A Clustering Approach to Linking Pavement Mix Design and Safety Outcomes for Pavement Management

Laura Thriftwood
Senior Data Scientist
WDM USA Limited

Ryland Potter
Vice President
WDM USA Limited

Isaac Briskin
Director of Analytics
WDM USA Limited

Paper prepared for the presentation at PV - Innovations in Pavement Management, Engineering and Technologies session of the 2023 TAC Conference & Exhibition, Ottawa, ON

Acknowledgements

The authors would like to acknowledge the Kentucky Transportation Cabinet (KYTC) for exploring the use of continuous pavement friction measurement in pursuit of its goal of preventing transportation-related fatal and serious injury crashes on Kentucky's roadway network. Special thanks are extended to the members of the KYTC team who helped analyze, review, and supplement WDM USA's data and whose expertise and involvement throughout the project has been vital - particularly Greg Garner and Amanda Dees.

Abstract

The goal of this ongoing research is to develop tools and techniques to design, deliver, and manage pavements for safety using continuous friction and macrotexture data. The technique outlined in this paper identifies and groups road segments by factors corresponding to the aggregate ingredient proportions of various pavement mix designs. The results of the analysis are used in a friction prediction model to better understand the impact of various elements of a mix design on in-service friction performance. In this paper, we will discuss data acquisition, preprocessing, clustering methods, results, and current and potential use cases.

Over the last three years, WDM USA collected approximately 40,000 lane miles of continuous pavement friction and macrotexture data on State-maintained roadways in the US state of Kentucky. The dataset includes annual measurements taken in both directions on all Interstate and Parkway roads and in one direction (cardinal or non-cardinal) on all State Primary and State Secondary roads. Road geometry features, continuous pavement friction, and texture values were obtained by WDM's SCRIM road survey vehicle. Road network features (e.g., AADT, speed limit) and aggregate mix design, pavement construction, and age/treatment information was provided by the Kentucky Transportation Cabinet (KYTC).

Preprocessing aggregate mix design and construction record data required significant manual review and consultation with subject matter experts. The team explored several aggregation methods to minimize information loss. Mix design information was assigned to each 0.1-mile (0.16-km) road segment in the WDM survey data.

The exploratory phase of the analysis used a K-Prototype (mixed variable type) algorithm. Aggregate mix design data was represented by a primary and secondary ingredient name and associated mix proportion value. Initial results of this analysis indicated that road geometry features dominated clustering behaviors, with aggregate mix ingredients having minimal impact.

In the subsequent phase of the analysis, aggregate mix design data were restructured to contain only numerical variables. The ingredients in each mix design were aggregated based on particle size and polish resistance, represented as a proportion of the overall design mix. The proportion of reclaimed asphalt pavement (RAP) in each mix was also represented. Various available clustering methods were evaluated, with the final analysis using the K-means clustering algorithm. Initial results showed clear cluster separation driven by particle size and polish resistance values.

The results of the clustering analysis were incorporated into a friction prediction model trained using the Extreme Gradient Boosting (XGBoost) library to predict a friction value represented as the Mean SCRIM Coefficient (MSC). Two methods of incorporating the results of the clustering analysis were compared to a model that did not use any of the results.

KYTC is using this analysis as an input to maintenance programming and pavement friction service level setting. Future implications also include enhanced pavement friction deterioration modelling, surface treatment selection, and safety outcomes through regular/preventative maintenance programming.

Introduction

The relationship between adequate road friction and safety outcomes is well-established. The development of a pavement friction management program is an important step in minimizing friction-related vehicle crashes by ensuring that pavement surfaces are “designed, constructed, and maintained to provide adequate and durable friction properties” [1]. Continuous Pavement Friction Measurement (CPFM) equipment, such as the SCRIM road survey vehicle, is an important tool to collect adequate road condition data for a road network. It is the method of friction measurement supported by the US Department of Transportation Federal Highway Administration (FHWA) over spot friction measurement devices like the Locked-Wheel Skid Trailer which “cannot safely and accurately collect friction data in curves or intersections, where the pavement polishes more quickly and adequate friction is so much more critical” [2]. Traditional methods of analyzing friction data combine crash history, road geometry, and road network features to identify areas of elevated friction-related crash risk. In addition to characterizing the relationship between friction and crashes, continuous pavement friction data can be utilized to understand the impact of pavement mix design characteristics on in-service friction performance.

A strong Pavement Friction Management Program (PFMP) is a key tool for road authorities to monitor and maintain a safe road infrastructure. Safe road infrastructure is one element described in the Safe System approach adopted in the US, modeled after the Vision Zero approach first developed in Sweden in 1997. The Safe Systems Approach operates under the recognition that humans make mistakes and that a road infrastructure should be designed and managed to anticipate and mitigate the impact of those mistakes [3]. In support of this goal, the FHWA has developed the Proven Safety Countermeasures initiative (PSCi) which provides a collection of 28 countermeasures and strategies that are proven to be effective in reducing roadway fatalities and serious injuries on all types of roadways. One of these countermeasures focuses on pavement friction management, recognizing friction as a “critical characteristic of a pavement that affects how vehicles interact with the roadway, including the frequency of crashes.” Particular attention is paid to locations with a higher friction demand with an aim toward preventing many roadway-departure, intersection, and pedestrian-related crashes [2].

There exists strong statistical evidence of the relationship between pavement frictional properties such as friction and macrotexture on road safety. Both friction and texture have a significant effect on predicting crash rates and are key elements in the development of Safety Performance Functions (SPFs) [4]. The primary means of road safety assessment at the network level relies on a retroactive analysis of crash history. Crash history is combined with features of the road network (e.g., grade, slope, curvature, speed limit, etc.) and measured friction and texture to develop SPFs. These SPFs make it possible to evaluate the effect of pavement friction changes on safety performance, which can then inform the cost effectiveness of pavement friction improvements.

An important element of friction performance not commonly considered in the development of SPFs is the pavement mix design. The aggregate mix ingredients, proportions, and volumetric characteristics described in the pavement mix design can influence the in-service performance and longevity. A better understanding of the specific characteristics of pavement mix design that influence in-service friction and texture performance can support long-term maintenance planning and can provide guidance for the development of new mix designs and surface treatment selections.

Literature Review

In addition to preventing crashes, an adequate friction level can reduce the severity of crashes that do occur. In a study that attempted to understand the influencing factors of crash injury severity at intersections in Wyoming, friction was found to have a substantial effect on intersection crash severity. The authors state that increasing friction numbers at intersections was found to mitigate crash injury severity, concluding that an increase in friction would "significantly decrease the risk of observing severe injuries and fatalities". [5]

Clustering analysis methods have previously been applied to asphalt pavement to identify climate regionalization, with a focus on ensuring good performance and service life [6]. The authors of the referenced study used K-Means clustering to divide asphalt pavement segments in the Liaoning Province in Northeast China into four climate zones using climate-related variables but did not include elements of the pavement mix design.

In another application, clustering was applied to a dataset of pavement distress characteristics in Rajasthan, India to support maintenance prioritization decisions under a limited-funding environment. K-Means clustering was used to group 100m (328 ft) segments of the road network using pavement distress parameters such as cracking, raveling, potholes, and patching. The authors concluded that this approach which used the results of the clustering analysis was necessary in effective decision-making for a pavement management system. [7]

Objectives

The goal of the work was to understand how road friction performance is impacted by pavement mix design characteristics. This paper presents a method of performing clustering analysis on features of a pavement mix design dataset and demonstrates the performance improvements gained from incorporating the results into a friction prediction model. It compares different methods of applying the results of a clustering analysis to improve the accuracy of a supervised machine learning algorithm. It will also discuss potential applications of this analysis for a pavement friction management program, project prioritization, and development of new pavement mix designs.

Data Collection

The pavement mix design data was provided by the Kentucky Transportation Cabinet (KYTC). The dataset contained a history of project contract details for the road network and details of the mix designs used in each project.

The road condition data was collected by WDM USA Limited during 2021 and 2022 using the Sideways-Force Coefficient Routine Investigation Machine (SCRIM). SCRIM vehicles have been used for continuous pavement friction measurement around the world, helping transportation agencies reduce crashes and save lives for over 50 years. SCRIM vehicles have been critical to the development of Pavement Friction Management Programs (PFMPs) since SCRIM's introduction to the United States in 2015.

SCRIM vehicles measure friction every 10 centimeters, or roughly 4 inches, using a freely rotating test wheel oriented at a 20-degree slip angle. Using instrumentation that simultaneously measures other data such as macrotexture, roughness, and road geometry, SCRIM vehicles provide data for analysis at a minimum 1-meter segment length. The typical SCRIM analysis dataset comprises either 26.4-foot or 33-

foot segments (8.05-meter and 10.1-meter, respectively), consisting of a SCRIM friction reading calculated using dynamic vertical load measurements, and Mean Profile Depth (MPD) or Sensor Measured Texture Depth (SMTD) as measurements of macrotexture. Additional data collected includes GPS location, cross slope, horizontal curvature, grade, surface temperature, tire temperature, and air temperature. WDM USA-operated SCRIM vehicles also include cameras that provide a forward-facing image for every data point [8].

Data Preparation

Selecting a Mix Design

Project contracts in the raw pavement mix design data were associated with multiple mix designs as each contract represents all work performed on a stretch of road, including in-lane and shoulder work. WDM USA worked with materials subject matter experts at KYTC to develop a hierarchy of rules for selecting the mix design most likely to have been used on the lane in which the SCRIM survey data was obtained. For example, when comparing two mix designs under a single contract, the mix design with a higher performance grade or polish resistance class was retained for analysis.

Aggregate Mix Ingredient Proportions

Each project contract included a list of the aggregate ingredients and associated proportions, with each set of mix design ingredient proportions summing to 100. To obtain meaningful results from a clustering analysis, the aggregate ingredients were categorized into the following five classes:

1. Coarse, polish-resistant (COARSE_PR)
2. Coarse, non-polish-resistant (COARSE_NON)
3. Fine, polish-resistant (FINE_PR)
4. Fine, non-polish-resistant (FINE_NON)
5. Reclaimed Asphalt Pavement (RAP)

These categories were developed in consultation with subject matter experts with the intention of describing aggregate ingredients in a simplified structure of key characteristics that impact friction performance – aggregate gradation and polish-resistance.

Additional Mix Design Features

In addition to the aggregate mix ingredient proportions, the raw pavement mix design data also contained volumetric information about each mix design that was incorporated into the subsequent friction prediction model. Features with a high proportion of null values, such as CT Index, Density, Hamburg Max Deformation Left/Right, and Hamburg Pass Number Left/Right, were excluded. Other features, such as Dust to Asphalt Ratio, were derived from the available data.

$$\text{Dust to Asphalt Ratio} = \text{\% Passing \#200 Sieve Size} / \text{\% Effective Binder}$$

To calculate the Coefficient of Curvature (CC) and Coefficient of Uniformity (CU), a nonparametric curve was fitted to the sieve size in millimeters (mm) and percent passing using locally weighted scatterplot smoothing for each mix design. Values were predicted for the sieve size in mm at 10, 30, and 60 percent passing. Linear interpolation between the two nearest data points was used to identify values for which the prediction could not be made based on the available data. These values were used to calculate the CC and CU using the following equations,

$$CC = \frac{D30^2}{D60 * D10}$$

$$CU = \frac{D60}{D30}$$

where D10, D30, and D60 are the sieve size in mm at 10%, 30%, and 60% passing, respectively.

When joining project contracts with the SCRIM survey dataset, the most recent contract for each 0.1-mile (0.16-km) segment was retained. The age of the pavement was initially determined based on the number of days between the contract completion date and the survey date, but close review of those contract completion dates revealed a lack of distinction between when the work was completed and when the contract was closed in the system. A 28-day lag was added to the age of the pavement based on a review of the forward-facing images from the survey vehicle to account for this potential delay, and rows were removed for any surveys that were taken prior to the installation of the mix design described in the contract.

The mix design data was joined to the survey data using the contract ID, resulting in the final analysis dataset of 63,247 rows representing approximately 6,122 miles (9,852 km) of Kentucky roadway.

Methodology

Review of Available Clustering Methods

Clustering is a method of unsupervised learning that can be applied to an unlabeled dataset with the goal of grouping similar datapoints based on their attributes. Unlike methods of supervised learning, there is no ground truth for the algorithm to classify or predict. Instead, the algorithm attempts to identify natural groupings within a dataset based on some determination of similarity among the elements. Ideally, data points within the same cluster are similar to each other and different from data points in other clusters. One of the most common industry applications of clustering analysis is customer segmentation - identifying groups of customers using data from purchase history, engagement behavior, and/or demographic traits. These customer segments can be used to make predictions about future behavior or to anticipate a reaction to a specific marketing campaign or outreach program. In this application, we apply the same logic to discover groupings of similar road segments of the KYTC road network based on elements of the mix design relating to the aggregate ingredient proportions. By first segmenting the network into individual clusters, we can apply individualized treatment or modeling tailored to each cluster under the assumption that cluster membership reflects differences in in-service behavior or performance.

The following clustering methods were considered for this application: K-Means, DBSCAN, Spectral, Mean Shift, Agglomerative Hierarchical, and Gaussian Mixture Modeling. Several important key features of the analysis were identified, against which each clustering method was evaluated.

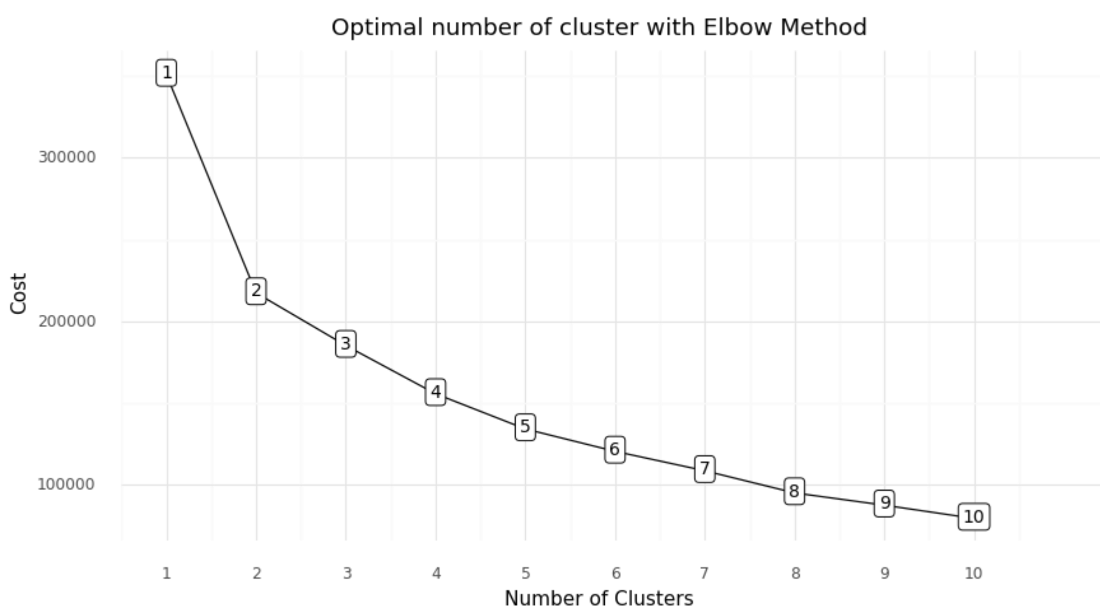
First, the algorithm needed to be efficient and scalable. Second, as the results would be incorporated in the subsequent friction prediction model, it was important to be able to control the number of clusters as too many clusters run the risk of overfitting. Third, the methodology used by the algorithm and the results of the analysis needed to be intuitive and easily interpretable to a non-technical audience. Additional traits that differentiate clustering methods that were not applicable for this dataset were not considered (e.g., sensitivity to outliers and the ability to handle highly dimensional data).

K-Means clustering was selected as the best method based on the data, the intended application, and a review of relevant literature. The algorithm is scalable, intuitive, and allows for manual selection of the number of clusters. K-means functions by dividing a set of samples into a designated number of separate clusters. These clusters have a described mean, called a centroid, which minimizes the within-cluster sum-of-squares. The algorithm first selects initial centroid values and assigns each datapoint to the nearest centroid/cluster. The mean/centroid of datapoints assigned to each cluster are calculated, providing new centroid values. The datapoints are again assigned to the nearest centroid from the latest iteration, new cluster centroids are calculated, and datapoints are reassigned as needed. This process repeats until the difference between the previous and current centroid values are below a specified threshold, indicating that the centroids are no longer moving, and the final cluster membership can be assigned to each datapoint. K-Means can be sensitive to the initial value selected for the cluster centroids, which can be mitigated using various methods. For example, using the k-means++ parameter can ensure initial centroid values are significantly distanced from each other. [9]

Selecting the Appropriate Number of Clusters

As the K-Means algorithm requires manual preselection of the number of clusters, the Elbow Method was employed to determine the optimal number of clusters. The total cost (cluster variance) was plotted at different numbers of clusters, ranging from one to ten. It was determined that three clusters were optimal for this dataset, after which no additional cost gains were realized by the inclusion of additional clusters, as seen in Figure 1.

Figure 1



The aggregate mix ingredient data was normalized, K-Means algorithm was applied, and a cluster membership was assigned to each datapoint.

Results of Clustering Analysis

A summary of the resulting cluster membership can be seen in Table 1.

Table 1

Cluster	Rows	Miles	Kilometers
0	10,629	1,037.5	1,669.7
1	22,873	2,195.8	3,533.8
2	29,745	2,888.7	4,648.9

A radial chart of the average value for each scaled feature is displayed for each cluster in Figure 2.

Figure 2

KEY

COARSE_NON: Coarse, non-polish-resistant
COARSE_PR: Coarse, polish-resistant
FINE_NON: Fine, non-polish-resistant
FINE_PR: Fine, polish-resistant
RAP: Reclaimed Asphalt Pavement

To support interpretation of the results of the analysis, a narrative description of the general attributes of road segments assigned to each cluster was developed.

- **Cluster 0** has similarities to Cluster 2 in that it contains road segments with higher proportions of non-polish-resistant materials, both coarse and fine. However, it contains little to no RAP content and is the smallest cluster in membership size.
- **Cluster 1** contains road segments with high proportions of polish resistant materials, both coarse and fine. It has a higher RAP content, like Cluster 2.
- **Cluster 2** contains road segments with higher proportions of non-polish-resistant materials, both coarse and fine. It contains the highest proportion of RAP content of all three clusters and is the largest cluster in membership size.

The results of the clustering analysis were also reviewed using descriptive statistics and visualization of the pairwise comparison of the features. The average raw values of some of the key features by cluster are given below in Table 2.

Table 2

Cluster	**0**	**1**	**2**
Mean Scrim Coefficient (MSC)	61.97	61.57	62.2
Mean Profile Depth (MPD)	0.84	0.8	0.81
Grade	0	-0.04	0.07
Cross Slope	2.07	2.12	2.06
Curve Radius	12726	15821.69	11789.11
International Roughness Index (IRI)	92.57	75.34	94
Average Annual Daily Traffic (AADT)	4198.7	13423.45	3933.67
Posted Speed	53.41	57.12	53.76
Lanes	1.11	1.64	1.1
Lane Width	10.52	11.37	10.21
Surface Width	3.05	6.52	2.47
Shoulder Width	5.18	7.62	4.49
Percent Passing %200 Sieve (P200)	6.21	5.65	6.22
Coefficient of Uniformity (CU)	19.74	19.86	21.74
Coefficient of Curvature (CC)	2.11	2.08	2.38
Binder	5.76	5.72	5.69
Effective Binder	5.13	5.03	5.14
Air Voids	3.97	3.8	3.82
Dust to Asphalt Ratio	1.21	1.12	1.21
Coarse, Non-Polish-Resistant	26.32	3.59	28.46
Coarse, Polish-Resistant	4.39	30.6	1.65
Fine, Non-Polish-Resistant	56.64	26.98	48.17
Fine, Polish-Resistant	11.63	24.49	6.08
Reclaimed Asphalt Pavement (RAP)	0.88	14.38	15.75
Pavement Age (days)	2476.65	1972.45	1817.38

Incorporating the Results into a Friction Prediction Model

The subsequent phase of the project explored methods and potential benefits of incorporating the results of the clustering analysis into a friction prediction model. The model used the Extreme Gradient Boosting (XGBoost) library to predict a friction value represented as the Mean SCRIM Coefficient (MSC). MSC is the measured friction value from the SCRIM survey that has been adjusted for speed, temperature, and seasonality.

XGBoost is a decision-tree based ensemble algorithm that uses a gradient boosting framework to improve performance. While similar in structure, it often outperforms traditional decision-tree algorithms like Random Forest, but with the added benefits of regularization techniques that reduce overfitting - a particular concern for this application.

XGBoost makes an initial prediction for each row of the data, then computes residuals based on the difference between the predicted value and observed value. A similarity score for the root leaf is calculated from the residuals and includes a regularization parameter. The tree is then split on one of the features, with new residuals and similarity scores calculated for each new leaf. To determine whether this split performs better than the root, the algorithm calculates the gain and compares it to the previous split (in this case, the root). A positive gain indicates a good split, a negative gain indicates a bad split. This process is repeated on the leaf nodes until it reaches the designated depth, after which the algorithm travels back up the tree to prune any splits that do not meet a designated gain threshold. Each leaf node receives a final output value based on the residuals for the variable on that leaf, including a regularization parameter.

With the full tree built, the algorithm makes new predictions for each row of the training data using the initial predicted value, output values of the leaves, and determined step rate. These values are compared to the observed values to calculate the residuals for the current iteration of the tree. Ideally, the residuals will be smaller than the residuals for the tree with a single root node. This tree-building process is repeated until the residuals are minimized or until the algorithm reaches the set number of iterations.

Feature Selection

The set of variables included in the friction prediction model were influenced by the availability of the data and a review of related work. A North Carolina State University study investigating how pavement overlays affect surface texture and friction characteristics provided important insight and guidance on the key mixture compositional factors that influence friction and texture modeling [10].

While XGBoost has some built-in feature selection capabilities based on the decision tree methodology, an analysis of collinearity was also conducted to identify any potential issues of multicollinearity between features. The variance inflation factor (VIF) was calculated for each predictor variable, quantifying how much the variance of each feature is inflated by any correlation with the other features. Generally, a VIF score of four or greater deserves investigation while a score exceeding ten indicates severe collinearity and warrants removal. Two features were removed from the model due to high VIF scores indicating collinearity concerns – Percent Passing #200 Sieve Size and Surface Width. The final set of features included in the model and their associated VIF scores are provided below in Table 3.

Table 3

Feature	VIF
Mean Profile Depth (MPD19)	1.296
Grade (GradeRaw)	1.006
Cross Slope	1.036
Curve Radius	1.292
International Roughness Index (IRI)	1.418
Annual Average Daily Traffic (AADT)	2.653
Lanes	3.145
Lane Width	1.975
Shoulder Width (ShoulderW)	2.619
Intersection	1.082
Coefficient of Uniformity (CU)	1.965
Coefficient of Curvature (CC)	1.796
Binder Content (BINDER)	2.235
Effective Binder Content (EFF_BINDER)	2.115
Air Voids	1.128
Dust to Asphalt Ratio (DA_RATIO)	1.515
Age of Pavement (age_days)	1.374

Hyperparameter Tuning

Given the concern regarding overfitting, special attention was given to the hyperparameters of the model. Overfitting can occur when a model is fit too closely to the training data. A model overfit to the training data will generalize poorly when attempting to make predictions using unseen test data. In this application, the model was trained with a small learning rate, restricted maximum tree depth, and used subsampling on the training data. A small learning rate limits the distance the algorithm steps toward the minimum of the loss function in a gradient descent algorithm. Restricting the maximum tree depth sets a hard limit on the number of splits in the final model, preventing it from reaching a depth too specific to the training data. Subsampling during the training process sets a proportion of the training data to randomly sample for each tree, improving the generalizability of the final model.

Initial Results

Three variations of the model were compared. The first model variation was trained on the set of features described in Table 3 and did not include any of the results of the clustering analysis performed on the aggregate mix ingredient proportions. The second model variation included cluster membership as a categorical variable. The third variation included separate models trained on each cluster. As mentioned, the primary concern when fitting the models to subsets of a dataset is overfitting. Overfitting is most apparent when the model performs well on the training data but performs comparatively poorly on the test data.

To assess the overfitting concerns, friction values were predicted for both the training and test data, with R-squared values displayed below in Table 4. The comparable performance between the training and test datasets does not indicate severe overfitting.

Table 4

	Model 1	Model 2	Model 3a	Model 3b	Model 3c	Model 3 (avg)
Training Data	0.55	0.56	0.79	0.73	0.63	0.72
Test Data	0.54	0.55	0.75	0.71	0.60	0.68

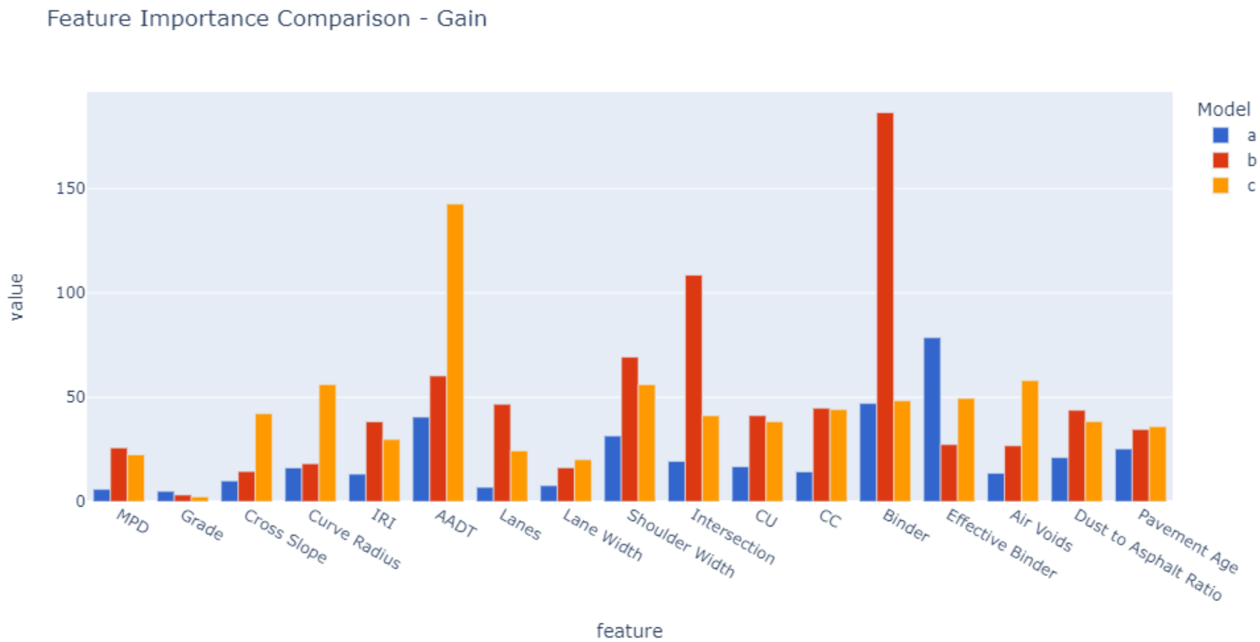
Additional metrics to evaluate the model performance were considered, including the Root Mean Squared Error (RMSE). The lowest RMSE values were achieved in Model(s) 3a-c.

Feature Importance

XGBoost provides several metrics to evaluate the relative importance of each feature in the final model – gain, cover, and weight. Gain refers to the relative contribution of each feature in improving the accuracy of the model at each split. Cover represents the relative number of observations related to each feature across all splits in the model. Weight (or Frequency) represents the relative number of times each feature occurs in the trees of the final model. For all three metrics, a higher score implies a higher relative value in the final prediction [11].

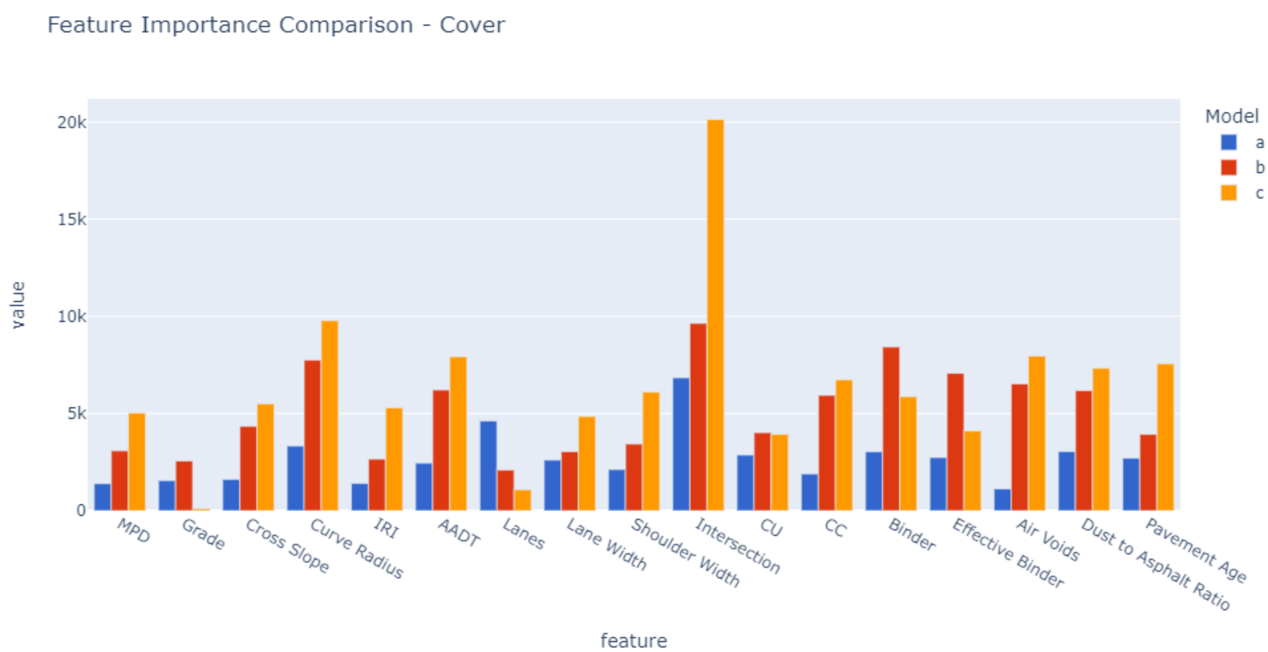
Features included in the friction prediction models were evaluated using all three relative importance metrics, though the most attention was given to the gain metric. While there were differences between the behavior of each of the three models trained on separate clusters, there were some features that had high relative importance scores in all three models. Of note is that Model 3B had a significantly higher relative importance for the binder content than the other two models. This model was trained on Cluster 1 which had the higher average proportion of polish resistant materials, both coarse and fine, than the other two clusters. AADT (traffic volume) had a significantly higher relative importance for the model trained on Cluster 2, which had the highest average RAP content of the three clusters. Model 3A showed the highest relative importance of effective binder content. A comparison of the feature importance scores for the three models is shown in Figure 3.

Figure 3



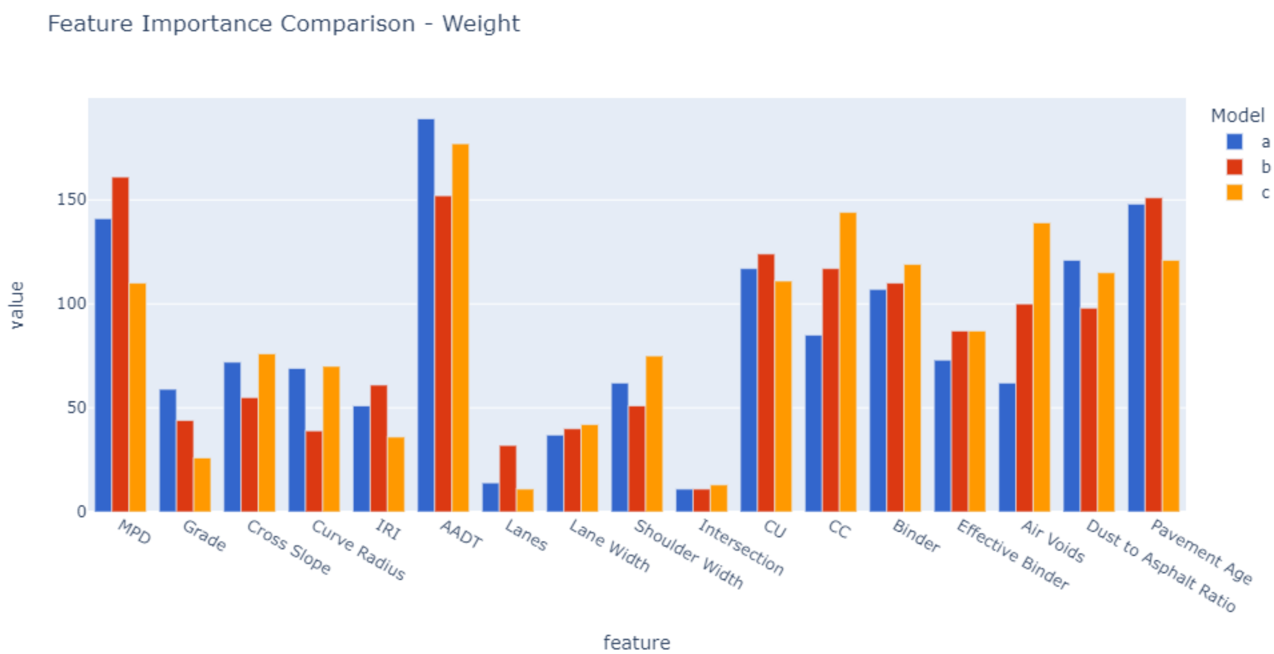
In terms of coverage, the relative importance of the Intersection feature was shared between all three models. Models 3b and 3c shared a higher importance for Curve Radius. A comparison of the feature importance scores for the three models is shown in Figure 4.

Figure 4



In terms of weight, all three models shared a relatively high importance of AADT, MPD (texture), and age of pavement. Model 3C showed a higher importance than the other models for the percent air voids and CC, but less importance for the pavement age. This model was trained on the cluster with the lowest average pavement age and lowest average traffic volume. A comparison of the feature importance scores for the three models is shown in Figure 5.

Figure 5

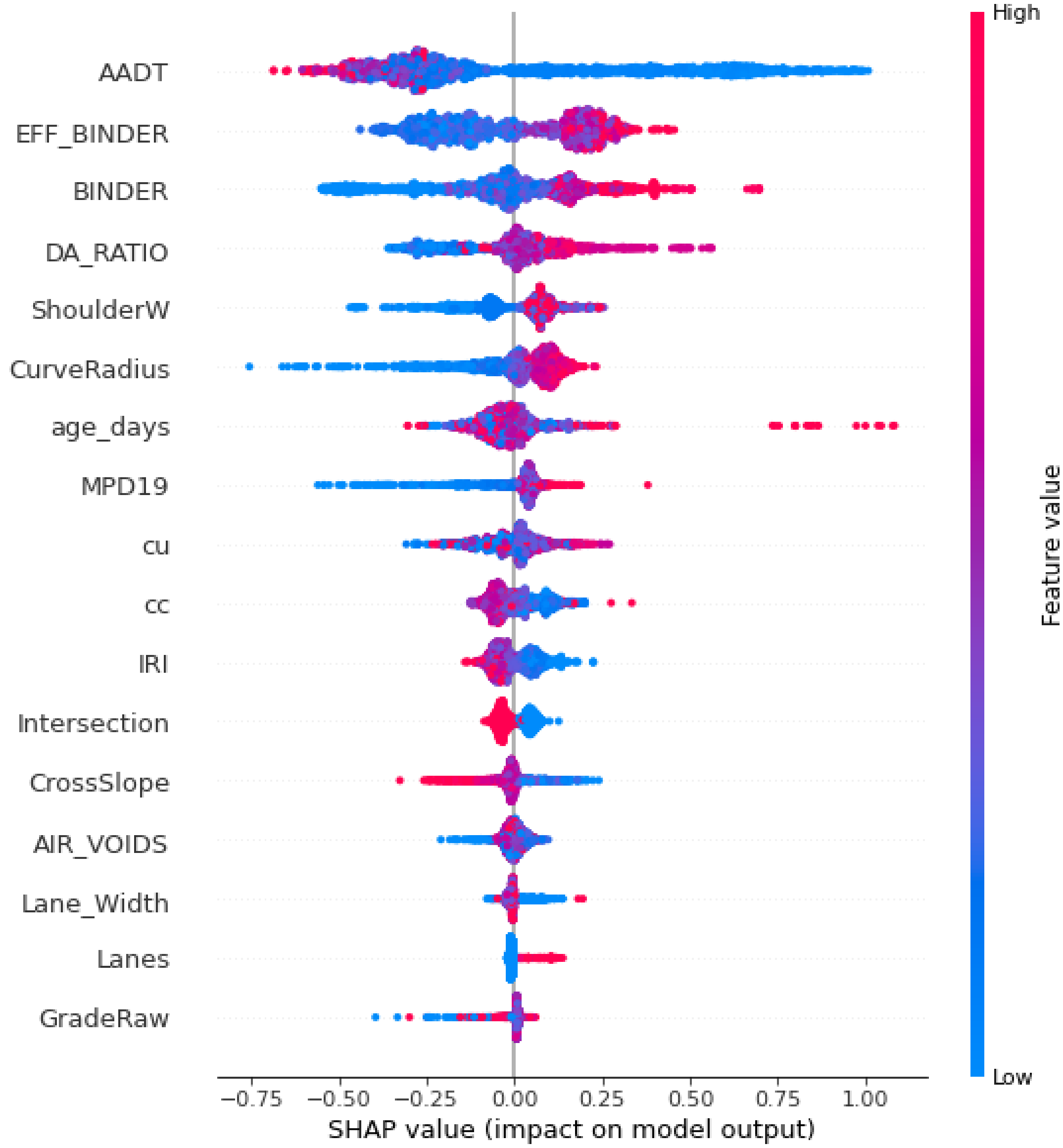


In addition to evaluating the relative importance metrics by each type of contribution, it is helpful to have a single metric that evaluates the relative global contribution of each feature across the entire model. As seen with the built-in metrics – gain, coverage, and weight – a feature might have a high importance in terms of what level it was used to split the data, but a low importance in terms of the contribution to the accuracy of the model. One way to accomplish this is by using the Shapley Additive Explanations (ShAP) approach where the model performance with or without each feature is used to evaluate its relative importance. ShAP calculates a local feature importance for every observation, then aggregates these values into a global feature importance metric that represents the "average marginal contribution of a feature value across all the possible combinations of features" [12]. The added benefit of this method is in the generation of a summary plot for each model that shows the distribution of the impact of each observation for the various features on the output, with the value encoded in red (high value) to blue (low value). Figures 6, 7, and 8 show the summary ShAP plots for each of the models, where the position of each datapoint along the x-axis shows the relative contribution to the predicted value of the final model output.

By comparing the relative impact of the various elements of a pavement mix design on the predicted friction value for the different clusters that would not be available for a model built on the entire unsegmented dataset. For example, we know that clusters 0 and 2 are similar, differing primarily in RAP content and pavement age. It is apparent from the ShAP plots that the relative importance of the effective binder content and percent air voids variables for each cluster is also significantly different. Cluster 0, with the oldest average age and the lowest RAP content, had a much higher relative importance for the effective binder content and low importance for percent air voids. Higher effective binder content contributed to a higher predicted friction/MSC value. In comparison, the impact of effective binder content on cluster 2 was much less significant on the predicted friction value while the percent air voids had a significant impact. One possible interpretation is that the percent air voids in a mix design is an important predictor early in the life cycle of a pavement, while effective binder content becomes more significant later in the life cycle.

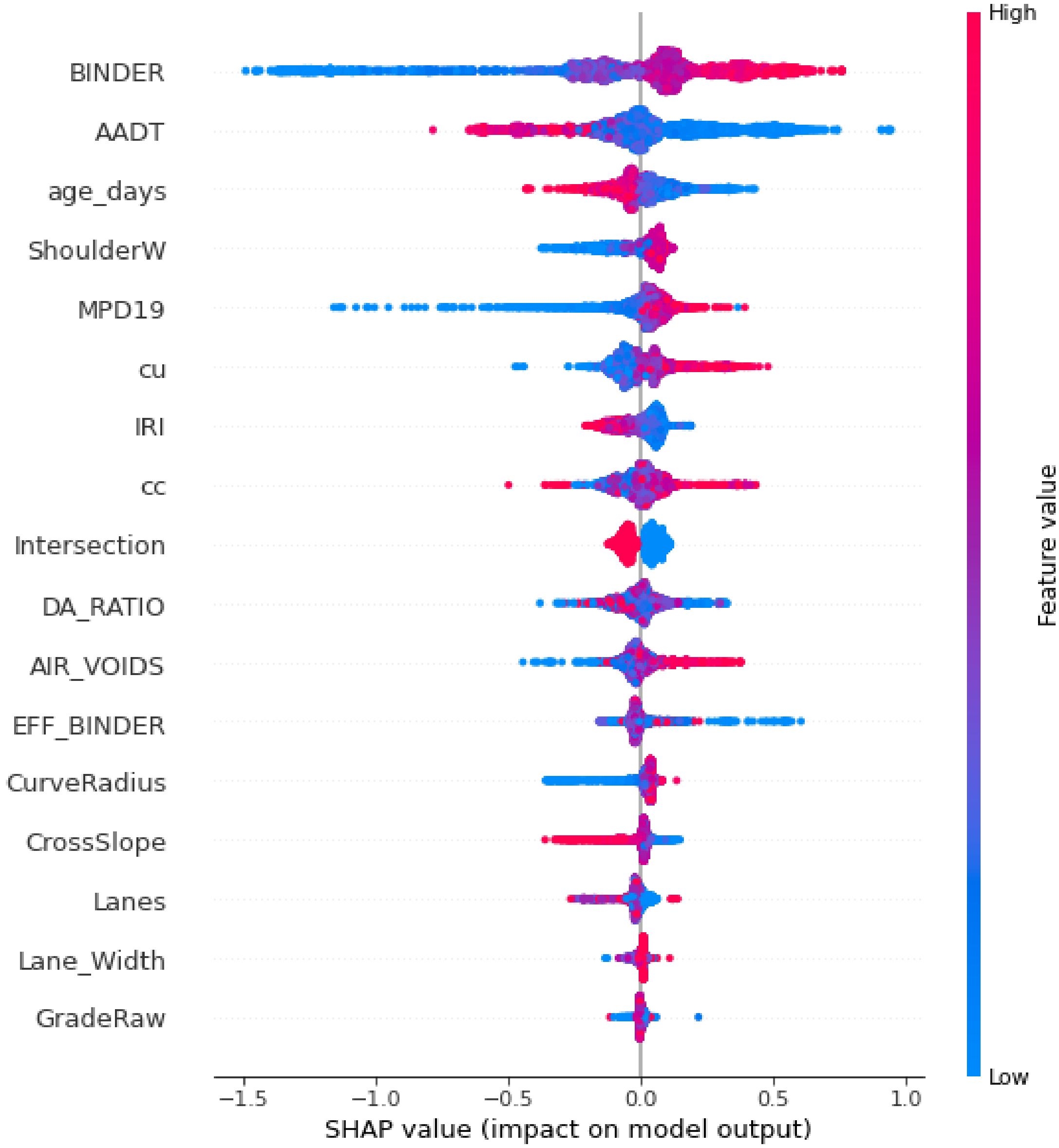
We can also see that for cluster 1, which contained road segments with a higher proportion of polish resistant materials and elements that describe “interstate-like” road segments (e.g., higher AADT, higher speed limit, more lanes, wider shoulders, etc.), the associated model relied more heavily on the percent binder content in the mix when predicting a friction value when compared to the models build on clusters 0 and 2.

Figure 6



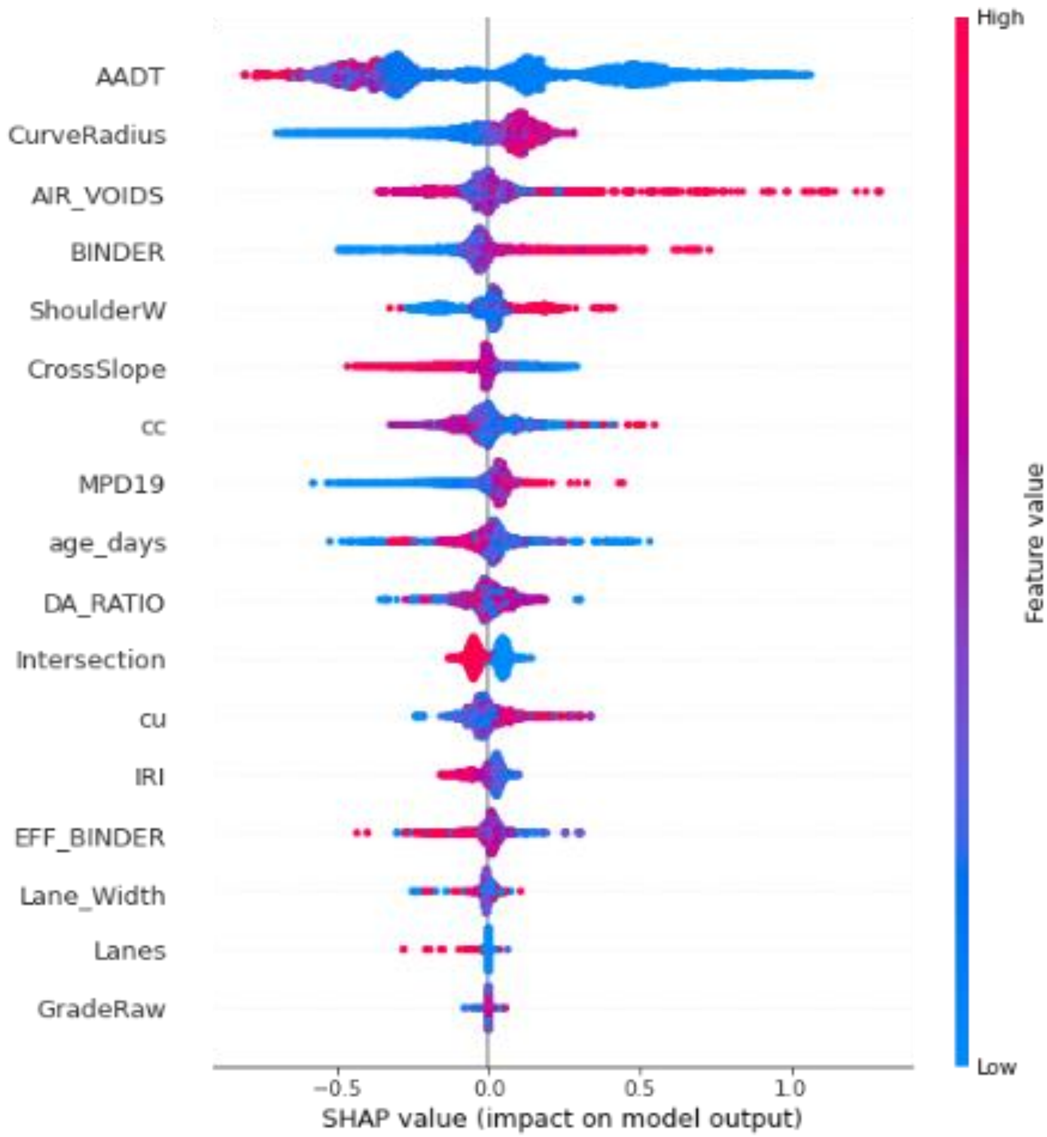
Model a
(Cluster 0)

Figure 7



Model b
(Cluster 1)

Figure 8



Model c
(Cluster 2)

Conclusions

Initial results suggest that the performance of a friction prediction model is improved when integrated with the results of a clustering analysis. While marginal performance improvements were achieved when cluster membership was included as a feature of a single model, more significant performance improvements were realized when cluster membership was used to determine subsets of the full dataset on which to build separate models. While the comparable performance between the test and training sets did not suggest an issue of severe overfitting, it remains an important consideration in future iterations of this work.

In addition to the improved performance of the models, the behavior of the features of each cluster in its associated model provides additional insight into the contribution of various pavement mix design elements to friction performance that would not be available from a model built on a single dataset. In addition to identifying significant differences between these models/clusters, this method also provides the means of identifying features which are important to all three models (i.e., traffic volume) which can support feature selection decisions in similar applications and analyses.

Discussion

The varying performance of pavement mix designs between clusters has the potential to support accurate modeling to predict the time to maximum friction value after initial application. Evaluating the full pavement friction life cycle requires an estimate of initial friction value, estimated maximum friction value, estimated time to maximum friction value, and estimated time to return to baseline friction value. A dataset that contained repeated measurements throughout the lifecycle would benefit from targeted analysis on subsets of the data identified through a similar clustering analysis. As indicated by the results of this project, the performance of a suite of friction prediction models is significantly improved when training separate models using the results of a clustering analysis. There is an opportunity for future work to determine whether a friction degradation model structured in a similar fashion would result in comparable improvements in performance.

In addition to contributions to a friction prediction model, the results of the clustering analysis have potential applications to the development of a pavement friction management program and maintenance project prioritization. By analyzing subsets of the road network as identified in the clustering analysis, road authorities can develop targeted intervention and treatment plans that consider the varying behavior of pavement sections belonging to different clusters. Clustering analysis can also be useful in identifying outliers, such as road segments that are performing outside the expected range when compared to other road segments belonging to the same cluster. This could provide a first step in investigating the root cause of in-service performance issues, such as characteristics of the mix design, variations in raw materials from a particular supplier/source, or other contributing factors not readily apparent in the raw data for the entire network.

Finally, the relative importance of the features included in the individual models can provide direction for the development of new pavement mix designs for testing, based on the general characteristics of the aggregate ingredient proportions. For example, an investigation into the contribution of RAP to the performance of a pavement mix design might initially focus on comparing mix designs with varying levels of air voids over the effective binder content.

References

1. Federal Highway Administration (FHWA). 2010. “Pavement Friction Management Technical Advisory”, T 5040.38, https://www.fhwa.dot.gov/pavement/t504038.cfm
2. US Department of Transportation. (n.d.). “Pavement friction management”. Pavement Friction Management | FHWA. https://highways.dot.gov/safety/proven-safety-countermeasures/pavement-friction-management
3. Federal Highway Administration (FHWA), “Zero Deaths and Safe System”. Retrieved 2023, from https://highways.dot.gov/safety/zero-deaths
4. Federal Highway Administration (FHWA). 2023. Flintsch, G., Izeppi, E. de L., McCarthy, R., Katicha, S., Persaud, B., Guo, F., Medina, A., Tobias, P., “Characterizing Road Safety Performance Using Pavement Friction”. Virginia Polytechnic Institute and State University.
5. Sharafeldin M, Farid A, Ksaibati K. 2022. “Investigating The Impact of Roadway Characteristics on Intersection Crash Severity”. 3(4):412-423. https://doi.org/10.3390/eng3040030
6. Yang, Yanhai & Qian, Baitong & Xu, Qicheng & Yang, Ye. 2020. Climate Regionalization of Asphalt Pavement Based on the K-Means Clustering Algorithm. Advances in Civil Engineering. 1-13. 10.1155/2020/6917243
7. Sandra, A.K., Sarkar, A.K. 2015. “Application of Fuzzy Logic and Clustering Techniques for Pavement Maintenance”. Transp. Infrastruct. Geotech. 2, 103–119. https://doi.org/10.1007/s40515-015-0021-z
8. Federal Highway Administration (FHWA), Office of Infrastructure Research and Technology, McLean, VA. Izeppi, Edgar de León, et al. 2019. “Locked-Wheel and Sideway-Force CFME Friction Testing Equipment Comparison and Evaluation Report”.
9. scikit-learn. 2007. 2.3. clustering. 2.3 Clustering. https://scikit-learn.org/stable/modules/clustering.html#k-means
10. North Carolina State University. North Carolina Department of Transportation, Underwood, S., Castorena, C., Goenaga, B., Rogers, P., 2022. “Evolution of pavement friction and macrotexture after asphalt overlay” 31–43.
11. Python API reference. Python API Reference - xgboost 1.7.5 documentation. 2022. [Retrieved October 2022], from https://xgboost.readthedocs.io/en/stable/python/python_api.html
12. Rathi, P. 2020. “A novel approach to feature importance - shapley additive explanations”. Medium. [Retrieved March 2023]. https://towardsdatascience.com/a-novel-approach-to-feature-importance-shapley-additive-explanations-d18af30fc21b