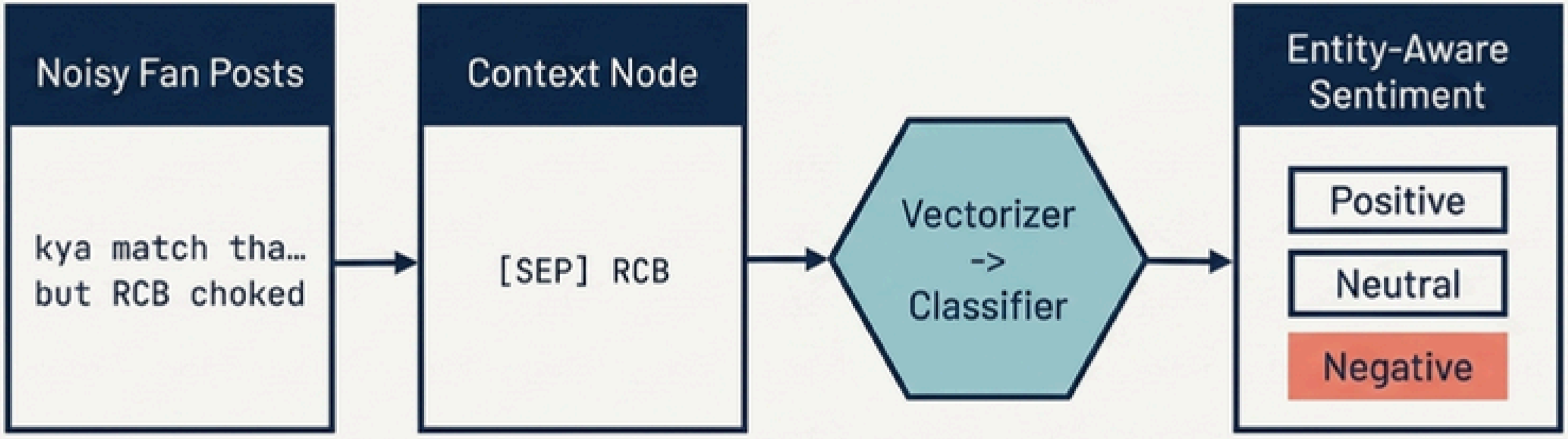
IPL Sentiment Pipeline Challenge

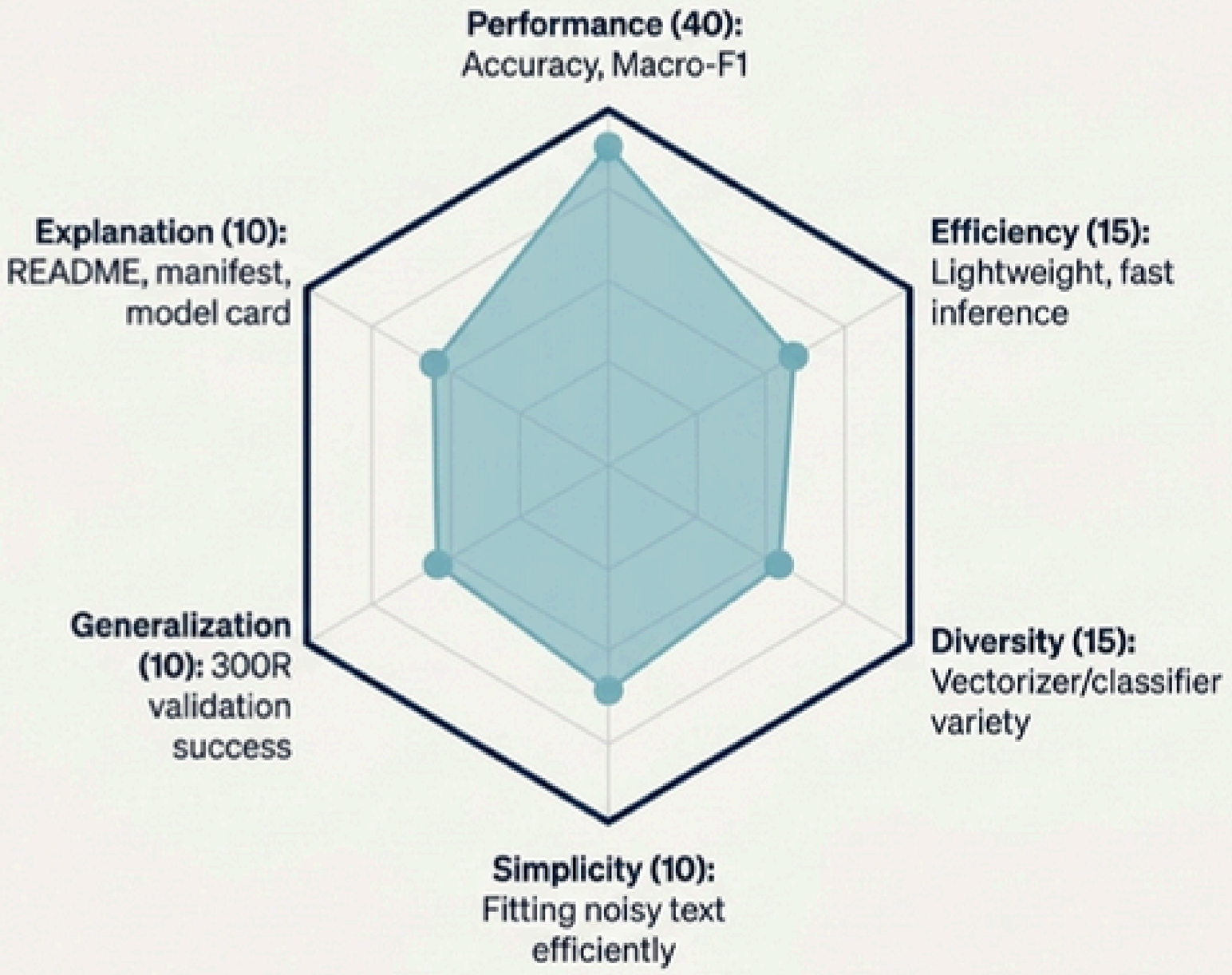
Submission Analysis & Industry Feedback

Cohort Audit // SAAI Applied ML Sprint

Entity-Aware Sentiment Analysis Pipeline



The Twist: Not just general sentiment. Models had to predict sentiment toward a specific IPL team, handling sarcasm, Hinglish, and entity confusion.



The Industry Standard MVP

An industry MVP requires balance. A heavy model with high accuracy but poor efficiency and missing explanations fails deployment.

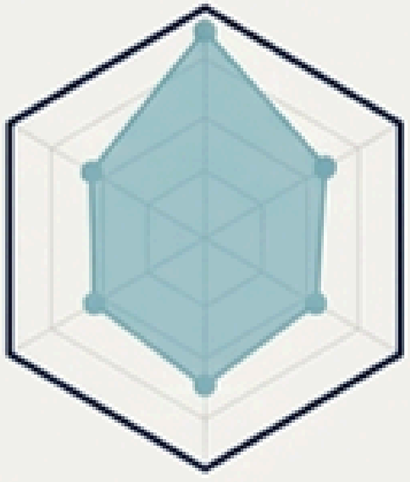
Total Cohort Evaluated: 11 Pipelines

RANK	PARTICIPANT	OVERALL SCORE	TITLE
1	Rehan Shaikh	92.0	Champion
2	Siddhant Agarwal	90.5	Runner-Up
3	Ayaan Shaikh	88.5	2nd Runner-Up
4	Sanvi A Reddy	84.5	
5	Rehaan Jain	82.0	
6	Aahan R. Lulla	70.0	
7	Rubaika Zaidi	66.0	
8	Sanvi	60.0	
9	Vanya Gupta	56.0	
10	Sagnik Sen	51.0	
11	Varchasv Biyani	27.0	

Rank 1 / Champion Profile

Rehan Shaikh

92 / 100



Pipeline: TF-IDF Char N-gram + Complement NB

Best Aspect: Practical Performance. Best 80R leaderboard result. Balanced simple models with high efficiency.

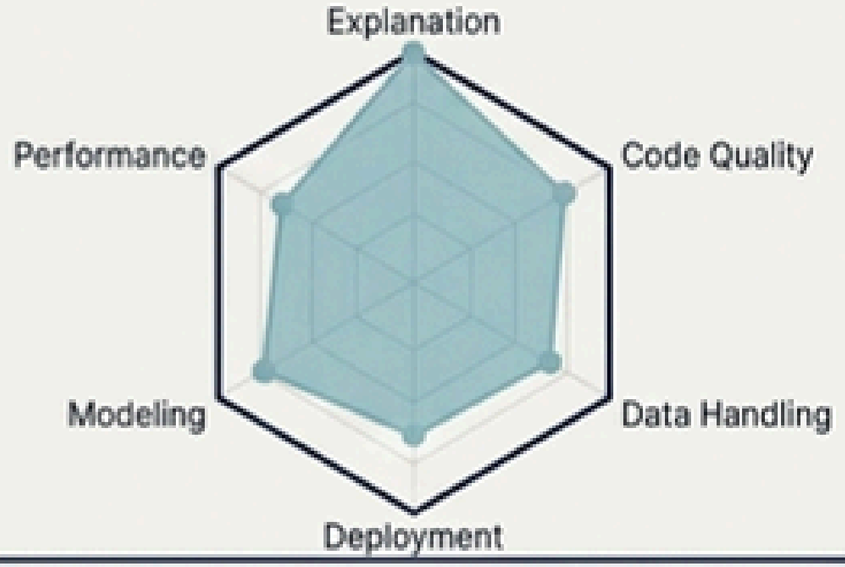
Integration Value: High. Fast, deployable, handles probability estimates well.

Needs Work: Architecture. Separation of config/inference logic from Jupyter notebooks is required.

Rank 2 / Spotlight Profile

Siddhant Agarwal

90.5/100



Pipeline: CountVectorizer (Binary) + Logistic Regression

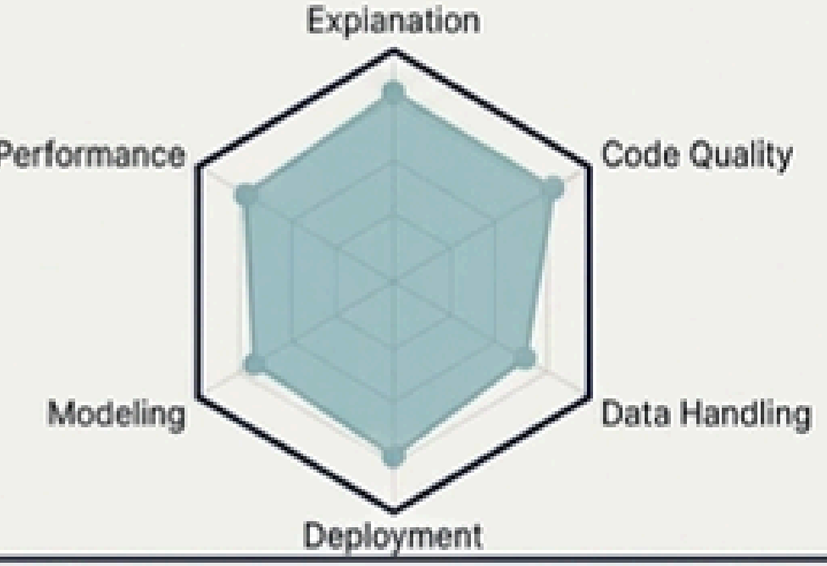
Best Aspect: Research & Engineering Discipline. Best repository structure and exceptional ML reasoning.

Highlight: Deep Understanding. Near-perfect 300R generalization; brilliant justification for avoiding heavy models.

Industry Gap: Packaging. Still relies on notebooks. Needs a standalone Python package for true deployment.

Ayaan Shaikh

88.5/100



Pipeline: TF-IDF Word 1-2 Grams + Logistic Regression

Best Aspect: Deployment Clarity. Submitted an industry-standard Model Card explaining intended use, limitations, and ethics.

Highlight: Clean 80R Submission. Highly efficient model choice, easy to serialize and integrate.

Critical Miss: Validation Gap. Failed to submit 300R validation results, leaving generalization untested.

The Challengers (Ranks 4-7)

Sanvi A Reddy | Score: 84.5

✓ Pro: Clean standardization & sensible classical ML comparisons.

✗ Con: Missing 300R generalization & weak error analysis.

Rehaan Jain | Score: 82.0

✓ Pro: Scalability thinking (HashingVectorizer) & strong raw scores.

✗ Con: Reliability issues; failed models silently left in output.

Aahan Ritesh Lulla | Score: 70.0

✓ Pro: Broad conceptual coverage & excellent selection criteria.

✗ Con: Critical artifacts deleted/missing from repository history.

Rubaika Zaidi | Score: 66.0

✓ Pro: Clear reporting of evaluated metrics.

✗ Con: Limited model diversity & missing full manifest.

The Challengers (Ranks 8-11)

Sanvi | Score: 60.0

- ✓ Pro: Excellent feature engineering ideas (emojis, cricket slang).
- ✗ Con: Metrics left as placeholders; incomplete final evidence.

Vanya Gupta | Score: 56.0

- ✓ Pro: High ambition and NLP curiosity.
- ✗ Con: Heavy-model spam (BERT/XGBoost) inappropriate for lightweight sprint.

Sagnik Sen | Score: 51.0

- ✓ Pro: Good awareness of Hinglish/sarcasm challenges.
- ✗ Con: Unverified claims with no accessible leaderboard evidence.

Varchasv Biyani | Score: 27.0

- ✓ Pro: Attempted workflow despite tooling limitations.
- ✗ Con: Format non-compliance (Word doc links instead of GitHub repo).

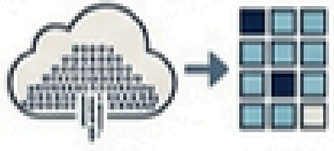
Cohort Diagnostic Matrix

The Good: Cohort Successes

- ✔ Classical NLP Mastery: Strong adoption of lightweight models (Naive Bayes, Logistic Regression) suited for sparse text.

- ✔ Hinglish Handling: Smart use of Character N-grams to bypass spelling variations.

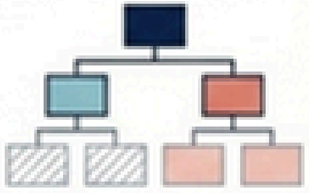
- ✔ Scalability Awareness: Experimentation with HashingVectorizers for Twitter-scale data.



The Gaps: Systemic Weaknesses

- ✘ 'Runs on my Colab' Syndrome: Heavy reliance on notebooks instead of deployable scripts.

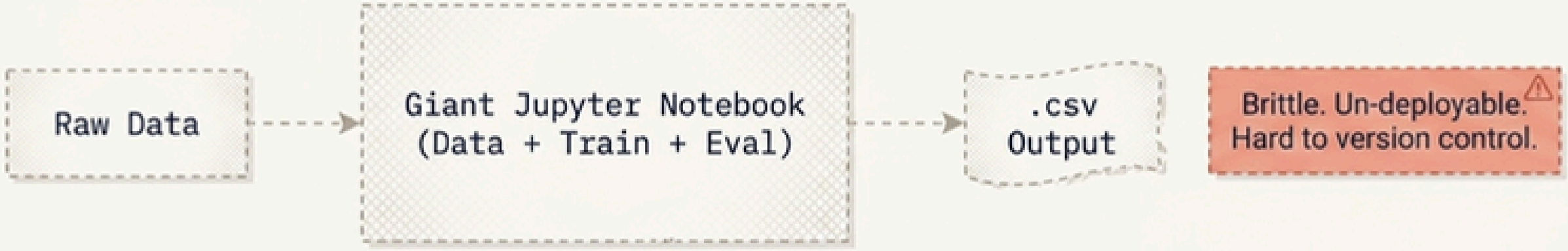
- ✘ Missing Confidence Scores: Outputs gave hard labels without probabilities.



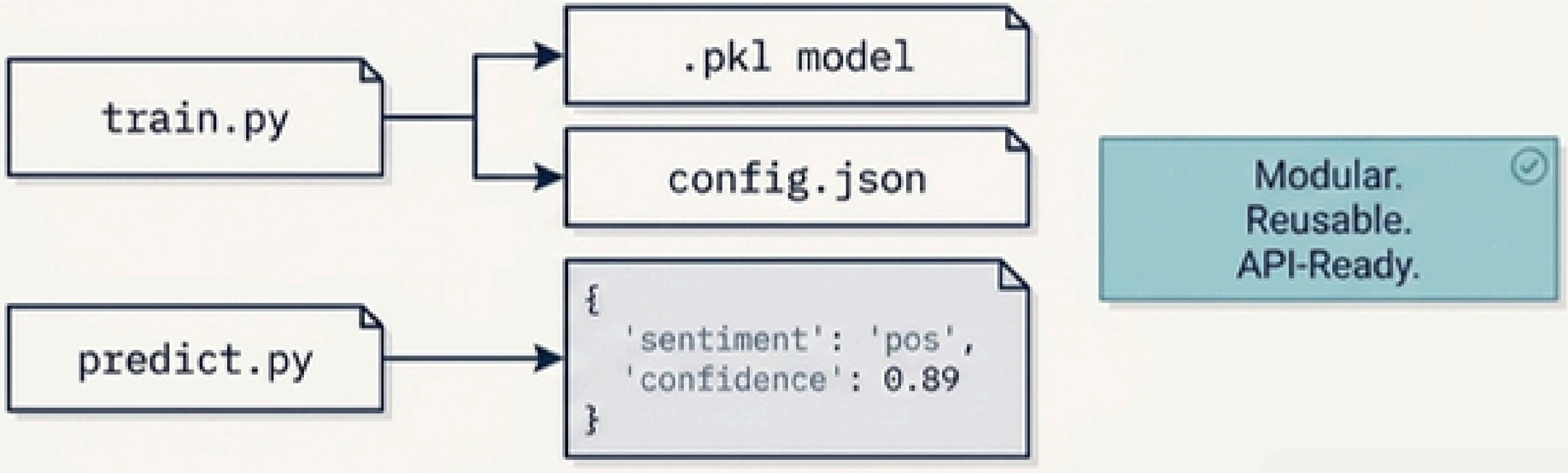
- ✘ Ghost Artifacts: Deleted histories, placeholder text, text, and unverified result files.

Pipeline Evolution: From Academy to Industry

Academy Habit



Industry Standard



Modular Architecture: The ‘twitter_sentiment_analyzer pipeline’

The Engine

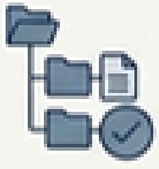
Source: Rehan Shaikh

Complement Naive Bayes + Char N-grams. Best practical accuracy and inference speed.

The Chassis

Source: Siddhant Agarwal

Repository structure, manifest, and validation logic. Industry-style organization.



The Manual

Source: Ayaan Shaikh

Comprehensive Model Card format. Clear ethical guidelines and intended use cases.

`twitter_sentiment_analyzer pipeline`

The Next Innings

☐	No standalone notebook submissions; code must be modularized.
☐	Must include a standalone `predict.py` inference script.
☐	Pipeline must return sentiment confidence scores, not just hard labels.
☐	Repository history must remain stable (no deleted evaluation artifacts).

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