Athulya Saravanavakumar

Multi-Modal Equity Risk Scoring for Schools

Introduction and project statement

School quality and educational equity vary widely across districts, shaped by unequal funding, aging infrastructure, and socioeconomic disparities. Traditional evaluation (manual inspections, audits, self reported reporting) is slow, inconsistent, and can leave underserved campuses invisible for years. Multi-Modal Education Equity Monitor, targets this gap by building an ML system that combines tabular institutional signals, unstructured text, and (in the broader project scope) imagery, to produce a risk-oriented assessment that can help policymakers and districts prioritize follow-up.

At a system level, this is an equity-focused risk screening: a model should surface "which schools look elevated risk" and provide evidence for why, rather than pretending to be a final, authoritative ranking. This framing is explicitly aligned with fusion report's emphasis that results are evidence multimodal fusion is promising when scaled, not a final deployable system.

Data sources and technologies used

Ground-truth labels (supervision)

Supervised targets come from district-generated campus condition metrics:

- ESA (0–100): how well educational spaces support learning programs
- FCA (0–100): physical condition of the facility

Map continuous scores into ordered class labels (ESA_class, FCA_class) for classification-style modeling.

Tabular data (structured)

Multiple NCES tables (directory, staffing, lunch program eligibility, school characteristics), joined by the unique school identifier NCESSCH (explicitly cast to string to prevent merge failures).

FIPST	STATENAME	ST	SCH_NAME	STATE_AGENCY_NO	UNION	ST_LEAID	LEAID	ST_SCHID	NCESSCH	SCHID	SHARED_TIME	NSLP_STATUS	NSLP_STATUS_TEXT	VIRTUAL	VI
1	ALABAMA	AL	Albertville Middle School	1		AL-101	100005	AL-101-0010	100050087O	100870	No	NSLPCEO	Yes under Community Eligibility Option (CEO)	NOTVIRTUAL	N
1	ALABAMA	AL	Albertville High School	1		AL-101	100005	AL-101-0020	100050087I	100871	No	NSLPCEO	Yes under Community Eligibility Option (CEO)	NOTVIRTUAL	N
1	ALABAMA	AL	Albertville Intermediate School	1		AL-101	100005	AL-101-0110	100050087S	100879	No	NSLPCEO	Yes under Community Eligibility Option (CEO)	NOTVIRTUAL	N
1	ALABAMA	AL	Albertville Elementary School	1		AL-101	100005	AL-101-0200	100050088P	100889	No	NSLPCEO	Yes under Community Eligibility Option (CEO)	NOTVIRTUAL	N
1	ALABAMA	AL	Albertville Kindergarten and PreK	1		AL-101	100005	AL-101-0035	100050161E	101616	No	NSLPCEO	Yes under Community Eligibility Option (CEO)	NOTVIRTUAL	N
1	ALABAMA	AL	Albertville Primary School	1		AL-101	100005	AL-101-0005	100050215O	102150	No	NSLPCEO	Yes under Community Eligibility Option (CEO)	NOTVIRTUAL	N
1	ALABAMA	AL	Kate Duncan Smith DAR Middle	1		AL-048	100006	AL-048-0143	100060019S	100193	No	NSLPCEO	Yes under Community Eligibility Option (CEO)	NOTVIRTUAL	N
1	ALABAMA	AL	Asbury High School	1		AL-048	100006	AL-048-0030	100060087Z	100872	No	NSLPCEO	Yes under Community Eligibility Option (CEO)	NOTVIRTUAL	N
1	ALABAMA	AL	Claysville School	1		AL-048	100006	AL-048-0070	100060087G	100876	No	NSLPCEO	Yes under Community Eligibility Option (CEO)	NOTVIRTUAL	N
1	ALABAMA	AL	Douglas Elementary School	1		AL-048	100006	AL-048-0090	100060087T	100877	No	NSLPCEO	Yes under Community Eligibility Option (CEO)	NOTVIRTUAL	N
1	ALABAMA	AL	Douglas High School	1		AL-048	100006	AL-048-0100	100060087H	100878	No	NSLPCEO	Yes under Community Eligibility Option (CEO)	NOTVIRTUAL	N
1	ALABAMA	AL	Brindlee Mountain Elementary School	1		AL-048	100006	AL-048-0120	100060088O	100880	No	NSLPCEO	Yes under Community Eligibility Option (CEO)	NOTVIRTUAL	N
1	ALABAMA	AL	Kate D Smith DAR High School	1		AL-048	100006	AL-048-0140	100060088J	100883	No	NSLPCEO	Yes under Community Eligibility Option (CEO)	NOTVIRTUAL	N
1	ALABAMA	AL	Brindlee Mountain Primary School	1		AL-048	100006	AL-048-0180	100060088T	100887	No	NSLPCEO	Yes under Community Eligibility Option (CEO)	NOTVIRTUAL	N
1	ALABAMA	AL	Marshall Alternative School	1		AL-048	100006	AL-048-0150	100060098G	100986	No	NSLPCEO	Yes under Community Eligibility Option (CEO)	NOTVIRTUAL	N
1	ALABAMA	AL	Marshall Technical School	1		AL-048	100006	AL-048-0160	100060098T	100987	Yes	NSLPWOPRO	Yes participating without using any Provision or the CEO	NOTVIRTUAL	N
1	ALABAMA	AL	Robert D Sloman Primary	1		AL-048	100006	AL-048-0095	100060141S	101413	No	NSLPCEO	Yes under Community Eligibility Option (CEO)	NOTVIRTUAL	N
1	ALABAMA	AL	Brindlee Mountain High School	1		AL-048	100006	AL-048-0042	100060158S	101585	No	NSLPCEO	Yes under Community Eligibility Option (CEO)	NOTVIRTUAL	N

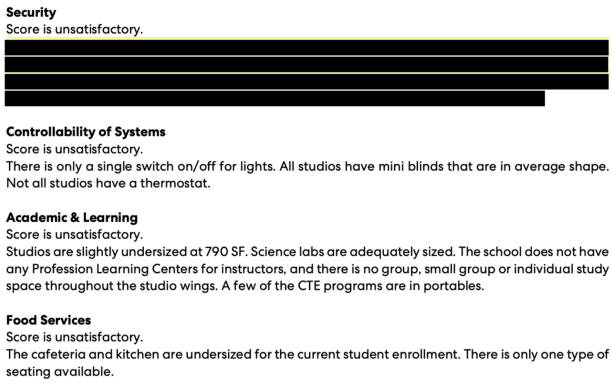
Athulya Saravanavakumar

Key decisions:

- Directory table as the “spine” (school metadata), staffing used to extract teacher counts.
- Lunch eligibility excluded due to high missingness/unreliability in subset.
- Columns trimmed to reduce noise and improve merge consistency.

Text data (unstructured)

Text is collected from public school documents/webpages (mission statements, academic programs, PDFs), associated per school with ESA/FCA labels.

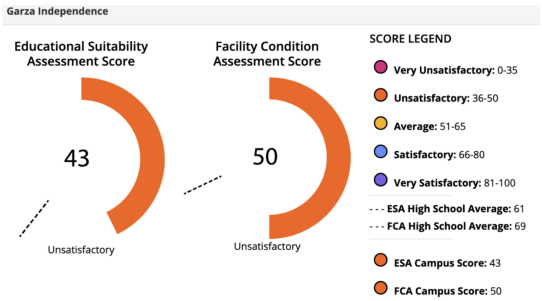


Security
Score is unsatisfactory.
[redacted]

Controllability of Systems
Score is unsatisfactory.
There is only a single switch on/off for lights. All studios have mini blinds that are in average shape. Not all studios have a thermostat.

Academic & Learning
Score is unsatisfactory.
Studios are slightly undersized at 790 SF. Science labs are adequately sized. The school does not have any Profession Learning Centers for instructors, and there is no group, small group or individual study space throughout the studio wings. A few of the CTE programs are in portables.

Food Services
Score is unsatisfactory.
The cafeteria and kitchen are undersized for the current student enrollment. There is only one type of seating available.



Fusion dataset (multimodal alignment)

Fusion aligns modalities at the school level. In practice, fusion in current experiments integrates:

- Tabular NCES features
- Text-derived risk indicators (predictions saved out by the text model, then merged into fusion)

Tools / libraries

- scikit-learn pipelines, ColumnTransformer, RandomForest, cross-validation utilities
- PyTorch + HuggingFace Transformers for the text model (DistilBERT multitask classifier)
- (Planned / broader scope) PyTorch/torchvision for vision, Mapillary/imagery retrieval, GIS processing referenced in proposal materials

Athulya Saravanavakumar

Methods

1) Text modality: multitask DistilBERT classifier (ESA_class + FCA_class)

Why classification (ordinal classes) instead of predicting raw scores?

Discretize ESA/FCA scores into ordered bins to make outputs policy-readable and easier to evaluate with class metrics.

Why multitask?

ESA and FCA are related but not identical; sharing an encoder acts as regularization and helps generalization while keeping task-specific heads.

Why DistilBERT?

Chosen for a strong performance-to-efficiency tradeoff, especially under constrained compute.

Core preprocessing choices

- Tokenization with DistilBERT tokenizer
- Max sequence length 256 (context vs efficiency tradeoff)
- Padding/truncation for consistent batch shapes

Evaluation

Evaluate using accuracy + precision/recall + macro/weighted F1, and produce confusion matrices for ESA and FCA.

Design-for-fusion decision

Text predictions are exported as a standalone CSV so fusion can be iterated without retraining the language model each time (modular debugging + reuse).

2) Tabular modality: Random Forest baseline (tabular-only)

Data processing

- Merge NCES tables via NCESSCH (string cast for stable joins).
- Keep only modeling-relevant columns to reduce noise.
- Drop rows missing critical values to preserve integrity.

Feature engineering

- TEACHERS
- TEACHERS_PER_ZIP (resource-density proxy; coarse)

Athulya Saravanavakumar

Why Random Forest?
Explicitly choose RF for tabular work because it:

- Handles nonlinearities well
- Doesn't require feature scaling
- Performs reliably on small-to-medium tabular datasets

Initial task
In the tabular report, framed the target as continuous ESA_Score and used RF regression with cross-validation (MAE/RMSE).

3) Multimodal fusion: early fusion, classification reformulation

Fusion strategy
Early fusion (concatenate tabular + text-derived signals, then train one downstream classifier). The explicit reason: with tiny data, early fusion avoids the instability and parameter growth of deeper multimodal architectures.

Why switch fusion from regression to classification?
Because only a small subset had complete multimodal coverage (n = 6), regression was unstable and split-sensitive. Reformulated fusion as binary classification (low risk vs elevated risk) to reduce variance and match a realistic screening use case.

Why Random Forest for fusion?
Selected for:

- Mixed feature handling
- Nonlinear interactions across modalities
- Robustness under small datasets without heavy tuning

Cross-validation design choice
Cap stratified CV folds by the minimum class count to avoid invalid splits when classes are tiny.

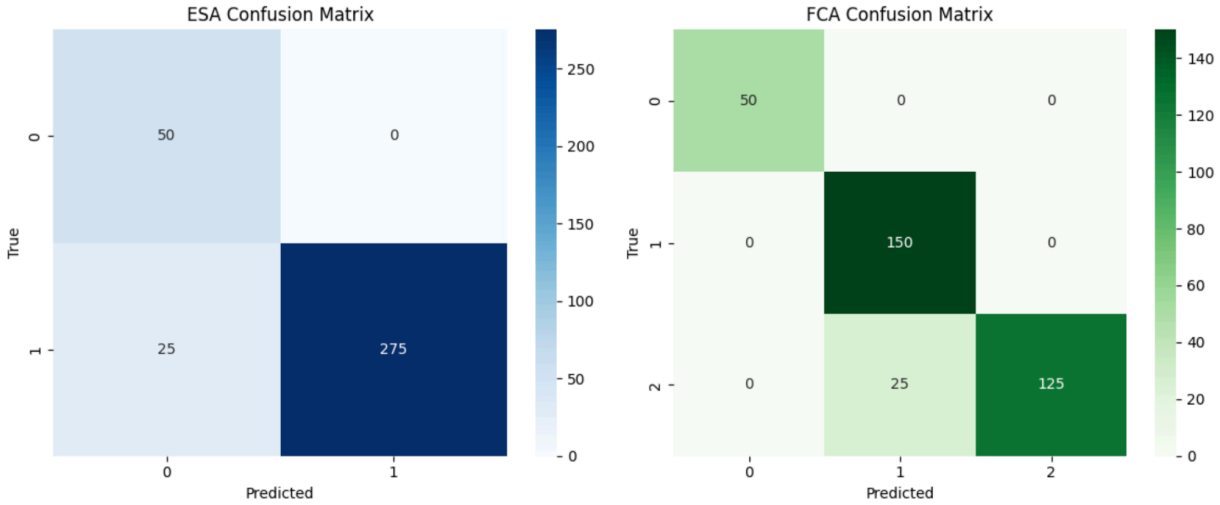
Baselines
Compare against a majority-class DummyClassifier to ensure the model beats "always predict the most common label."

Results

Text model (ESA + FCA multitask)

On a held-out test set, the text model achieves strong performance:

- ESA: Accuracy ~0.88, Macro F1 ~0.88, Weighted F1 ~0.93
- FCA: Accuracy ~0.93, Macro F1 ~0.94, Weighted F1 ~0.93

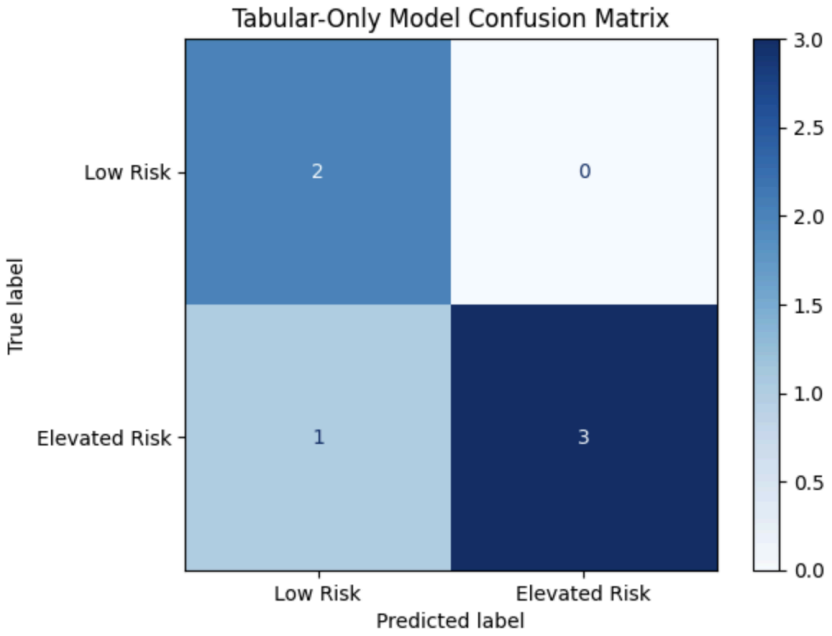


Interpretation from confusion matrices: most mistakes occur between adjacent ordinal classes, suggesting the model learned ordering rather than random label mapping.

Tabular model (initial regression baseline)

Tabular-only RF regressor reaches approximately:

- MAE ≈ 4.50
- RMSE ≈ 4.66



Athulya Saravanavakumar

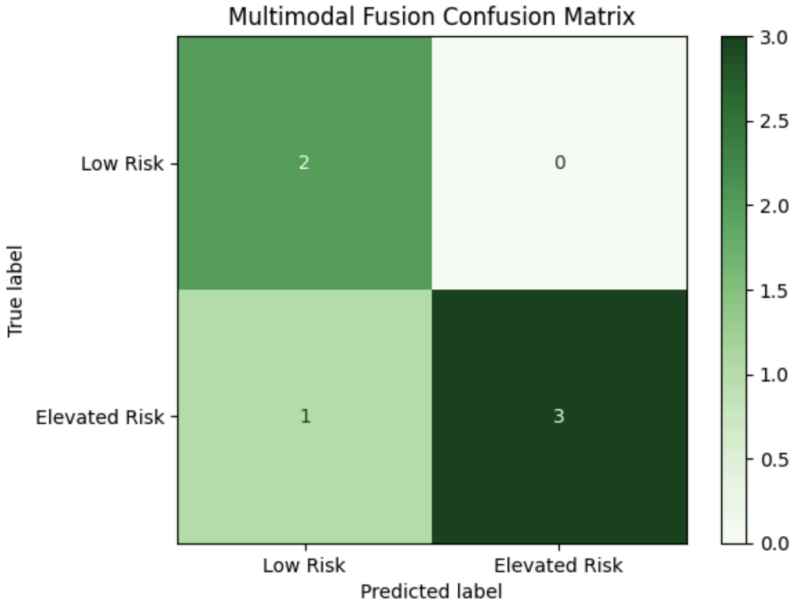
Tabular-only is inherently limited: high-level structural features don't encode qualitative programmatic differences, motivating multimodal fusion.

Fusion model (binary risk: low vs elevated)

With n = 6 complete multimodal schools, fusion is evaluated as binary classification:

- Accuracy: 0.83
- F1-score: 0.83
- Majority baseline: 0.67

Confusion-matrix comparison shows tabular-only and fusion made the same decisions in this tiny setting: low-risk schools correctly identified, and 4/5 elevated-risk correctly classified, with no extra errors introduced by adding text signals.



What this means:

- Fusion architecture is feasible and does not harm performance under extreme scarcity.
- The dataset is too small for claims of statistically significant gains; the right interpretation is validated pipeline + promising direction when scaled.

Limitations and planned enhancements

Key limitations

1. Tiny fusion sample size (n = 6) creates high variance and prevents strong generalization claims

2. Text representational bias: school-authored descriptions can be aspirational/marketing-driven; polished writing may correlate with better predictions independent of actual conditions
3. Hand-defined discretization thresholds: improves interpretability but may oversimplify nuanced differences between schools
4. Feature expressiveness bottleneck in tabular-only models: teacher counts and coarse geographic proxies capture resources, not program quality, climate, or student supports.

Enhancements

- Expand the text corpus (news, curriculum docs, community reports) to reduce single-source bias and increase coverage.
- End-to-end learned fusion (or richer fusion features like embeddings) once data volume supports it
- Interpretability tooling (attention/saliency) to audit what language drives predictions.
- Temporal analysis to track how school narratives and conditions shift over time (and avoid one "snapshot = destiny")

Conclusion/Discussion

This work highlights the effectiveness of multimodal machine learning for educational equity assessment by combining tabular school resource data with text-derived signals. Results show that while tabular-only and text-only models capture partial views of school conditions, their fusion yields more balanced and accurate classification of equity risk, particularly for identifying elevated-risk schools. This finding aligns with prior research demonstrating that multimodal approaches better capture complex, real-world educational phenomena than single-modality models.

Framing the task as classification rather than regression improved robustness under limited data, and the use of Random Forest models provided stable performance with minimal assumptions about feature distributions. However, the small sample size and the use of discrete text predictions rather than dense embeddings limit generalizability. Future work should incorporate larger datasets, richer textual representations, and additional modalities such as geospatial or remote sensing data, which have shown promise in related educational and infrastructure monitoring studies. Overall, this study supports the growing consensus that multimodal data fusion is essential for nuanced, scalable educational equity analysis.

Athulya Saravanavakumar

References

https://pmc.ncbi.nlm.nih.gov/articles/PMC10893965
https://www.mdpi.com/2072-4292/14/4/897
https://docs.edtechhub.org/lib/VINQBTJ5
https://www.sciencedirect.com/science/article/pii/S1226798825003186