The goal of our project is to compare different classical models for traffic forecasting in different situations.

Dataset

For the dataset, we take the PEMS-BAY, which includes data from 325 sensors in San Jose in the first half of 2017. To avoid public holidays, we used 13 weeks from February 26th to May 20th. We split them into 10 weeks for the training part and 3 weeks for the validation. For the test dataset, we took the data for the same sensors but, in 2022, we used 5 weeks in this dataset.

Models

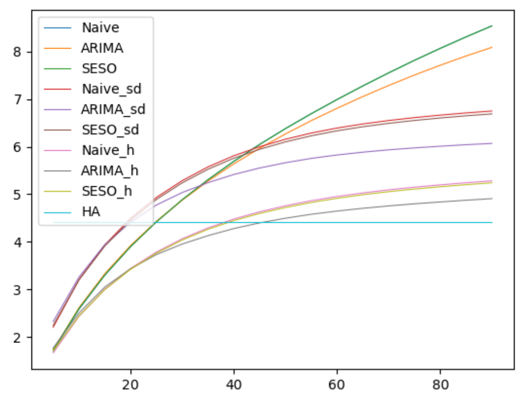
We chose three models, Naïve (predicting the previous value), ARIMA with auto order and Exponential Smoothing with an optimized parameter. We trained each model on three datasets: the base dataset, the dataset of seasonal differences and the dataset of the deviances from the historical average. Also, we include the historical average itself, which is the average speed of this sensor during this timeslot among all training weeks. So, we got 10 models in total.

Forecasting

After fitting the models on the training data, we used them to predict the future traffic speed. We were forecasting up to 18 steps, from 5 to 90 minutes into the future, for the validation dataset and then did the same for the testing one.

Comparing the models

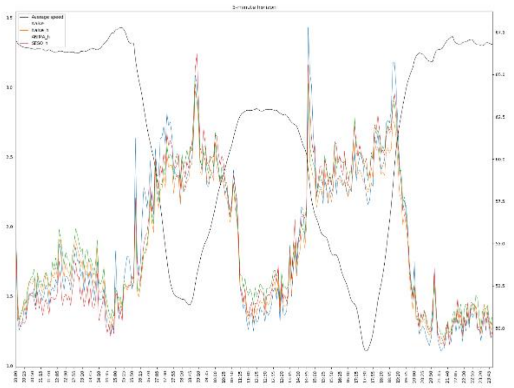
To compare the performance of different models, we used the root mean square errors of the actual and the predicted values. Here is the graph, where on the x-axis there is the forecasting horizon and on the y-axis there is the average root mean square error:

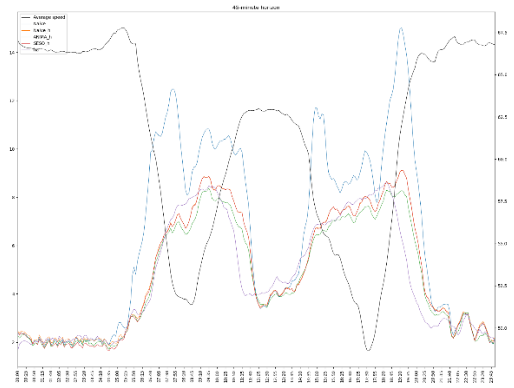


	5	15	30	45	60	90
Naive	1.70	3.30	4.89	6.05	6.99	8.54
ARIMA	1.72	3.33	4.88	5.97	6.81	8.09
SESO	1.73	3.30	4.88	6.05	6.99	8.54
Naive_sd	2.21	3.94	5.29	6.00	6.39	6.75
ARIMA_sd	2.32	3.94	5.04	5.55	5.83	6.07
SESO_sd	2.24	3.91	5.24	5.95	6.33	6.69
Naive_h	1.67	3.00	4.06	4.63	4.95	5.28
ARIMA_h	1.77	3.05	3.95	4.40	4.65	4.91
SESO_h	1.71	3.00	4.04	4.60	4.91	5.24
HA	4.41	4.41	4.41	4.41	4.41	4.41

We can notice that when we forecast 5 minutes into the future, most models are having similar performance and for 10-15 minutes, the exponential smoothing model, trained on the historical average dataset is slightly better than the others. From 20 minutes and further, the ARIMA model, trained on the historical average dataset is turned out to be the best one. Starting 50 minutes, all models lose to the historical average, so we can say that this is the point where it begins to be useless to use the current data to predict the traffic using classical models and the historical data is enough.

We also computed the average error for each time throughout the weekday and the weekend.





Here we can notice that all models are having more or less similar performance on the 5 minutes horizon, but on the 45 minutes horizon, ARIMA outperforms other models and the Naïve model is significantly worse, especially during the beginning and the end of the rush hours.

Conclusion

We conclude that for very short-term predictions, classical models don't give a significant advantage in comparison to the base model. However, for a more distant future, they allow us to improve the prediction, especially models trained on the dataset with the deviance from the historical average. For very long term predictions, these models lose to the predicting historical average. The difference is more significant during rush hours, especially at the beginning and the end of them.

There are several different ways to continue this work. First, one can broaden research to the different models. Also, one can test these models in other cities. Also, it is interesting to look at the behavior of these models during public holidays or non-regular special days (i.e., game days).