

**Predicting the Structural Evaluation Rating of a Bridge to Understand
Improvement and Reconstruction Requirements**

Devasya Agarwal

10/28/25

Materials Science + Machine Learning

Abstract

The structural evaluation rating of a bridge is a critical indicator of its condition and reconstruction urgency, traditionally determined through on-site inspections. In this study, we aim to predict this rating using machine learning models trained on readily available bridge attributes, such as year built, structural type, and span length. Using the California subset of the National Bridge Inventory dataset, we performed thorough data preprocessing, feature selection through correlation analysis, and extensive data visualization. We then trained and evaluated multiple machine learning models, including logistic regression, random forest, and XGBoost on the grouped structural evaluation classes. The best-performing models, Random Forest and XGBoost, both achieved approximately 65% accuracy. This research demonstrates the potential of machine learning to assist in prioritizing bridge maintenance and reconstruction efforts.

1. Introduction

Bridges play a vital role in national infrastructure by supporting the movement of people and goods. However, many U.S. bridges are aging and at risk of structural deterioration, making early detection of poor conditions essential. Traditional assessments are time- and resource-intensive, often relying on human inspections. With the rise of open infrastructure datasets and machine learning, predictive modeling offers a complementary tool for evaluating bridge condition.

This project uses the National Bridge Inventory dataset from data.gov, focusing on bridges in California (state code 24). The dataset includes 6,741 bridge records and 123 variables. We selected between 11 and 20 input features based on their correlation with the target variable: Structural Evaluation Rating (STRUCTURAL_EVAL_067). The objective is to use machine learning models to predict bridge condition categories using features that are easier and quicker to collect. The workflow included data cleaning, feature selection, visualization, and training of various classification models to evaluate their effectiveness in identifying bridges in need of maintenance or reconstruction.

2. Dataset

We used the 2024 release of the National Bridge Inventory (NBI) dataset from the U.S. Department of Transportation, filtered to include only bridges located in California. This subset consists of 6,741 entries and 123 features. Many columns were dropped due to excessive missingness, lack of relevance, or redundancy.

The target variable was STRUCTURAL_EVAL_067, a categorical rating of structural adequacy. Ratings were grouped into four classes:

- 1.0 (poor condition)
- 5.0 and 6.0 (average condition)
- 9.0 (good condition)

Classes like 0.0, 2.0-4.0, and '*' were excluded or grouped due to low frequency. The resulting dataset had 5,390 valid entries.

3. Methodology and Model Development

Our approach began with structured data cleaning, including:

- Selecting variables with strong correlation to the target.
- Dropping or imputing missing values.
- Grouped the target variable (STRUCTURAL_EVAL_067) into four broader classes for simpler classification.
- Encoding categorical variables using `pd.get_dummies()`.
- Scaling numeric features using `StandardScaler()`.
- Splitting the data into an 80/20 training-test set using `train_test_split()`.

Exploratory data analysis included histograms, boxplots, violin plots, and a heatmap of the top 20 features most correlated with the target. These visualizations helped uncover relationships between predictors and the structural rating, guiding feature selection for model building.

We trained and evaluated five classification models to predict the structural condition class of each bridge.

Logistic Regression

Logistic Regression served as the baseline model, chosen for its interpretability and ease of implementation. It estimates the probability of class membership as a weighted function of input variables. Trained with 1,000 iterations for convergence, the model achieved ~60% accuracy. It performed better on well-maintained bridges (ratings 6 and 9) but struggled with lower-rated ones (1 and 5), indicating that the relationships in the data are not purely linear. The confusion matrix revealed frequent misclassification between adjacent rating groups.

Random Forest

Random Forest, an ensemble of decision trees, significantly outperformed the baseline with a 64-65% accuracy. It effectively captured non-linear relationships and feature interactions. Hyperparameter tuning with `GridSearchCV` improved performance, with optimal parameters: `n_estimators=250`, `max_depth=None`, `min_samples_leaf=4`, `min_samples_split=2`, and `max_features='sqrt'`.

This configuration helped reduce overfitting while maintaining diversity across trees. Random Forest was particularly accurate for high-condition bridges, though it unpredicted low-related ones due to class imbalance.

XGBoost

XGBoost, a gradient-boosted tree model, further improved classification performance. The default model yielded 62% accuracy, which increased to 65% after hyperparameter tuning with 5-fold cross-validation. The best parameters were: `n_estimators=75`, `max_depth=7`, `learning_rate=0.1`, `subsample=0.8`, `colsample_bytree=0.8`.

XGBoost handled misclassified examples well, offered good feature importance weights, and outperformed Random Forest on minority classes. It offered a strong balance between predictive power and computational efficiency.

Support Vector Classifier (SVC)

A support vector classifier (SVC) was used to capture non-linear decision boundaries using a radial basis function (RBF) kernel. The model achieved ~60% accuracy, similar to logistic regression but with better balance between classes. It was particularly precise for well-rated bridges but struggled with minority classes. While less interpretable and harder to tune, SVC served as a useful benchmark to evaluate kernel-based performance.

Neural Network (MLPClassifier)

The neural network model, configured with four hidden layers of 100 neurons each, achieved ~59% accuracy. It aimed to capture deeper, non-linear feature interactions. However, limited dataset size and class imbalance reduced its effectiveness. While the model trained stably after 500 iterations, additional layers or neurons did not improve performance. Neural networks underperformed compared to tree-based models for this tabular dataset.

4. Results and Discussion

Model comparison highlighted that tree-based ensemble methods (Random Forest and XGBoost) performed best across all metrics. Both achieved ~65% accuracy, with strong precision, recall, and F1-scores. Their ability to model non-linear relationships and handle diverse feature types gave them a clear edge.

Model	Precision	Recall	F1 Score	Accuracy
Logistic Regression	0.58	0.6	0.57	0.6
Random Forest	0.64	0.64	0.63	0.64
Random Forest TUNED	0.65	0.65	0.64	0.65
XGB Classifier	0.62	0.62	0.62	0.62
SVC	0.6	0.6	0.58	0.6
Neural Network	0.58	0.59	0.58	0.59
XGB Classifier TUNED	0.65	0.65	0.63	0.65

Logistic Regression and SVC, though easy to implement, plateaued at ~60% accuracy. The Neural Network model slightly underperformed, limited by dataset size and class imbalance. Overall, the confusion matrices showed that adjacent class misclassification was common, suggesting soft boundaries between condition groups.

These results reinforce the strength of ensemble tree models in structured engineering datasets. Even with moderate accuracy, they offer practical insight into structural condition trends and potential deterioration.

5. Conclusion and future studies

This study demonstrated the potential of machine learning to support structural bridge assessments by predicting the Structural Evaluation Rating using readily available attributes. Focusing on the California subset of the National Bridge Inventory, we developed and compared five machine learning models. Among these, Random Forest and XGBoost achieved the highest accuracy of 65%, outperforming both linear models and neural networks. These results indicate that tree-based ensemble methods are especially well-suited for structured engineering datasets, as they effectively capture complex, non-linear relationships.

Rather than replacing traditional field evaluations, these models can serve as complementary tools, flagging potentially high-risk structures during early design stages for further review and helping prioritize maintenance decisions at scale. Their ability to offer insights could improve efficiency in large-scale infrastructure management systems.

Future work should focus on expanding the dataset to include bridges from other states and more recent inspection cycles to improve the models' generalizability. Incorporating temporal features such as historical trends, repeated evaluations, or age-based degradation could further boost predictive accuracy. Additionally, adjusting for class imbalance, particularly for underrepresented, low-rated bridges, through techniques such as class weighting or oversampling may improve performance on minority classes. Overall, this research provides a foundation for using data-driven methods in infrastructure maintenance, demonstrating the value of machine learning in supporting decision-making for structural evaluation and reconstruction planning.