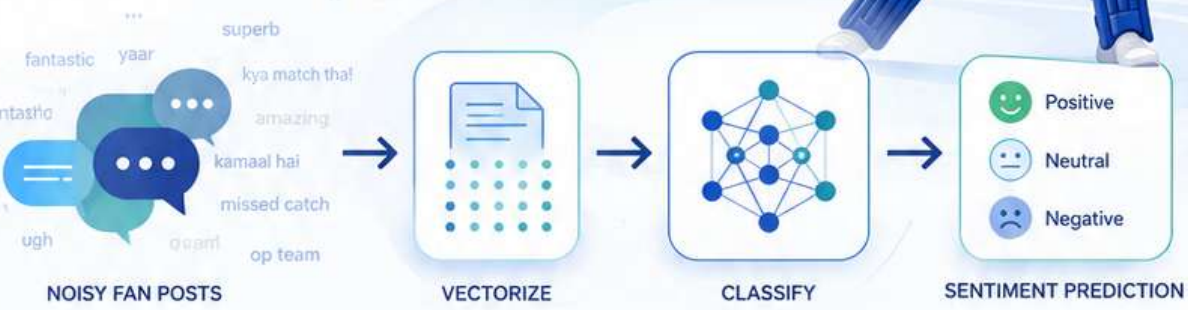
SAAI Research · Applied ML Sprint

IPL Sentiment Pipeline Submission Challenge

Official Submission Brochure



1 Challenge Statement

Build a machine learning pipeline that predicts sentiment toward an IPL team keyword from noisy IPL fan posts. Each row contains a tweet or statement, a subject keyword, and a sentiment label. Your task is to explore vectorizers and classifiers, compare pipeline combinations, select final components, and justify your choices.

2 What Students Are Building

Vectorizer

Converts text into numerical features.

Examples: CountVectorizer, TF-IDF, character TF-IDF, HashingVectorizer.

Classifier

Uses features to predict sentiment.

Examples: Naive Bayes, Logistic Regression, LinearSVC, SGDClassifier.

Justification

Explain why the chosen components fit noisy Indian English, Hinglish, typos, and IPL fan language.

3 Submission Snapshot

Task: IPL team sentiment analysis

Input: subject_keyword + "[SEP]" + tweet

Output: positive / negative / neutral

Submit: GitHub repository URL

Must include: Notebook, JSON, README, results, pkl if created

4 Official Resources

Dataset Repo:
github.com/HimanshuKhale/ipl_sentiment

Colab Notebook:
colab.research.google.com/drive/1ppNtfC5jKyLTrWz-Fhp9HHcvTPnOTdDV

Final submission is the GitHub repository URL containing all required artifacts.

SAAI Research · Official Submission Rules

Submission Rules & GitHub Format

1 Required Rules

- Upload all artifacts to GitHub.
- Share only the GitHub repository URL as the final submission.
- Notebook must run from top to bottom.
- Keep the subject keyword in model input.
- Include explanations and justifications.

2 Disallowed Practices

- No screenshots-only submission.
- No missing README.
- No unexplained pickle files.
- No train-test leakage.
- No heavy-model spam without justification.

3 GitHub Repository Submission Format

Create one repository for your challenge submission. Put every artifact inside this repo and share the repo URL. The repository must be accessible to evaluators. Public is easiest; if private, grant evaluator access before the deadline.

4 Recommended Repository Name

```
ipl-sentiment-pipeline-challenge-<your-name-or-team>
```

5 Recommended Folder Structure

```
ipl-sentiment-pipeline-challenge/
├── README.md
├── notebooks/
│   └── completed_pipeline_selector_notebook.ipynb
├── manifests/
│   └── submission_manifest.json
├── results/
│   ├── results_80R.csv
│   └── results_300R.csv
├── models/
│   ├── best_model.pkl
│   └── model_card.md
├── explanations/
│   ├── vectorizer_classifier_justification.md
│   ├── error_analysis.md
│   └── final_reflection.md
├── requirements.txt
└── sample_predictions.md
```

Every required artifact should live inside one GitHub repository.

Final submission = the repository URL.

SAAI Research · Submission Essentials

Artifacts, Evaluation & Final Checklist

1 Required Artifacts

Required Artifacts	Required?	Purpose
README.md	Required	Explains project, task, dataset, approach, how to run, and conclusion.
Completed notebook	Required	Shows the full experiment and must run top to bottom.
submission_manifest.json	Required	Structured summary of selected vectorizers, classifiers, parameters, and hypothesis.
results_80R.csv	Required	Results from the starter 80-row experiment.
results_300R.csv	If used / recommended	Results from the 300-row round.
.pkl model files	If created / recommended	Saved best model or demo pipeline.
model_card.md	If pkl submitted	Explains what the pkl contains, input format, metrics, and limitations.
justification markdown	Required	Explains why selected vectorizers and classifiers were chosen.
error_analysis.md	Recommended	Shows where the model made mistakes and what was learned.
requirements.txt	Required	Lists libraries needed to reproduce the notebook.
sample_predictions.md	Recommended	Shows sample inputs and predictions.

2 Evaluation Parameters

Model performance — 40%
Accuracy, Macro F1, Weighted F1, precision, recall, confusion matrix.

Efficiency — 15%
Training time, prediction time, file size, local-compute suitability.

Diversity — 15%
Variety of vectorizer families and classifier families.

Simplicity and suitability — 10%
Appropriate model choice for noisy IPL text.

Generalization — 10%
Works on larger or unseen datasets.

Explanation quality — 10%
README, justification, error analysis, final reflection.

3 Final Submission Checklist

- [x] Completed notebook included
- [x] submission_manifest.json included
- [x] results_80R.csv included
- [x] results_300R.csv included if used
- [x] README included
- [x] requirements.txt included
- [x] pkl files documented if submitted
- [x] explanations and justifications included
- [x] Repository is accessible to evaluators

Final submission:
Share the GitHub repository URL containing all required artifacts.

https://github.com/your-username/ipl-sentiment

You are not only submitting a model — you are submitting a machine learning strategy.