FORECASTING TRAFFIC VARIATIONS IN PRESENCE OF EVENTS

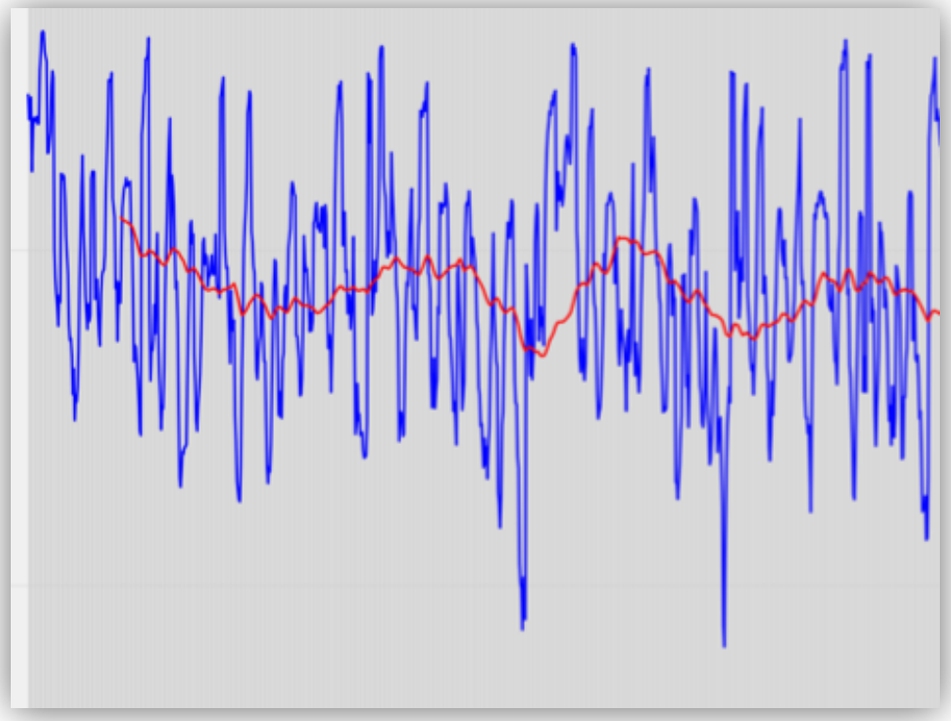
10504172 - Fossati Nicola
10544315 - Montesi Daniele
10640083 - Zaccaro Espedito

POLITECNICO
MILANO 1863

PREPROCESSING

- Big Datasets from which we had to extract only the useful measurements
- Available data about Weather, Events, Road Characteristics
- Main approach was to merge all the meaningful information in one smaller dataset
- Distances dataset used for retrieving weather information
- Splitting the dataset when processing in order to overcome computer memory limits

1. Weekly moving average of avg. speeds in normal conditions

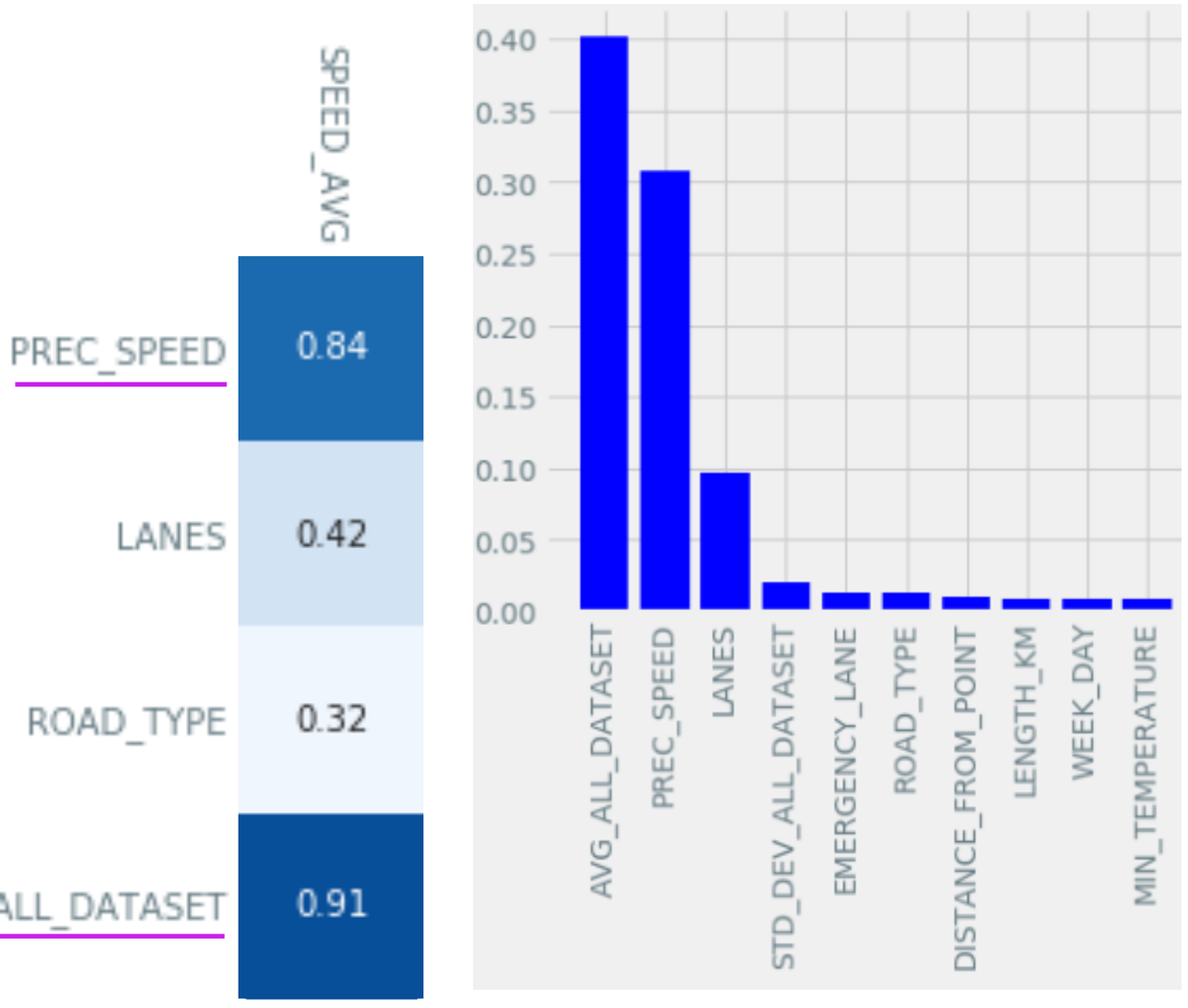


2. Weekly moving average of avg. speeds in presence of events

DATA EXPLORATION

POLITECNICO MILANO 1863

- New, more correlated features
- Handling the 15 minutes interval
- Extracting Numerical and Categorical features from Timestamps
- Time of the day
- Periodicity
- Obtaining a baseline



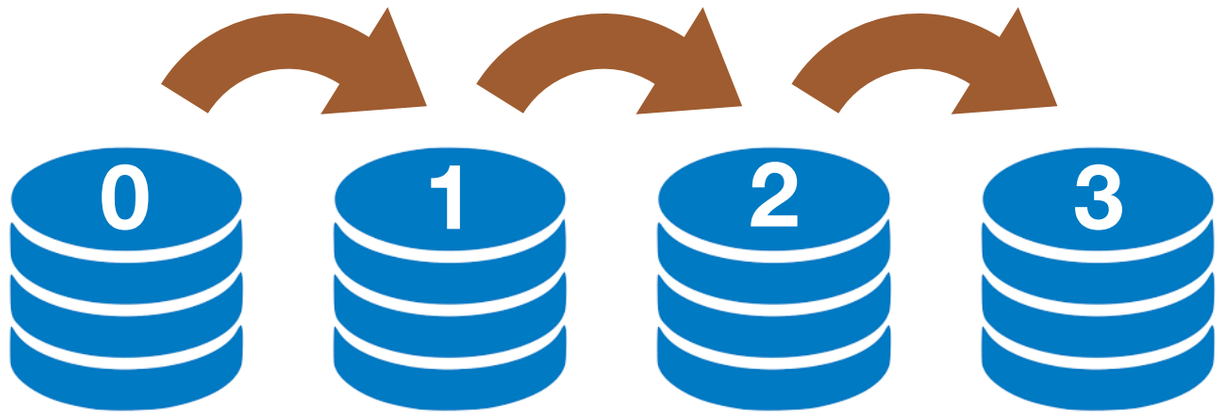
3. Correlation with the target variable and the most promising features
3.1. feature ranking

THE MODEL

POLITECNICO MILANO 1863

- Gradient Boosting Regressor (Catboost)
- Two approaches attempted
- First it predicts average speeds whitin 15 minutes from the event's beginning
- Then it uses the obtained speeds for predicting the next quarter, and so on
- Basically, we use the same model four times
- It behaves at its best with these parameters:
  - Learning Rate = 0.04
  - 1000 predictors
  - Max depth = 9
  - Regularisation Lambda = 3.68

CatBoost



Prediction Steps

SUMMARY & FURTHER IMPROVEMENTS

POLITECNICO MILANO 1863

- Mean average error = 6.11 - Estimated in Km/h from the real values in the validation set

- Major improvements involved an accurate preprocessing

- Further improvements - more data, more features,

MAE:

10.53 Km/h → **6.46 Km/h** → **6.11 Km/h**

XGBoost Baseline | CatBoost With feature engineering | With F.E. + Tuning