Bridges at Risk: Coastal vs. Inland Environmental Impacts on Lifespan and Maintenance

Team-1
Saivarun Tanjore Raghavendra
Mohammed Tareq Sajjad Ali
Suraj Poldas
Mano Harsha
Vasishta Chandala

Under the Guidance
Dr. Lindi Liao
Big Data Essentials (AIT-614) Sec:001
George Mason University

Abstract

Wherever the transportation network is concerned, bridges are considered the most crucial assets; their condition is usually very much dependent on geographical and environmentalfactors. This analysis deals with a dataset containing detailed attributes, such as Bridge Condition, Structure Age, Main Span Material and Design, Bridge Length, and Environmental Factors, including Average Temperature, Humidity and Total Precipitation. Such comparative studies are done to understand the variation in bridge conditions across coastal and inland areas. Furthermore, statistical models are utilized in the research study to determine linear correlations between such environmental variables as temperature fluctuations, precipitation, and humidity with the structural condition of bridges. The study identifies environmental stressors and certain structural design decisions as contributors to the aging and maintenance needs of bridges. The results provide valuable information to transportation agencies and city planners in support of data-driven policies for prioritizing maintenance and resource allocation and infrastructure resilience. The study integrates environmental data into bridge condition assessments to support long-term, sustainable infrastructure management practices and further resilient transportation system development.

Introduction

The bridge is a very crucial structure in transportation that connects communities and forms a catalyst for economic growth. A bridge's structural service life will be determined by design, material, traffic load, and environmental conditions. Climate, geography, and forces from all quarters make these factors determining performance over time that needs incessant monitoring and maintenance.

The environmental stressors such as temperature, precipitation, etc; differ between coastal and inland regions. Consequently, these elements generate different bridge deterioration rates. The coastal region has very unique weather patterns, while inland areas have temperature extremes that further accelerate structural aging.

This research investigates these factors, considering the division of coastal versus inland bridges. First, a complete dataset of structural and environmental features was preprocessed to ensure their quality and consistency. Bridges were clustered using the K-means clustering method according to condition-related features in order to gain insight into patterns, while an integrated model of Random Forest, XGBoost, and Neural Networks was developed for risk level prediction of bridges.

The results are intended to help transportation agencies in resource allocation, priority setting for maintenance, and policy formulation to improve resilience and sustainability of bridge infrastructure.

Methodology

The research examines the discrepancy in the conditions of bridges related to the region, whether coastal or inland, and examine how the conditions vary with regard to such environmental exposure as temperature, precipitation, humidity. Accordingly, this chapter presents a multi-step approach that combines data preprocessing, statistical analysis, clustering, and predictive modeling. The following sections provide a detailed description of the methodology adopted for this study.

1. Data Preprocessing

Data preprocessing is a very important step in preparing a dataset for further analysis. This is a phase at which raw data is cleaned and transformed to make the data quality and suitability for modeling assured. In this study, several columns in the provided dataset pertain to structural and environmental conditions of bridges.

Preprocessing includes:

Missing Value Handling: The data was checked for missing values, and depending on the nature of the data (numerical or categorical), the data was imputed with the relevant technique: using mean/ median in the case of numerical data, and mode for categorical data.

Pre-processing of data: Numerical features are standardized-think bridge age, traffic volume, and environmental variables (temperature, precipitation, etc.)-to put them on the same scale. This especially holds with clustering and predictive modeling where these attributes take importance.

Categorical Encoding: Categorical variables like bridge material or location had to be encoded using a technique such as one-hot encoding to convert them into numerical formats suitable for machine learning algorithms.

Feature Selection: Only relevant features were selected according to the research questions, such as bridge condition ratings, traffic data, environmental factors (e.g., temperature andprecipitation), and finally geographical information (e.g., latitude and longitude).

2. Geographical Segmentation

Geographical segmentation was done based on the locations of the bridges, which distinguished between coastal and inland area bridges. Based on the latitude and longitude coordinates provided in this dataset, the following categorization of bridges in two broad groups was done:

Coastal bridges are those that are located in regions very close to the coast and are subjected to the special environmental conditions typical for the coastal areas, like high humidity.

Inland bridges are those that are more interior and may be subject to various environmental stresses, including extreme temperature fluctuations and reduced humidity.

This research will categorize these data into two groups and study how those environmental factors and geographic locations affect differently the condition of the bridges, whether they are in the coastal or inland area.

3. Statistical Analysis

Statistical analysis was done to determine how bridge conditions can be related to environmental factors. In other words, the project aimed to study the relationships between temperature, precipitation, humidity and other environmental stresses with respect to the structural condition of bridges. With regard to the above requirements, following has been done:

Correlation Matrix: The correlation matrix is developed to visually inspect the interaction between different environmental parameters, i.e., average temperature, total precipitation, and average relative humidity, which affect the conditions of bridges. This helped in defining which parameters significantly affect the structural integrity and longevity of bridges.

Comparison by Groups: The statistical tests incorporate t-tests or ANOVA in order to compare the bridge conditions in the coastal versus inland regions. This would quantify whether the structural conditions change significantly with environmental influences and would give insight into deterioration patterns at a regional scale.

4. Cluster Analysis

K-means clustering is one of the clustering techniques that have been used to identify how the volume of traffic and environmental factors interactively affect the conditions of the bridge. Clustering simply groups bridges into distinctive classes based on similarity in features such as volume of traffic, environmental factors, and condition ratings. The major steps followed in clustering are highlighted below:

Feature Selection for Clustering: ADT, Average_Temperature, Total_Precipitation, and bridge condition rating would provide the feature variables on which clustering would be performed. These variables were selected because they are likely to capture most of the important environmental and traffic-related variations in bridge condition.

k-means Algorithm: Cluster Analysis Several clusters of bridges were segmented using the k-means algorithm. The number of clusters could be determined by using methods that search for an optimal value based on the intrinsic structure of the data, such as the elbow method or silhouette analysis.

Identification of Clusters: After determination of the number of clusters, interpretation of identified clusters is done to comprehend the characteristics of those clusters. Example:

Cluster 0: consists of highly utilized bridges under extreme environmental conditions. Bridges within the classes of urban areas with high vehicle loads and intense weather were classified as high-risk bridges.

Cluster 1: Older bridges with moderate usage and less environmental stress found in suburbs or rural areas-as medium-risk.

Cluster 2: Newer bridges with low usage and low environmental stress that were normally situated in mild climates were grouped into the lower-risk bridges.

This clustering methodology allows for the detection of traffic and environmental factor-related patterns in bridge conditions, while using such information to enable maintenance effort prioritization and resource allocation.

5. Predictive Modeling

An ensemble modeling approach is adopted for this task, which predicts the risk level of a bridge using structural and environmental features. Several machine learning algorithms have been combined in this approach, allowing the extraction of complementary strengths from each algorithm to provide robust predictions. The models incorporated:

Random Forest: It is a tree-based algorithm which is resilient to overfitting and has the potential to handle complex interaction between variables. It had been used to identify the important features and modeled bridge risk.

XGBoost - It is a gradient boosted decision tree algorithm, considered to possess high predictive power with efficiency in handling large data. Used here for boosting the predictive accuracy of the model.

Neural Networks: Deep learning that captures complex nonlinear relationships between features. A neural network model was implemented with the other algorithms to assess how well it could predict bridge condition.

The final ensemble model was trained on historical bridge data including the age of bridges, volume of traffic, environmental conditions, and condition ratings. The model was then applied to cross-validation to evaluate the performance of the model that generalizes well on unseendata. Feature importance analysis will be conducted to determine what factors-traffic, temperature, or humidity are most influential in predicting the risk associated with bridges.

6. Model Evaluations

Predictive models were evaluated through clustering analysis and machine learning algorithms that predict the factors affecting bridge conditions and their risk. In the evaluation, the Random Forest, XGBoost, and MLP Neural Network models are applied with different performance optimizations. Additionally, a clustering analysis was carried out in order to identify distinct clusters of bridges due to their environmental and structural conditions.

Clustering Evaluation

The clustering analysis groups bridges based on the main characteristics of traffic volume, age of the bridge, and environmental stressors. It had achieved a Silhouette Score of 0.87, which means well-defined, quite distinct clusters. The result indicates that the model had successfully identified separate classes of bridges with similar characteristics. In further understanding these clusters, PCA was applied. This was clearly indicated in the 2D PCA representation, where it showed how the environmental and bridge conditions contributed to cluster separation. Inclusion of a 3D PCA plot will enhance the interpretation that shall comprehensively show the role played by features such as traffic volume and precipitation in defining the clusters.

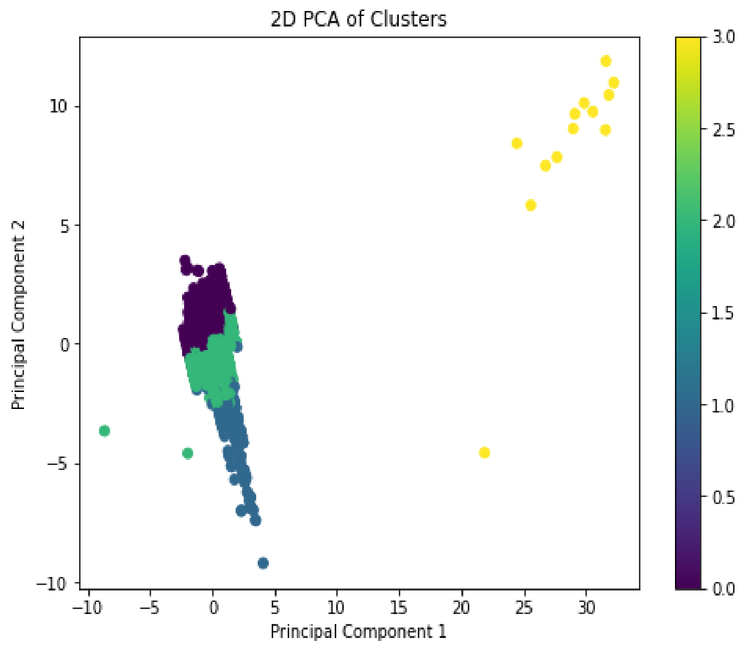

Fig1. 2D PCA visualization of data clusters

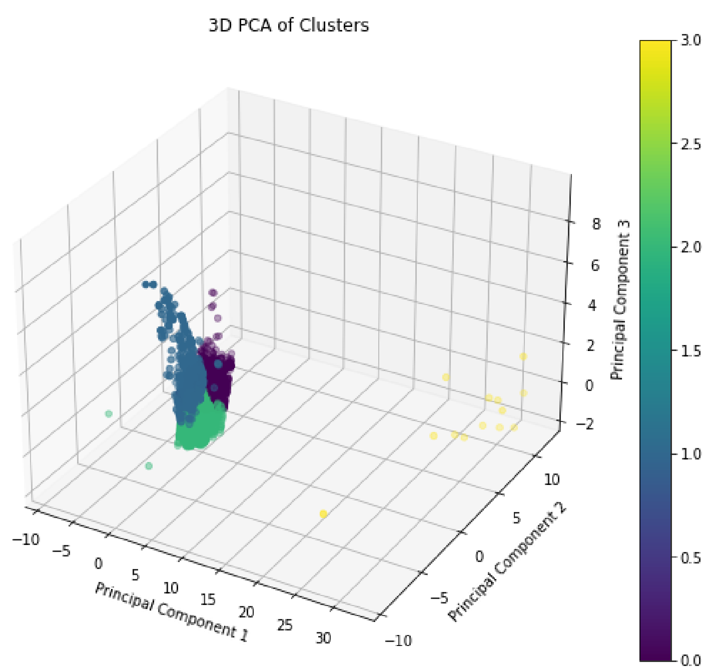

Fig2. 3D PCA visualization of data clusters

Predictive Model Evaluation

Several predictive models were trained and evaluated for bridge risk prediction based on structural and environmental features. These models include **Random Forest**, **XGBoost**, and **MLP Neural Network**.

- The Random Forest model, with 300 estimators, max depth of 30, and min_samples_split of 2, achieved a best score of 0.776. This model identified Bridge Age and Average Daily Traffic Volume as key features, performing well in detecting high-risk bridges with balanced accuracy across classes.
- The XGBoost model, tuned with learning_rate=0.3, max_depth=3, and 100 estimators, achieved a best score of 0.7771 (approximately 77.71% accuracy). It demonstrated strong predictive power, especially in handling class imbalance and distinguishing between "Poor" and "Good" bridges. This model was particularly effective in identifying high-risk bridges.
- The MLP Neural Network, optimized with hidden_layer_sizes=(100,), alpha=0.001, max_iter=1000, and solver='adam', performed moderately with a best accuracy of 68.47%. Although the model captured complex relationships, particularly how temperature influences bridge condition, its performance lagged behind that of Random Forest and XGBoost.

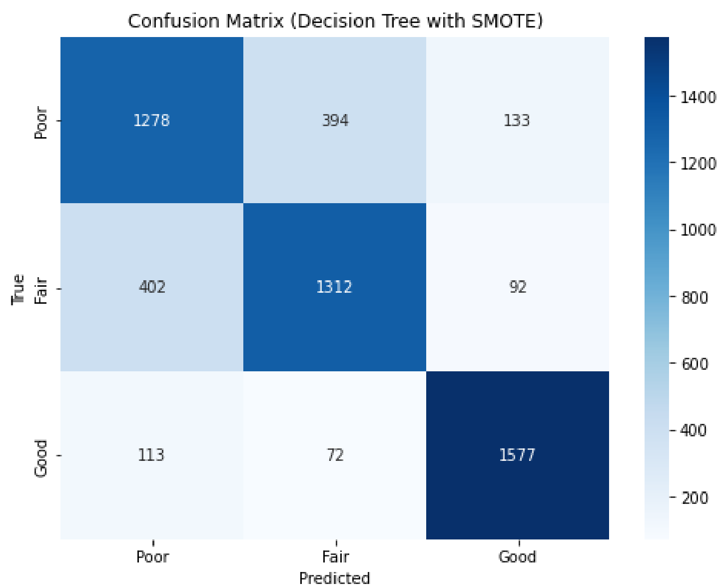

Fig3. Model performance evaluation with SMOTE

Data Preprocessing and Evaluation Metrics

The dataset underwent several preprocessing steps, including Label Encoding for the target labels ("Poor", "Fair", "Good") and the application of SMOTE (Synthetic Minority Over-sampling Technique) to address class imbalance. The dataset was split into training and testing sets using an 80-20 split for validation.

Confusion Matrices and classification reports were used to evaluate the models. All models performed well, particularly in identifying high-risk bridges. High precision and recall for the "Poor" class ensured that bridges requiring urgent maintenance were flagged accurately, enabling effective resource allocation.

7. Results Interpretation

With the completion of clustering and predictive modeling, the respective results were interpreted in detail to identify useful insights from them. The key patterns identified from clustering were used to find the environmental and traffic characteristics for bridges of different risk levels. The results of the predictive models were verified for their accuracy and then used forpredicting conditions of bridges in different regions. Further,

Coastal vs. Inland: A comparative analysis of the conditions among the coastal versus inland bridges was done to understand the impact of geographical location on structural health.

Environmental Stressors: The relationship of the environmental stressors, such as temperature and precipitation, with bridge condition was assessed to identify high-risk factors.

These analyses yield actual recommendations for transportation agencies, enabling them to target activities of maintenance and mitigation according to the risk factors identified.

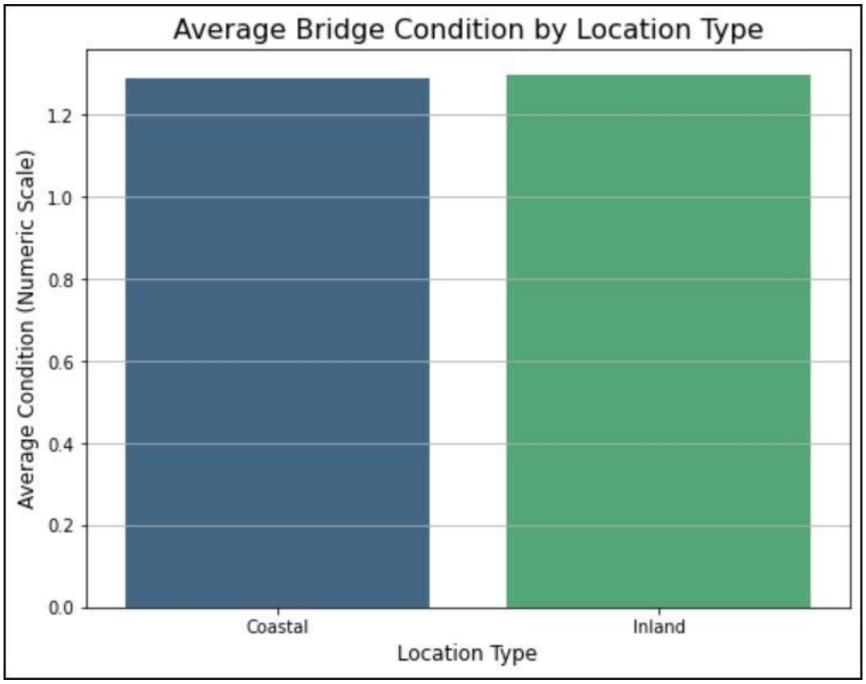

Fig4. Average bridge condition by location type

Based on the analysis of the dataset, we calculated the average bridge condition for both coastal and inland regions. The numeric average condition score for coastal bridges was 1.297, while for inland bridges, it was 1.286. This indicates that coastal bridges tend to have slightly better overall conditions compared to inland bridges. A bar plot visually reinforces this observation, showing that the inland bridges have a slightly lower average condition score compared to the coastal ones. This could imply that inland bridges may face more maintenance challenges over time. Further analysis could involve breaking down the condition score by individual factors, such as age and environmental conditions, to gain deeper insights into maintenance needs.

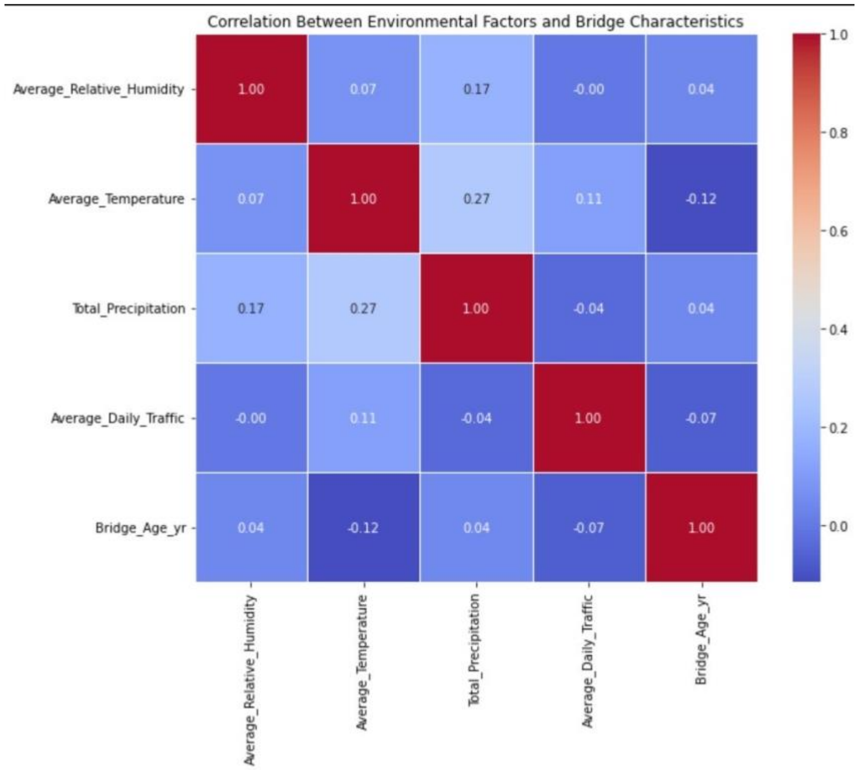

Fig5. Correlation between environmental factors and bridge characteristics

Summary of Correlations Between Variables:

Average Relative Humidity shows weak correlations with other variables: a slight positive correlation with Total Precipitation (0.17) and minimal relationships with Average Temperature (0.07), Average Daily Traffic (-0.00), and Bridge Age (0.04).

Average Temperature has a moderate positive correlation with Total Precipitation (0.27) and a weak correlation with Average Daily Traffic (0.11), but no significant correlation with Bridge Age (-0.12).

Total Precipitation shows moderate positive correlations with Average Temperature (0.27) and Average Relative Humidity (0.17), but weak correlations with Average Daily Traffic (-0.04) and Bridge Age (0.04).

Average Daily Traffic has weak correlations with Average Temperature (0.11) and Bridge Age (-0.07), with negligible correlation to other environmental factors.

Bridge Age has weak positive correlations with Average Temperature (0.04) and Relative Humidity (0.04), but no significant correlations with other variables.

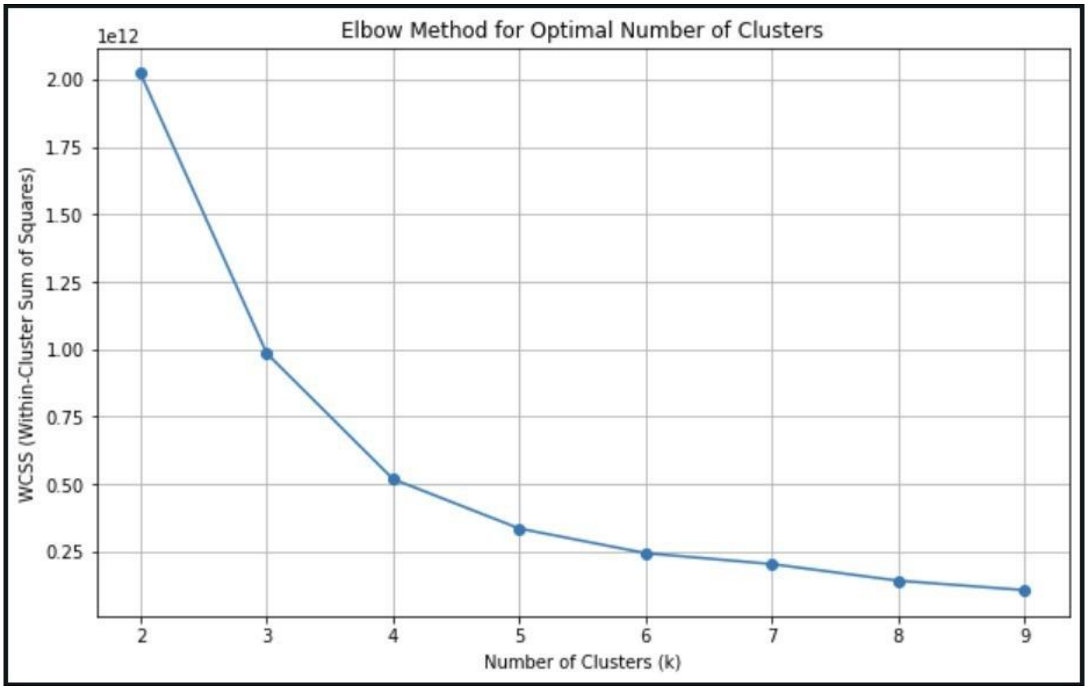

Fig6. Elbow method for optimal cluster number

```
Silhouette Score: 0.87
Cluster 0 Center: [2316.8021544    51.4459605    14.07466495 1147.28862694]
Cluster 1 Center: [2.32908806e+04 4.43019031e+01 1.45911824e+01 1.13102065e+03]
Cluster 2 Center: [1.62670494e+05 5.17294118e+01 1.50665204e+01 1.13114865e+03]
Cluster 3 Center: [6.56235252e+04 4.63943662e+01 1.51361255e+01 1.12979324e+03]
```

Fig7. Silhouette score and cluster centroids

Elbow Method and Cluster Centers Insights:

Elbow Method:

The optimal number of clusters is 4, as indicated by the sharp drop in WCSS followed by a slower decrease, suggesting diminishing returns in cluster variability.

Cluster Centers:

Cluster 0: Represents bridges with lower values in traffic, age, temperature, and precipitation, indicating lower risk and maintenance needs.

Cluster 1: Shows moderate values for environmental factors and bridge characteristics,suggesting moderate risk.

Cluster 2: Has higher values, especially in temperature and precipitation, indicating high risk or more frequent maintenance needs.

Cluster 3: Displays the highest values, suggesting very high risk or very high maintenance needs.

These insights can aid in predicting maintenance costs, with higher-risk clusters (like Cluster 3) requiring more frequent and costly maintenance.

Discussion

In our analysis, the clustering of these features brings out clear distinctions between the bridges with respect to their usage, environmental exposure, and age, all of which dictate their maintenance and risk status. The clustering, which was informed by the Elbow method, allowed us to group the bridges into four categories based on their conditions and environmental variables.

Cluster 0 (Lower Risk): It includes bridges with low traffic, age, and extreme environmental conditions. Such bridges are probably in a situation where wear and tear on them are less, and thus the needs for maintenance are also lower. From a practical standpoint, such bridges may require less frequent inspections and repairs, which eventually lead to cost savings in the long run.

Cluster 1-Moderate Risk: Bridges in this cluster are essentially exposed to moderate traffic and environmental factors. Such bridges might be outside big cities into suburban or rural areas with minimal extreme weather conditions. This, therefore, means that the upkeep schedules for such bridges would need to be higher than for Cluster 0 but not as heavy as for Cluster 2 or 3.

Cluster 2 (High Risk): Bridges in this cluster are highly exposed to environmental stressors, suchas fluctuations in temperature and precipitation. These variables increase risks and the likelihoodof more frequent maintenance. The bridges in this category may be coastal or high-altitude, where extremes of weather accelerate the deterioration process. More proactive strategies of maintenance and budgeting may thus be needed for these bridges.

Cluster 3-Very High: These bridges are subjected both to high volumes of traffic loads and severe attack from environmental conditions. This results in a very high degree of risk, necessitating more frequent inspections, repairs, and increased maintenance costs accordingly.

Bridges classified under this category would then be located in urban areas with very heavy traffic and harsher environmental conditions.

In terms of the prediction of maintenance needs, our findings point to the importance of taking into account both traffic loads and environmental factors when resources are allocated to bridge management. While traffic does not in all cases strongly correlate with maintenance need, it is clear that the environmental stressors of temperature fluctuations, precipitation, and humidity are of primary importance in determining the bridge's structural health over time. Therefore, incorporating environmental data into bridge monitoring systems can help agencies prioritize resources and reduce the risk of unexpected failures.

Our results as a whole indicate that an integrated model for maintenance of bridges, taking environmental and traffic-related variables into account, is likely to improve decision-making and assist in optimizing maintenance budgets. This would allow the forecast not only of when bridges are likely to need repairs but also of the most cost-effective ways to extend their lifespan according to each particular risk profile. Further research might consider other environmental variables; for example, saltwater surrounding coastal bridges, or look at how such clusters may change when more data becomes available.

Acknowledgment

We would like to thank Dr. Lindi Liao, our professor, for her guidance and support throughout this project. Her valuable advice and expertise made a great difference for us during work. A special thanks goes to Ravi Rachuri, our teaching assistant, who is always available for us to address various technical questions in and has been of immense help regarding support for the project. We also would like to recognize George Mason University for its resources and support in facilitating the conduct of this research project. We would also like to express our deep appreciation to the Federal Highway Administration, for their invaluable data that we needed in our work. Finally, we would like to thank our family and friends for their continued love and support through this process. This indeed has been a great learning experience for the project,and we value all the help received from everyone mentioned above.

Conclusion

This project aimed to understand the contribution of various factors (such as but not limited to location - coastal vs. inland - and environmental conditions) to the condition of the bridges. We extracted relevant patterns from bridges, traffic, and weather data that indicate how these factors affect the condition and maintenance needs of bridges.

For these, clustering techniques were used to group bridges into different risk categories based on their characteristics and environmental exposure. A model was also developed that can predict the risk level of bridges to prioritize maintenance tasks.

These results underscore how much this project does or can reveal about temperature and rainfall as major factors in the condition of bridges. Such insights can serve transportation agencies in better decision-making on how to manage their bridge inventories and further enhance bridge safety and service life.

Further insight into possible extensions could be provided by an in-depth look at other external factors, such as coastal erosion or soil conditions, making our predictions even more accurate and thus assisting in more effective bridge maintenance strategies.

References

Data - LTBP InfoBridge. (n.d.). https://infobridge.fhwa.dot.gov/Data

Ph.D., J. M., & Kavlakoglu, E. (2024, August 15). What is ensemble learning?. IBM. https://www.ibm.com/topics/ensemble learning#:~:text=Ensemble%20learning%20is%20a %20machine,than%20a%20single%20model%20alone.

American Institute of Steel Construction (AISC). "Steel Bridge Design

Handbook. https://www.aisc.org/nsba/design-and-estimation-resources/steel-bridge-design-handbook/.

Correlations Between Sets of Data for Investigating Bridge Conditions in Virginia. (2024). ScholarSpace. https://scholarspace.library.gwu.edu/concern/gw_etds/5x21tg31d

Neumann, J. E., Price, J., Chinowsky, P., Wright, L., Ludwig, L., Streeter, R., Jones, R., Smith, J. B., Perkins, W., Jantarasami, L., & Martinich, J. (2014). Climate change risks to US infrastructure: impacts on roads, bridges, coastal development, and urban drainage. Climatic Change, 131(1), 97–109. https://doi.org/10.1007/s10584-013-1037-4