STREETLIGHT InSight

StreetLight AADT 2019 Methodology and Validation White Paper

United States

Version 2.0
MAY 2020

Table of Contents

Introduction: Why We Developed an AADT Metric

Annual Average Daily Traffic (AADT) volumes are the core input to many important parts of the transportation practice. With new federal government requirements put in place by the Moving Ahead for Progress in the 21st Century Act (MAP-21), local and state government agencies must increase reporting of performance metrics for planning and funding allocation. This in turn has increased the need for a quick, easy, cost-effective AADT measure and its extension, Vehicle Miles Traveled, or "VMT." Modeling AADT counts leads to inaccuracies, as does temporary count expansion. Additionally, temporary counts are cumbersome and expensive. As a result, StreetLight Data developed an accurate AADT for over four million miles of urban and rural roadway in the U.S. that outperforms industry-standard accuracy targets and can be used by industry practitioners for traffic impact studies and more.

What's New in AADT 2019 V2

StreetLight's AADT 2019 V2 metric is an update to our previously released AADT 2019 V1 metric. Some highlights and improvements of the new metric are:

- Training data was captured from 48 states across the U.S., an improvement from the 29 states used in the 2019 V1 model.
- Training data was increased to 10,785 permanent counter locations across the U.S., an improvement from 6,029 used in the 2019 V1 model.
- A "hybrid" model was implemented, combining a random forest model with two linear regression models at the high and low tails. This is meant to improve performance on low and high-volume roads, and better handle extremes.
- Selected features were adjusted across hybrid models to allow for improved performance.
- Overall estimation accuracy has significantly improved from prior versions, especially on lower-volume roads.

Defining Target Accuracy

Our first question was to define what was "accurate enough" for our AADT. We had to balance the need for accuracy with the need to build an algorithm that could calculate an AADT very quickly while being compatible with the StreetLight InSight® transportation analytics platform. This meant that the algorithm had to be computationally elegant and scalable.

Our goal was to develop AADT estimates that are more accurate than a temporary count expansion. A paper by researchers by the University of Texas found that the absolute error for temporary count expansion ranges from 4.9 - 83.2%. The mean absolute percentage errors range from 11.5 - 18.5%, depending on rural/urban divide and functional class[1]. Ranges were not available by AADT class, so we set our overall target stretching from 20% for lower AADT bins to 12% for higher AADT bins. The equation for mean absolute percentage error (MAPE) is shown below.

$$MAPE = \left(\frac{1}{n}\right) * \sum_{i=1}^{n} 100 * \left| \frac{AADT_{Estimate(i)} - AADT_{Permanent\ counter(i)}}{AADT_{Permanent\ counter(i)}} \right|$$

We also wanted our metrics to be as accurate as AADT estimated from a model. There are more sources on acceptable error in this body of work, as summarized by the Travel Modeling Improvement Program (see Figure 1). All the sources evaluate error in terms of Root Mean Square Error (RMSE) as a percent of AADT. The equation is:

$$RMSE\ as\ \% = 100 * \frac{\sqrt{\left(\frac{1}{n}\right) * \ (Traditional\ AADT - Big\ Data\ AADT)^2}}{\left(\frac{1}{n} * \sum_{i=1}^{n}(Traditional\ AADT)\right.}$$

As you can see in Figure 1, below, many DOTs generally agree on the acceptable RMSE. We took the average of acceptable RMSEs for each AADT range as the target for our results (we merged the 0 - 1,000 and 1,000 - 2,500 ranges due to low availability of permanent counters). The targets for RMSE as % of AADT range from 21% for high AADT to 47% for lower AADT bins. Note – just because the RMSE values are higher than the MAPE does not mean the RMSE-conveyed results are “worse” – they are simply different metrics for cross validation, and RMSE mathematically imposes a heavier penalty on errors.

[1] See Table 2 in: Gadda, S., A. Mangoon, and K. Kockelman. *Estimates of AADT: Quantifying the Uncertainty,* 11th World Conference on Transport Research, Berkeley CA, 6-24-2007 to 6-28-2007.

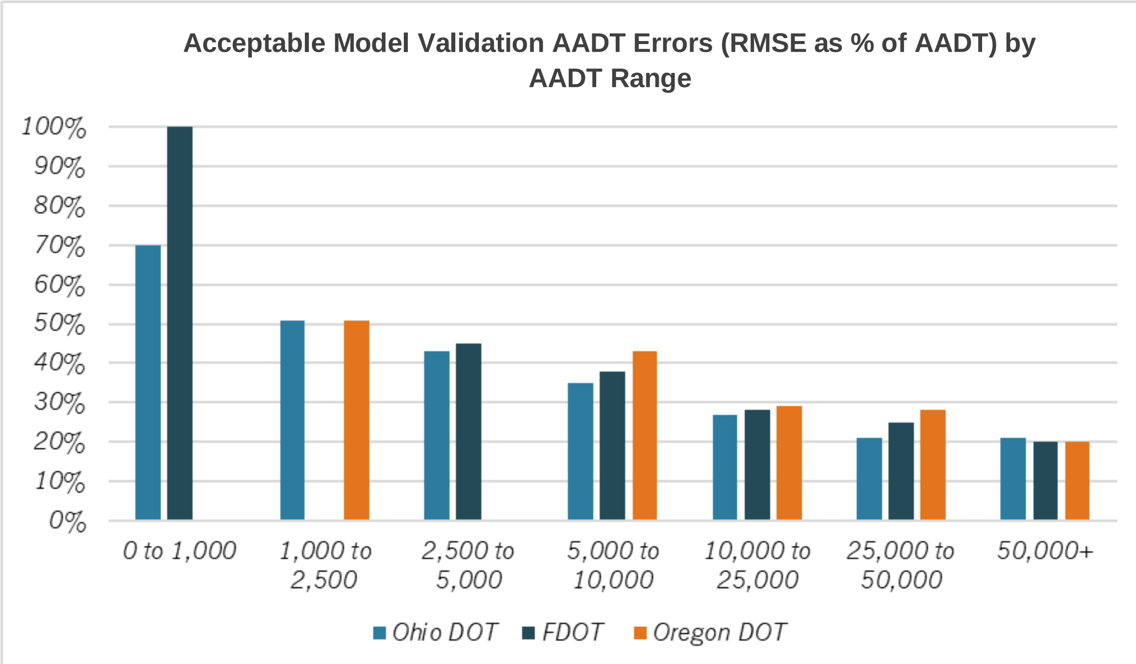

Figure 1: From Travel Model Improvement Program, Travel Model Validation and Reasonableness Checking Manual Second Edition*, 2010.*

Calculating the Accuracy of StreetLight's AADT 2019 V2 Metric for the U.S.

We trained our model on a set of 10,785 permanent AADT counters with outliers removed from 48 states across the U.S. To evaluate our model performance, we used a technique called "leave-one-state-out cross validation," training our model on a set of counters from 47 different states. The model was tested 48 times, each time blind to the permanent counters of a different test state. This allows us to estimate the expected performance of our model for novel roads across the country.

As shown in Figure 2, the actual and estimated AADT values through the cross validation are highly correlated with an extremely high R^2 and the fitted line is very close to 45-degree line, which indicates that the performance of our model is excellent without bias.

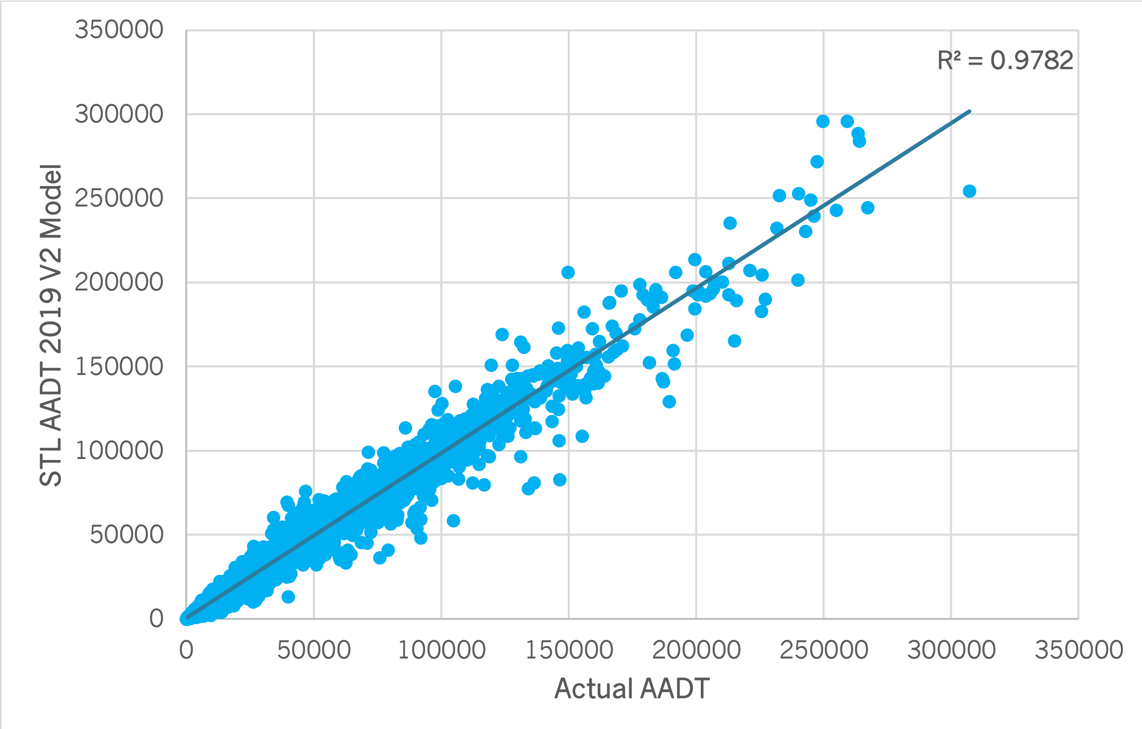

Figure 2 – StreetLight AADT 2019 V2 for test data compared to permanent counter AADT (U.S.).

Tables 1a and 1b show the MAPE and RMSE of StreetLight's AADT 2019 V2 compared to target errors.

AADT Range	# of Segments	Target Abs. Error	StreetLight Algorithm Mean Abs. Error	Delta to Target (positive means "better than target")
50,000+	1286	12%	8.25%	3.75%
25,000 - 49,999	1540	16%	9.46%	6.54%
10,000 - 24,999	2578	20%	10.21%	9.79%
5,000 - 9,999	1787	20%	13.32%	6.68%
2,500 - 4,999	1498	Not available - too few in comparison paper. See Table 1b.	16.33%	NA
0 - 2,499	2017		24.92%	NA

Table 1a: Key results from StreetLight's 2019 V2 national validation test – MAPE (U.S.).

AADT Range	# of Segments	Target RMSE as % of Average AADT	StreetLight Algorithm's RMSE as % Average AADT	Delta to Target (positive means "better than target")
50,000+	1286	21%	12.64%	8.36%
25,000 - 49,999	1540	22%	13.71%	8.29%
10,000 - 24,999	2578	25%	13.76%	11.24%
5,000 - 9,999	1787	29%	17.78%	11.22%
2,500 - 4,999	1498	36%	21.18%	14.82%
0 - 2,499	2017	47%	29.29%	17.71%

Table 1b: Key results from StreetLight's 2019 V2 national validation test – RMSE (U.S.).

Table 1c below shows the percentile of percent errors compared to actual AADT. The 5th and 95th percentiles form a confidence interval that illustrates how extreme the percent errors can be on two tails, as well as how much over- or under- estimation might be occurring in the model. This breakdown is valuable in that it provides more information on the range of expected errors, rather than just averages as provided by the MAPE. For example, the 95th percentile bin tells us that for roads with AADT above 50,000, we would expect only 5% of results to have an error greater than 16.89%. Overall, we are pleased with these results and see that most categories are unbiased.

AADT Range	# of Segments	5th Percentile	50th Percentile	95th Percentile
50,000+	1286	-18.63%	-0.67%	16.89%
25,000 - 49,999	1540	-18.61%	0.84%	20.42%
10,000 - 24,999	2578	-21.59%	0.97%	22.32%
5,000 - 9,999	1787	-26.51%	3.07%	29.46%
2,500 - 4,999	1498	-30.76%	5.19%	37.17%
0 - 2,499	2017	-27.96%	13.34%	66.54%

Table 1c: Key results from StreetLight's 2019 V2 national validation test – percentile breakdown of percent error (U.S.).

Our AADT 2019 V2 metric includes a range with 90% confidence for each individual outcome to locate the true AADT value. Our confidence range (also known as Prediction Interval) is an estimate of an interval in which the estimated AADT value will fall in 90% probability. In Table 1d, we summarize the 90% confidence ranges for different road sizes.

AADT Range	# of Segments	Mean Lower Boundary of 90% Confidence	Mean Upper Boundary of 90% Confidence	Actual AADT Within Confidence Range
50,000+	1286	-17.50%	20.58%	89.19%
25,000 - 49,999	1540	-18.75%	24.58%	90.06%
10,000 - 24,999	2578	-21.13%	27.23%	91.12%
5,000 - 9,999	1787	-26.12%	37.80%	92.00%
2500 - 4,999	1498	-30.36%	48.34%	91.72%
0 - 2,499	2017	-38.64%	65.74%	91.52%

Table 1d: Key results from StreetLight's 2019 V2 national validation test – 90% confidence range (U.S.).

In addition to far exceeding our desired targets for all classes of roads, which indicates that our model performs well on both small and large roads, our AADT 2019 V2 metric is able to provide confidence ranges for AADT values. Our confidence interval estimates show that the percentages of real-world AADT values that fall within the upper and lower confidence boundaries of our AADT metrics are high, all above 89%. This supports the validity of our confidence range across different road types. We welcome opportunities to conduct additional validation studies for customers who are interested in the results for their specific locales.

Comparing Cross-Validation Results Among AADT 2018, 2019 V1, and 2019 V2

When comparing our previous AADT metrics, we see that our AADT 2019 V2 model performs better than both the 2019 V1 model, and the 2018 model. In the table below, we compare the cross-validation results of the RMSE values among AADT 2018, 2019 V1, and 2019 V2. Through the comparison, it can be seen that AADT 2019 V2 improves results across all road sizes, and particularly improves results on lower volume roads with AADT values below 5,000. This improvement in 2019 V2 is due to increased availability of high-quality counter data in 2019, as well as an updated modeling methodology.

AADT Range	Target RMSE as % of Average AADT	Delta to Target (positive means "better than target")		
		AADT 2019 V2	AADT 2019 V1	AADT 2018
50,000+	21%	8.36%	3.76%	5.18%
25,000 - 49,999	22%	8.29%	5.16%	3.92%
10,000 - 24,999	25%	11.24%	6.95%	5.88%
5,000 - 9,999	29%	11.22%	6.52%	7.58%
2,500 - 4,999	36%	14.82	7.03%	11.96%
0 - 2,499	47%	17.71%	5.93%	15.85%

Table 2: Cross validation results of the RMSE metric among AADT 2018, 2019 V1, and 2019 V2 (U.S.).

Data Sources Used in Our AADT 2019 Metric

Our AADT 2019 blends together the following data sources to create our best prediction of Annual Average Daily Traffic:

Input 1: Location-Based Services Trips Data

Location-based services (LBS) data are created by smartphone applications providing a service that depends upon on a device's geographic location in the physical world — for example, shopping apps, weather apps, or dating apps. We use algorithmic processing techniques to link these data points into trips. We sampled trips throughout 2019 in order to create the best model possible for the 2019 calendar year. Our 2019 LBS sample size increased by 2-4x from 2018.

Input 2: Navigation GPS Trips – Commercial

The navigation-GPS data we use is created by connected commercial vehicles. Our data set is tagged by vehicle type: heavy-duty commercial vehicle, or medium-duty commercial vehicle. Since roads vary heavily in the share of commercial trucks (and in the share of medium vs. heavy duty trucks), having a combination of data sources from commercial vehicles is critical.

Input 3: U.S. Census Data

We normalized our LBS trips using the U.S. Census. Normalizing is an important step to adjust a sample that is not perfectly distributed. In short, if ten devices in our sample "live" on a block with 100 people, each of those devices is scaled up by a factor of 10. If ten devices "live" on a block with 50 people, each is scaled by a factor of 5. This adjusts for variation in geographic distribution, which is correlated with demographic factors like income. We also looked at the population density near the road in question (which is a proxy for identifying a road as rural or urban), as well as employment and income data pertinent to the surrounding area. To better understand this method, please read the analysis on our website, "Evaluating Location-Based Services Data for Transportation: Is Our Sample Representative?"

Input 4: Open Street Maps Data

We included features commonly extractable from Open Street Maps (OSM) such as road geography, speed limits, number of lanes, availability of parking, road classification, and other factors. We know all OSM features are not always available for every road. Our algorithm is factored to adjust to a different set of coefficients if no OSM feature data is available. We also use the OSM to "lock" a trip to a route by connecting pings along the most viable network path a vehicle can take.

Input 5: Weather Data

We included data on precipitation and temperature to account for areas that have extreme weather events (like snow storms) on a regular basis and might experience different travel patterns as a result.

Input 6: Training and Testing AADTs Using Permanent Loop Counters

We researched extensively to find well-cleaned permanent counter data. We wanted our data to be spread across the U.S., between small and large roads, urban and rural. The biggest challenge was finding permanent counter data for small rural roads. The following maps and charts show the locations of the 10,785 counter data points in the U.S. that we used to develop our algorithm. We were able to include count locations in all U.S. states, with the exception of Alaska, Hawaii, and Washington D.C.

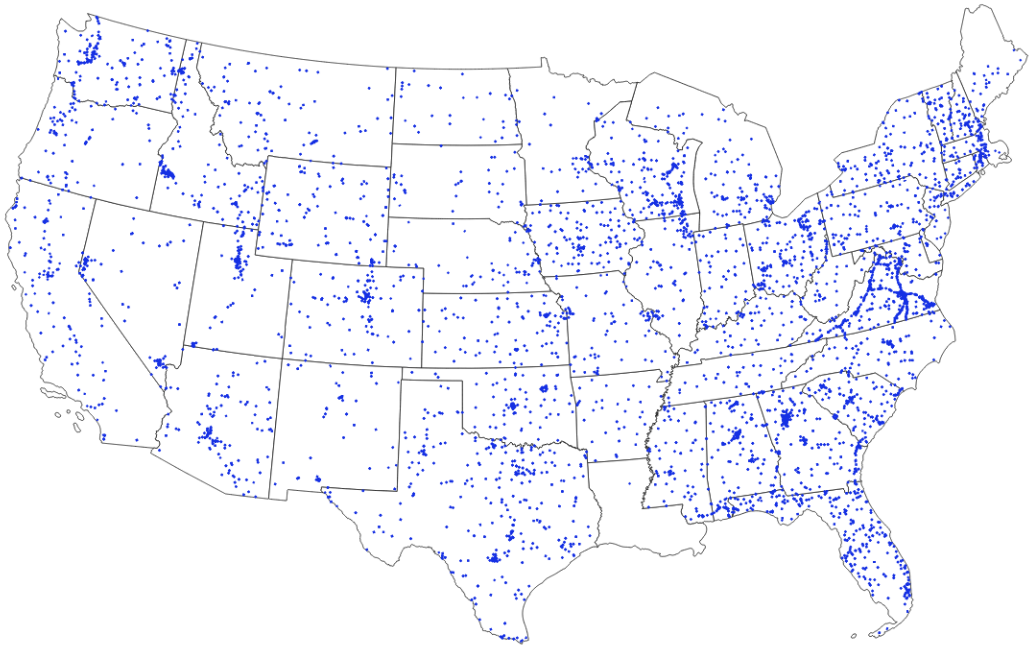

Figure 3: Map of all permanent counters in the U.S. used for training and testing the AADT 2019 V2 metric.

AADT Range	# Permanent Counters U.S.
50,000+	1286
25,000 - 49,999	1540
10,000 - 24,999	2578
5,000 - 9,999	1787
2500 - 4,999	1498
0 - 2,499	2017

Table 3: AADT range spread for all permanent counters used for training and testing the AADT 2019 V2 metric.

Selecting and Testing the Algorithm

We weighed dozens of different algorithmic approaches when developing our AADT 2019 V2 metric. In this section, we will provide an overview of our major decisions.

First, we checked to see how our normalized LBS trips, which comprise our strongest and largest data set, were correlated with AADT. The results are shown below. As you can see, the correlation is fairly strong. As such, the remainder of the task was to use machine learning to reduce error and improve correlation (compared to scaling to counts from LBS trips alone).

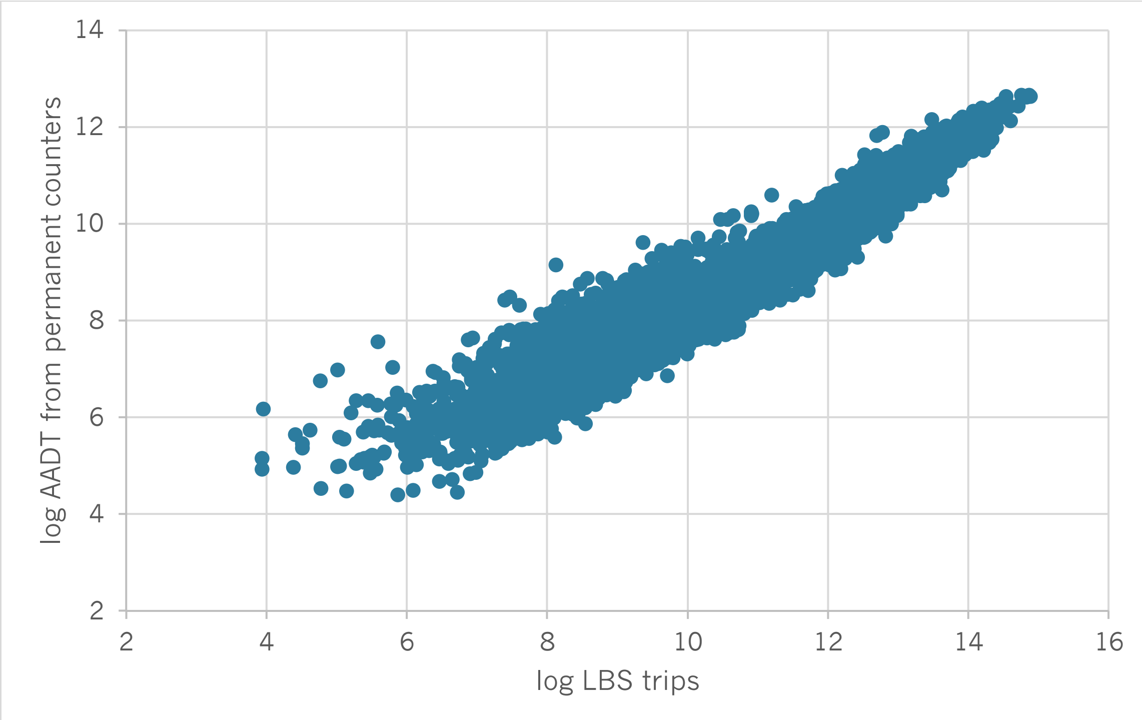

Figure 4: Correlation of population-normalized LBS to permanent loop counter data – the single best predictor value. The rest of the machine-learning work aimed at improving these results.

The choice came down to two options for machine learning techniques: Ordinary Least Squares (OLS) and random forest.

First, we tried OLS, a multivariate equation framework with machine learning. The benefit of a multivariate regression technique is that it is easier to explain, as it is more-or-less building a classic y = mx + b style equation. The disadvantage, as we found, is that the results were not as accurate as we wanted and the model was prone to throwing outliers.

We also tried a random forest model (for more information, see our summary in Box 1).

Box 1: The Random Forest Model

This approach is similar to a decision tree, but it uses several decision trees. For example, let's predict whether or not a patient entering an emergency room is at high risk. A decision tree may look like this: If age is over 50, blood pressure is over 150, and temperature is above 100F, then the patient is high risk. That's a decision tree. It is very interpretative, but does not have much predictive power alone. Random forest uses a lot of decision trees (say, an ensemble), where each tree is a little bit different from the others. When a new patient arrives, we take the majority vote of the decision tree ensemble to get a final result.

The different trees use random samples of observations and subset of features to train. For example, instead of considering age, blood pressure, and temperature, we may train one tree with age and blood pressure, another with blood pressure and temperature, another with age and temperature, and so on if we had more features. The key is that the trees become a bit different (less correlated), so when we average the results, we get a "diverse" answer. The idea behind this model is that a bunch of poor decision makers put together in a room to form a committee will start making better decisions. If each decision maker comes with a different perspective, that creates better results.

The benefit of a random forest model for AADT estimation is that it is more accurate: In particular, it did a far better job of handling unusual roads (such as small ones or ones with extremely high commercial traffic near ports/warehouses). In the end, we think that the accuracy and algorithmic robustness for unusual roads and outliers were more important.

In previous versions of our metric, we relied exclusively on the random forest model to predict our AADT values. With our 2019 AADT V2 metric, we explored the benefits of using a hybrid-model approach, where the bulk of the results are predicted with a random forest model, but the tail ends of the model are predicted using linear regression. After outlier removal, 10,785 locations were trained with the random forest model, 198 of those locations were used for the lower linear regression, and 337 of those locations were used for the upper linear regression.

This hybrid model provides a few benefits. First, it allows us to predict data beyond the range of permanent counters in our training data set. This means that if the highest AADT value in our permanent counter data set was 200,000, we'd still be able to predict values above that threshold. Second, we found that the random forest model alone struggled with performance on lower volume roads. Having a separate linear model specifically for these lower volume roads improved our prediction capabilities, as shown earlier in the paper. Figure 5 below illustrates this new framework, where the lower linear regression model was trained on sites with AADT under 400, and the upper linear regression model was trained on sites with AADT above 100,000.

Figure 5: Structure of the hybrid model approach where a random forest model is combined with two linear regression models of the upper and lower tails. The model can predict values between 0 and infinity.

The next step was to decide which features (input variables) to include in our random forest and linear regression models. We tested hundreds of combinations. We wanted the most accurate results, but we wanted our algorithm to be scalable to anywhere in the U.S., and be computationally efficient.

In the end, we built three models that relied on a combination of 23 features captured from different data sources. We strive to avoid over-fitting of Big Data by throwing far too many features into a machine-learning model. This may make initial results look very good, but it also prevents the approach from scaling well outside of the research setting.

After we made the algorithm and feature choices for our model, we performed several types of cross validation. First, we ran the model several times, each time randomly designating different zones as “training set” and “test set.” If the model output is very similar for each of these runs (or “folds”), it is a good model.

Next, we performed another set of folds where we pulled out each state in turn as the “test” set. This was meant to mimic a new StreetLight InSight® user’s experience because we trained our model on a set of data from a group of states. We wanted to know whether our model would perform well if someone ran AADTs in a new location. The answer again is yes, we showed a strong match of our results to permanent counter data.

We did find that the state-based folds had more variation than the random folds. This means that we may be able to further improve results by tweaking the algorithm for an individual state or region. We are look forward to collaborating on this approach with customers who want to use our AADT metric at large scale.

Running AADT 2019 V2 in StreetLight InSight

If you have a StreetLight InSight account with AADT enabled, choose to create a new analysis. Click the “Create Analysis” button under the AADT analysis option. Name your analysis, select the zone sets covering the roads of interest, and choose the AADT year “2019.” Then click “Confirm Analysis” to begin processing.

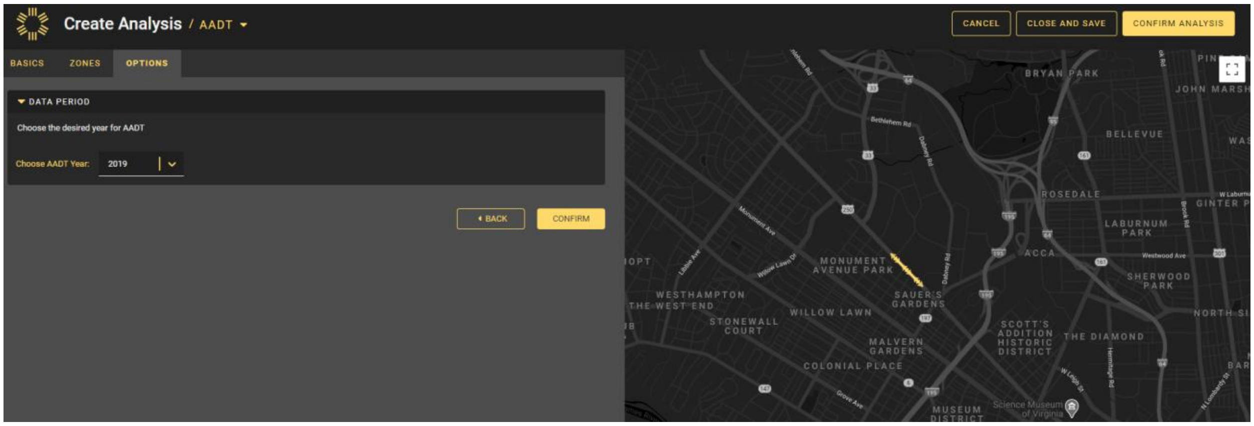

Figure 7: AADT 2019 V2 as represented in the StreetLight InSight® software platform.

Our Next Steps

In future StreetLight InSight releases, we will work to improve processing speed so that AADT analyses complete more quickly within the platform. In addition, we will continue conducting validation and improvement studies on our AADT metrics.

About StreetLight Data

StreetLight Data, Inc. pioneered the use of Big Data analytics to help transportation professionals solve their biggest problems. Applying proprietary machine-learning algorithms to over 100 billion location data points every month, StreetLight measures multimodal travel patterns and makes them available on-demand via the world's first SaaS platform for mobility, StreetLight InSight®. From identifying sources of congestion to optimizing new infrastructure to planning for autonomous vehicles, StreetLight powers more than 3,000 projects every month.


STREETLIGHT
DATA