

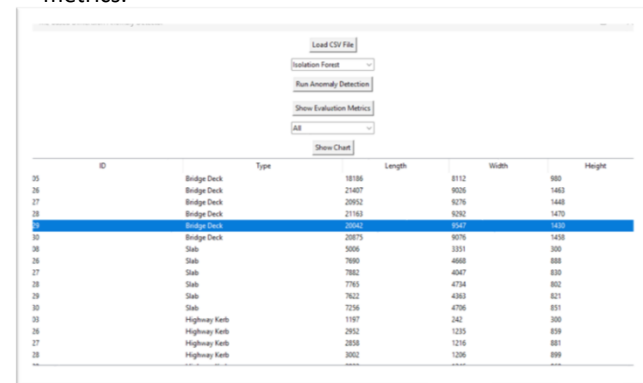
DEVELOPING A MACHINE LEARNING TOOL FOR DIMENSIONAL QA IN BIM AND DIGITAL TWIN WORKFLOWS

Introduction

Dimensional errors in Building Information Modelling (BIM) can lead to very costly design and construction mistakes. Traditional quality assurance (QA) methods often use manual rule-based checks, which can be slow and limited. This project investigates how unsupervised machine learning can automatically detect dimensional anomalies in civil engineering models, without needing predefined rules. A user-friendly tool was developed to support early anomaly detection, improving efficiency and helping engineers make informed decisions during the design phase.

Aims & Objectives

- Detect dimensional anomalies using unsupervised ML.
- Compare Isolation Forest and K-Means clustering methods.
- Build a simple graphical user interface (GUI) for engineers.
- Evaluate results using real test users and quantitative metrics.



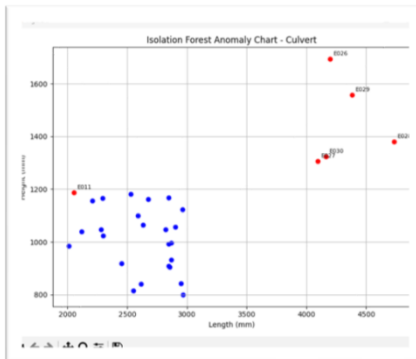
ID	Type	Length	Width	Height
25	Bridge Deck	18186	8112	980
26	Bridge Deck	21407	9026	1463
27	Bridge Deck	20942	9276	1448
28	Bridge Deck	21163	9282	1470
29	Bridge Deck	20643	9147	1430
30	Bridge Deck	20875	9076	1438
36	Slab	3351	389	889
37	Slab	7690	4668	888
27	Slab	7882	4047	830
28	Slab	7783	4734	802
29	Slab	7822	4363	821
30	Slab	7256	4796	851
35	Highway Kerb	1197	242	300
27	Highway Kerb	2952	1233	859
27	Highway Kerb	2938	1236	881
28	Highway Kerb	3002	1206	889

Methodology

- Developed in Python using Scikit-learn, Pandas, Matplotlib, and Tkinter.
- Geometry exported from iTwin-style BIM models (CSV format).
- ML models detect outliers in component dimensions.
- Evaluation using F1 score, precision, and recall.
- User testing conducted with practicing engineers.

How the Tool Works

1. Load geometry data (CSV).
2. Select ML model (Isolation Forest or K-Means).
3. Run anomaly detection.
4. Filter results by component type.
5. View flagged anomalies in tables or charts.



Model	Type	Precision	Recall	F1 Score
Isolation Forest	Bridge Deck	0.67	0.8	0.73
Isolation Forest	Slab	0.83	1	0.91
Isolation Forest	Highway Kerb	0.83	1	0.91
Isolation Forest	VRS	0.83	1	0.91
Isolation Forest	Wall	0.83	1	0.91
Isolation Forest	Culvert	0.83	1	0.91
K-Means	Bridge Deck	1	1	1
K-Means	Slab	1	1	1
K-Means	Highway Kerb	1	1	1
K-Means	VRS	1	1	1
K-Means	Wall	1	1	1
K-Means	Culvert	1	1	1

Testing

- The tool was tested using both synthetic datasets and feedback from five professional engineers.
 - Evaluation metrics included F1 score, precision, and recall for both ML models.
 - Engineers tested the full interface, assessed usability, and identified improvement areas.
 - Based on feedback, features were added like:
 - Filtering by element type
 - Visual comparison of ML models
 - Enhanced anomaly chart clarity
- The testing confirmed the tool's potential for real-world QA workflows and guided final refinements.

Key Findings

- Isolation Forest flagged more borderline outliers, useful for early QA.
- K-Means provided clearer group-based anomalies.
- Engineers preferred the clarity of the GUI and valued visual outputs.
- The tool helped identify outliers that would be missed using rules alone.

Conclusion

This project shows that unsupervised machine learning can support QA in digital design workflows. It bridges the gap between academic research and practical engineering, offering a tool that integrates into existing workflows and supports better design decisions.

Future Work

- Integrate with live BIM platforms (e.g., iTwin API).
- Support cloud-based scaling for large models.
- Expand ML methods (e.g., Autoencoders, DBSCAN).
- Allow user feedback to retrain models over time.

