

(NLP) Natural Language Processing Project

Project Title:

Building a Text Classification Pipeline – Word Embedding Exploration

Description:

This project introduces participants to the practical implementation of **Natural Language Processing (NLP)** techniques for text classification.

Participants will design an end-to-end pipeline that transforms raw text into meaningful numerical representations using **word embedding**, followed by training classification models to predict text categories.

Project Overview:

This project focuses on building a complete **text classification pipeline** using Natural Language Processing (NLP) techniques.

You will explore how raw textual data is transformed into numerical representations and how **word embeddings** help models understand semantic relationships between words.

The project covers the full NLP workflow including:

- Text preprocessing
- Feature extraction
- Word embedding exploration
- Model training
- Performance evaluation

By the end of this project, participants will gain practical experience in implementing real-world NLP pipelines used in sentiment analysis, topic classification and document categorization.

Project Reference Link:

Note: Please carefully review the complete project details available at:

<https://lopsided-stinger-061.notion.site/Assignment-Building-a-Text-Classification-Pipeline-Word-Embedding-Exploration-1f6c7b58ffa98016bbe0fb955859ed99?pvs=143>

Evaluation Rubric: (100 Marks)

Area	Criteria	Marks
Understanding of Problem	Clear definition of objective and dataset	10
Text Preprocessing	Cleaning, tokenization, normalization	15
Feature Engineering	BoW and TF-IDF implementation	15
Word Embedding Exploration	Embeddings, visualization, explanation	20
Model Development	ML and DL model implementation	20
Model Evaluation	Metrics and comparison	10
Results Interpretation	Analysis and insights	5
Conclusion & Future Scope	Summary and improvements	5
Total		100

Project Deliverables

Participants are required to submit:

- **GitHub Repository (Compulsory)**

- Complete project code
- Clean and well-organized folder structure
- Code Notebook
- README file including:
 - Project overview
 - Dataset description
 - Steps to run the project
 - Results summary

- **Jupyter OR Google Colab Notebook (.ipynb)**

- Clean structure
- Well-commented code
- Proper markdown explanations

- **Dataset source link**

Submission Guidelines:

- Notebook must run without errors
- Code should be well implemented
- Proper explanations are mandatory
- Late submissions may be penalized

Learning Outcomes:

After completing this project, participants will be able to:

- Build NLP text classification pipelines
- Apply word embedding techniques

- Compare traditional and deep learning models
- Evaluate NLP models effectively