

# Jigsaw - Agile Community Rules Classification: Detailed Task Tracker

Team of Five

July 26, 2025

## Detailed Task Tracker for Jigsaw Competition

This tracker outlines all tasks for the Jigsaw - Agile Community Rules Classification competition, covering Data Exploration, Preprocessing, Model Development, Evaluation, and Submission phases. Update the tracker weekly during team meetings to track progress, mark status (Not Started, In Progress, Completed, Blocked), and add notes for blockers or dependencies. Implement this tracker in a digital tool (e.g., Google Sheets, Notion, Trello) for real-time collaboration, and compile this LaTeX document for a PDF reference.

Table 1: Detailed Task Tracker

Task ID	Phase	Task Description	Assignee	Start Date	Deadline	Status
DE-01	Data Exploration	Download and inspect train.csv, test.csv, sample_submission.csv	Member 1	07/26/2025	07/28/2025	Not Started
DE-02	Data Exploration	Analyze distribution of rule_violation labels	Member 1	07/28/2025	07/30/2025	Not Started
DE-03	Data Exploration	Visualize subreddit and rule distributions	Member 1	07/28/2025	07/30/2025	Not Started
DE-04	Data Exploration	Analyze rule-specific patterns in comments	Member 2	07/28/2025	07/30/2025	Not Started
DE-05	Data Exploration	Generate word clouds for comment text	Member 2	07/29/2025	07/31/2025	Not Started
DE-06	Data Exploration	Check for data imbalances and missing values	Member 1	07/30/2025	08/01/2025	Not Started
DE-07	Data Exploration	Write EDA report summarizing findings	Member 1, 2	07/31/2025	08/02/2025	Not Started

Table 1: Detailed Task Tracker

Task ID	Phase	Task Description	Assignee	Start Date	Deadline	Status
PR-01	Preprocessing	Clean comment text (remove URLs, special characters)	Member 3	08/03/2025	08/05/2025	Not Started
PR-02	Preprocessing	Apply tokenization, lowercasing, stop-word removal	Member 3	08/04/2025	08/06/2025	Not Started
PR-03	Preprocessing	Apply lemmatization to comment text	Member 3	08/05/2025	08/07/2025	Not Started
PR-04	Preprocessing	Create one-hot encodings for subreddit	Member 4	08/03/2025	08/06/2025	Not Started
PR-05	Preprocessing	Develop rule embeddings using BERT	Member 4	08/04/2025	08/07/2025	Not Started
PR-06	Preprocessing	Incorporate positive/negative examples as features	Member 3	08/06/2025	08/08/2025	Not Started
PR-07	Preprocessing	Split train.csv into 80/20 train/validation sets	Member 3	08/07/2025	08/09/2025	Not Started
PR-08	Preprocessing	Create preprocessing pipeline in Kaggle Notebook	Member 3, 4	08/08/2025	08/10/2025	Not Started
MD-01	Model Development	Implement Logistic Regression with TF-IDF	Member 1	08/11/2025	08/15/2025	Not Started
MD-02	Model Development	Implement Naive Bayes with Bag-of-Words	Member 1	08/11/2025	08/15/2025	Not Started
MD-03	Model Development	Fine-tune BERT model on train.csv	Member 2	08/11/2025	08/20/2025	Not Started
MD-04	Model Development	Fine-tune RoBERTa model on train.csv	Member 2	08/15/2025	08/25/2025	Not Started
MD-05	Model Development	Implement DistilBERT for efficiency	Member 2	08/20/2025	08/28/2025	Not Started
MD-06	Model Development	Develop prototypical networks for few-shot learning	Member 3	08/15/2025	08/25/2025	Not Started
MD-07	Model Development	Implement MAML for unseen rules	Member 3	08/20/2025	09/01/2025	Not Started

Table 1: Detailed Task Tracker

Task ID	Phase	Task Description	Assignee	Start Date	Deadline	Status
MD-08	Model Development	Build ensemble model (BERT, RoBERTa, Logistic)	Member 4	08/25/2025	09/05/2025	Not Started
MD-09	Model Development	Optimize hyperparameters (grid/random search)	Member 5	08/25/2025	09/10/2025	Not Started
MD-10	Model Development	Validate models on local validation set	Member 5	09/01/2025	09/15/2025	Not Started
EV-01	Evaluation	Submit initial predictions to public leaderboard	Member 5	09/16/2025	09/18/2025	Not Started
EV-02	Evaluation	Analyze public leaderboard feedback	Member 1, 2	09/18/2025	09/20/2025	Not Started
EV-03	Evaluation	Conduct error analysis (false positives/negatives)	Member 1, 2	09/20/2025	09/25/2025	Not Started
EV-04	Evaluation	Fine-tune BERT/RoBERTa based on feedback	Member 2	09/20/2025	09/30/2025	Not Started
EV-05	Evaluation	Fine-tune ensemble weights	Member 4	09/25/2025	10/01/2025	Not Started
EV-06	Evaluation	Test robustness on synthetic unseen rules	Member 3, 4	09/25/2025	10/05/2025	Not Started
EV-07	Evaluation	Submit optimized predictions to leaderboard	Member 5	10/01/2025	10/10/2025	Not Started
FS-01	Final Submission	Generate predictions on test.csv	Member 5	10/11/2025	10/15/2025	Not Started
FS-02	Final Submission	Format predictions as submission.csv	Member 5	10/15/2025	10/17/2025	Not Started
FS-03	Final Submission	Verify submission file content	All Members	10/17/2025	10/19/2025	Not Started
FS-04	Final Submission	Submit final predictions to Kaggle	Member 5	10/19/2025	10/21/2025	Not Started
FS-05	Final Submission	Document final approach and results	Member 5	10/15/2025	10/23/2025	Not Started

## Instructions for Use

- **Update Weekly:** During team meetings (e.g., every Monday), update the Status and Notes columns.
- **Status Options:** Not Started, In Progress, Completed, Blocked.
- **Notes:** Log blockers, dependencies, or progress details (e.g., "Awaiting Kaggle API setup").
- **Digital Implementation:** Import this table into Google Sheets, Notion, or Trello for real-time updates. Use conditional formatting (e.g., red for overdue tasks) and set up notifications for deadlines.
- **GitHub Integration:** Link tasks to GitHub commits or issues for code-related tasks (e.g., MD-01 to MD-10).
- **PDF Reference:** Compile this LaTeX document using PDFLaTeX (e.g., via Overleaf) for a downloadable PDF.

## Collaboration Tools

- **Google Sheets:** Create a spreadsheet with this table, share with edit access, and use filters for phase/assignee views.
- **Notion:** Use a table or Kanban board, assign tasks, and embed links to Kaggle Notebooks or GitHub.
- **Trello:** Create lists for each phase, add task cards with assignees and due dates, and use labels for status.
- **GitHub Projects:** Integrate with your repository, create a project board, and link tasks to code commits.
- **Slack/Discord:** Set up a channel for daily updates and reminders.

## Tips to Stay on Track

- Schedule 30-minute weekly meetings to review the tracker and resolve blockers.
- Assign a backup team member for critical tasks in case of delays.
- Use automation (e.g., Google Apps Script, Notion reminders) to alert for upcoming deadlines.
- Celebrate phase completions (e.g., Data Exploration by 08/02/2025) to maintain team morale.
- Monitor Kaggle Notebook runtime limits (12 hours CPU/GPU) and optimize tasks (e.g., use DistilBERT for efficiency).