

Veridion Challenge — One-page Report

Generated: 2025-10-19T15:52

Summary (auto_evaluation_summary.txt):

Rows: 9492
Coverage: 100.00%
Avg labels/company: 3.00
Scores — mean: 0.794, median: 0.780, std: 0.112

Top labels (by count):

Label	Count	Avg Score
Agricultural Equipment Services	726	0.7941955922865014
Food Processing Services	636	0.7979795597484276
Accessory Manufacturing	485	0.7574680412371134
Sheet Metal Services	420	0.8244190476190477
Travel Services	396	0.825540404040404
Plastic Manufacturing	395	0.8003189873417721
Furniture Manufacturing	388	0.8252242268041238
Confectionery Manufacturing	379	0.7631240105540896

Notes & Next Steps:

- 1) Prune noisy, high-count labels with low avg_score.
- 2) Choose an operational threshold from topk_threshold_candidates.csv (95%/99% policy).
- 3) Keep small human-labeled gold set (~200-300 rows) for precision@k validation and calibrator training.
- 4) Publish only scripts, README and final annotated CSV to GitHub; remove large embeddings and caches.