

World Conference on Transport Research - WCTR 2023 Montreal 17-21 July 2023

Forecasting traffic flow by vehicle category on a major highway impacted by road maintenance works

Eline A. Belt^{a,c}, Lili Yao^b, Xiaoyue Li^b, Nan Wang^b, Varun Shekhar^b, Thomas Koch^{a,b},
Elenna R. Dugundji^{a,b,*}

^aCentrum Wiskunde & Informatica, Science Park 123, 1098 XG Amsterdam, The Netherlands

^bMassachusetts Institute of Technology, 1 Amherst St., Cambridge, MA 02142, USA

^cVrije Universiteit Amsterdam, De Boelelaan 1105, 1081 HV Amsterdam, The Netherlands

Abstract

In this paper we look at the forecasting of traffic flow on a major highway in the Netherlands impacted by road maintenance works, examining the effects of lane closures on intensities per vehicle category. We apply several forecasting methodologies such as Prophet, Harmonic Regression, Seasonal Autoregressive (SAR), and Seasonal Autoregressive Integrated Moving Average (SARIMA) and compare them against a seasonal naive baseline model. We observe that SARIMA performs better than other models across all forecasting metrics for all sensors. This is mainly because of its capability of capturing linear trends and seasonality. There is also an opportunity to further improve the forecast accuracy of the SARIMA model by incorporating holiday information as seen in the Prophet model. Overall, the analysis showed that road works and holidays are two features that have more influence on traffic flow, which should be considered as main factors when carrying out future road plans. If multiple areas are affected, the K-means model can be adopted effectively to cluster the sensors into groups to minimize traffic disruption.

© 2024 The Authors. Published by ELSEVIER B.V.

This is an open access article under the CC BY-NC-ND license (<https://creativecommons.org/licenses/by-nc-nd/4.0>)

Peer-review under responsibility of the scientific committee of the 16th World Conference on Transport Research

Keywords: Highway traffic forecasting; SAR; SARIMA; Prophet; Harmonic regression; Road maintenance; Vehicle category

1. Introduction

Road network maintenance is a pervasive activity across countries that not only troubles every commuter on a micro level but also indirectly impacts economic activities. As much as we would like to avoid the inconvenience brought by road maintenance, we foresee more and frequent maintenance events scheduled in the future because of urbanization and climbing disposable income. Considering the broad impact on people's daily life and the economy, the government should carefully plan each roadwork. Leveraging massive data points collected from road sensors and applying machine learning methodologies, it is likely to optimize roadwork to minimize the impact on citizens and businesses. With five important European seaports in the Netherlands, numerous trucks and cars are on the road days and nights to transport commodities, causing unavoidable road construction projects to maintain a strong and well-

* Corresponding author.

E-mail address: elinebelt1@hotmail.com

designed road network. The Ministry of Infrastructure, responsible for traffic management and road design, would like to understand the impact of road maintenance activities on traffic flow, including the main and underlying road network, and hear recommendations and advice for future road work plans.

2. Literature Review

The problem of traffic forecasting has been studied before. There are many different types of models to create these forecasts; statistical and Bayesian models, neural networks, and hybrid models which combine both statistical methods and neural networks.

Some often used statistical methods are Autoregressive Integrated Moving Average (ARIMA) and a seasonal version thereof, Seasonal ARIMA (SARIMA). [Yu and Zhang \(2004\)](#) use a switching ARIMA model because conventional switching models do not describe the fluctuations of a traffic flow adequately. [Lin et al. \(2009\)](#) use a space-time ARIMA model to not only model the temporal aspects of the traffic flow but also the spatial aspects. [Kumar and Vanajakshi \(2015\)](#) propose a SARIMA model for short-term traffic prediction which needs only limited input data, as this is sometimes scarce.

More recently, different types of neural networks are used for this problem. [Zhao et al. \(2017\)](#); [Ma et al. \(2020b\)](#); [Awan et al. \(2020\)](#) use models based on long short-term memory (LSTM) networks. Graph neural networks are also used often because they utilize the spatial information of a road network, oftentimes a graph convolutional neural network is used, as shown by [Bogaerts et al. \(2020\)](#); [Diao et al. \(2019\)](#); [Mallick et al. \(2020\)](#); [Guo et al. \(2020\)](#).

Hybrid approaches of combining statistical methods with neural networks are also used to optimize forecasting accuracy. [Yang et al. \(2021\)](#) use an ARIMA-BPNN (Back Propagation Neural Network) which is optimized using Simulated Annealing (SA). This model can improve traffic prediction accuracy by using the linearity of ARIMA and the non-linearity of the BPNN. [Ma et al. \(2020a\)](#) use an approach they call NN-ARIMA, in which a Multi-Layer Perceptron (MLP) is used to identify the pattern of the traffic flow and ARIMA is used to process the residuals of the MLP to identify location-specific traffic features.

3. Case Study

3.1. Data Understanding

The data revolves around observations captured by sensors placed across highways, particularly in the regions of the country containing the Port of Amsterdam, Rotterdam and Schiphol Airport. This study will focus on sensors located on the A4 highway.

For our study, we are considering data from April and May 2021. The maintenance of the road was split into two phases, phase 1 and phase 2. Phase 1 consisted of the complete closure of the road, and during phase 2, the road was partially closed. This study will focus on phase two of the maintenance, which took place for three weeks starting from 5 AM on May 10, 2021, through 5 AM on May 31, 2021. Our study will analyze data and compare observations in this period with that in April to see whether the maintenance activity had any significant impact. The impact is studied in terms of average intensity, which is the average number of vehicles and average speed of the vehicles that a sensor had captured in the given measurement period. Vehicles observed by the sensors are placed in one of five different length classes, to simplify the analysis, we then combine these classes into two distinct categories. Table 1 shows these vehicle length classes and their corresponding category.

Table 1: Vehicle length measured by the sensor and the corresponding category.

Vehicle Length Measured (meters)	Category
1.85 <length <= 2.4	Small Vehicles
2.4 <length <= 5.6	Small Vehicles
5.6 <length <= 11.5	Trucks
11.5 <length <= 12.2	Trucks
length >12.2	Trucks

The data also consists of latitude and longitude information of the sensor location. Using the starter code we received, we were able to narrow in on the three sensors that were located on the A4 highway where the roadworks took place. The following three sensors will be the focus of our study:

- GEO0B_R_RWSTI358271
- GEO0K_K_RWSTI358317
- GEO0K_K_RWSTI358270

After this section, these sensors will be referred to as sensor1, sensor2, and sensor3, respectively. The data has five lane categories, namely: main carriageway, connecting carriageway, parallel carriageway, entry slip road, and exit slip road. The lane types would be helpful in identifying the preferred or most utilized lane types during the maintenance period compared to the non-maintenance period.

3.2. Data Validation

Before we began our exploratory analysis, we validated the data received in the following way:

- The initial dataset size is about 25M rows and 44 columns with 2788 unique sensors.
- We dropped 16 columns that could be ignored as per the data dictionary instructions.
- We identified that fields, 'quality_indicator_speed' and 'quality_indicator_intensity' had missing values for about 95% and 50% of total records respectively. These fields could be safely ignored.
- At any given point in time, a sensor would capture the vehicular movement in only one direction. Hence the 'direction' field is not considered for our analysis and more than 50% of the data does not have values for this field.
- 'Location name' field is ignored as we have more accurate information on sensor locations based on latitude and longitude.

3.3. Data Analysis

3.3.1. Statistical Analysis

Table 2 presents a preliminary understanding of the distribution of numeric data. From this table, we see that the mean average intensity is 126, and the median is 6 with a standard deviation of 276. This indicates a possible skewness in the data. To validate this, a density plot of average intensity was created as shown in Figure 1. As we can see, the data is positively skewed suggesting that the majority of observations lie between 0 and 100.

3.3.2. Regression Analysis & Heat Map Description

Reading previous data and supporting materials, we chose the avg_intensity variable to reflect the impact of the road works on the traffic flow. In other words, avg_intensity will be the dependent variable in the predicting model. We fit a linear regression model to check the correlation between the three variables: avg_intensity, avg_speed, and number_of_traffic_lanes. Although the result shows the statistical significance of two independent variables, the coefficient shows there is no strong correlation among these three. We also need to consider other factors. We created a heat map as shown in Figure 2 to double-check the correlation relationship between these three variables. The result matches the regression result. These three are not linearly related. The relationships among these three variables will be explored in future research.

3.3.3. Traffic Pattern Deep Dive

We explored the seasonality of sensors. To begin with, we considered one sensor - sensor1 - and explored average traffic flow by the same hour of the week (i.e., 168 data points for weekday-hour combination). Weekday late afternoons (3-6 PM) are the most peak hours, followed by weekday mornings (8-10 AM). Non-peak hours usually happen between 0-5 AM. This insight could be used for planning small maintenance work that requires blockage of the road for a few hours while minimizing impacts.

Furthermore, we also managed to show a breakdown of vehicles by lane type and the number of lanes, this is shown in Figure 3. The main carriageway is the busiest, followed by the parallel and connecting carriageway. Most vehicles

Table 2: Statistical inference of numerical columns.

	count	mean	std	min	25%	50%	75%	max
n_observations_intensity	24,774,506	41	141	0	0	0	3	7,293
n_observations_speed	24,774,506	51	154	0	0	0	8	2497
observation_period_intensity	24,774,506	57	14	0	60	60	60	60
observation_period_speed	24,774,506	42	25	0	11	60	60	60
data_error_speed	24,774,506	0	0	0	0	0	0	1
data_error_intensity	24,774,506	0	0	0	0	0	0	1
avg_intensity	24,774,506	126	271	-1	0	6	93	7293
avg_speed	24,774,506	58	44	-1	-1	73	93	646
quality_indicator_speed	936,860	34	28	0	0	47	47.5	94.1
quality_indicator_intensity	12,884,836	77	21	0	61	75	97	100
number_of_traffic_lanes	25,079,680	1	1	1	1	1	2	4
measurement_period	25,079,680	60	0	60	60	60	60	60

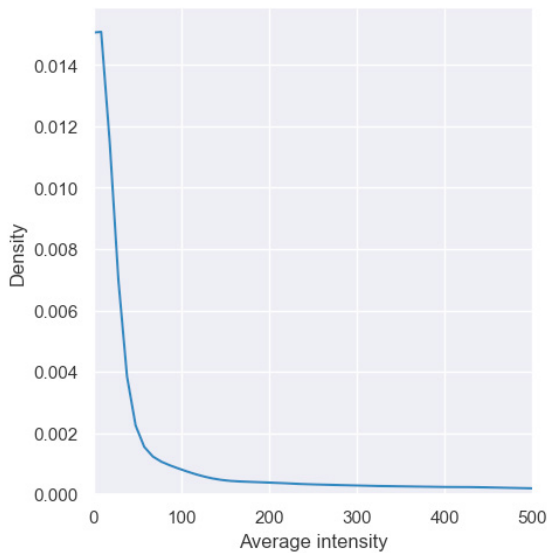


Fig. 1: Density plot of the average intensity.

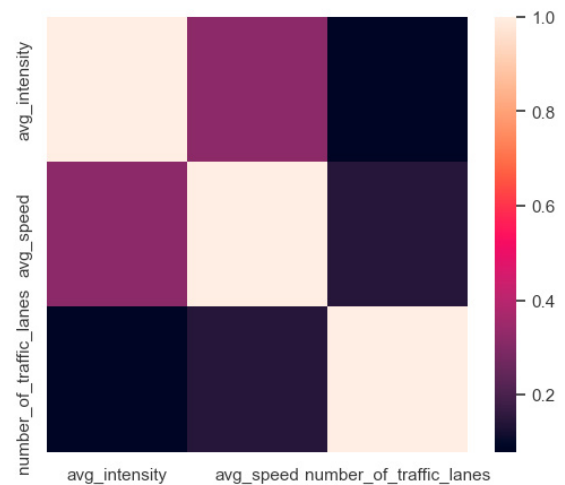


Fig. 2: Heatmap of average intensity, average speed, and traffic lane.

travel along lanes 4, 3, and 2, suggesting that drivers tend to drive in the center. Analyzing traffic during holidays indicated lesser vehicles on holidays compared with usual.

3.3.4. Analysis of A4 Highway Sensors

We identified that there are a total of 69 sensors along A4 Highway and subsequently narrowed them down to five that were impacted by the maintenance in May 2021. Of these five sensors, we will focus on the three impacted sensors in the north part of the roadworks, which are the sensors mentioned in Section 3.1. These sensors are shown in Figure 4.

Zooming in on sensor1, for example, the data suggests that the sensor is on the entry slip road, and oversees 1 lane. Figure 5 shows the impact of road work on traffic flow. It is unclear whether traffic flow was impacted when phase 2 road work was carried out. Neither small vehicles (blue on the primary y-axis) nor trucks (orange on the secondary y-axis) show an obvious reduction during maintenance. The details will be tested in hypothesis testing later on.

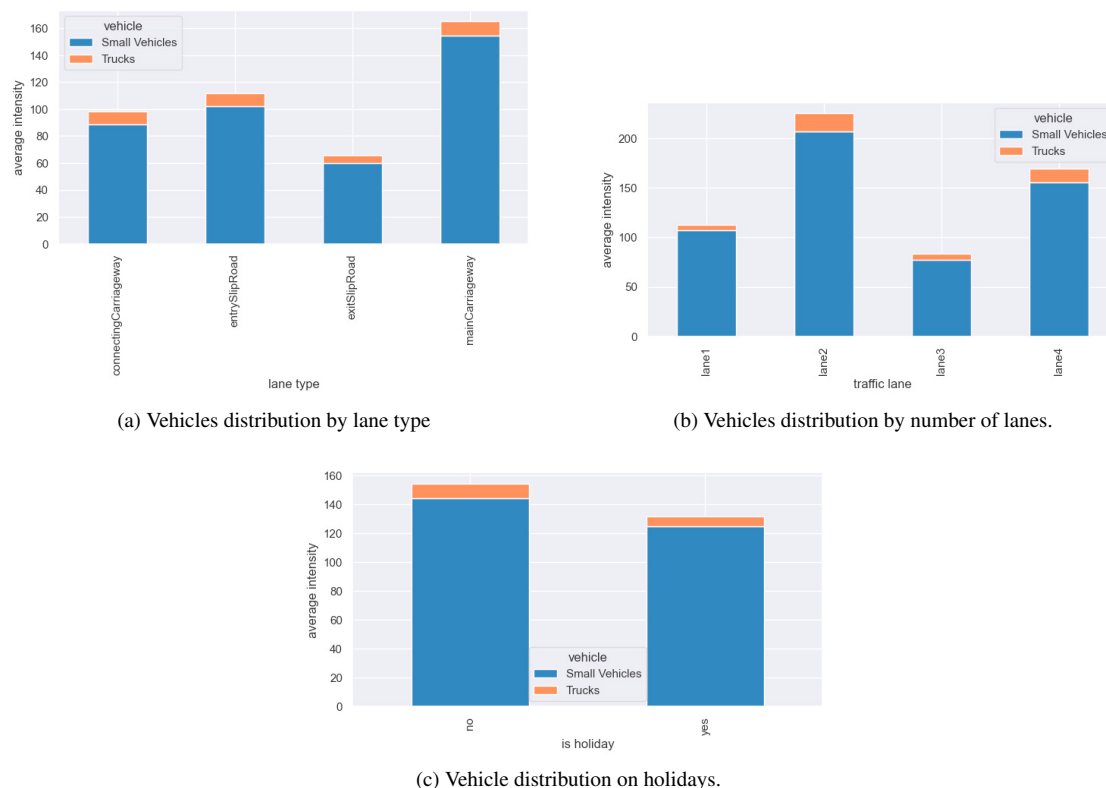
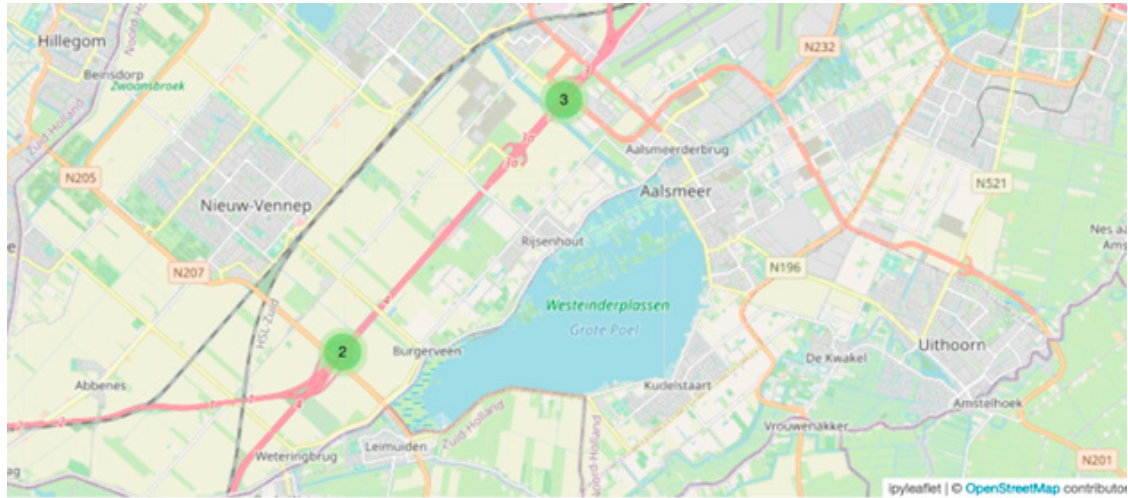


Fig. 3: Distribution of vehicles per lane type, traffic lane, and on holidays.

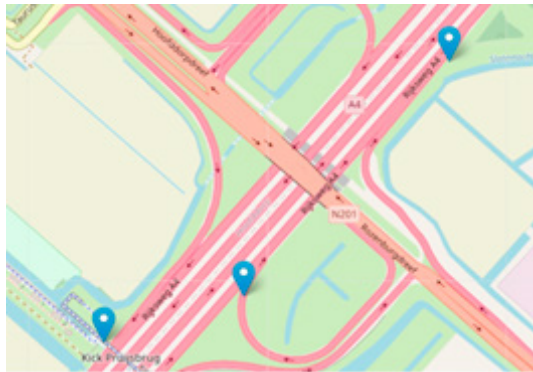
3.4. Hypothesis Testing

We created five hypotheses to understand the road maintenance impact from different facets and conducted tests on four of them. The reason that we were unable to perform testing on the fourth hypothesis - wide vs. narrow lane - is that no existing feature in the dataset indicates the width of each lane before and after roadwork. These hypotheses are:

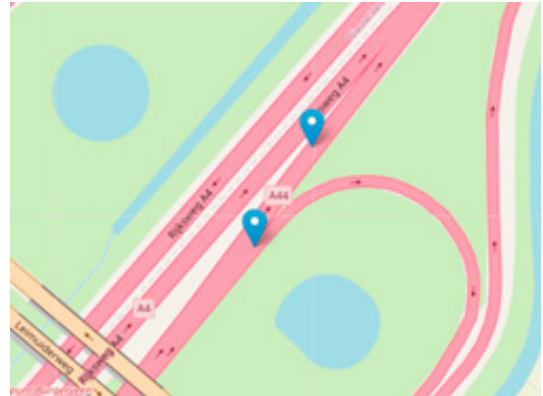
1. Comparison of traffic intensity before and after maintenance
 - H0: The mean of traffic intensity before maintenance is the same as after maintenance.
 - H1: The mean of traffic intensity before maintenance is different from after maintenance.
2. Comparison of traffic intensity before and after maintenance by vehicle type (small vehicles and trucks)
 - H0: For vehicle type i , the mean traffic intensity before maintenance is the same as after maintenance.
 - H1: For vehicle type i , the mean traffic intensity before maintenance is different from after maintenance
3. Comparison of traffic intensity on holidays and non-holidays after maintenance started
 - H0: There is no difference in mean traffic intensity between holidays and non-holidays.
 - H1: There is a difference in mean traffic intensity between holidays and non-holidays.
4. Comparison of trucks on 'wide' lanes vs. 'narrow' lanes to find whether trucks use only the designated lanes
 - H0: The traffic intensity of trucks on 'narrow' lanes is zero (indicating that trucks only use the 'wide' lanes they should use).
 - H1: The traffic intensity of trucks on 'narrow' lanes is not zero (indicating that trucks do not only use the 'wide' lanes they should use).
5. Spatial impact of road work on A4 road to the deviations
 - H0: The mean traffic intensity on adjacent roads (such as N207, N201) are the same during and outside of the A4 road work period.



(a) Sensors impacted the maintenance in May 2021.



(b) Locations of the three impacted sensors in the north.



(c) Locations of the two impacted sensors in the south.

Fig. 4: Overview of sensors impacted by the maintenance in May 2021.

H1: The mean traffic intensity on adjacent roads (such as N207, N201) are not the same during and outside of the A4 road work period

Additionally, we added one more hypothesis looking into one specific A4 sensor1, with the following null hypothesis: the roadwork did not reduce the traffic passing through sensor1.

4. Methods

4.1. Principal Component Analysis and Unsupervised Machine Learning (K-means)

We choose to use principal component analysis (PCA) and unsupervised machine learning as the validation approach method.

4.1.1. PCA

The data we obtained uses different dimension features to record the current environment when the data is obtained. These features influence our target - avg_intensity. We want to find which feature has the most significant influence on the traffic flow. Therefore, we choose the PCA model.

PCA is a mathematical dimensionality reduction method that uses an orthogonal transformation to convert a series of possibly linearly related variables into a set of linearly uncorrelated new variables, also known as principal com-

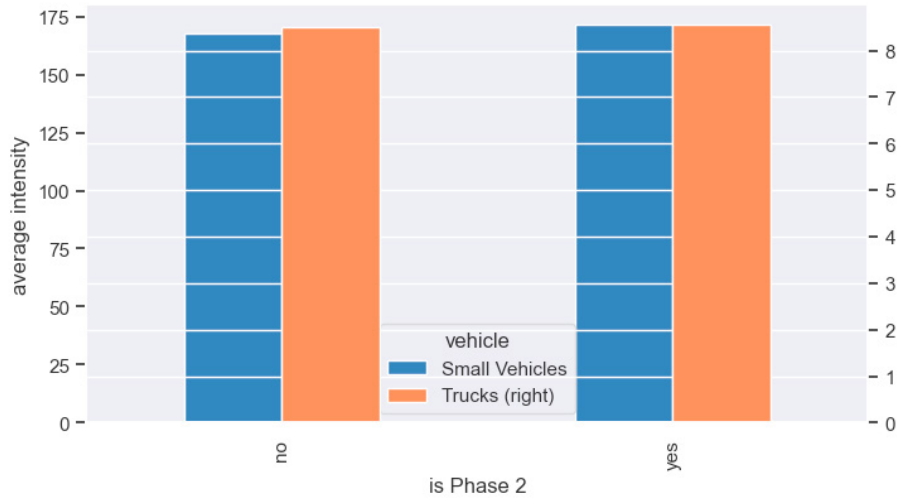


Fig. 5: Overall impact of maintenance on sensor1.

ponents, so as to use the new variables to display the data in a smaller dimension feature. The principal components are linear combinations of the original variables, the number of which is not more than the original variables. After the combination, it is equivalent to obtaining a batch of new observation data. The meaning of these data is different from the original data, but contains most of the characteristics of the previous data, and has a lower dimension, which is convenient for further analysis.

4.1.2. K-means

The k-means algorithm is one of the most classic algorithms in the field of unsupervised learning. It is used for scenarios where the category attributes in the dataset are unclear. It is hoped that objects with similar characteristics can be identified or automatically classified through data mining. We want to use k-means to dig out sensors with similar characters and group them accordingly for future analysis and forecasting.

The calculation method of k-means is to divide the sample set into k clusters according to the distance between samples for a given dataset. Let each point within the cluster be as close as possible, and the distance between the clusters should be as large as possible.

4.2. Time Series

4.2.1. Seasonal Naive Model

The Seasonal Naive model is similar to the Naive model. For the seasonal naive model, each forecast will be equal to the value in the same season from the previous cycle. It is more suitable than the Naive model, given the observed high seasonality of the dataset.

4.2.2. Prophet Forecasting Model

The prophet forecasting model by [Taylor and Letham \(2018\)](#) is a decomposable time series model, it uses trend, seasonality, and holidays as its main components and is similar to a generalized additive model ([Hastie and Tibshirani, 1987](#)). The prophet model is defined by the following equation:

$$y(t) = g(t) + s(t) + h(t) + \epsilon_t, \quad (1)$$

where $g(t)$ models the trend, $s(t)$ models the seasonality, and $h(t)$ models the effects of holidays. ϵ_t is the error term, which is assumed to be normally distributed.

4.2.3. Harmonic Regression

Harmonic regression uses periodic oscillations to make forecasts. The forecasts made with this model will follow a sinusoidal wave function. It regresses the dependent variable against independent variables which are time periods in terms of sine and cosine components such that sine and cosine phases match the seasonality cycle of actual data.

4.2.4. Seasonal Autoregressive Model

An Autoregressive (AR) model bases its forecasts of the output variable on its own historic values. The Seasonal Autoregressive (SAR) model only uses historic values that are a certain seasonal time period away.

4.2.5. Seasonal Autoregressive Integrated Moving Average

The Seasonal Autoregressive Integrated Moving Average (SARIMA) is a seasonal variant of the Autoregressive Integrated Moving Average (ARIMA) model. It consists of the following components: a non-seasonal and seasonal Autoregressive polynomial term, a non-seasonal and seasonal Moving Average part, and a non-seasonal and seasonal difference operator.

5. Results

5.1. Hypothesis Testing

We tested the four hypotheses that we could test and the results are shown in Table 3. All the four null hypotheses are rejected, suggesting there is an impact of roadwork on traffic flows in total and by vehicle type for all sensors in scope including adjacent roads, and there is a difference between traffic flows on holidays and non-holidays. These were also observed from the bar charts described before.

Table 3: Results of hypothesis testing.

	Null Hypothesis	p-value	Result
1	The mean of traffic intensity before maintenance is the same as after maintenance.	0.001	Reject null hypothesis
2	Comparison of before and after maintenance traffic intensity by vehicle type of vehicle (small vehicles and trucks).	small vehicles: 0.031 trucks: 0.002	Reject null hypothesis
3	There is no difference in mean traffic intensity between holidays and non-holidays after maintenance started.	0	Reject null hypothesis
5	The mean traffic intensity on adjacent roads (such as N207, N201) are the same before and after the A4 roadwork period.	0.027	Reject null hypothesis
6	The mean traffic intensity on the A4 road are the same before and after the A4 roadwork period.	0.562	Fail to reject null hypothesis

Additionally, we added one more hypothesis looking into one sensor1, and the result failed to reject the null hypothesis. This result is against intuition but in line with our preliminary results, which suggests that roadwork did not reduce the traffic passing through sensor1. The result for sensor1 indicates that the current roadwork plan on this specific sensor is effective in terms of minimizing traffic impact. This hypothesis testing can be replicated on individual sensors or clusters of sensors of interest (linking to K-means clustering).

5.2. PCA and Unsupervised Machine Learning (K-means)

5.2.1. PCA

Again, we selected sensor1 as an example for PCA, and number_of_traffic_lanes, is_phase, and is_holiday as the three features to analyze. Categorical variables were converted to dummy variables in this step. The new combined

Table 4: Number of components VS. Explained variance ratio.

Number of components	Explained variance ratio
1	0.58
2	1.00
3	1.00

Table 5: Weights of the original features in the combined features.

Combined Features	Number_of_traffic_lines	Is_Phase2_yes	Is_Holiday_yes
PC1	0.000	0.707	0.707
PC2	0.000	-0.707	0.707
PC3	1.000	0.000	0.000

features from the PCA model are called PC1, PC2, PC3, etc. The explained variance ratio corresponding to the number of components is as in Figure 6. The exact numbers are shown in Table 4.

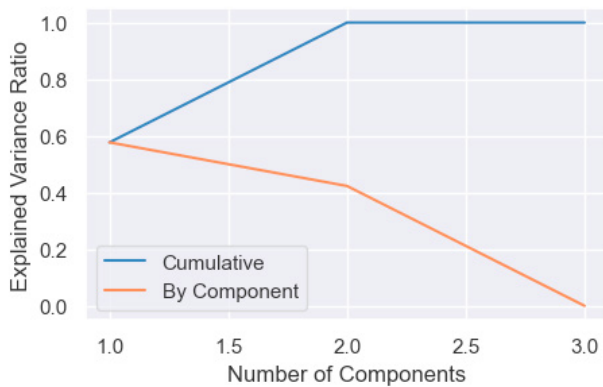


Fig. 6: Number of components in PCA VS. the explained variance ratio.

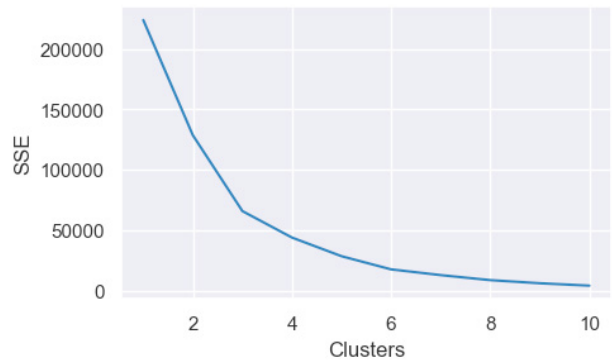


Fig. 7: Elbow method result for the number of clusters.

Through the results in Figure 6 and Table 4, it is obvious that the two components are able to explain all the data variance. Then we drilled down to look at the proportions of three original features in the new combined features, as shown in Table 5.

The result shows that in PC1 and PC2, *is_phase2_yes* and *is_holiday_yes* are the two features with more weights, which means they have more influence on *avg_intensity*. So, *is_phase2_yes* and *is_holiday_yes* are critical features for traffic flow. This result is consistent with intuition as well as our earlier analysis.

5.2.2. K-means

The first thing we need to do to create the clusters is to find a suitable number k , which is the number of clusters which will be created.

We used the Elbow Method to help us find k . Most numerical features were used as the characters for clustering. We run the algorithm for different values of k ($k = 1$ to 10) and plot the k values against the Sum of Squared Errors (SSE). The result is as Figure 7. This figure shows that When $k < 3$, the curve drops rapidly; when $k > 3$, the curve

tends to be stable. We consider inflection point 3 as the optimal value of k by the elbow method. Therefore, we divided our data into 3 clusters. We can check the distribution of 3 sets of data under different features by needs.

5.3. Time Series Models

In order to showcase the different methods of time series analysis, we again selected the 3 sensors sensor1, sensor2, and sensor3 as examples and made predictions using five methods - Seasonal Naïve, Prophet, Harmonic Regression, SAR, and SARIMA. Analysis was performed on two case sets of data: analysis of large vehicles and analysis of all vehicles. The results for sensor1 are shown. The predicted values of average intensity are plotted and a few metrics, RMSE, SMAPE, and R^2 values, are employed to evaluate how well the models fit.

5.3.1. Seasonal Naïve Model

Firstly, the Seasonal Naïve method is used as a benchmark. We first ran the model with all vehicles and obtained an RMSE value of 140.35, a SMAPE value of 22.21, and an R^2 score of 74.1%.

Despite the good level of fitness based on SMAPE value and R^2 score, there are a few obvious deviations as seen in Figure 12, mainly due to holidays falling on weekdays that the models failed to capture.

As we are also interested in the flow of large vehicles, we filtered the dataset and further ran the Seasonal Naïve model on large vehicles only, obtaining an RMSE value of 18.05, a poor SMAPE value of 55.64, and an R^2 score of 51.90%. The results shown in Figure 8 have similar failures as Figure 12, where holidays are not captured.

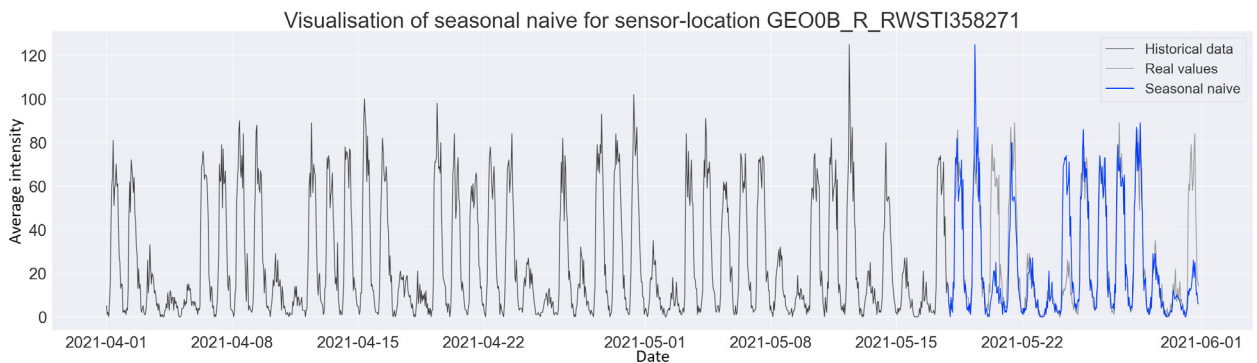


Fig. 8: Seasonal naive prediction on large vehicles for A4 sensor1.

With Seasonal Naïve as the baseline, we then employed four other models to make predictions and compare the results.

5.3.2. Prophet Model

The Prophet model was built with additional regressors aiming to capture more variables that may increase the prediction accuracy. Six regressors; holidays, roadwork, weekdays, and weather on the day (including temperature, precipitation, wind speed) were added as regressors. The RMSE value was calculated as 219.23, the SMAPE value of 63.74, and the R^2 score of 47.12%. The Prophet model does not lead to better RMSE, SMAPE, and R^2 values than the Seasonal Naïve model, suggesting that adding more features may not necessarily improve the forecast accuracy. Figure 9 is the one-week prediction made by the Prophet model.

However, despite the worse metric compared to Seasonal Naïve, the Prophet model has one advantage - it is able to capture the holiday impacts correctly - whereas Seasonal Naïve failed to as discussed in Section 5.3.1, due to the addition of holiday regressors.

We further extended the prediction period from one week to two weeks with the prediction line shown in Figure 10. The accuracy is better than the one-week Prophet prediction. However, compared to Seasonal Naïve, all RMSE, SMAPE, and R^2 values are worse.

Similarly, we applied a two-week Prophet forecast model on large vehicles only and obtained Figure 11. The RMSE, SMAPE, and R^2 scores are 19.46, 86.93, and 37.34% respectively, all worse than the Seasonal Naïve model. It matches the results obtained with all vehicles.

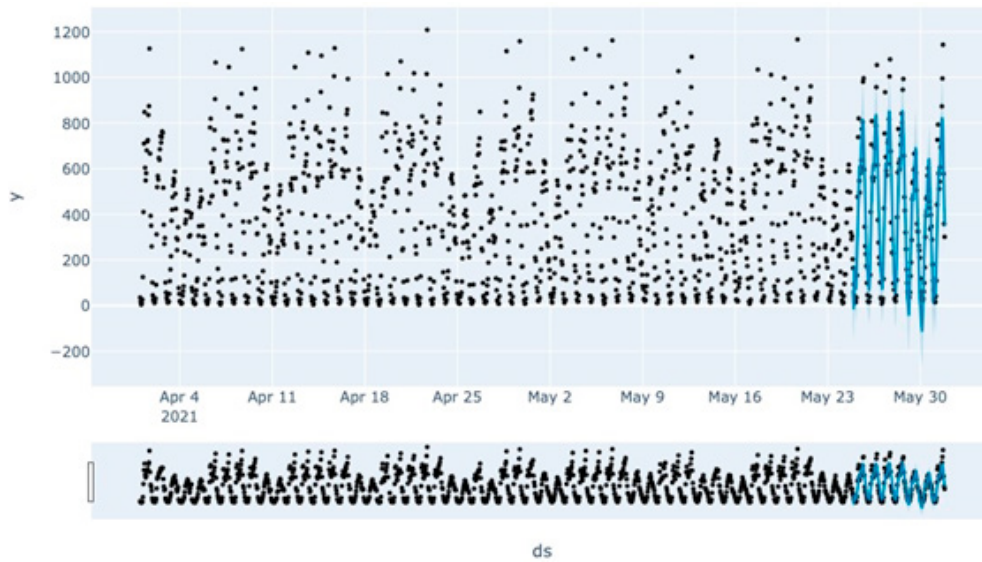


Fig. 9: One-week prophet prediction on all vehicles for A4 sensor1.

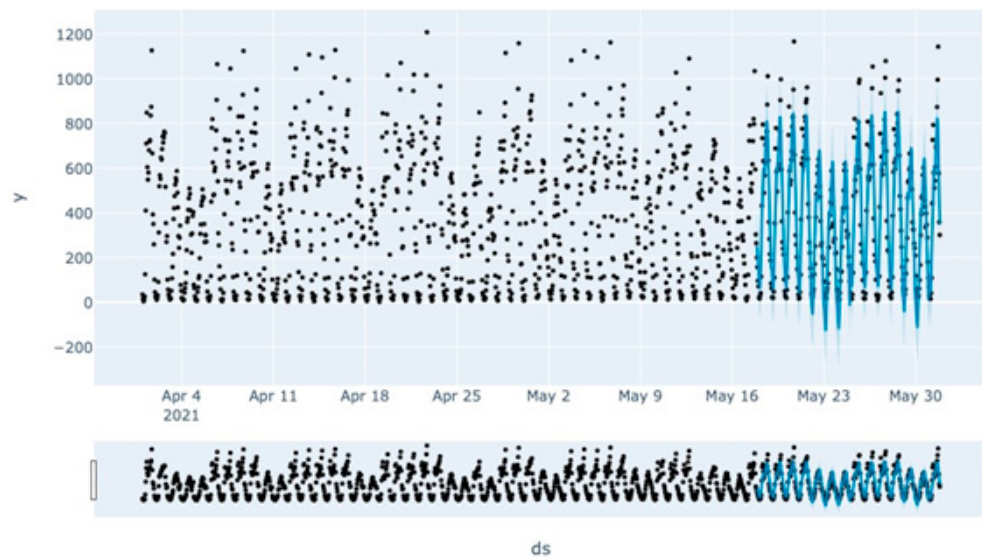


Fig. 10: Two-week prophet prediction on all vehicles for A4 sensor1.

5.3.3. Harmonic Regression, SAR, and SARIMA Models

Case 1: Analysis of Large Vehicles:

For harmonic regression, first, we plot average intensity across time periods. This indicates seasonality every week or every 168 hours as expected.

We obtained sine and cosine components of time periods for each observation. We then obtained the predicted average intensity by regressing the actual average intensity against time periods defined by sine and cosine components. Although the long-term trend is not well captured, we find that harmonic regression captures the seasonality fairly well as shown in Figure 12.

To address the problem of not capturing long-term trends in harmonic regression, we explored the SAR model. In this model, instead of using the sine and cosine components of time periods as features, we used previous time periods

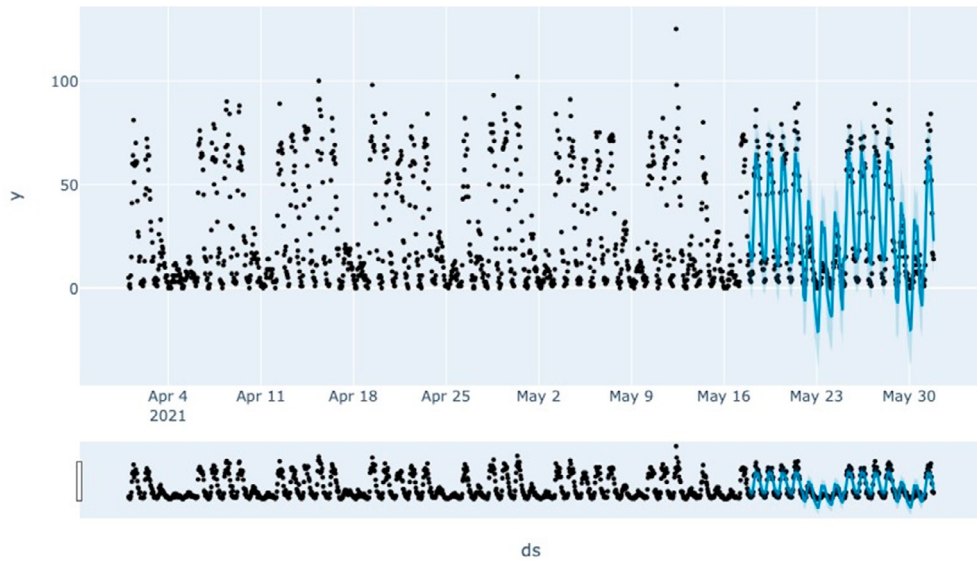


Fig. 11: Two-week prophet prediction on large vehicles for A4 sensor1.

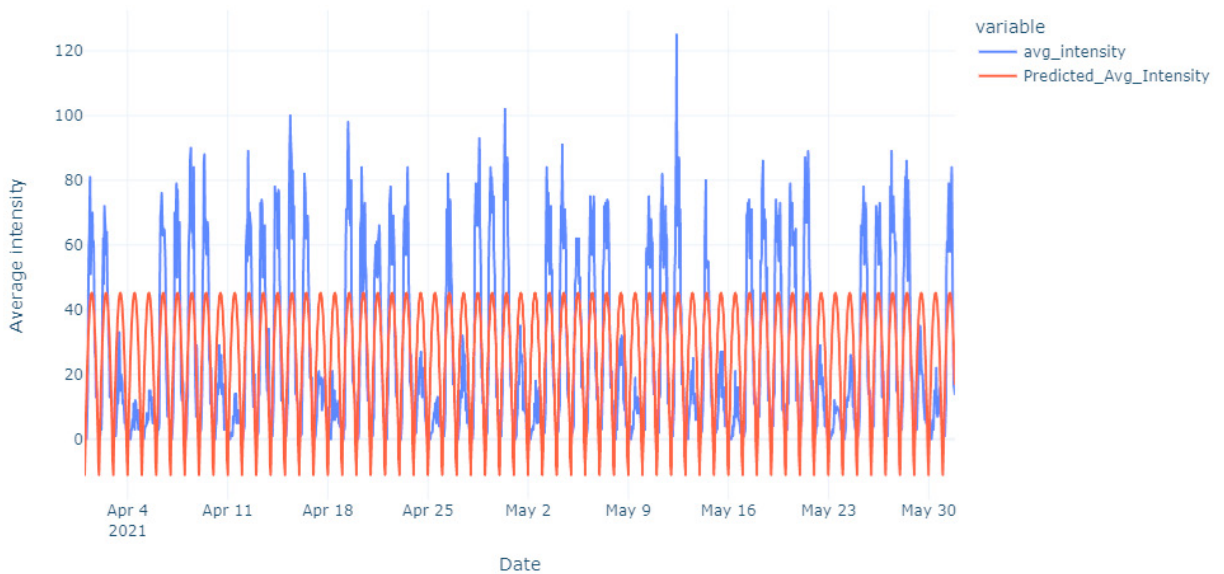


Fig. 12: Harmonic Regression of A4 sensor1 for large vehicles.

as the predictor. Since seasonality is seen at a weekly level, we split the feature matrix (X) and target vector (y). The lag value will be 168 hours (one week). Both target vector and feature matrix should have the same dimensionality. SAR model has captured long-term trends and seasonality as shown in Figure 13. But it failed to capture holiday impacts as the Naïve model.

Lastly, there could be cases where long-term trends do not change not only according to seasonality cycles but may also vary at weekly or hourly levels. Towards this, we explored the SARIMA model. In this model, the feature matrix is a combination of sine-cosine components and moving averages. For this data, we calculate averages at hourly levels

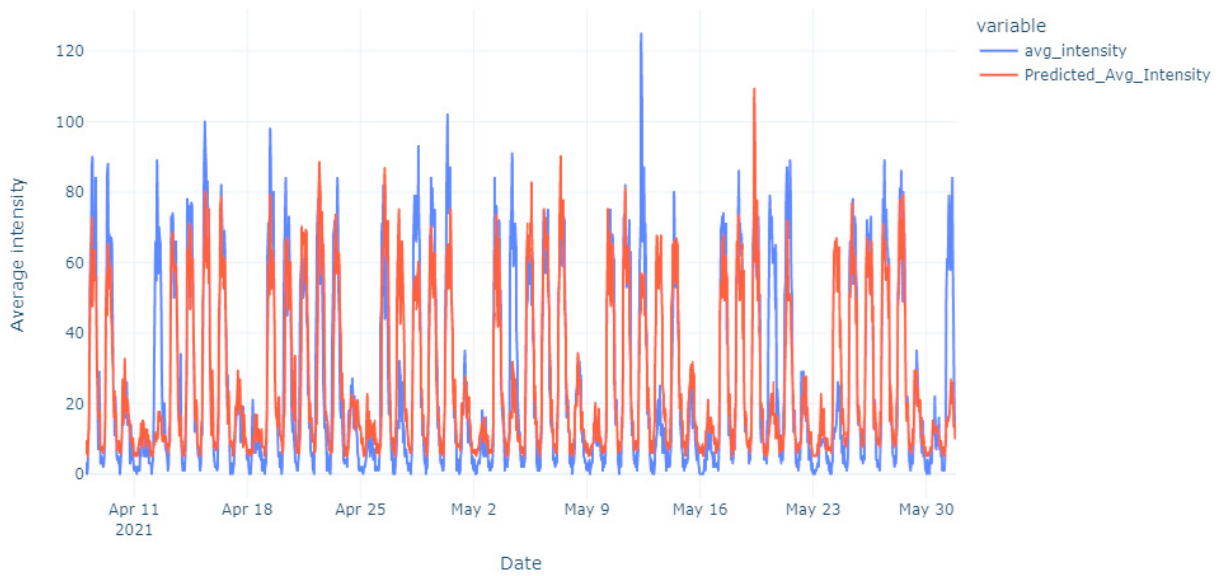


Fig. 13: SAR model for A4 sensor1 for large vehicles.

for each observation along with capturing trigonometric features (Sine and Cosine components). As seen in Figure 14, SARIMA tries to capture long-term trends a little better at time periods where the SAR model could not capture well. For example, on April 12, SARIMA performs a better forecast in capturing the trend compared to the SAR model on the same date.

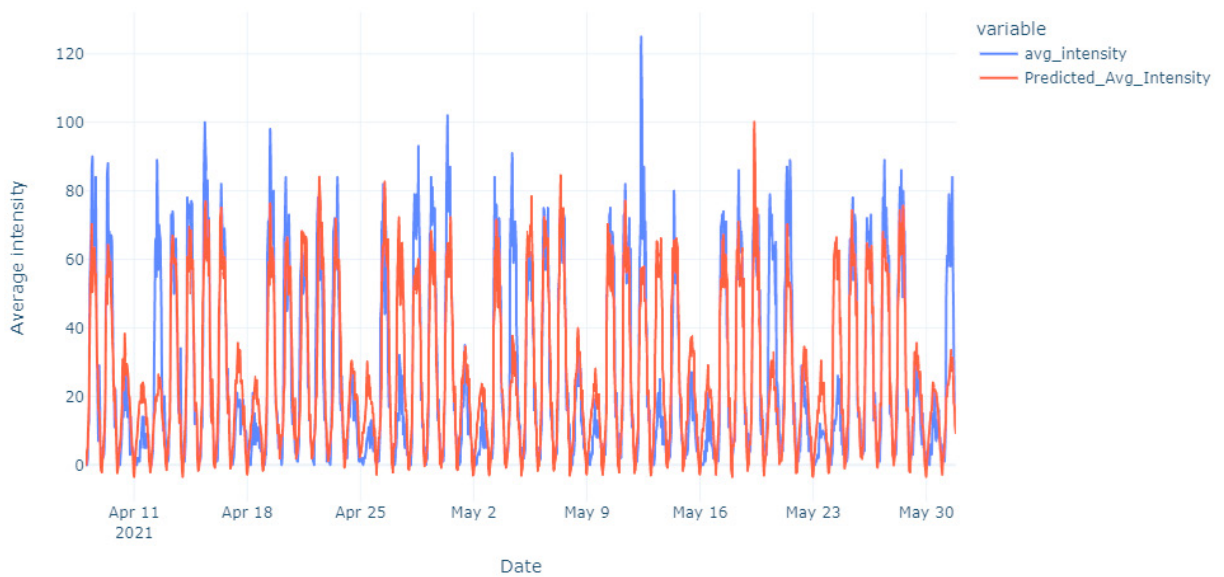


Fig. 14: SARIMA for A4 sensor1 for large vehicles.

The procedure was repeated for the other two A4 sensors to test the robustness of the models. After building these models, we compared their forecast accuracy using RMSE, SMAPE, and R^2 as metrics. Tables 6, 7, and 8 summarize the forecast accuracies for different metrics for large vehicles for all five models and all three sensors. From these tables we observe that SARIMA performs better than other models across all forecasting metrics for all sensors. This is mainly because of its capabilities of capturing linear trends and seasonality. There is also an opportunity to further improve the forecast accuracy of the SARIMA model by incorporating holiday information as seen in the Prophet model.

Table 6: Forecast RMSE of three A4 sensors with different models on large vehicles.

Sensors	Seasonal Naïve	Prophet	Harmonic Regression	SAR	SARIMA
sensor1	18.05	19.46	19.88	15.27	14.45
sensor2	22.41	38.77	38.03	27.56	26.24
sensor3	16.83	17.64	17.52	13.96	13.16

Table 7: Forecast SMAPE of three A4 sensors with different models on large vehicles.

Sensors	Seasonal Naïve	Prophet	Harmonic Regression	SAR	SARIMA
sensor1	55.64	86.93	90.42	57.28	66.06
sensor2	41.54	81.72	85.99	45.98	52.51
sensor3	42.96	79.54	80.76	48.16	54.60

Table 8: Forecast R^2 of three A4 sensors with different models on large vehicles.

Sensors	Seasonal Naïve	Prophet	Harmonic Regression	SAR	SARIMA
sensor1	51.90%	37.34%	44%	67.40%	71%
sensor2	47.83%	38.43%	41.40%	69.60%	72.40%
sensor3	46.59%	37.72%	45%	66%	69.70%

Case 2: Analysis of All Vehicles

In case 2, We repeated all analyses performed in the case on all vehicles. Here as well, the SARIMA model outperformed other models across. Tables 9, 10, and 11 summarize the forecast metrics for all vehicles for all five models and all three sensors, and Figures 15 to 17 show the example for sensor1.

Table 9: Forecast RMSE of three A4 sensors with different models on all vehicles.

Sensors	Seasonal Naïve	Prophet	Harmonic Regression	SAR	SARIMA
sensor1	51.90%	37.34%	44%	67.40%	71%
sensor2	47.83%	38.43%	41.40%	69.60%	72.40%
sensor3	46.59%	37.72%	45%	66%	69.70%

Table 10: Forecast SMAPE of three A4 sensors with different models on all vehicles.

Sensors	Seasonal Naïve	Prophet	Harmonic Regression	SAR	SARIMA
sensor1	22.21	63.74	62.11	38.52	45.17
sensor2	23.27	60.45	61.76	33.42	36.6
sensor3	23.39	54.05	56.49	37.44	43.33

Table 11: Forecast R^2 of three A4 sensors with different models on all vehicles.

Sensors	Seasonal Naïve	Prophet	Harmonic Regression	SAR	SARIMA
sensor1	74.17%	47.12%	61%	77.70%	81%
sensor2	66.84%	50.78%	49.50%	74.60%	77.30%
sensor3	76.41%	61.68%	63%	76%	80.00%

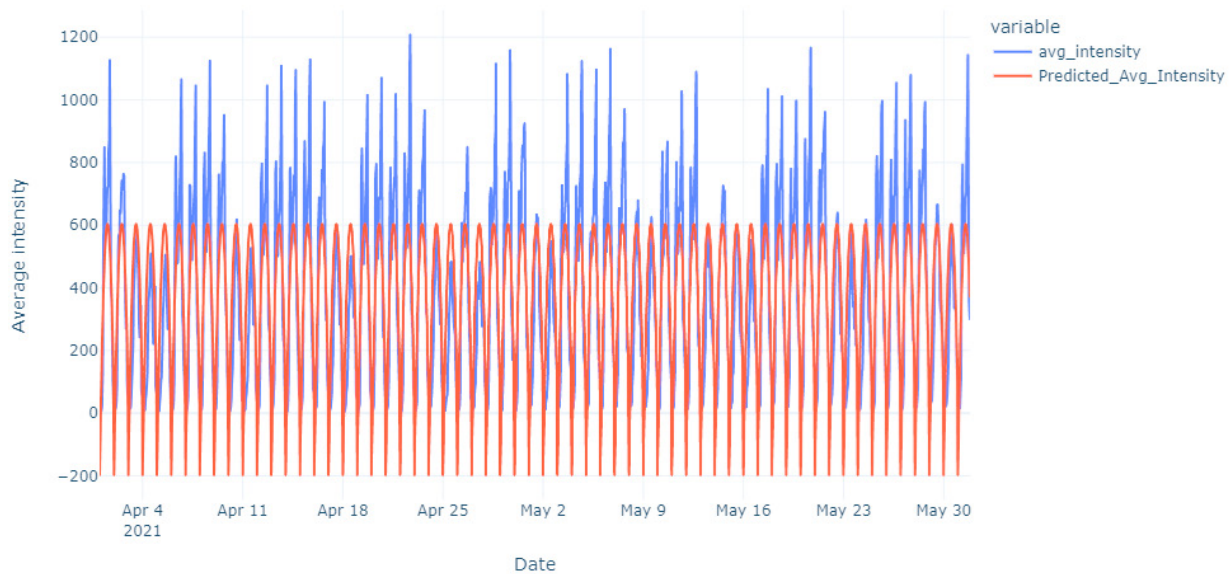


Fig. 15: Harmonic regression model for A4 sensor1 for all vehicles.

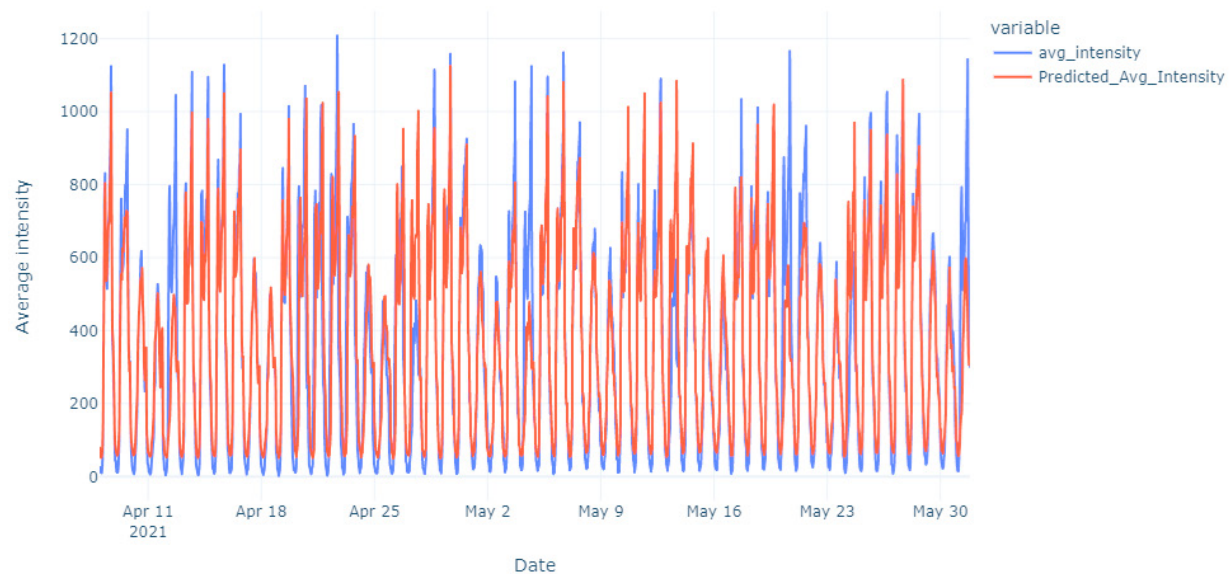


Fig. 16: SAR model for A4 sensor1 for all vehicles.

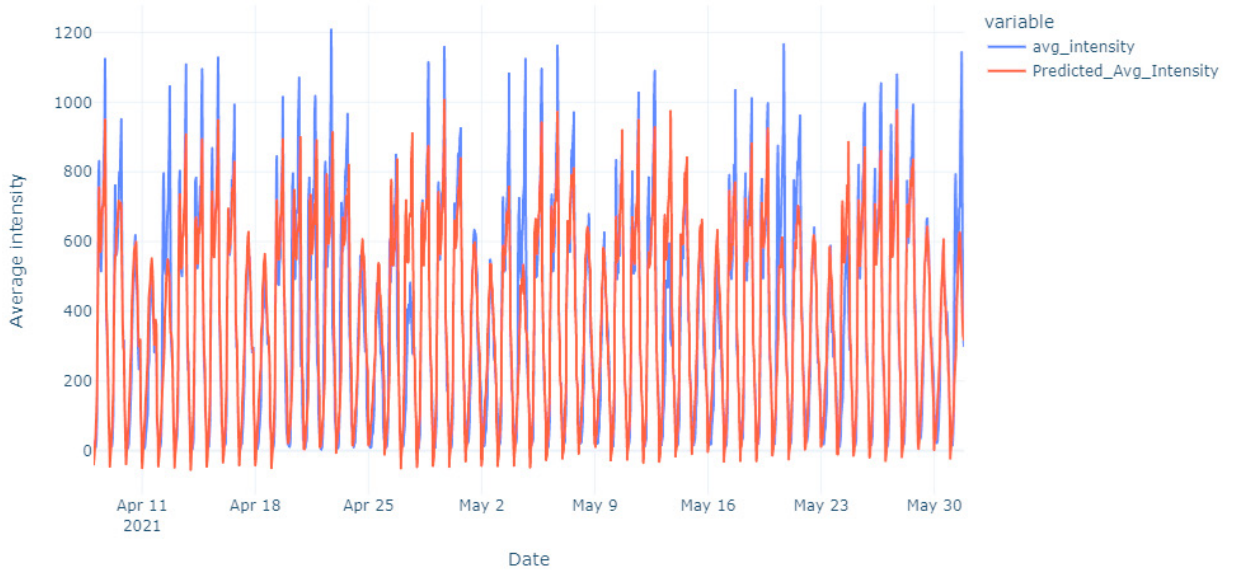


Fig. 17: SARIMA model for A4 sensor1 for all vehicles.

6. Conclusion

Overall, the analysis showed that road maintenance and holidays are two features that have more influence on traffic flow, which should be considered as main factors when carrying out future road plans. If multiple areas are affected, the K-means model can be adopted effectively to cluster the sensors into groups to minimize traffic disruption. Despite the advantages and disadvantages seen in various time series models, SARIMA performs the best and is recommended to be adopted for traffic flow prediction purposes. Our analysis is performed over 3 sensors along the A4 highway, but the analysis can be extended for other selected sensors going forward.

Finally, since our analysis shows that maintenance activity had a significant impact on traffic intensity, except for sensor1, the Ministry of infrastructure could implement the following actions for effective road maintenance activities:

- Encourage motorists to travel on parallel transit lines, use pooling, and use alternate modes of transportation.
- Raise continuous public awareness of upcoming road maintenance work to reduce traffic through work zones.
- Use forecasting to understand the impact of work zones on traffic flow and congestion to help for better planning road maintenance activities.
- Use materials and methods that expedite construction time to reduce the frequency of future maintenance.

7. Future Study

As part of the future work, our time series analysis could be extended to implement Lasso Regression which will aid in shrinking the coefficients for those input variables that do not contribute much to the prediction task. This penalty allows some coefficient values to go to the value of zero, allowing input variables to be effectively removed from the model, providing a type of automatic feature selection. Neural network models could also be implemented, as research shows promising results for their forecasting abilities.

Acknowledgements

This research has been conducted in collaboration with the Netherlands Ministry of Infrastructure and Environment. We would like to thank the Master program of Supply Chain Management at the Massachusetts Institute of Technology Center for Transportation and Logistics for giving us the opportunity to work on the research during the course module SCM.256 - Data Science and Machine Learning for Supply Chain Management.

References

- Awan, F.M., Minerva, R., Crespi, N., 2020. Improving road traffic forecasting using air pollution and atmospheric data: Experiments based on lstm recurrent neural networks. *Sensors* 20, 3749.
- Bogaerts, T., Masegosa, A.D., Angarita-Zapata, J.S., Onieva, E., Hellinckx, P., 2020. A graph cnn-lstm neural network for short and long-term traffic forecasting based on trajectory data. *Transportation Research Part C: Emerging Technologies* 112, 62–77.
- Diao, Z., Wang, X., Zhang, D., Liu, Y., Xie, K., He, S., 2019. Dynamic spatial-temporal graph convolutional neural networks for traffic forecasting, in: *Proceedings of the AAAI conference on artificial intelligence*, pp. 890–897.
- Guo, K., Hu, Y., Qian, Z., Liu, H., Zhang, K., Sun, Y., Gao, J., Yin, B., 2020. Optimized graph convolution recurrent neural network for traffic prediction. *IEEE Transactions on Intelligent Transportation Systems* 22, 1138–1149.
- Hastie, T., Tibshirani, R., 1987. Generalized additive models: some applications. *Journal of the American Statistical Association* 82, 371–386.
- Kumar, S.V., Vanajakshi, L., 2015. Short-term traffic flow prediction using seasonal arima model with limited input data. *European Transport Research Review* 7, 1–9.
- Lin, S.L., Huang, H.Q., Zhu, D.Q., Wang, T.Z., 2009. The application of space-time arima model on traffic flow forecasting, in: *2009 International Conference on Machine Learning and Cybernetics*, pp. 3408–3412. doi:[10.1109/ICMLC.2009.5212785](https://doi.org/10.1109/ICMLC.2009.5212785).
- Ma, T., Antoniou, C., Toledo, T., 2020a. Hybrid machine learning algorithm and statistical time series model for network-wide traffic forecast. *Transportation Research Part C: Emerging Technologies* 111, 352–372.
- Ma, Y., Zhang, Z., Ihler, A., 2020b. Multi-lane short-term traffic forecasting with convolutional lstm network. *IEEE Access* 8, 34629–34643.
- Mallick, T., Kiran, M., Mohammed, B., Balaprakash, P., 2020. Dynamic graph neural network for traffic forecasting in wide area networks, in: *2020 IEEE International Conference on Big Data (Big Data)*, IEEE. pp. 1–10.
- Taylor, S.J., Letham, B., 2018. Forecasting at scale. *The American Statistician* 72, 37–45.
- Yang, H., Li, X., Qiang, W., Zhao, Y., Zhang, W., Tang, C., 2021. A network traffic forecasting method based on sa optimized arima-bp neural network. *Computer Networks* 193, 108102.
- Yu, G., Zhang, C., 2004. Switching arima model based forecasting for traffic flow, in: *2004 IEEE International Conference on Acoustics, Speech, and Signal Processing*, IEEE. pp. ii–429.
- Zhao, Z., Chen, W., Wu, X., Chen, P.C., Liu, J., 2017. Lstm network: a deep learning approach for short-term traffic forecast. *IET Intelligent Transport Systems* 11, 68–75.