Reddit Comment Rule Violation Checker

Project Plan

"Building an NLP-based system to detect hate speech, spam, and harassment automatically on Reddit."

Group: 25GK23

Om Sanjeev Ronte (2302CS04)
Chaitanya Saagar (2301CS77)
Siddharth Anand (2301CS95)
Akshat Jha (2301AI53)
Aditya Prakash (2302CS11)

Final Submission: 20 October (Tentative)

Contents

1 Project Summary

Objective: The goal of this project is to develop an NLP-powered classification system capable of automatically identifying Reddit comments that violate community guidelines (e.g., hate speech, harassment, offensive language, or spam).

Approach:

The project follows a structured pipeline consisting of:

- **Data Preparation** – Collecting, cleaning, and annotating Reddit comments.

- **Preprocessing & Feature Engineering** – Applying NLP techniques such as tokenization, stopword removal, lemmatization, embeddings, and engineered linguistic features.

- **Modeling** – Experimenting with both classical ML models (e.g., Logistic Regression, SVM) and advanced deep learning architectures (e.g., RNNs, Transformers such as BERT).

- **Evaluation** – Using metrics such as accuracy, F1-score, precision/recall, and confusion matrices; complemented with error analysis and explainability methods.

- **Post-processing** – Incorporating rule-based checks to refine predictions and reduce false positives/negatives.

- **Documentation** – Preparing comprehensive reports and presentation slides highlighting methodology, findings, and results.

Deliverables:

- A clean, well-labeled dataset.

- A robust preprocessing and feature-engineering pipeline.

- Trained ML/DL models with comparative performance benchmarks.

- Evaluation results with detailed error analysis and interpretability insights.

- An integrated violation detection system ready for deployment.

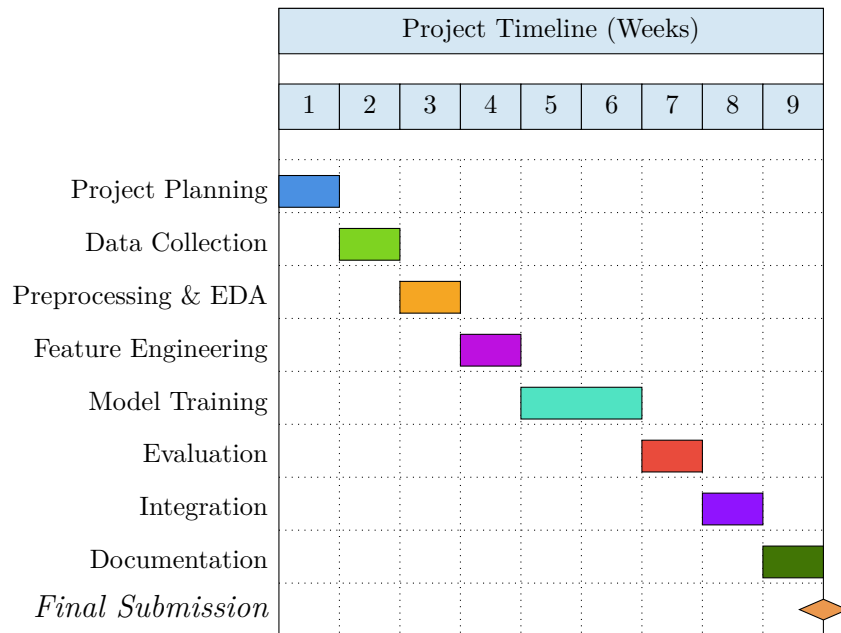- Final documentation and presentation materials.

2 Team Roles

Member	Role	Responsibilities
Om Sanjeev Ronte 2302CS04	HLD & Documentation Lead	Prepares system design, workflow diagrams, documentation, and manages coordination.
Chaitanya Saagar 2301CS77	Data Preprocessing & EDA Lead	Collects Reddit data, cleans, preprocesses, performs EDA, and creates data splits.
Siddharth Anand 2301CS95	Feature Engineering Lead	Handles embeddings (TF-IDF, BERT), feature selection, and builds baseline ML models.
Akshat Jha 2301AI53	Model Training Lead	Trains ML/DL models, performs hyperparameter tuning, and manages experiments.
Aditya Prakash 2302CS11	Evaluation & Post-Processing Lead	Defines metrics, conducts error analysis, handles rule-based layers, and explainability.

3   Project Timeline

Week	Timeline	Task	Responsible
Week 1	Aug 24 – Aug 31	Project planning, literature review, setup GitHub & tools	All
Week 2	Sep 1 – Sep 7	Data collection & annotation plan	2301CS77
Week 3	Sep 8 – Sep 14	Preprocessing pipeline & EDA report	2301CS77
Week 4	Sep 15 – Sep 21	Feature engineering & baseline ML models	2301CS95
Week 5	Sep 22 – Sep 28	Deep learning model selection & initial training	2301AI53
Week 6	Sep 29 – Oct 5	Model tuning & validation	2301AI53
Week 7	Oct 6 – Oct 12	Evaluation, error analysis, rule-based layer	2302CS04, 2302CS11
Week 8	Oct 13 – Oct 17	Integration & testing	2301CS95, 2301AI53, 2302CS11
Week 9	Oct 18 – Oct 19	Documentation & presentation prep	2302CS04
Final	Oct 20	Final Submission	All

4 Gantt Chart

Project Timeline (Weeks)								
1	2	3	4	5	6	7	8	9

Project Planning
Data Collection
Preprocessing & EDA
Feature Engineering
Model Training
Evaluation
Integration
Documentation
Final Submission

Week Mapping:

- Week 1: Aug 24 – Aug 31 (Project Planning)

- Week 2: Sep 1 – Sep 7 (Data Collection)

- Week 3: Sep 8 – Sep 14 (Preprocessing & EDA)

- Week 4: Sep 15 – Sep 21 (Feature Engineering)

- Week 5: Sep 22 – Sep 28 (Model Training - Phase 1)

- Week 6: Sep 29 – Oct 5 (Model Training - Phase 2)

- Week 7: Oct 6 – Oct 12 (Evaluation)

- Week 8: Oct 13 – Oct 17 (Integration)

- Week 9: Oct 18 – Oct 20 (Documentation & Final Submission)