the recent research domains in the field of Machine Learning. It is the process of transforming unstructured textual content [3] into structured data for simple analysis.

**Algorithm 2 Basic Genetic Algorithm**

1. Generate random population Pr of n chromosomes.
2. Evaluate fitness function 𝑓𝑓(𝑥) for each chromosome 𝑥 in population 𝑃𝑟.
3. Select two parent chromosomes from the population Pr according to their fitness value.
4. Perform Crossover operation.
5. Mutate new offspring.
6. Place new offspring in the new population 𝑃𝑛𝑒𝑛𝑛
7. Determine the fitness value of 𝑃𝑛𝑒𝑛𝑛.
8. Repeat the steps 3-7 until the termination condition is satisfied.

***REFERENCES***

* 1. Guan, R., Zhang, H., Liang, Y., Giunchiglia, F., Huang, L. and Feng, X., 2020. Deep feature-based text clustering and its explanation. *IEEE Transactions on Knowledge and Data Engineering*, *34*(8), pp.3669-3680.
  2. Bouveyron, C., Celeux, G., Murphy, T.B. and Raftery, A.E., 2019. Model-based clustering and classification for data science: with applications in R (Vol. 50). Cambridge University Press.
  3. Zong, C., Xia, R. and Zhang, J., 2021. Text data mining (Vol. 711, p. 712). Singapore: Springer.

# \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

**\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\* Additional Information**

# Style & Format

1. Use A4 page size (8.27 x 11.69 in./21 x 29.7 cm) with Normal Margins (Top: 2.54 cm, Bottom: 2.54 cm, Left: 2.54 cm, Right: 2.54 cm) as in MS Word. All text including footnotes and illustrations must appear within this area.
2. No paragraph indent is required. Use 1.5 line spacing for all text.
3. Use Times New Roman Font for all text. Bold Face or italic text can be used for emphasis only.

|  |  |
| --- | --- |
| Report Title | 18 pt bold title case |
| Chapter Title | 18 pt bold italic title case |
| Section Heading | 16 pt bold title case |
| Subsection Heading | 14 pt bold normal |
| Body text | 12 pt normal, justified to both margins |
| Figure / Table Captions | 12 pt normal |
| Footnote text | 9 pt normal |

1. Each chapter must start on a new page. See the format of the chapter on page no. 1.
2. Avoid inline equations in the sentences. Provide proper number to each equation.
3. As far as possible figures (particularly block diagrams, simple line drawings, simple graphs) and all tables must be computer generated. Try to use your own images. Consult your guide about correct use of e-resources.
4. The maximum limit is 30 pages for the body of the report (i.e., excluding cover page, certificate page, abstract, contents and all other lists, acknowledgment, and annexure etc.). Reverse printing (both sides) is encouraged to save paper.
5. Only two copies of the report need to be submitted. Tape (Blue) binding with transparent plastic sheet is required.
6. **Plagiarism checking is mandatory. Consult your guide for instructions on plagiarism check according to the institute rule. Use iThenticate software to**

**compute similarity index. The plagiarism report must be attached to the summer internship report.**

1. **The marking scheme shall provide weightage to the presence of the following components of the report.**
   1. **Proper literature reviews.**
   2. **An articulate algorithm of flow of process (code snippets, algorithms, DFDs etc.).**
   3. **Proper reporting of the results obtained.**
   4. **Well annotated data set descriptions.**
   5. **Clear and precisely labelled figures.**
   6. **Topic specific references.**

**These guidelines are to be followed even if a student interns in a corporate organization.**

**\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\***

**\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\***