Athulya Saravanakumar

Multi-Modal Education Equity Monitor: A Machine Learning System for Identifying School Infrastructure and Resource Inequities

Project Proposal: Multi-Modal Education Equity Monitor

1. Introduction & Problem Statement

School quality varies widely across districts, often due to unequal funding, aging infrastructure, and socioeconomic disparities. Current evaluation methods—like inspections or self-reported data—are slow and inconsistent, leaving underserved schools unidentified for years.

Goal: Build a system that uses **images, tabular data, and text** to estimate an **Equity Risk Score** for each school and generate a short explanation of why the school might be at risk. This can help districts and policymakers identify where additional support is needed.

2. Data Sources

- **Street-level images** from OpenStreetCam/Mapillary (school building conditions).
- **NCES (National Center for Education Statistics)** datasets: funding, enrollment, test scores, student–teacher ratios, free/reduced lunch rates.
- **U.S. Census/ACS data:** neighborhood income and broadband access.
- **School inspection reports** and district facility audits (PDFs/text).

3. Methods / Techniques / Technologies

- **Vision Models:** CNNs (ResNet/EfficientNet) to assess building condition and potential overcrowding indicators.
- **Tabular Models:** XGBoost/Random Forest to analyze funding levels, performance metrics, and demographics.
- **Text Models:** BERT-based models to extract issues (e.g., maintenance, safety) from inspection reports.
- **Multi-Modal Fusion:** Combine image, tabular, and text outputs into one **Equity Risk Score**.
- **LLM Explanation:** Use a small language model (Flan-T5 or similar) to generate a short summary explaining the score.

Athulya Saravanakumar

4. Project Deliverables

1. **Vision model** evaluating school infrastructure from images.
2. **Tabular model** predicting resource and performance risk.
3. **Text extraction model** identifying key issues from inspection reports.
4. **Combined multi-modal Equity Risk Score** for each school.
5. **Short explanation generator** summarizing why the school is high/medium/low risk.
6. **Final report** describing data, methods, experiments, and results.