The Impact of Smart Roads' Adaptive Speed Limit on Road Safety

Pierre-Elie Bélouard - UC Berkeley SafeTREC
CSCRS report

December 10, 2018

Abstract

According to a study conducted by the Royal Society for the Prevention of Accidents, 25 % of the fatal crashes in the United Kingdom are caused by innapropriate speeds. Namely: speeds below the legal speed limit, but speeds which are not appropriate to the weather, traffic or visibility conditions. So-called adaptive (or dynamic) speed limits is a promising new technique which aims to prevent these crashes by adapting in real-time the speed limit to the road conditions.

This report reflects the research work lead during the Fall 2018 semester on the topic of adaptive speed limit by Pierre-Elie Bélouard, a CSCRS fellow and UC Berkeley graduate student. Pierre-Elie carried out research under the supervision of Dr. Offer Grembek, Co-Director of UC Berkeley, Safe Transportation Research and Education Center. This report is an intermediary step in the research project, which will be continued during the Spring 2019 semester.

The first part of this report recalls figures showing the economical importance of road safety. It introduces the concept of adaptive speed limit and features a short state of the art on the concept. The second part is devoted to a theoretical definition of the adaptive speed limit. After time, space and speed-discretization, we introduce the concepts of density of risk and explain how we could theoretically derive the adaptive speed limit by inverting the risk curve. This method could unfortunately not be implemented in practice, due to the difficulties to have an accurate estimation of the density of risk on a given road portion. Another approach is therefore chosen and explained in the third part of the report. We apply machine learning technique to learn design speed from features. The machine learning model, which uses the technique of random forest, demonstrates promising performances. The next step will be to extend this model to dynamic features.

Acknowledgements

Funding for this project was provided by UC Berkeley Safe Transportation and Research Education Center (SafeTREC) and the Collaborative Sciences Center for Road Safety (CSCRS), a U.S. Department of Transportation-funded National University Transportation Center led by the University of North Carolina at Chapel Hill's Highway Safety Research Center.

Contents

1 Introduction of the problem of adaptive speed limit. 6

1.1 The cost of road safety 6

1.2 Smart motorways in the UK: one of the first implementations of adaptive speed limits 6

2 A theoretical model: derivation of the adaptive speed limit by the inversion of the risk curve 7

2.1 Framework for adaptive speed limit 7

2.1.1 Space discretization and static maximal speed limit . . . 7

2.1.2 Dynamic speed limit vector as a function of the conditions parameters vector 7

2.1.3 Time discretization 7

2.1.4 Speed limit discretization 8

2.2 How to keep the drivers informed about the adaptive speed limit? 8

2.3 Relevant conditions parameters 8

2.3.1 Static parameters 9

2.3.2 Dynamic parameters 9

2.4 Design an efficient speed limit function 10

2.4.1 The microeconomic approach 10

2.4.2 The constraint approach 10

2.5 The risk density approach and its application to adaptive speed limit 11

2.5.1 Definition of the risk density 11

2.5.2 Risk density function on a given road portion at a given time 13

2.5.3 One way to derive Adaptive Speed Limit from the knowledge of the risk density 14

3 Application of machine learning techniques to estimate the design speed of highway portions 15

3.1 Context 15

3.1.1 Link to adaptive speed limit 15

3.1.2 Previous works 15

3.1.3 Main idea of machine learning techniques 16

3.2 About the data 16

3.2.1 Origin of the data 16

3.2.2 Objective: a regression to estimate the design speed . . . 16

3.2.3 Features selection 17

3.2.4 Features engineering 18

3.2.5 How to deal with missing data 18

3.3 Split of the data in two sets 20

3.4 Random forests in action 20

3.4.1 Principle of the random forest algorithm 20

3.4.2 Implementation of the method 20

3.4.3 Tuning of the hyperparameters 20
3.4.4 Importance of the different features 21
3.4.5 Results 21
3.4.6 Comments about the difficulties to correctly classify low design speeds 23

1 Introduction of the problem of adaptive speed limit.

1.1 The cost of road safety

According to [5], the total economic cost of motor vehicle crashes in the United States was worth **$242 billion in 2010**. This represents the present value of lifetime economic costs for **32,999 fatalities**, 3.9 million non-fatal injuries, and 24 million damaged vehicles. This cost represents 1.6 percent of the Gross Domestic Product of the United States of America.

1.2 Smart motorways in the UK: one of the first implementations of adaptive speed limits

The **United Kingdom** has been one of the first countries in the world to experiment adaptive speed limits. Adaptive speed limit are, along with the possibility to drive on the hard shoulder during rush hours and the massive introduction of sensors, the main innovative aspects of the program named **"UK smart motorway"**. This project was initially developed to reduce critical travel time during rush hours on UK's busiest highway sections. This program now appears to be a success, from a travel time as well as from a road safety point of view.

- The first smart motorway in the UK was introduced in 2006 (M42, in the West Midlands)
- In 2018, there were 100 miles of smart motorways in the United Kingdom.
- The estimated cost of turning a traditional motorway into a smart one is 5-15 millions pounds per kilometer.
- Average travel time was reduced by 27 % on M26 after it was turned to a smart motorway in 2010. In the meantime, the number of road traffic incidents has been reduced by half.

2 A theoretical model: derivation of the adaptive speed limit by the inversion of the risk curve

2.1 Framework for adaptive speed limit

In this part, we care about implementing adaptive speed limit on a road of length L, between a point A and a point B. The road between these two points is called the **road section** of interest.

Hypothesises: We assume that:

- There aren't any crossroads on this portion.
- The **conditions are uniform** along the portion. This includes weather, as well as traffic and visibility conditions.

2.1.1 Space discretization and static maximal speed limit

The road portion is divided into N elementary segments. Each segment i has a length l_i. All road segments do not necessarily have the same length. Every segment is assigned a given static speed limit $\bar{v}_i$. This speed limit is set once and corresponds to the *speed limit under ideal conditions*. We denote by $\bar{V}$ the column vector (size n) of the maximal static speed limits.

2.1.2 Dynamic speed limit vector as a function of the conditions parameters vector

Let t_0 be fixed. Let us now assume that we have access to the vector giving the relevant parameters $C(t_0)$. Our goal is to find a relevant dynamic speed limit vector at time t_0 $V(t_0)$. This vector should be smaller than the maximal static speed limits vector:

$$0 \leq V(t_0) \leq \bar{V}$$

where the inequality is a coefficientwise inequality.

Moreover, $V(t)$ will be calculated as a function of the parameters vector:

$$V(t) = f(C(t))$$

2.1.3 Time discretization

We will not be looking for an explicit continuous expression of the adaptive speed limit vector on the road portion V, but rather re-estimate it on a regularly basis. Let us now introduce some time interval δ. For instance, we could choose: $\delta = 1$ min. At time $t_k = k \times \delta$, the adaptive speed limit will be recalculated:

$V_{k\delta} = f(C(k\delta))$. Between $k\delta$ and $(k+1)\delta$, the adaptive speed limit will be constant and equal to $V_{k\delta}$:

$$\forall t \in [k\delta, (k+1)\delta[, V_t = V_k = f(C(k\delta))$$

2.1.4 Speed limit discretization

The speed limit has to take discrete values. At a first glance, we might want him to take integer values:

$$v_i(C) \in \mathbb{N}_+$$

But we should not forget that the speed limit is designed for human driver. A human driver will make almost no difference between 36 and 37 miles per hour speed limits. In addition, speedometer are not able to display the exact speed of the vehicle, but only an approximation of it. For all these reasons, we choose to restrict ourselves to **speed limit why are divisible by 5 miles per hour**. The set of admissible speed limits becomes therefore therefore:

$$S = \{V \in \mathbb{R}_+^n, \forall i = 1, ..., n, 0 \leq v_i \leq \bar{v}_i, \exists k_i \in \mathbb{N}, v_i = 5k_i\}$$

The set of admissible speed limits is therefore **finite**. Let us denote by $\mathbb{F}_\mathbb{i}$ the set of feasible speed limits for the road portion i.

2.2 How to keep the drivers informed about the adaptive speed limit?

This adaptive speed limit may be communicated to the drivers by two different ways:

- **Adaptive road signs**, displaying the speed limit in real time. Such adaptive road signs are located at the beginning of each segment.
- Adaptive speed limit could be displayed in real time on the driver's dashboard. **Vanet systems** could be used to transmit the information to the car from the adaptive road sign. See [1] for more informations on how to use VANET systems to achieve such communications.

Let us now lead some investigations in order to choose both a relevant set of condition parameters and a relevant adaptive limit function $f(C(t))$.

2.3 Relevant conditions parameters

Let's now look for some relevant quantitative parameters reflecting the road conditions. The suggested list is neither exhaustive nor static; new relevant parameters could be added to it during the pace of the research project.

Figure 1: Adaptive road signs on smart motorways in the United Kingdom

Conditions Parameters or features? In this part, we choose to use the word *conditions parameters*, since we feel that it reflects better the physical reality of road safety. We will then do some machine learning and will therefore use the word *features* instead. The reader of this report should keep in mind that *features* and *conditions parameters* are synonyms.

Below is a non-exhaustive list of features which could be relevant in the process of determining the adaptive speed limit.

2.3.1 Static parameters

- Road width
- Road bend radius
- Quality of asphalt

2.3.2 Dynamic parameters

Weather parameters

- History of precipitations
- Current intensity of precipitations
- Visibility

- Luminosity
- Speed of wind gust
- Temperature

Traffic parameters

- Number of different vehicle families (e.g.: cars, trucks, coaches, motorbikes)
- Typical behavior for every vehicle family (e.g.: cars drive at the maximal speed limit, trucks drive at 40 mph, etc.)
- Flow and density for every vehicle family

2.4 Design an efficient speed limit function

This section explores two ways of expressing the dynamic speed limit as a function of the condition parameters vector.

2.4.1 The microeconomic approach

In this first approach, the adaptive speed limit on one given segment is defined as **the speed limit vector achieving the highest utility**. The utility is a function of both speed and the conditions parameters vector. It takes into account road safety and travel time. Both travel time and risk of injury have a negative effect on the utility.

$$V^*(t) = max_{V s.t. 0 \leq V \leq \bar{V}}(U(C(t), V))$$

2.4.2 The constraint approach

In this second approach, the adaptive speed limit is determined as **the maximum speed limit satisfying a set set of security constraints**. The first constraint is a static one: the adaptive speed limit should not be greater than the speed limit under ideal conditions. The other constraints are dynamic and take one or more conditions parameters into consideration. One dynamic constraint should for instance be a stopping constraint.

Advantage of the constraint approach The constraint approach has two main advantages:

1. It allows easy and **quick computation** of the adaptive speed limit
2. It has a huge **pedagogical advantage**: exhibiting one binding constraint is sufficient to justify the adaptive speed limit.

Difficulty On the other hand, the main difficulty of this method is to find a relevant set of constraints. They should be determined by complex physical equations.

Remark The constraint approach is a particular case of a microeconomic approach. For one given road segment i, if we denote by n_c the number of constraints and $S(a)$ the admissible set for the constraint a, the constraint problem could be written as following:

$$max(\frac{1}{2 \times \bar{v}_i} + \sum_{a=1}^{n_c} \mathbb{1}_{\{v_i \in S(a)\}})$$

under the constraint: $0 \leq v_i \leq \bar{v}_i$

2.5 The risk density approach and its application to adaptive speed limit

This subsection introduces the concept of risk density. It then explains how the problem of adaptive speed limit could be reduced to the one of the estimation of the risk density function. This risk approach is one example of a microeconomic approach. It is similar to many classical approaches in actuarial sciences and quantitative finance. The biggest challenge, which makes this method difficult to apply in practice, is to accurately estimate the risk density function.

2.5.1 Definition of the risk density

Let us consider a small road portion of size dx. Let us now assume the following points:

- The conditions C on this small portion do not change at all during a very long time T (infinite).
- Many drivers go through this road portion. They all have the same behaviour and travel at the same speed v.

On this road portion, between 0 and T, we will observe some crashes. Let P be the cumulated cost caused by all these crashes. This cost includes both physical and properties damages. Let N be the number of vehicles going through this portion during the interest period.

Risk density The risk density under the conditions (C, v) is defined by the ratio of the cumulated cost of the observed crash over the product of the number of vehicles which have been through the road portion and the length of the road portion dx:

$$r(C, v, x) = \frac{P}{N} \times \frac{1}{dx}$$

Unit A relevant unit to express the density of risk is *USD per vehicle per million kilometers*. Assuming that, based on different criteria, we assign a price to every human death, the density of risk could be expressed in terms of equivalent death per vehicle per million kilometers.

Order of magnitude of the density of risk on the French network

- In France, the **total cost related to traffic injuries** is estimated to be 32.8 billion euros in 2015 (source: INSEE, the French National Institute of Statistics).
- During the same year, the **total distance traveled on French roads** is estimated to be 500 billion kilometers. This roughly corresponds to an average of $10,000$ kilometers per adult per year.
- The **average number of passengers per vehicle** in France is equal to 1.6.
- We could derive from these three figures an order of magnitude of the average risk density on the French road network:

$$\bar{r} \sim 100,000 \text{ €per million kilometers per vehicle}$$

- A 2013 study ordered by the French Prime Minister Jean-Marc Ayrault estimated that a fatal injury causing one death has a cost of 3 million euros for the society. If we choose this figure as a reference for the cost of a human life, we could have an order of magnitude of the value of the density of risk:

$$\bar{r} \sim 0.03 \text{ equivalent death per million kilometers per vehicle}$$

Risk and average risk on a road portion Given the conditions C and the speed limit v on a road portion $[AB]$ of length L at a given time t, we could compute the total risk on the portion by integrating the risk density over x:

$$R([AB]) = \int_{[AB]} r(C, x, v) dx$$

The **average risk density** $\hat{r}$ is the ration of the total risk on the road portion over the length of the road portion:

$$\hat{r}([AB]) = \frac{R([AB])}{L}$$

2.5.2 Risk density function on a given road portion at a given time

Back to the adaptive speed limit framework

- We now assume that the risk density on each road portion is constant and equal to the average risk density on the portion.
- Moreover, we assume that, during the time frame of interest, the conditions do not change. The conditions are summarized by a feature vector C.

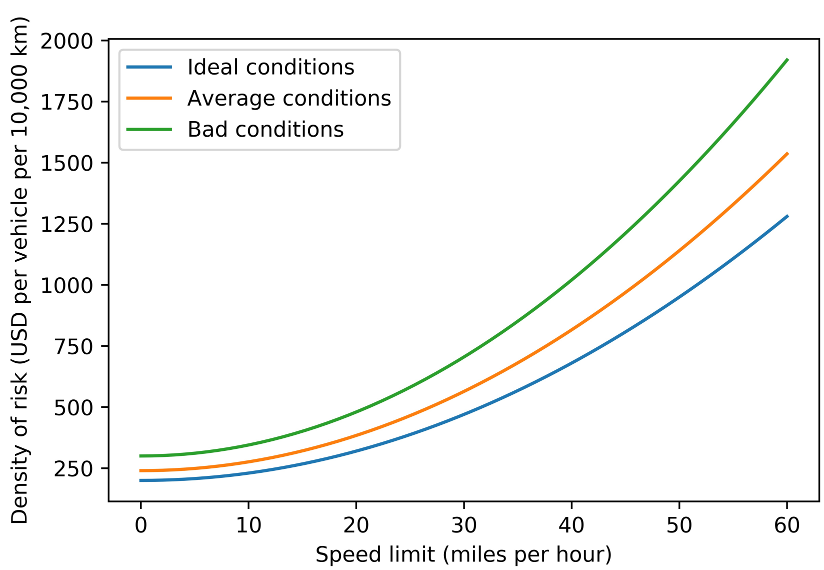
The **risk density function at time t on a given road portion** is the function which associates to every speed limit on the road portion the corresponding average risk density:

$$r(v) = r(C(t), v)$$

Properties of the risk density function Let $[AB]$ be a road portion. Let t be a fixed time. The risk density function on the portion $[AB]$ at time t is:

- positive
- increasing
- convex

Some examples of possible risk density functions: The shape of the curve is assumed. We have chosen parabolic curves.



2.5.3 One way to derive Adaptive Speed Limit from the knowledge of the risk density

Impact of a change of the conditions on the risk density function The risk density function depends on the conditions parameters. We assume that some changes in the conditions will not change the shape of the curve. But it will shift the curve. For a given speed limit, the risk density will be higher under bad conditions than under ideal conditions.

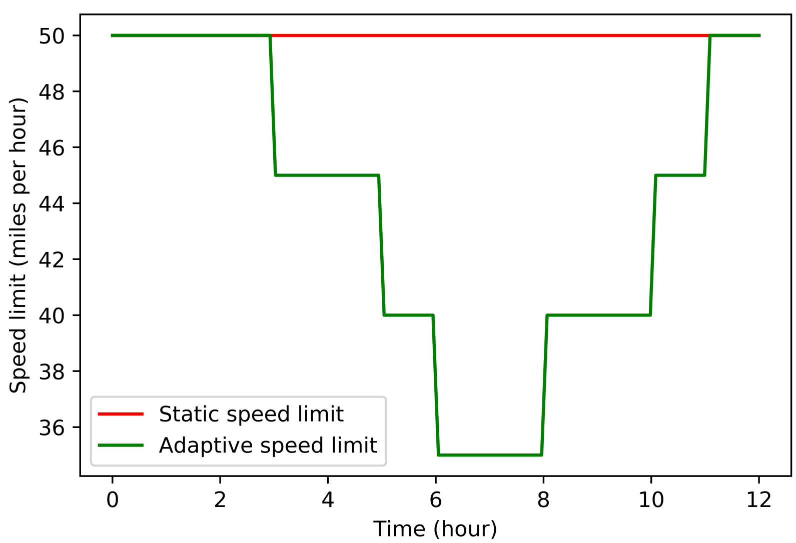
Adaptive Speed Limit by inverting the risk density function One simple way to derive an adaptive speed limit from the density curve is the following one:

1. Fix one risk level threshold r^*
2. Choose the greatest speed limit of the set of feasible speed limit that achieves a density not greater than the threshold:

$$v_{adaptive}(i,t) = max(v \in \mathbb{F}_i \text{ s. t. } r(C(t), v) \leq r^*)$$

This method to obtain the speed limit is called **inversion of the risk density function**.

Remark Some parallels could be drawn between this method to derive the adptive speed by inverting the risk density function and classical methods in financial mathematics that deal with risk. Further investigations may deserve to be lead on this topic.



3 Application of machine learning techniques to estimate the design speed of highway portions

3.1 Context

3.1.1 Link to adaptive speed limit

Derive the adaptive speed limit from the risk curve During the first part of our research project, we carried out theoretical research on road safety. We introduced the notion of density of risk under given conditions for a given speed limit on a given road segment. We then defined the risk curve which states the evolution of density of risk as a function of the speed limit. Finally, we showed that a good adaptive speed limit could easily be derived from the knowledge of the risk curve.

The previous method is not usable in practice. Indeed, it appears if not impossible at least highly complicated to estimate the risk curve with enough accuracy to derive a relevant adaptive speed limit. We explored the possibility of drawing the risk curve using simulations. But root causes of accidents are rarely known, making an accurate crash simulator barely implementable.

A new idea to empirically estimate the adaptive speed limit Rather than evaluating the density of risk, we now try to directly deduce the adaptive speed limit from the parameters. In this part, we focus on the static estimation of a good speed limit (i.e. we only take static road features into consideration). The main assumption is that the design speed is adequatly chosen. In other words, we assume that the design is the maximal speed achieving a reasonable level of risk under ideal conditions. We have therefore trained a machine learning model which estimates the design speed of a highway segment by analyzing the road static features. The next step (to be investigated during the next semester) will then be to extend this model to dynamic features (weather, traffic, visibility conditions, ...) in order to derive an adaptive speed limit from the conditions parameters.

3.1.2 Previous works

The idea of using **machine learning** to road safety is not new. Most of the attempts, like [6] have been focused on the **pattern extraction from crash data**. The approach chosen in this research project is relatively new: we have indeed not worked on crash data, but rather on technical data to estimate the design speed (i.e. the speed that, under good conditions, achieves a given level of risk). Since our machine learning model is not trained on crash data, which by nature reflect rare events and noisy and volatile, it is more robust. The drawback is that our model does not give us the risk level as a function of the speed limit, but only the speed limit which is supposed to achieve a reasonable risk level.

3.1.3 Main idea of machine learning techniques

Machine learning is the practical implementation of statistical learning. Two fundamental tasks in machine learning are classification (assigning a label given features) and regression (finding an approximation of a function of the parameters). In this research project, we will focus on the use of supervised machine learning techniques to **estimate the design speed** of a road segment given some technical characteristics (features) of the road portion.

Some widely-used regression machine learning techniques include:

- Random Forest regression.
- Support-Vector Machine (SVM) Regression.
- Logistic regression.

3.2 About the data

3.2.1 Origin of the data

In this research work, we are using **TASAS highway data**. TASAS is a huge data base which contains information about Californian motorways. Motorways are divided into segments. The database contains informations about $60,000$ **highway segments**, including:

- the location (latitude, longitude, length, elevation, ...) of the highway portion
- its design speed
- Technical informations: hard shoulder, type of surface, date of construction, width of the medium separation, ...
- Specific informations: presence of a toll plaza, rural/urban environment, ...

3.2.2 Objective: a regression to estimate the design speed

Every road portion has a design speed. This speed, which ranges between 30 and 70 miles per hour, is determined after technical studies. It corresponds to the *maximal appropriate speed under ideal conditions on this portion.*The design speed is set to achieve a certain level of risk and therefore **ensure reasonably safe journey** under ideal conditions to travelers who respect the design speed. Our objective is to build a classifier which will take in input the technical characteristics of the road segment (these characteristics are called the *features* in machine learning) and predict the design speed of the road portion.

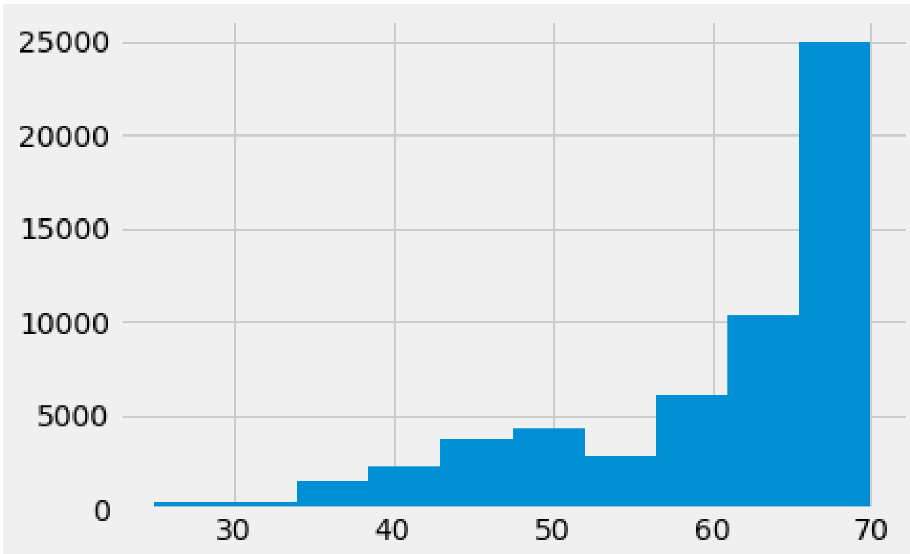


Figure 2: Histogram of the design speed of our data set

Remarks

- The design speed may differ from the true (legal) speed limit the user has to respect. The true legal speed limit may not be greater than the design speed on the road portion. In some cases, it might be smaller, because of external reasons that do not affect road safety such as noise constraints or political choices.
- Even tough we try to build a classifier which is as reliable as possible, a 100% accuracy is impossible to achieve. A possible method to evaluate the quality of the design speed predicted by the model is to *analyze the driver's responses*.

3.2.3 Features selection

Here is the list of the 18 features we selected to train our model:

- Surface type: Concrete, Oiled Earth-Gravel, Bridge Deck, Unpaved
- Number of lanes
- Outside shoulder total width
- Outside shoulder treated width
- Date of construction of the road

- Travel way width
- Interior total width
- Interior treated width
- Date of construction of the median barrier
- Type of the median barrier
- Width of the median barrier
- Type of the population (Rural/Urbanized/Urban)
- Presence of a toll plaza
- Presence of a forest
- Other specific features

Remarks: In the data set, characteristics are given for both sides of the road (left side and right side). We assumed that values were almost equal for both sides of the road and only used the values corresponding to one side (the left one, incidentally).

3.2.4 Features engineering

Once the relevant characteristics are selected for the model, some features need to be **preprocessed** before being used for the training of the model:

- In the data base, the construction date is a string (format: "YYYY/DD/MM"). We only keep the year and convert it to an integer.
- Some parameters, such as the terrain or the surface type, are non-numerical. We use the pandas function get-dummies which creates additional boolean (0/1) features and remove the non-numerical feature.

This preprocessing phase is necessary, since machine learning models only deal with numerical features.

3.2.5 How to deal with missing data

It appeared that the data set was of **high quality** and only a couple of thousands rows had some missing values. Since the number of missing values (about one thousand) is negligible compared to the size of the data, we chose to remove all road portions with at least one missing value. The pandas method dropna() is therefore convenient.

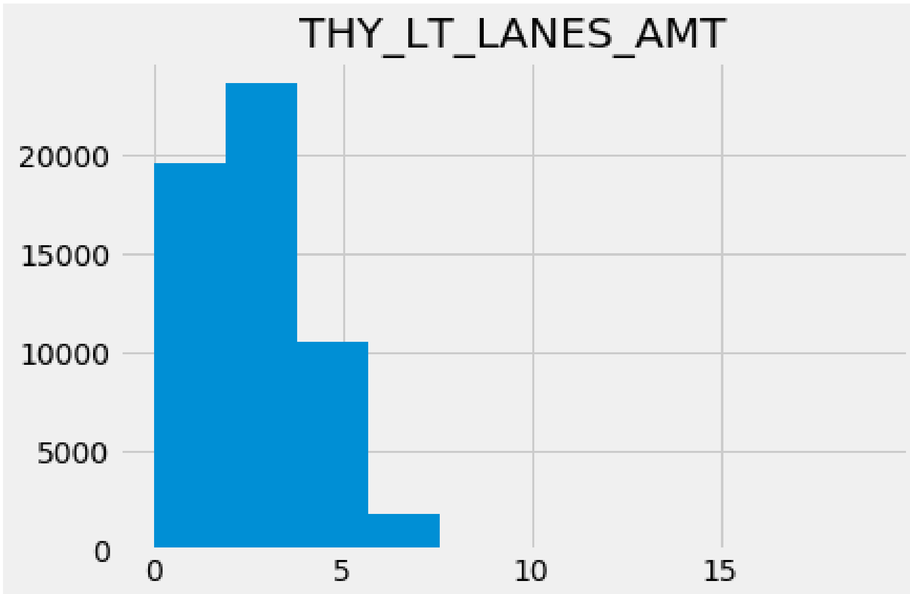


Figure 3: Histogram showing the repartition of the number of lanes

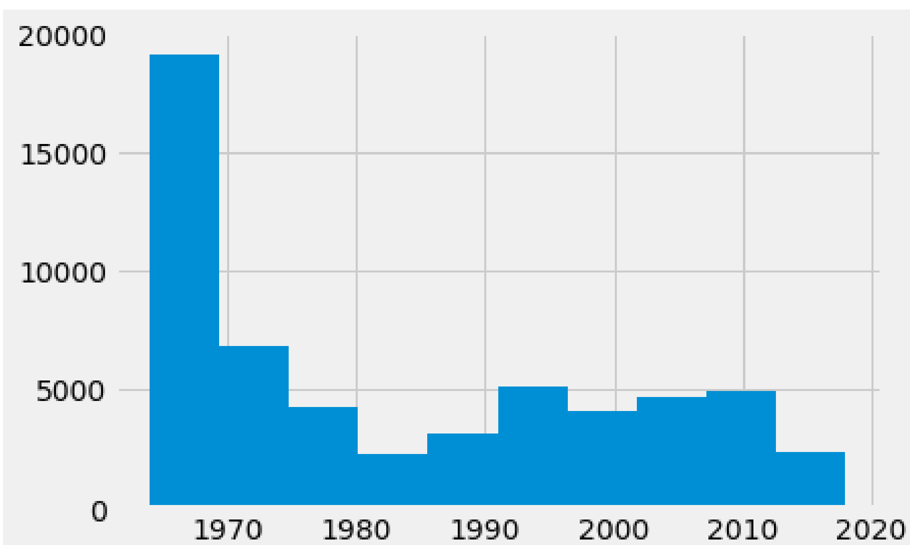


Figure 4: Histogram of the construction year

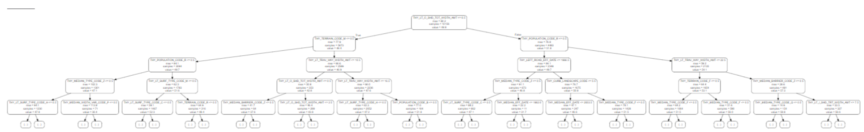


Figure 5: An example of a (truncated) regression tree

3.3 Split of the data in two sets

Our data set contains 56,727 rows. We randomnly split this data set into two subsets:

- One **training data set** (75 % of the rows)
- One **testing data set** (25% of the rows)

For this purpose, we use the function train-test-split of the machine learning open source library scikit learn.

3.4 Random forests in action

3.4.1 Principle of the random forest algorithm

Random forests is a powerful machine learning technique. It relies on the construction of many classification trees. A regression tree is a weak classifier. The random forest algorithm builds many different classification trees (in our case, 1,000) using different features and trained on different (randomnly selected) subsets of the training data. The value predicted by the model is the average of the values predicted by all decision trees.

3.4.2 Implementation of the method

We use the implementation of the random forest algorithm in the open source machine learning library **scikit learn**.

3.4.3 Tuning of the hyperparameters

In random forests, the main hyperparameter to tune is the **number of classification trees to build**. In our case, we chose to build 1,000 classification trees. Increasing the number of trees does not improve the performance (neither does it degrade it) and extends the training time. Decreasing the number of trees accelerates the training phase but slightly decreases the overall performance. **1,000 trees therefore seem to be a good compromise**.

```
Variable: THY_LT_I_SHD_TRT_WIDTH_AMT Importance: 0.39
Variable: THY_LEFT_ROAD_EFF_DATE Importance: 0.08
Variable: THY_MEDIAN_WIDTH_AMT Importance: 0.08
Variable: THY_MEDIAN_EFF_DATE  Importance: 0.07
Variable: THY_LT_TRAV_WAY_WIDTH_AMT Importance: 0.05
Variable: THY_LT_O_SHD_TOT_WIDTH_AMT Importance: 0.04
Variable: THY_LT_O_SHD_TRT_WIDTH_AMT Importance: 0.04
Variable: THY_TERRAIN_CODE_M   Importance: 0.03
Variable: THY_POPULATION_CODE_R Importance: 0.03
Variable: THY_LT_I_SHD_TOT_WIDTH_AMT Importance: 0.02
Variable: THY_LT_SURF_TYPE_CODE_H Importance: 0.02
Variable: THY_MEDIAN_BARRIER_CODE_Z Importance: 0.02
Variable: THY_LT_LANES_AMT     Importance: 0.01
Variable: THY_CURB_LANDSCAPE_CODE Importance: 0.01
Variable: THY_MEDIAN_TYPE_CODE_F Importance: 0.01
Variable: THY_MEDIAN_TYPE_CODE_G Importance: 0.01
Variable: THY_MEDIAN_TYPE_CODE_H Importance: 0.01
Variable: THY_MEDIAN_TYPE_CODE_Z Importance: 0.01
Variable: THY_MEDIAN_BARRIER_CODE_E Importance: 0.01
Variable: THY_TERRAIN_CODE_F   Importance: 0.01
Variable: THY_TERRAIN_CODE_R   Importance: 0.01
Variable: THY_POPULATION_CODE_B Importance: 0.01
Variable: THY_POPULATION_CODE_U Importance: 0.01
Variable: THY_TOLL_FOREST_CODE Importance: 0.0
```

Figure 6: The selected features sorted by decreasing importance

3.4.4 Importance of the different features

The higher a feature appears in a decision tree, the higher the role it plays. Once the model is trained, it is therefore interesting to have a look at the importance of the different features. Here are the results:

3.4.5 Results

Finally, we apply the trained model on the testing set and compare the design speed predicted by the model to the true design speed. In our case, we obtain an accuracy of 95 % with mean deviation between the predicted value and the true value of 2.38 miles per hour. This is significantly better than the naivest classifier (which always predicts the mean of the design speed), which has a deviation of 7.84 miles per hour.

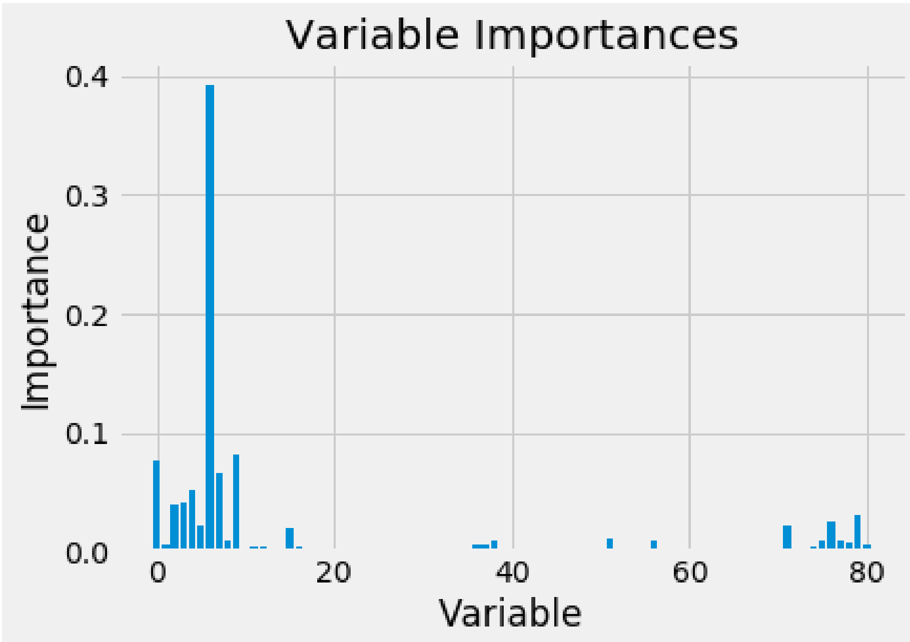


Figure 7: The relative importance of the different features

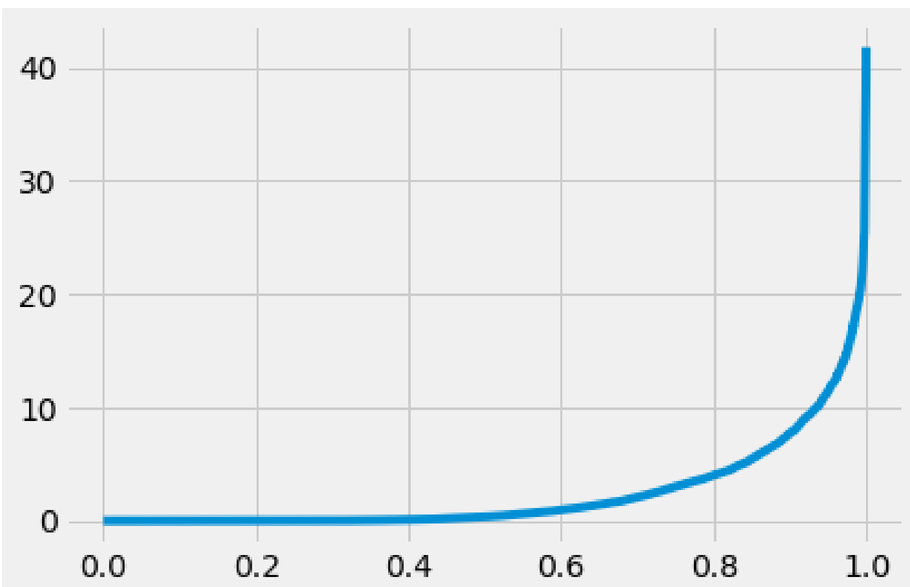


Figure 8: Cumulated density function of the classification error

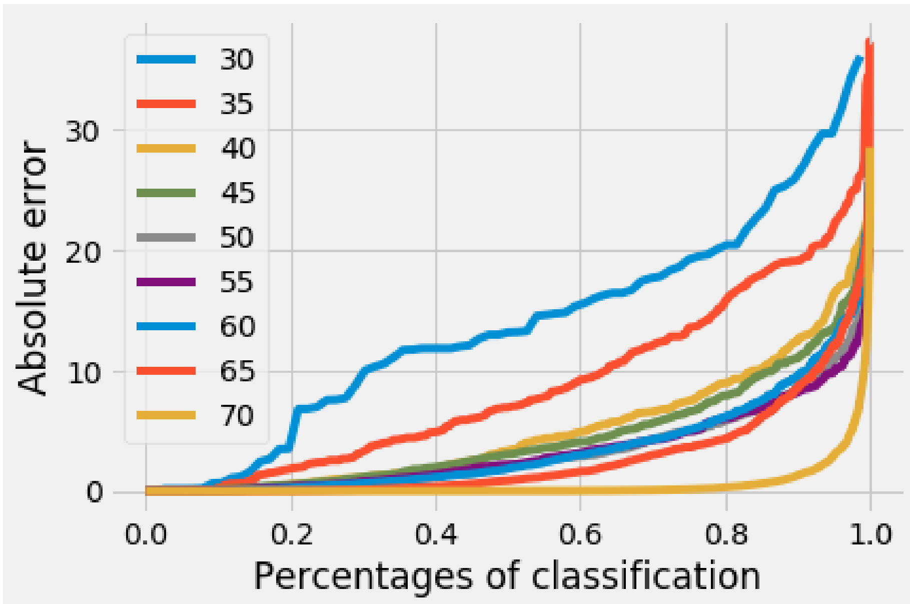


Figure 9: Cumulated density function of the classification error grouped by true design speed

3.4.6 Comments about the difficulties to correctly classify low design speeds

If we study the results of our method, we could remark that our regressor has outstanding performances for the 70 mph road segments and very good performances for the 40-65-mph road segments. Results are much less convincing for the 30- and 35-mph road segments. This poor performance could be explained by two causes:

- **The presence of hidden features**: 30- and 35-mph speed limits may be justified by a cause which is not expressed in the features.
- **The lack of data for 30-and 35-mph road segments**: very few 30- and 35-mph road segments are present in the training set. This could explain why our model performs poorly for this road segments.

References

[1] Arafat J. Al-Dweik, Matthew Mayhew, Radu Muresan, *Using Technology to Make Roads Safer: Adaptive Speed Limits for an Intelligent Transportation System*, IEE Technology Magazine, March 2017

[2] The Royal Society for the Prevention of Accidents, Great-Britain, *Inappropriate speed factsheet*, June 2018

[3] Sandy Balkin, Keith Ord, *Assessing the Impacts of Speed-Limit Increases on Fatal Interstate Crashes*, Journal of Transportation and Statistics, January 2001

[4] G.S. Donald, *Adaptive speed limit signs for road safety*, Australian Federal Office of Road Safety, 1994

[5] US Department of Transportation, *The Economic and Societal Impact Of Motor Vehicle Crashes*, National Highway Traffic Safety Administration

[6] Guangyuan Pan, Liping Fu, Lalita Thakali, *Development of a global road safety performance function using deep neural networks*, International Journal of Transportation Science and Technology, September 2017