



Statens vegvesen

Norwegian Public Roads
Administration



Statens vegvesen

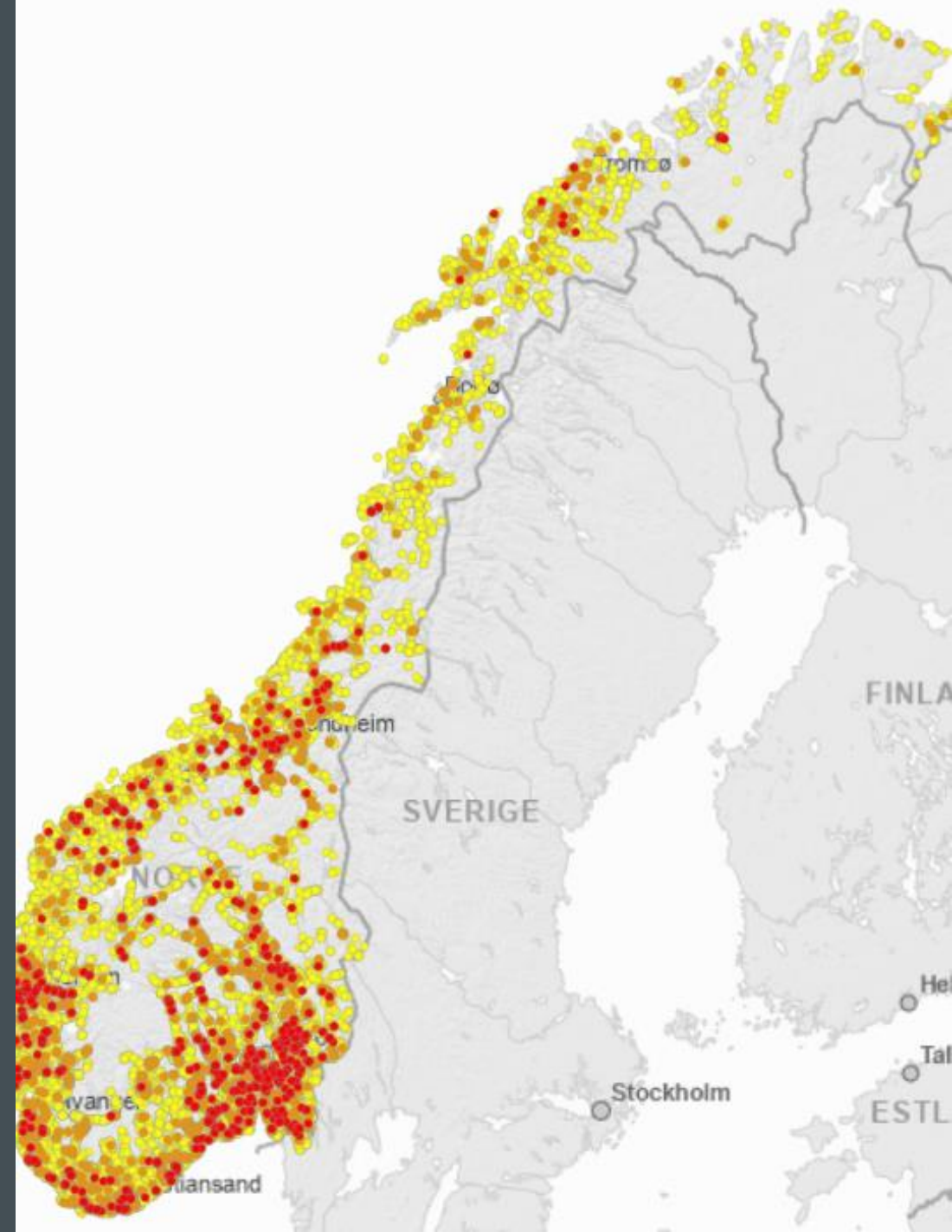
Norwegian Public Roads
Administration

Crash Risk Curves

Rank curves by **relative risk**

Machine-learning based decision support

By Preben Drangsholt



Agenda

- -The core philosophy
- -How our data is structured
- -How we define analytical units
- -How we connect facts
- -Methodology and results
- -Evaluation strategy
- -Closing remarks

Background and motivation

- Serious accidents are:
 - Rare
 - geographically dispersed
 - rarely repeat at the same location
- property-based decision support

From network data to relational analysis

- Historically road segments
 - Navigation
 - Routing
 - visualization
- Tabular analysis
 - Joining many data sources
 - Statistical modeling
 - Machine learning

Why two models

- One model for State roads
 - Larger information space
- One model for County roads
 - 13x more curves

Why curves

- Curves represent a disproportionate share of serious accidents.
- They are relatively short
- They are discrete
- Targeted interventions are often feasible

Data foundation

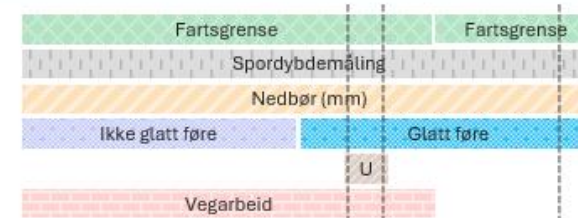
Data type	Original source
Curve radius	ViaPPS*
Crossfall (superelevation)	ViaPPS*
Elevation	NVDB Road points
Injury crashes (by severity)	TRULS
Speed limits	NVDB
Guardrail types	NVDB
Road markings	NVDB
AADT (Annual Average Daily Traffic)	NVDB
Road width	NVDB
Road point metadata	NVDB + DIA (“Road points”)
Traffic signs	NVDB

Road points as the core data structure



1				
2				
3				
4				

Fakta og attributter:



Veglenke:

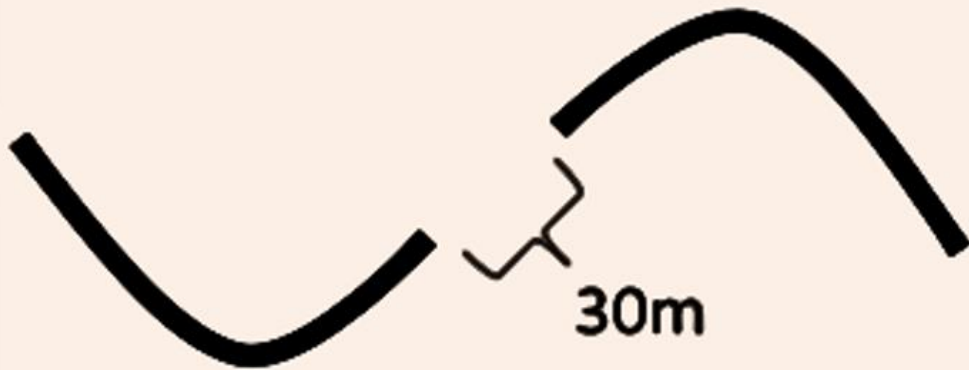


Vegpunkter langs lenke:



Defining curves as analytical units

From 2 curves...



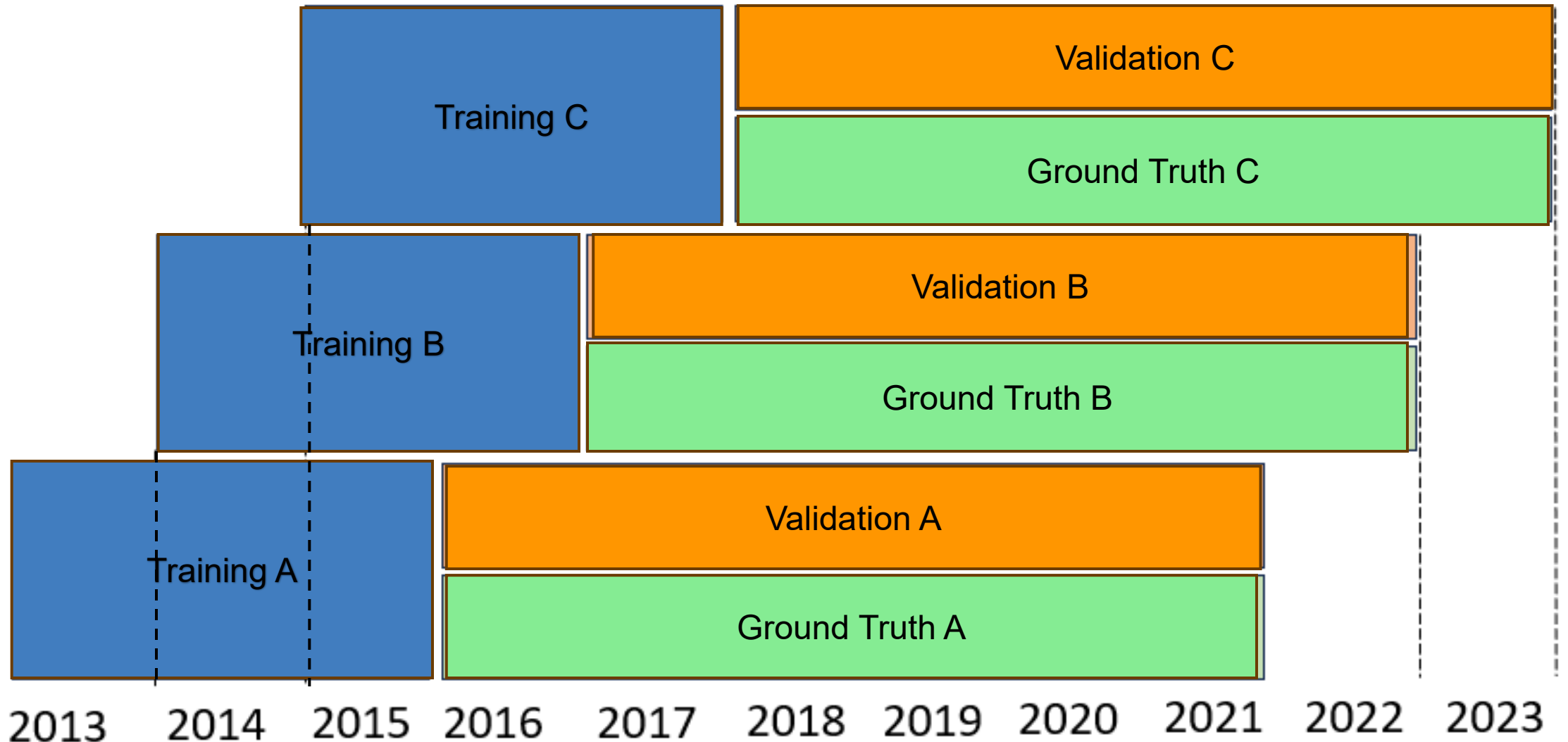
To 1 longer curve



Modeling philosophy and architecture

- Two layer optimization
- Road geometry
- Driver behavior
- Vehicles

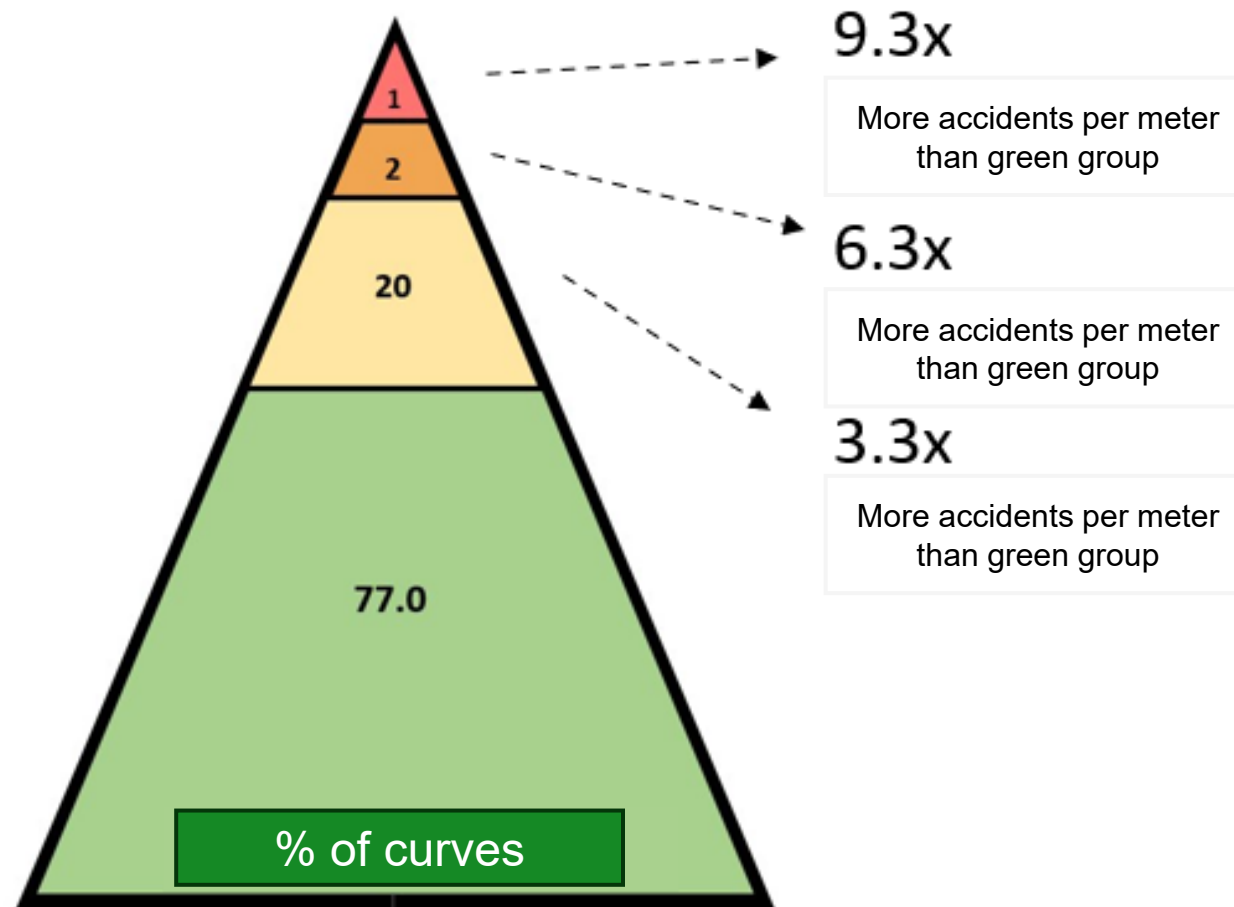
Experimental setup and evaluation strategy



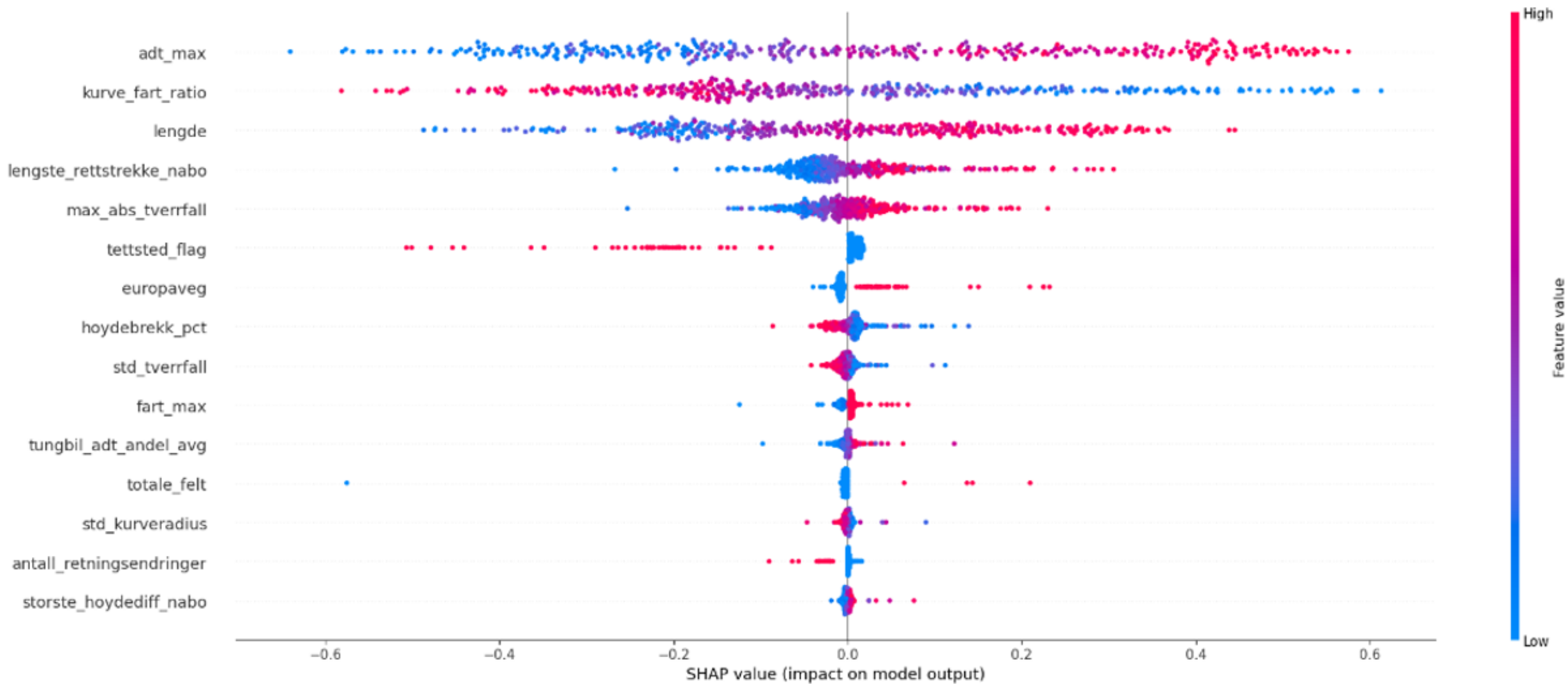
Evaluation metrics and optimization goals

- ▶ AUC
- ▶ Recall
- ▶ Positive rate
- ▶ Category consistency
- ▶ Geographic spread

Risk category and results



Step	Description
Imputer	Replaces missing values with the mean
Variance filtering	Removes variables with no variance
Outlier handling	Caps extreme values (above the 99.9th percentile) by replacing them with the 99.9th percentile value
Min–max scaling	Scales values to the range [0, 1]
Power transformation	Applies a power transformation to variables (Yeo–Johnson)
Clustering	Creates 30 K-means++ clusters of the variables and applies target encoding based on y_train
Variance filtering	Removes variables with no variance (re-applied after transformations)
Feature agglomeration	Creates new aggregated features by merging variables with similar correlation structure
Decorrelation	Iteratively removes variables that correlate ≥50% with stronger variables but have lower mutual information than the strongest univariate drivers
Int casting	Rounds values to two decimals and casts to integers by multiplying by 100. This reduces overfitting and prevents re-identification of specific curves through memorization of highly precise floating-point values
Logistic regression	A relatively “fuzzy” model, focused on capturing overall trends rather than fine-grained details. Parameters: C = 0.1, tol = 0.001, max_iter = 10 000. Class imbalance is partially corrected, adjusting from a 1:10 to a 1:5 ratio



What the model does and does not

■ The model does not:

- Predict individual accidents
- Claim that an accident will occur at a specific curve next year
- Recommend specific countermeasures

■ The model does:

- Rank curves by **relative risk**
- Identify locations with higher-than-expected accident occurrence given traffic volume
- Detect surprising risk rather than raw accident frequency

Future improvements

- **Optimal sight distance**, derived from lidar point clouds and line-of-sight calculations
- **Driver behavior classification**, using external speed and volume distributions from commercial providers

Closing remarks

- This model does not replace professional judgment.
It strengthens and focuses it.
- By sorting first and judging second, we can direct expert attention to locations where it is most likely to improve safety.



Statens vegvesen

Norwegian Public Roads
Administration