Bridging the Gap

A New Approach to Infrastructure Modeling and Visualization

Brian Luu
bluu7@gatech.edu

Georgia Tech College of Computing

Travis Cole
tcole41@gatech.edu

Benjamin Boughton
bboughton3@gatech.edu

Evan Jones
evanjones@gatech.edu

Jacob Shalkhauser
jshalkhauser@gatech.edu

Mernegar Dorgoly
mernegar@gatech.edu

Introduction & Our Motivation

There are over 617,000 bridges in the United States. Each is paramount to ensuring the reliable and efficient transit of goods, services, and individuals. Despite each bridge's critical function, a significant portion are in a state of disrepair; this issue is not widely recognized in part due to the lack of non-proprietary solutions for transparent exploration of bridge data. Unfortunately, the most recognized tool for bridge data exploration, "InfoBridge," lacks the clarity and flexibility expected of modern web apps. This gap in technology motivated us to build a user-friendly application to help others better understand existing bridge data and modern deterioration modeling approaches. As such, we combined weather, traffic, and bridge data to build a modeling and visualization framework that improves upon current solutions. This new tool takes the form of an interactive, Flask-based web application that offers simple access to bridge infrastructure data, visualizations, and predictive models.

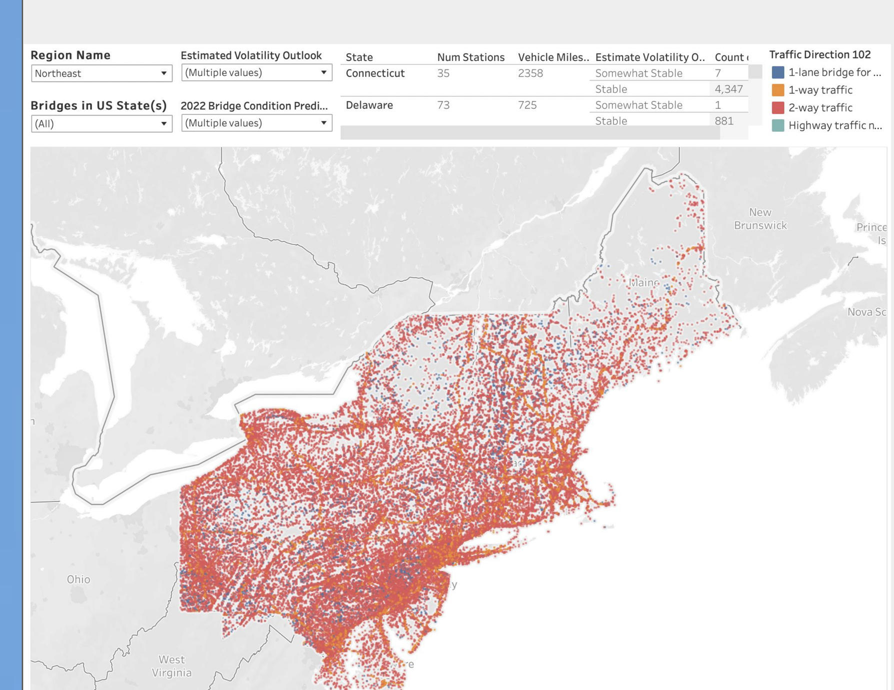
Data Approach & Unification

Our application is supported by three primary data sets: bridge information from the Federal Highway Administration spanning back to 1993, monthly historical weather data from 1990 to Present from the National Oceanic and Atmospheric Administration, and monthly traffic volume data from 2002 to 2020 from the Federal Highway Administration. After being downloaded manually, these datasets were cleaned and loaded into our AWS MySQL database via Python scripts. Using cross-reference tables, we joined these datasets together based on state name, nearest weather station, or other fields. Bridge condition predictions were loaded into MySQL via Python and joined via the unique structure number of each bridge. In total, this data amounts to 6,509,127 rows or roughly 1.9 GiB of disk space.

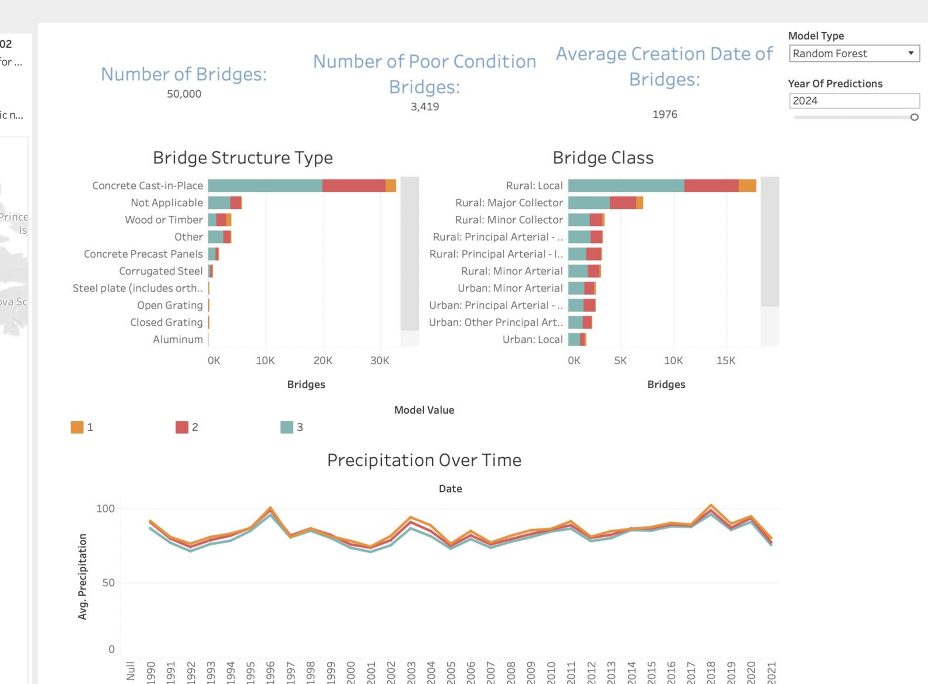
Visualization Approach & Evaluation

To display the data we used the popular data visualization tool Tableau, utilizing their self hosting option on Tableau Public which allowed us to publish the dashboards and embed them within the flask application. We built two dashboards, the first one is a detailed map that provides information at the granularity of the individual bridges and allows the user to view information about each bridge such as the average traffic in the state, average precipitation from the nearest weather station, what material the bridge is made of, the results of model predictions for the condition of the bridge for the previous year, current year and next year as well as other important information. The second dashboard is a higher level view that allows the user to interact with the model results and select a year for the predictions as well as a model type and view how the predictions change for dimensions including bridge class and structure type. Displaying the data through these visualizations allows users to easily navigate a complex data set consisting of over 600,000 bridges and view key features about those bridges in an intuitive manner. Our approach improves upon the existing solution by providing an easy to navigate interface and incorporating the traffic and weather data to provide a more comprehensive view of the bridges.

To evaluate our approach to visualization we had 10 users compare the Infobridge portal and our Flask app, and then rate them across qualities including: clarity, ease-of-use, and functionality. Users rated the Flask app an average of 4.5 out of 5, with the highest marks being in functionality and ease-of-use. The InfoBridge portal was rated an average of 2.5 out of 5, primarily due to low marks in functionality.



Dashboard 1



Dashboard 2

Modeling Experiments & Evaluation

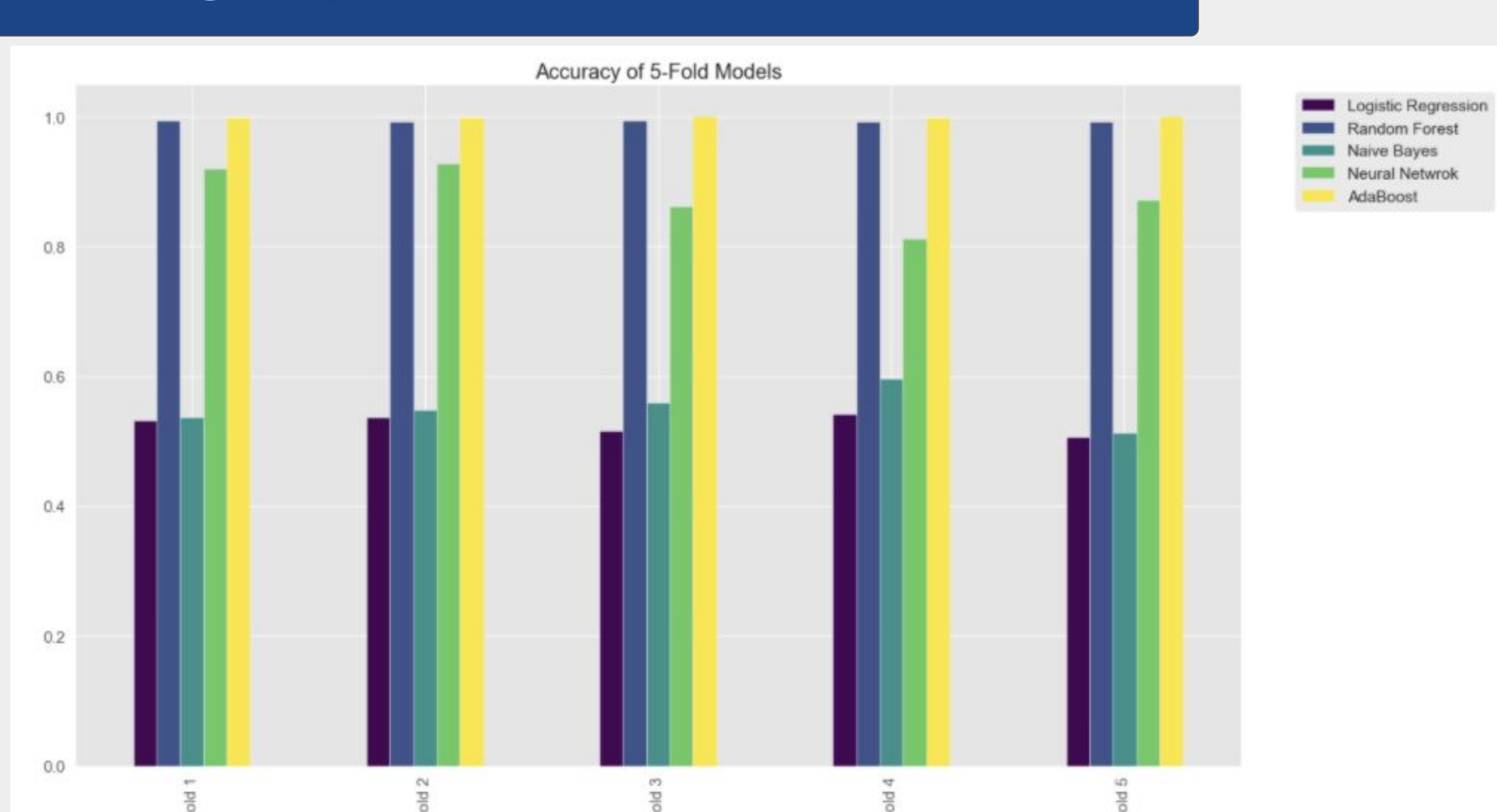


Figure 1: Cross-Validation Accuracy of Each of the 5 Model Classes Tested on the Southeast U.S.

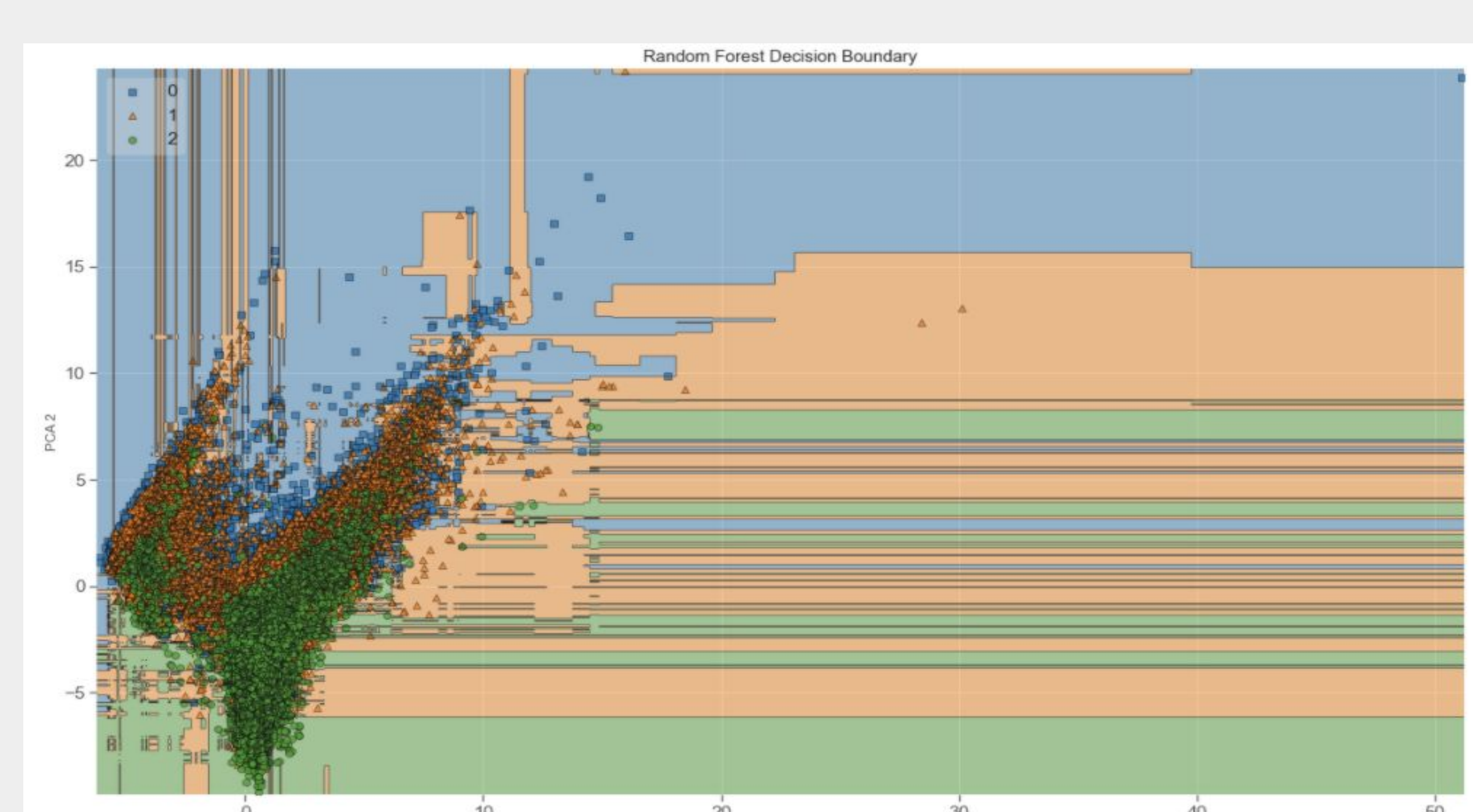


Figure 2: Random Forest Decision Boundary with PCA Dimensionality Reduction for the Southeast Region

Through 5-fold cross-validation model selection and subsequent hyper-parameter optimization of five leading modeling approaches for the categorical prediction task of future bridge condition (Poor, Fair, or Good) we were able to achieve results that significantly improve upon the present state-of-the-art models in two key aspects. Firstly, unlike the existing body of academic research on the task, our models leverage both additional and novel datasources, and generalize beyond a single state's bridge inventory to the regional level. Secondly, our production random forest and adaboost models significantly improved the existing benchmark categorical accuracy metric achieved by Heng Liu and Yunfeng Zhang [1] using convolutional neural networks from 85% to 98.8% and 91.4% respectively. Additionally, neural network models improved upon this existing accuracy benchmark by a modest 3%. As shown in the cross-validation performance plot above (upper left), the modeling performance of more complex model structures such as random forest, adaboost, and neural networks proved far superior to more interpretable, but less complex structures such as naive Bayes and logistic regression. This highlights both the difficulty and complexity of the decision boundary necessary to model the data. As an example of this complexity, the random forest decision boundary chart above (upper right) with dimensionality reduction was included to demonstrate some of the data and resulting classifier's unique structure.

[1] Y.-H. Huang, "Artificial neural network model of bridge deterioration," Journal of Performance of Constructed Facilities, vol. 24, no. 6,pp. 597–602, 2010. [Online]. Available: https://ascelibrary.org/doi/abs/10.1061/%28ASCE%29CF.1943-5509.0000124