



Evaluation and aggregation of pay-as-you-drive insurance rate factors: A classification analysis approach



Johannes Paefgen^a, Thorsten Staake^{b,c}, Frédéric Thiesse^{d,*}

^a ITEM-HSG, University of St. Gallen, Dufourstrasse 40a, 9000 St. Gallen, Switzerland

^b MIS and Energy-Efficient Systems, University of Bamberg, An der Weberei 5, 96047 Bamberg, Germany

^c Information Management, WEV G 217, Weinbergstrasse 58, 8092 Zurich, Switzerland

^d IS Engineering, University of Wuerzburg, Josef-Stangl-Platz 2, 97070 Wuerzburg, Germany

ARTICLE INFO

Article history:

Received 15 October 2012

Received in revised form 26 April 2013

Accepted 3 June 2013

Available online 12 June 2013

Keywords:

Location data

Actuarial decision-making

Pay-as-you-drive insurance

Classification analysis

Logistic regression

Neural networks

Decision trees

ABSTRACT

Vehicle sensor data enable novel, usage-based insurance premium models known as 'Pay-As-You-Drive' (PAYD) insurance, but pose substantial challenges for actuarial decision-making because of their inherent complexity and volume. Based on a large real-world sample of location data from 1572 vehicles, the present study proposes a classification analysis approach that addresses (i) the selection of predictor variables, (ii) the presence of class skew and time-variant prior distributions, and (iii) the suitability of classifier scores as an aggregated actuarial rate factor. Using raw location data, we derive a set of 15 predictor variables that we use to train and compare logistic regression, neural network, and decision tree classifiers. We find that while neural networks exhibit superior classification performance, logistic regression is better suited from an actuarial viewpoint in several ways. In sum, our results clearly demonstrate the potential of high-resolution exposure data for reducing the complexity of PAYD insurance pricing in practice.

© 2013 Elsevier B.V. All rights reserved.

1. Introduction

'Ubiquitous Computing' [54], 'Pervasive Computing' [47], 'Things that think' [27], 'Ambient Intelligence' [1], 'Silent Commerce' [21] – a plethora of novel terms presage the arrival of a paradigm shift in information processing. Common to all these concepts is the shared vision of a future world of everyday physical objects equipped with sensors and networking capabilities, whose quantity is expected to surpass the number of current Internet infrastructure entities by orders of magnitude [13]. From a management perspective, the appeal of this vision is grounded in the hope of closing the gap between real-world objects and their digital counterparts in the network, which may ultimately lead to a state of 'real world awareness' [29] of enterprise information systems. Ubiquitous wireless sensor technologies hold the unique potential of providing firms with a continuous stream of fine-granular and timely information on physical events within the organization and beyond [2].

A current example of this development is the increasing availability and commercial usage of vehicle sensor data. Today's road vehicles are equipped with an abundance of sensors that provide information on their location, situation, and state. Location information in particular, and the vehicle trajectories derived from it, have become an important cornerstone of applications such as road infrastructure optimization [43] and resource planning strategies in a vehicle fleet [46,50]. In this paper, we consider the use of vehicle location information in actuarial decision-making, a concept frequently referred to as 'Pay-As-You-Drive' (PAYD) insurance [17]. The idea underlying PAYD insurance contracts is to utilize location data in order to re-calculate car insurance premiums at periodic intervals – for example, on a monthly basis – based on the policyholders' individual driving patterns. The hope among insurers is that PAYD insurance will allow for tariffs that (i) more closely reflect actual risk exposure of road vehicles, and (ii) are adaptive over time, thus providing policyholders incentives to minimize risk. With PAYD insurance, information asymmetry between insurers and policyholders is reduced, which mitigates adverse selection and moral hazard [10]. For instance, prior studies have shown that respective tariffs have a significant impact on safe driving behavior among young drivers [7]. Furthermore, substituting conventional lump-sum premium payments with flexible rates brings about a better allocation of mobility-related costs. PAYD insurance has therefore also been associated with macroeconomic benefits such as insurance affordability, higher consumer surplus, improved traffic safety, and reduced externalities [36]. However, in order

* Corresponding author at: Information Systems Engineering, Julius-Maximilian University of Wuerzburg, Josef-Stangl-Platz 2, 97070 Wuerzburg, Germany. Tel.: +49 931 3180789; fax: +49 931 3181268.

E-mail addresses: johannes.paefgen@unisg.ch (J. Paefgen), thorsten.staake@uni-bamberg.de (T. Staake), frederic.thiesse@uni-wuerzburg.de (F. Thiesse).

to reap the alleged benefits of PAYD insurance, providers need to overcome the substantial challenge of adapting actuarial decision-making to incorporate vehicle sensor data, which differs substantially in complexity and volume from established variables.

Multivariate tariff models in motor insurance commonly follow a two-step procedure in the determination of premiums. First, actuaries define a set of discrete tariff classes. These classes are inferred either directly from categorical variables or by applying bounded-interval rules on continuous variables [3,25,30]. Pricing variables (*rate factors*) include gender or age of policyholders, among others. Secondly, for each tariff class, actuaries estimate the expected loss per policy over a certain period, referred to as the *pure premium*, based on distribution characteristics of historical claims data [16]. Tariff classes should accordingly exhibit significant variations in the corresponding pure premiums. Depending on the amount of claims data available to an insurer, there is a trade-off between the number of differentiating tariff classes and the accuracy of premium estimation within these classes.

Collecting vehicle sensor data allows insurers to generate various risk-related variables such as mileage, time of day, and the type of road a vehicle traverses. Since each additional rate factor increases the dimensionality of an actuarial table and multiplies the previous number of tariff classes by its number of categories or intervals, insurers must carefully consider which of these variables constitute suitable rate factors in PAYD insurance. An important precondition for PAYD insurance is hence determining suitable levels of data aggregation that combine several variables. Ideally, one might derive a one-dimensional aggregated variable that adds only one further dimension to actuarial tables. Such an aggregated variable may also be a suitable substitute for rate factors that have been ruled discriminatory in the wake of recent changes in insurance regulation in several regions. However, the existing body of literature on the topic of sensor data-derived rate factors is still very limited. Despite the rapidly growing interest among actuarial practitioners, there appears to be no consensus on how to approach the problem of PAYD insurance variable assessment and aggregation. Against this backdrop, we address this research gap by a classification analysis approach to rate factor aggregation. For the purpose of model building and evaluation, we build upon an extensive data sample gathered under real-world conditions from 1572 vehicles. Our results confirm the value of sensor data to the insurance business and show that the performance of different classification approaches varies considerably in the PAYD insurance scenarios.

The remainder of the paper is organized as follows. In the next section, we canvass key aspects of state-of-the-art actuarial decision-making in motor insurance and review related work on rate factors obtained from vehicle sensor data. Section 3 details the development of classifiers using logistic regression, neural networks, and decision trees. Sections 4 and 5 present the process of data collection and discuss the results from the empirical evaluation. The paper closes with conclusions and limitations to our work, and outlines suggestions for further research.

2. Theoretical background

2.1. Conventional motor insurance rate factors

In order to differentiate the risk of insurance policies, actuaries use a set of rate factors to separate policies into groups (i.e., tariff classes). The construction of tariff classes is ultimately a clustering task [3]. Each tariff class corresponds to a certain combination of rate factor categories or intervals in the case of continuous rate factors. For each tariff class, actuaries analyze historical claims data to arrive at a reliable estimate of the corresponding pure premium, that is, the minimum required payment per policy to cover the expected losses from its class [16,36]. An important distinction is usually made between claim frequency, claim type, and claim amount as different dependent variables to be estimated [24]. For each of these variables,

actuaries estimate a non-linear regression model from the historical claims in a tariff class. For a detailed review of specific regression techniques used to model the different dependent variables, we refer the reader to the review by Lord & Mannering [37].

Once the estimates for frequency, type, and amount distributions within a tariff class become available, actuaries calculate pure premiums as the expected claims value for a given period and tariff class. In automotive insurance, different coverage types exist such as liability, personal injury protection, protection from underinsured accident counterparties, and comprehensive insurance [26]. As our research is concerned with rate factors that define tariff classes rather than the calculation of pure premiums within these classes, we disregard coverage type in our analysis without a loss of generality. For each tariff class, pure premiums for a specific coverage type are independently calculable from historical claims of that type.

Actuarial decision-making in conventional motor insurance incorporates a broad range of established rate factors. These are roughly divisible into driver-related and vehicle-related predictor variables, all of which are typically obtained from a potential policyholder before the conclusion of an insurance contract through administering a questionnaire. Which specific rate factors actuaries choose for this questionnaire depends on an insurer's specific business policies as well as insurance regulations. Possible driver-related criteria include age, gender, nationality, and family status, while vehicle-related factors typically comprise the model and make of the car, its registration date, and cubic capacity [25,45]. While driver-related factors are typically more indicative of differences in claim *frequencies* within a tariff class, vehicle-related factors are more relevant to the estimation of claim *amounts* as they represent the residual value of an insured vehicle. Annual mileage constitutes a separate type of rate factor that is also referred to as the *exposure* of a vehicle [57]. However, while empirical evidence suggests that mileage is highly relevant for predicting accident risk [4,11], it is typically difficult to obtain correct mileage values directly from policyholders. Researchers have found that vehicle owners at times underreport their annual mileage, which may be attributed at least to some extent to elevated insurance rates associated with higher annual mileage [55].

2.2. PAYD insurance rate factors and vehicle sensor data

The rationale behind the PAYD insurance concept is that objective data on the actual usage of a vehicle – usually location information – allow for a more refined differentiation of premiums, which is also adaptive to changes in insured risk over time. Today, several tariffs require policyholders to inform the insurer if their annual mileage is significantly different from that of the previous year. However, PAYD insurance considers a much broader variety of variables and is consequently regarded as more objective than self-reported data. From another perspective, usage-based variables are also important as a substitute for established rate factors in insurance. For example, the European Court of Justice has ruled that from December 2012 at the latest, unisex premiums are mandatory for newly concluded insurance contracts (i.e., gender is to be disregarded as a rate factor) [52]. Similar provisions are likely for other factors such as nationality. The omission of such rate factors considerably impedes insurers' ability to differentiate between risks. PAYD insurance rate factors are potentially a nondiscriminatory alternative, and insurers who use them can more adequately price policies and gain competitive advantage.

Despite the impact of PAYD insurance on insurers' business practices and the substantial amount of PAYD insurance policies contracted by pioneering insurers, there are only few published studies that specifically address the relationship between actual vehicle sensor data and accident risk. Jun et al. [31,32] report an analysis of location data collected from 167 vehicles over a 14 month duration, 26 of which were involved in accidents during the study period. The authors derive mileage and velocity variables as derivatives of vehicle position after Kalman filter

smoothing, and they test for significant differences between accident-free and accident-involved vehicles. They find that the relatively small subsample of accident-involved vehicles exhibits higher velocities at most times, but this is inconsistent across road types and daytimes. Furthermore, violations of posted speed limits also appear more frequently in this group, as do hard deceleration events determined by a threshold on changes in velocity.

Toledo et al. [53] combine several such factors in an aggregated risk index variable. In an empirical study, 191 vehicles (belonging to a single commercial fleet) are equipped with a device that continuously records sensor data over the duration of several months. A regression analysis confirms a positive, significant relationship between the aggregated risk index and crash counts, albeit with low goodness of fit (R^2 between .056 and .064). The authors do not provide any specifics on the identification of driving maneuvers and conditions that are included in this variable. They address the relevance of an aggregated measure for fleet management and insurance purposes. In particular, they emphasize its suitability for feedback to drivers by allowing for a more comprehensible performance evaluation.

A comprehensive attempt to understand antecedents of road accidents that are identifiable from vehicle sensor data was undertaken by the US National Highway Traffic Safety Administration [18]. The study included driver characteristics such as attention and aggressiveness of driving. It focused on unobtrusive collection of vehicle data to achieve a high external validity. However, its contribution is largely of a conceptual nature, as it develops a qualitative classification scheme of relevant influence factors on driving safety and discusses descriptive statistics along these classes. No inference statistics are provided. Furthermore, the sensorization of vehicles is very elaborate, including five channel video, and presumably not transferable to the case of PAYD insurance.

These and other empirical studies typically suffer from small sample sizes – around 100 vehicles – and often rely on proxies of accident risk due to the low overall probability of accidents in typical vehicle populations. Conceptual contributions have suggested further risk-related variables retrievable from positioning data such as the number of trips, the length of individual trips, and the amount of driving in regions of elevated accident risk [40], but these are speculative in nature and do not provide hard evidence. The lack of empirical studies using samples of sufficient size and resolution may be attributed to technical and organizational challenges and in particular to the reluctance of insurance companies to publish data from existing PAYD insurance systems. This has hampered the advancement of knowledge in this relevant field of research in the past.

3. Research approach

Our research builds upon a dataset that overcomes previous limitations of both size and resolution. In our analysis, we draw upon established risk factors inferred by traffic safety researchers from observed-accident statistics. Besides vehicle mileage, these factors include different times of day [41], days of the week, vehicle velocities, and driving locations. The influence of such factors on accident risk is typically estimated by count regression models that compare accident frequencies under varying external conditions [37], for example, on a specific road type. In recent years, vehicle sensor data has made similar variables available on the level of individual road users. In combination with the corresponding drivers' mileage, these variables allow for the computation of relative exposure under specific situations and thus for rate factors suitable to PAYD insurance schemes. However, it is evident that a multitude of such variables exists, which requires a method that aids actuaries in the process of variable selection and combination.

Data mining methods are an established instrument in actuarial decision-making that supports the structuring of information and the discovery of new risk-relevant patterns in data [3]. We propose a data mining approach to aggregating variables obtained from vehicle sensor data – specifically, GPS-based location data – into a scalar rate factor. In particular, we apply classification analysis to reduce an n -dimensional set of predictor variables representing driving patterns to a one-dimensional variable (see Fig. 1). We hypothesize that *classifier scores* [20] are a suitable candidate for an aggregated PAYD insurance rate factor. Classifier scores represent a measure of certainty in the assignment of a predicted class label. In classifier training, we consider the differences in driving patterns between accident-involved and accident-free vehicles in a real-world sample of location data, which corresponds to a binary classification problem.

For tariff classification, which lies beyond the scope of this paper, actuaries may employ the aggregated PAYD rate factor in the same way as conventional scalar rate factors such as age or self-reported mileage. As stipulated in the introduction, tariff classes are obtained as combinations of categorical variables. In order to categorize scalar rate factors, intervals are therefore defined depending on an insurer's specific policy portfolio. These may be chosen, for example, such that tariff classes are of roughly equal volume.

In devising our methodology, we follow the general guidelines for classification analysis in a decision support context as proposed by Shreve et al. [48]. The present study focuses on the comparative evaluation of (i) different sets of predictor variables and (ii) different classifiers.

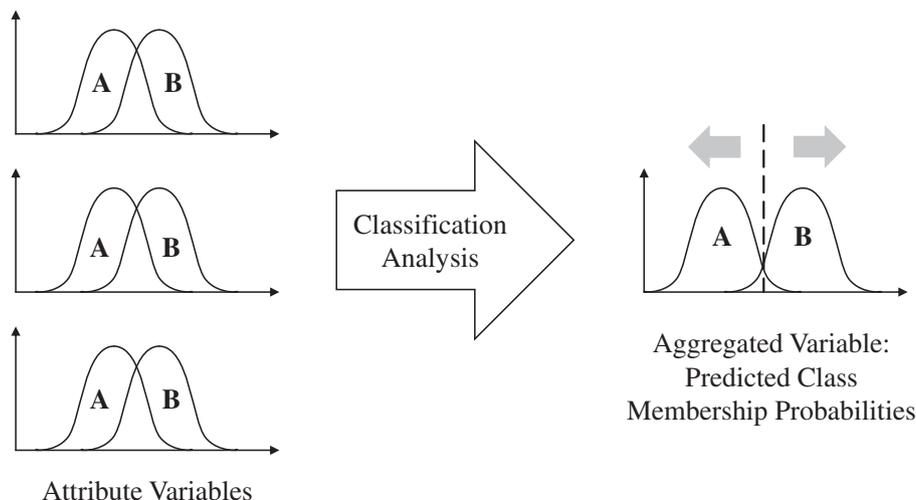


Fig. 1. Reduction of dimensionality of decision problem through classification analysis.

The relative importance of different variables that are extractable from location data may vary significantly. A comparison of classifier performance between different variable subsets from a sample allows us to assess variable subsets according to their predictive relevance. Besides different variable subsets, we also consider varying prior probabilities [48], in this case the ratios of accident-involved to accident-free vehicles in the sample. This is imperative, as insurers do not know the a priori distributions of insured risks within a portfolio. Classifiers should thus be insensitive to sample selection bias (i.e., the dependency of certain classifier parameters on prior distributions of classes in the sample) [33,60]. Moreover, depending on the observation period, the number of accidents in a vehicle population is typically only a fraction of its overall size, resulting in considerable class skew. We therefore investigate the effects of asymmetric priors on classifier training and the robustness of classification models against changes in class ratios over time.

We also compare the performance of different types of classifiers, a common theme in decision support research [9,44]. However, we remark that predictive performance is not sufficient as a selection criterion for classifiers in the context of PAYD insurance rate factor aggregation. The proposed use of classifier scores requires an analysis of their distributional characteristics and their suitability for the discrimination of tariff classes, that is, the formation of categorical variables in a subsequent step based on an actual policy portfolio.

In summary, our intent is to

- provide a decision tool for the selection of predictor variables from vehicle sensor data,
- examine the effect of asymmetrical and time-variant class distributions on prediction performance, and
- assess the resulting classifier scores with respect to their use as a novel, aggregated rate factor.

We also take into account the comprehensibility and structural logic of the resulting classification models from an expert perspective [38], and we outline relevant implications for the DSS community as well as for insurance professionals. In a larger sense, we hope that by providing empirical evidence for the usefulness of vehicle sensor data to the insurance business, we increase managerial awareness of the topic and motivate more insurers to investigate the impact of PAYD insurance on their business practices [8]. Moreover, as we expect the topic of ubiquitous sensing to gain significant traction in the decision support domain in the future, we hope our work can serve as an anchoring point and will spur future research in the community.

4. Data and methodology

4.1. Sample description

We obtain data from the database of a major European PAYD insurance provider that currently comprises more than 1.0 M vehicles. Each vehicle is equipped with an on-board unit that includes a GPS sensor and wireless transmission capabilities. During vehicle operation, position updates are carried out every few seconds and aggregated on the device level to reduce costs of transmission and storage. For aggregation, the system calculates traveled distance from incremental position updates and generates new data entries every 2000 m. In addition to a vehicle's latitude and longitude, data points consist of a time stamp, ignition status of the vehicle, and driven distance since the previously generated data point. This distance can in some cases exceed the 2000 m interval if no position update is available for some time, for example owing to signal obstruction. Through straightforward computations, we extended raw data points to include the elapsed time since the last update, which in turn allowed us to compute the average velocity for the previously driven distance. In addition, the system infers a road type indicator from data point locations, which distinguishes urban roads, extra-urban roads, and highways. Start and end locations

of vehicle trips are available from data points generated upon changes of the vehicle ignition status (i.e., engine start and switch off).

Since the database was not accessible to us in its entirety due to its size and privacy restrictions, we extracted data in a randomized sampling procedure. We obtained reference samples for “high” and “low” risk driving patterns as follows: we randomly drew a sample of 600 vehicles that had accidents in 2008, which contained six months of location data prior to the accident event. We used stratified sampling to achieve an even distribution of accident events over the year, so that one-twelfth (i.e., 50 vehicles) had accidents in the same month. By sampling an equal number of accident events for each month, we hoped to eliminate the effect of seasonal variations on accident frequencies in our analysis. No location data beyond an accident event were included in the sample, as previous work has reported strong variations in driving patterns in the aftermath of an accident [39]. As a baseline, we furthermore randomly drew a second sample of 1000 vehicles from the data pool with twenty-four months of available location data without accident-involvement throughout this period, spanning from July 2007 to June 2009. For privacy reasons, no driver particulars of any kind were included in the sample.

A number of vehicles were eliminated from both samples due to the following reasons:

- Further accident events in the six-month observation periods of accident-involved vehicles that would affect vehicle usage,
- Errors in data recording or storage that rendered the resulting log files impossible to process,
- Failure of GPS sensors over prolonged periods, so that no location data was available for certain vehicles while ignition status indicated vehicle use, and
- Non-continuous GPS readings that resulted in excessively long traveled distances.

These instances were identifiable without doubt and resulted in a reduction of the accident-involved sample by 17 vehicles to 583 and of the accident-free sample by 16 vehicles to 984. No further elimination of outliers was undertaken, since we argued that due to the large sample size, their effect on the analytical results was negligible. Both samples combined cover approximately 45.7×10^6 km driven distance in 1.0×10^6 h of vehicle operation.

We acknowledge that the ratio of the two classes does not represent typical accident frequencies in policy populations, which are on the order of 1:10 over a one-year period and correspondingly lower for the one-month tariff periods conceivable for PAYD insurance premiums. We deliberately chose a biased sample to focus on the objective differences between groups and achieved a sufficient resolution of within-class differences. However, as stated in the previous section, we specifically analyzed the effect of class skew and varying prior probabilities on classification performance.

4.2. Predictor variables

For classification analysis, we derive predictor variables from raw location data. By accumulating traveled distances over all data points of a vehicle, we calculate individual overall mileage M . Besides mileage, we consider four different categories of situational exposure: time of day (T), day of the week (D), road type (R), and the average velocity (V) with which a vehicle was operated between two data points. We discretize these features into a fixed number of intervals, where each interval corresponds to a predictor variable that captures the amount of exposure accumulated under specific situational parameters. Time of day is divided into four variables T1–4 that capture vehicle operation at nighttime, daytime, as well as early and late evenings. Literature suggests that within these periods accident frequencies can be assumed stable [48], which we confirmed for our data in a comparison of class means across a high-resolution time grid. Two further predictor variables capture exposure on weekdays and weekends (W1–2), while

another variable triplet distinguishes the three road types (R1–3). We separate the last variable group, average velocity, into five intervals of 30 km/h increments V1–5, where the fifth interval comprises mileage accumulated at velocities larger than 120 km/h (i.e., V5 has only a lower bound). Table 1 gives an overview of these variables and the respective groupings.

Values of all situational predictor variables are determined as the mileage accumulated within the specified situational interval. We thus allocate a traveled distance to several variables at the same time (e.g., a particular road type on a particular day), which causes the resulting predictor variables to exhibit a high degree of collinearity. This is an undesired property of predictors in most classifiers [33] and is particularly cumbersome for logistic regression [12]. We therefore normalize all variables by dividing them by the vehicle's total mileage, converting it into the fraction of exposure accumulated under the specified condition

$$\bar{X}_v = \frac{X_v}{M_v}, \quad (1)$$

where v is a unique vehicle identifier. For the example of a normalized predictor variable T1, a value of .1 therefore implies vehicle operation between midnight and 5 am in 10% of total mileage. As observation periods are not consistent for the two subsamples of accident-involved and accident-free vehicles, and are even subject to minor variations within a subsample, we furthermore normalize the overall mileage M in a subsequent step by dividing it by the observation period of a vehicle. This operation yields rather small values of the normalized M between 0 and 0.01. Thus, we rescale the variable so that its maximum value over all vehicles in the dataset takes the value of 1. Finally, we compute an additional variable MLN by taking the *natural logarithm* of variable L instead of rescaling. This is motivated by the observation that the distribution of M appears to follow a log-normal distribution, which is consistent with previous findings in the analysis of driving patterns [42]. Table 2 provides descriptive statistics for the entire set of predictor variables for each group separately, where “Q” denotes a quartile and the second quartile is equivalent to the median.

4.3. Classification models

Based on recommendations for classifier selection made by Kiang [33], our analysis considers three types of classifiers, namely logistic regression, neural networks, and decision trees. Logistic regression is a straightforward extension of conventional linear regression that allows for binary dependent variables and hence suits a two-class classification problem. It employs the linear predictor function $g = \beta_0 + \sum \beta_i X_i$ as the argument of a non-linear logistic function, thus implying a binomial

Table 1
Predictor variables.

Grouping	Variable description	Label
Time of day	Exposure fraction between midnight and 5 am	T1
	Exposure fraction between 5 am and 6 pm	T2
	Exposure fraction between 6 pm and 9 pm	T3
	Exposure fraction between 9 pm and midnight	T4
Day of week	Exposure fraction on weekdays except Friday	D1
	Exposure fraction on weekends with Friday	D2
Road type	Exposure fraction on urban roads	R1
	Exposure fraction on highways	R2
	Exposure fraction on extra-urban roads	R3
Average velocity	Exposure fraction between 0 and 30 km/h	V1
	Exposure fraction between 30 and 60 km/h	V2
	Exposure fraction between 60 and 90 km/h	V3
	Exposure fraction between 90 and 120 km/h	V4
	Exposure fraction above 120 km/h	V5
Mileage	Overall exposure 30-day average	M
	Overall exposure 30-day average, LN-transformed	MLN

Table 2
Descriptive statistics of sample data.

Variable	Accident-free vehicles				Accident-involved vehicles			
	Mean	1st Q	2nd Q (median)	3rd Q	Mean	1st Q	2nd Q (median)	3rd Q
T1	0.080	0.007	0.022	0.082	0.061	0.010	0.032	0.083
T2	0.788	0.721	0.807	0.876	0.758	0.683	0.785	0.865
T3	0.093	0.031	0.089	0.141	0.117	0.059	0.114	0.164
T4	0.039	0.009	0.027	0.054	0.063	0.015	0.041	0.089
D1	0.590	0.511	0.592	0.677	0.622	0.550	0.620	0.697
D2	0.410	0.323	0.408	0.489	0.378	0.303	0.380	0.450
R1	0.319	0.177	0.292	0.432	0.313	0.180	0.275	0.406
R2	0.320	0.141	0.294	0.477	0.372	0.218	0.367	0.541
R3	0.361	0.214	0.349	0.482	0.315	0.192	0.289	0.420
V1	0.171	0.100	0.152	0.217	0.137	0.084	0.118	0.172
V2	0.265	0.193	0.259	0.330	0.247	0.164	0.226	0.306
V3	0.315	0.190	0.266	0.393	0.275	0.179	0.232	0.304
V4	0.142	0.066	0.139	0.197	0.182	0.123	0.175	0.232
V5	0.107	0.002	0.058	0.166	0.159	0.030	0.129	0.260
M	0.163	0.072	0.107	0.176	0.287	0.184	0.259	0.342
MLN	0.715	0.652	0.705	0.770	0.816	0.776	0.821	0.858

distribution. Maximum likelihood optimization estimates the coefficients β_i for given training data. The output of the resulting logistic regression function in Eq. (2) is interpreted as the probability of event occurrence given a predictor variable vector \mathbf{X} [12]. This value corresponds to the classifier score and for classification purposes, a simple threshold applied on it yields the predicted class labels

$$\hat{P}(a = 1|\mathbf{X}) = \frac{e^g}{1 + e^g}. \quad (2)$$

For a second classifier, we consider a multi-layer perceptron with a single hidden layer and a sigmoid activation function [43,56]. Depending on the specific sigmoid used, this type of neural network corresponds to a “stacking” of logistic regression operators in the form of the right hand side in Eq. (1). In each node, the weighted sum of continuous node outputs from the previous layer (or the predictor variables) is the argument of the activation function, where the coefficients β_i correspond to the weights on the edges connecting nodes between layers [19]. The continuous output of the last node corresponds to the classifier score assigned to a specific case by the neural network. We choose the number of nodes in the hidden layer dynamically as $N / 2 + 2$, where N is the number of input variables of the classifier. Therefore, at least three nodes will be contained in the hidden layer. The comparison of neural networks with logistic regression is a common theme in decision support systems, and previous work has established that the more complex model structure of neural networks often outperforms logistic regression in classification tasks [9,34,51].

Decision trees as the third classifier considered in our analysis separate cases in a sample according to simple, one-dimensional thresholds applied on predictor variables. The resulting partitions, or leaves of the decision tree, exhibit a certain class ratio on the basis of which classification decisions are made. Decision trees have been used in an actuarial context to determine an appropriate discretization of continuous rate factors [3]. Their classifier score output corresponds to the within-leaf odds and is thus piecewise constant, i.e., all cases in a leaf share identical classifier scores. For all three classifiers, sample size should be appropriately large in order to incorporate 15 predictor variables without significant deterioration of performance [14,48].

4.4. Model evaluation

We train logistic regression, neural network, and decision tree classifiers on different subsets of the predictor variables (see Table 1) and under varying sample compositions. For ease of comparison, we employ

identical subsets of predictor variables across models [6]. To achieve generalizable results and avoid model over-fitting, we employ a 10-fold cross validation setup as depicted in Fig. 2. For the comparison of different prior distributions (i.e., class ratios), we consider all cases of the minor class and randomly sample cases from the other class to achieve the given ratio before applying cross validation.

We assess the predictive performance of classifiers based on three different metrics: accuracy, F1 measure, and area under the ROC curve (AUROC). Although more advanced test series for the statistical validation of classifier performance exist [15], we argue that an in-depth analysis of individual classifier performance does not add substantially to the gist of the present study. Accuracy, or hitrate, is presumably the most basic performance measure in classification analysis. It denotes the overall percentage of correctly predicted labels of both classes. By convention, we refer to the minority class (i.e., accident-involved vehicles in the original sample) as the positive class. Accuracy is calculated as

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}, \tag{3}$$

where true positives *TP* and true negatives *TN* denote correctly classified vehicles, respectively, while false positives *FP* and false negatives *FN* denote the incorrectly classified cases. Accuracy as a performance criterion has well-known shortcomings particularly for skewed class distributions, where it rewards model bias [48]. We therefore also include the F-measure as the harmonic mean of within-class accuracies – also referred to as sensitivity for the positive and specificity for the negative class – in our analysis:

$$\text{F-measure} = \frac{2TP}{2TP + FP + FN}. \tag{4}$$

Accuracy and F-measure assess the absolute classification performance by comparing predicted with observed class labels. In that sense, they can in fact be said to measure different sides of the same coin, and exhibit acknowledged shortcomings [23]. We thus consider the area under the receiver operating characteristic curve (AUROC) as an additional performance measure that takes the entire range of classifier scores produced by a classification model into account and is a measure of their discriminatory power in terms of a ranking of cases in a sample, instead of the resulting binary class labels [20,49]. The ROC curve is calculated by applying a shifting threshold to the monotonically increasing classifier scores and plotting *TP* versus *FP* values for the resulting discrete classification in a two dimensional plane. In the case of a perfect classifier, a classifier score threshold exists that separates classes without overlap and its AUROC thus equals one.

Recently, problematic assumptions implicit in AUROC analysis were pointed out by Hand [28] for between-model comparison. Though we report AUROC scores for all classification runs, we therefore limit our interpretation to within-model performance comparison.

Although AUROC analysis gives an indication of how well a classifier is capable of separating two classes by their classifier score ranking, it is not a suitable criterion for the similarity of these rankings between two classifiers. Two classifiers may exhibit equal AUROC performance yet disagree to a large extent with respect to the relative positions of cases in a ranking. Following this line of thought, we also consider Spearman's rank correlation coefficient ρ [22] in our analysis. For this purpose, case rankings are inferred from classifier scores and the coefficient is computed as

$$\rho = \frac{\sum_i (c_i^A - \bar{c}^A)(c_i^B - \bar{c}^B)}{\sqrt{\sum_i (c_i^A - \bar{c}^A)^2 \sum_i (c_i^B - \bar{c}^B)^2}}, \tag{5}$$

where c^A and c^B denote the classifier score ranks according to classifiers A and B, respectively, from which the average rank is subtracted. If two classifiers yield scores that correlate with a high value of ρ , they would produce similar tariff classes in an actuarial table.

5. Results and discussion

5.1. Choice of predictor variable set

We evaluated classifier performance on eleven variable subsets. The first four subsets each correspond to a specific group of situational variables, namely time of day (T), day of week (D), road type (R), and average velocity (V). The subsets TD and RV combine two variable groups in which interaction effects are particularly likely. On the one hand, the daytime-dependent accident risk is supposedly different on weekdays than on weekends, for example, due to commuter traffic. On the other hand, road type presumably affects the riskiness of a recorded average velocity. Velocity levels considered appropriate on a highway can be much more questionable on urban roads. The subset TDRV combines all fourteen situational predictor variables. In addition, the mileage exposure variable M and its natural transform MLN each form an additional subset of their own. Finally, we merge all available variables in the remaining two subsets, again without (TDRV + M) and with logarithmic transform (TDRV + MLN) of the exposure variable.

Fig. 3 provides results for classification performance. The T and V subsets exhibit a better performance than the D and R subsets across all three classifiers. The F-measure is generally low and deteriorates

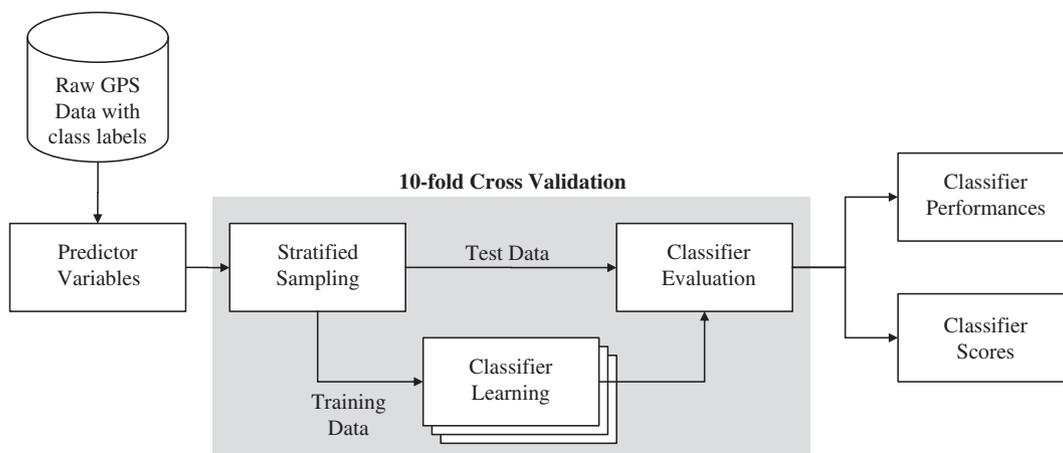


Fig. 2. Validation methodology.

to zero for the day of week variables, indicating that all cases were labeled as negative (accident-free) by classifiers. This still results in approximately 65% of overall cases labeled as correct, which explains the constant hitrate across all models as well as the minor overall variations in hitrate for the first four variable subsets. As expected, the combination of time of day with weekday improves the results, which does not hold for the combination of road type and velocity. In the majority of instances, neural network is the best performing classifier; however, the relative differences between classifiers vary substantially with the considered score. The single-variable subset M generally performs better than the combination of other variables TDRV, except for logistic regression where it has lower values of both accuracy and F-measure. The logarithmic transform of M yields an improvement in performance of the logistic regression classifier of 3% accuracy and 15% F-measure, which we attribute to the better linear separability of the transformed predictor. With respect to AUROC, however, no improvement is observable. In the full models, this effect is still evident and includes the AUROC value. The TDRV + MLN subset results in the best performing classifiers except for the accuracy of the neural network classifier, which is slightly lower than for the TDRV + M subset. The highest achieved performances were 82% accuracy, 75% F-measure, and an AUROC of 0.89.

Performance metrics, and in particular the AUROC of 0.89, can be considered high given the fact that no data regarding the driver was incorporated in the models. Ideally, we would contrast these with the classification performance for conventional rate factors, but such data is seldom published for confidentiality reasons. Also, in actuarial studies, empirical analyses of accident frequencies and rate factors typically report statistical measures that cannot be directly translated into performance criteria common in the Machine Learning domain (e.g. [5,59]). One possible explanation for the observed performance might be that our model is overfitted to the data. However, as we adhered to a cross validation methodology, we can exclude such issues. Thus, we argue that the used predictor variables in fact do explain accident risk to a significant extent – which is the general premise of PAYD insurance. An additional explanation might be that our analysis is strictly limited to the risk of accident occurrence, and precludes other dependent variables such as accident type or claim costs. These may be much better explained by driver and vehicle characteristics than by sensor data, but lie outside the scope of this paper.

5.2. Class skew and changes in class distribution

The full valid sample comprises 984 accident-free and 583 accident-involved vehicles. To investigate the robustness of the chosen classification models under class skew, which we expect in realistic vehicle populations, we evaluate their performance for nine different class ratios. Next to the class ratio of 984 to 583 (1.688:1) in the original sample, further samples with the ratios 1:2, 1:1, 2:1, 5:1, 10:1, and 20:1 were generated. We trained classifiers on these new samples again using 10-fold cross validation and the full set of predictor variables with logarithmically transformed mileage (TDRV + MLN).

Fig. 4 depicts the resulting performance scores for the three classifiers. Since accuracy and F-measure reward bias and are therefore unsuitable for comparing models trained on skewed class distributions, we report only AUROC values. Logistic regression exhibits the most stable performance in terms of AUROC, bounded between .88 and .91. For the other two classifiers, performance has a higher dependency on class skew, but remains in corridors of .04 (neural network) and .06 AUROC points width. We conclude that for decision trees, class skew has the strongest effect on the ability to discriminate cases. We attribute the apparent differences in AUROC for class ratios 1:2 and 2:1 to the decreased sample size in the former scenario, which falls below 1000 cases and may thus violate the criteria given in [48].

Imbalanced samples result in an overestimation of the majority class and thus introduce classifier bias. To substantiate this line of thought, Table 3 reports the coefficients and intercepts for logistic regression classifiers for different levels of class skew. The constant term displays a much stronger deviation (from 1.04 to -4.454) across class ratios than any variable coefficient, so that the observed improvement in accuracy evidently results to a major extent from this influence. From an actuarial perspective, rate factors should not be susceptible to bias and rather represent the objective differences between cases, thus also mitigating the issue of time-variant class ratios in an insurer's portfolio. Therefore, we consider the logistic regression classifier suitable for rate factor aggregation if the intercept is either omitted from the model or appropriate corrections [35] are applied by actuaries. The assessment of bias in the other two classifiers, neural network and decision tree, is more complex due to their model structure. Training on each of the nine samples resulted not only in differences in class ratios in decision tree leaves, for

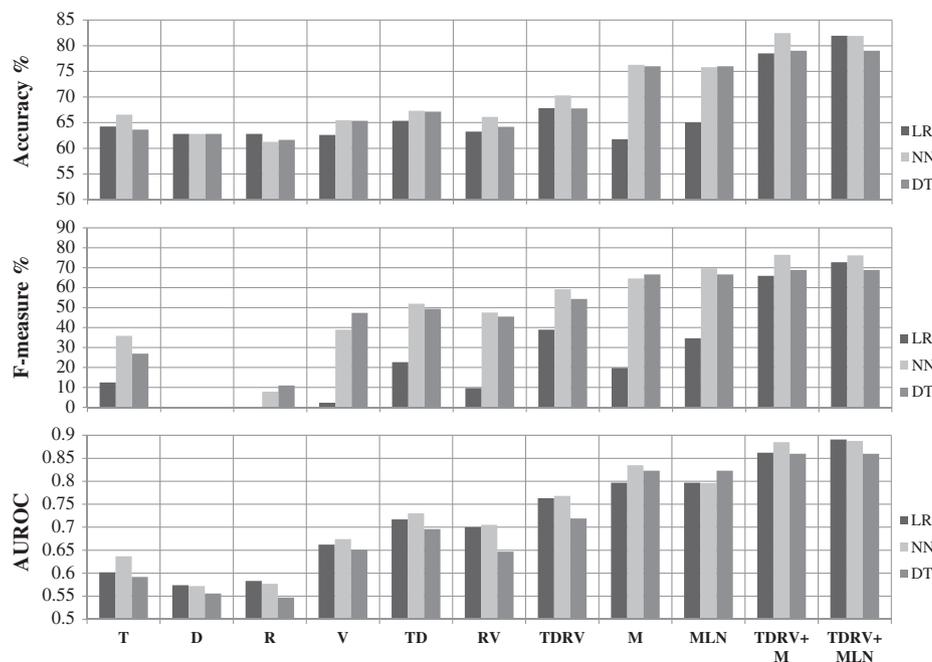


Fig. 3. Prediction accuracy, F-measure and AUROC for different predictor variable subsets.

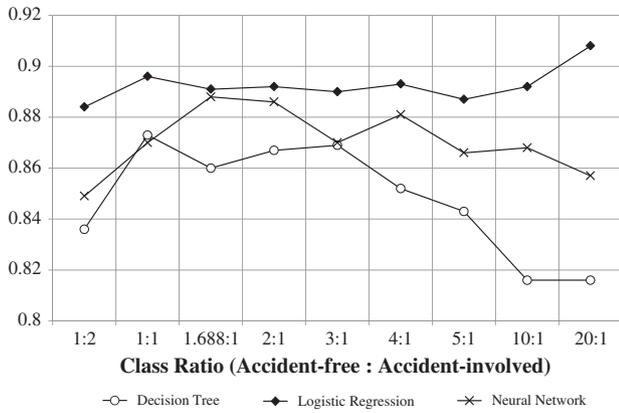


Fig. 4. AUROC of the classifiers for varying observed class ratios with variable set TDRV + MLN.

instance, but in an entirely different tree structure. It is hence questionable whether tariff classes obtained from an aggregated rate factor are stable if the underlying class ratios used for model learning change over time. The general opaqueness and difficult interpretability of neural network classifiers in particular is a common problem in decision support systems [33] and renders them useless in many applications where classifier outputs require justification [38].

5.3. Analysis of classifier scores

For three predictor variable subsets (TDRV, MLN, TDRV + MLN) and the full sample, Fig. 5 depicts classifier score histograms of the three classifiers, with light color representing accident-free case labels and dark color representing accident-involved ones. These scores can be interpreted as the classifier’s certainty in the prediction of the respective classification label. The closer the values are to the binary class label, i.e., zero or one, the more certain is the assignment of that class label. For an explanation of how classifier scores are derived, the reader may refer to Section 4.3. The distributions in the histograms are distinct for each classifier (note the varying scales on the vertical axes). For logistic regression, classifier scores follow a comparatively steady and continuous pattern. While the negative class has a prominent peak and exhibits skew towards zero, classifier scores of cases in the positive class are distributed more evenly. For the neural network classifier, both classes show oppositely skewed classifier scores, and

the TDRV set of predictors has a common third peak in the middle of the scale. The highest performing neural network classifier clearly separates classifier scores towards the bottom and top end of the scale with an unstructured overlap interval in the middle. The confidence value distribution for the decision tree classifier follows a third, also distinctive pattern. Here, we attribute the peaks in the histogram to the rule-based structure of decision trees. Close class ratios in adjacent leaves – which correspond to similar classifier scores – also result in several leaves being depicted as one histogram bin. Distances between leaves on the classifier score scale appear to be somewhat arbitrary. For the derivation of actuarial tariff classes from these classifier scores, the logistic regression output appears the most suitable. It distributes cases evenly, whereas the neural network classifier distinguishes two case groupings, albeit with only minor performance improvement as discussed above. The decision tree output already takes the form of a categorical variable due to the piecewise constant classifier scores. However, the criteria according to which these categories are formed differ from tree classifier to tree classifier, and it is difficult to justify the abrupt changes in class assignment associated with a slight change in only one predictor variable.

Table 4 contains rank correlations between the classifier scores depicted in Fig. 5. Not surprisingly, we found the lowest accordance between models trained on the TDRV versus the MLN predictor set. Complete agreement (i.e., an identical ranking of cases) exists only between the logistic regression and the neural network classifier for the MLN set. The decision tree output appears to deviate more from the other two models than differ from each other. Although distributions for the TDRV + MLN predictor set are very different between logistic regression and neural network, they agree on case ranking with a correlation of .876, which suggests that they differ more in the distance between cases than in their actual order. As may be expected intuitively, we always observe high correlations between models built on the same set of attributes (see the shaded table cells along the middle diagonal).

6. Conclusion

This paper presented a classification analysis approach to actuarial decision-making with vehicle sensor data in Pay-As-You-Drive (PAYD) insurance. We used logistic regression, neural network, and decision tree techniques to predict individual accident risk from location data obtained from a large real-world sample of vehicles. Our first objective was to provide a decision tool for the selection of predictor variables derived from unstructured location data. A comparison of several variable subsets revealed that mileage exposure and four groups of hypothesized situational factors all improve classification performance of the three classifiers. Results of a comparative performance evaluation were ambiguous and depended on the used metric. For the full variable set, neural network exhibited superior performance in terms of accuracy and F-measure, while logistic regression achieved the highest AUROC value. Vehicle mileage was the strongest single predictor variable; its predictive power was further improved, particularly for logistic regression, by applying a logarithmic transformation. To meet our second objective, we examined the effect of class skew and changes in class distribution. While we observed constant AUROC levels even for excessive class skew, all three models were susceptible to bias to some extent. However, its mitigation appears the most feasible for logistic regression. Lastly and most importantly, we investigated the use of classifier scores as a new scalar rate factor that combines various predictor variables into a one-dimensional rate factor. Our analysis revealed substantial differences in the distribution of classifier scores between classifiers and in the relative ranking of cases by classifiers in spite of almost equal predictive performance. While decision tree output appears more suitable as a categorical variable, the allocation of cases to these categories is somewhat arbitrary considering that minor changes in

Table 3 Logistic regression weights β_i and intercept for varying class ratios.

Predictor variable	Class ratio (accident-free:accident-involved vehicles)						
	292:583 (1:2)	583:583 (1:1)	984:583 (1.688:1)	984:492 (2:1)	984:197 (5:1)	984:98 (10:1)	984:49 (20:1)
T1	-0.431	-0.416	-0.484	-0.710	-0.848	-0.847	-0.927
T2	0.115	0.163	0.060	-0.255	-0.455	-0.382	-0.667
T3	0.324	0.344	0.233	0.071	0.078	0.183	-0.068
T4	0.428	0.499	0.488	0.324	0.191	0.181	0.119
D1	0.125	0.147	0.149	0.172	0.307	0.376	0.218
D2	-0.126	-0.148	-0.149	-0.172	-0.307	-0.375	-0.217
R1	1.071	0.979	1.081	1.143	0.837	0.418	0.390
R2	0.508	0.420	0.602	0.685	0.253	-0.303	-0.383
R3	0.530	0.481	0.614	0.670	0.282	-0.120	0.026
V1	-0.071	-0.096	-0.237	-0.327	-0.144	0.120	0.220
V2	0.050	0.094	-0.023	-0.124	-0.066	-0.462	-0.487
V3	-1.070	-1.166	-1.339	-1.517	-1.320	-1.428	-1.372
V4	0.029	-0.038	-0.184	-0.295	-0.226	-0.254	-0.238
V5	-0.172	-0.195	-0.367	-0.495	-0.225	-0.323	-0.235
MLN	2.318	2.683	2.750	2.754	2.723	2.638	2.808
Intercept (const.)	1.04	-0.14	-1.055	-1.31	-2.749	-3.664	-4.454

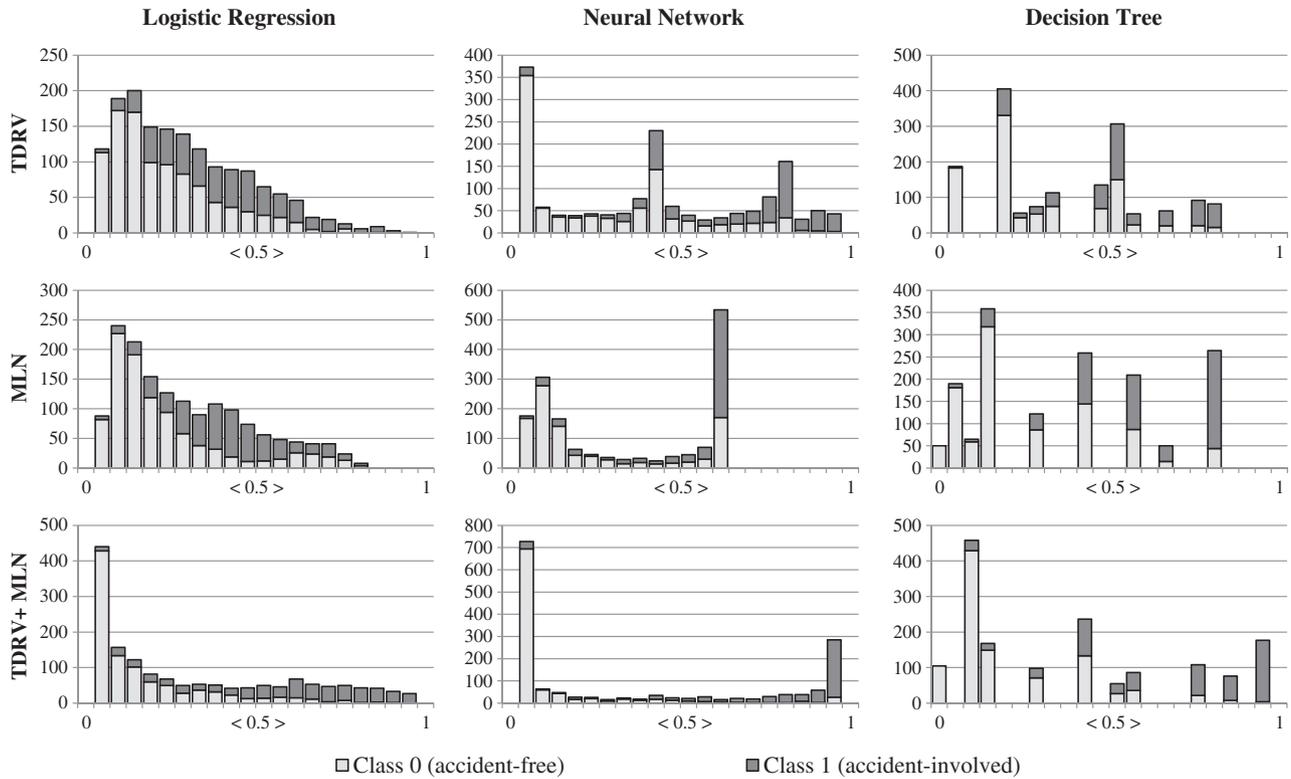


Fig. 5. Histograms of predicted classifier score distributions for different classifiers and variable sets. Bar coloring indicates true (known) class membership, vertical axes denote case counts in histogram bins.

data or training parameters result in different decision rule patterns. Logistic regression may be the most suitable model due to relatively uniformly distributed classifier scores and a high degree of interpretability by insurance professionals.

In the interpretations of our results, some methodological and conceptual limitations should be taken into account. These also point to opportunities for future research. Although we selected the three classifiers in our analysis with care and due to their specific characteristics, other models should be applied to complete the picture. This would involve the combination of different classifiers to further improve predictive performance [58]. Our analyses should also be replicated with datasets of different composition and provenance in order to validate our findings. Where available, other variables should extend the predictor set and yield further improvements to predictive performance. In a more general view, our method is principally adaptable to other insurance business sectors where PAYD insurance tariffs can be

conceived as well as to non-insurance applications that make use of vehicle sensor data, for example, for the pricing of car sharing or in fleet management.

From the perspective of insurance professionals, our analysis has confirmed the very promising potential of vehicle sensor data in predicting policyholders' accident risk. Even if current expectations on the part of practitioners are elevated, both actuaries and insurance IT executives are well advised to carefully evaluate the impact that the informational value of vehicle sensor data may exert on their business. As our results have shown, classification analysis is a valuable tool for deciding on specific sets of predictor variables. The degree of relevance of individual predictors with regard to accident risk may serve as a guideline for the inclusion or exclusion of variables considered too costly to acquire or too intrusive on privacy. Consumers have expressed strong privacy concerns regarding their whereabouts, for instance, such that the use of location data may inhibit the adoption of PAYD insurance. Moreover, the method applied in the present paper is also applicable to further variables derived from different forms of sensor data.

If successful, PAYD insurance is likely to have a considerable impact on the interaction of insurers with hundreds of millions of customers. It may also motivate the collection and processing of such data for other purposes in both business and public administration. Not least, the strong interdependencies between mileage, accident risk, and carbon emissions may further contribute to the popularity of “green” PAYD insurance tariffs if governments decide to stimulate their adoption via regulatory incentives. We therefore encourage fellow researchers in the decision support domain to investigate the mechanics of PAYD insurance and its impact on the insurance sector, where it may very well prove to be a major driver of change in the near future.

Table 4 Rank correlation (Spearman's rho) between classifier scores.

Variable subset classifier model	TDRV			MLN			TDRV + MLN			
	LR	NN	DT	LR	NN	DT	LR	NN	DT	
TDRV	LR	1.000	.802	.690	.392	.392	.451	.677	.645	.561
	NN		1.000	.685	.395	.395	.446	.621	.716	.571
	DT			1.000	.356	.356	.415	.539	.573	.515
MLN	LR				1.000	.864	.772	.610	.745	
	NN					1.000	.864	.772	.610	.745
	DT						1.000	.765	.668	.804
TDRV + MLN	LR							1.000	.876	.838
	NN								1.000	.789
	DT									1.000

References

[1] E. Aarts, R. Harwig, M. Schuurmans, Ambient intelligence, in: P. Denning (Ed.), *The Invisible Future – the Seamless Integration of Technology in Everyday Life*, McGraw-Hill, New York, NY, 2001, pp. 138–144.

- [2] G. Allmendinger, R. Lombreglia, Four strategies for the age of smart services, *Harvard Business Review* 83 (10) (2005) 131–145.
- [3] C. Apte, E. Grossman, E.P.D. Pednault, B.K. Rosen, F.A. Tipu, B. White, Probabilistic estimation-based data mining for discovering insurance risks, *IEEE Intelligent Systems* 14 (6) (1999) 49–58.
- [4] R.A. Bailey, L.J. Simon, Two studies in automobile insurance ratemaking, *Astin Bulletin* 1 (4) (1960) 192–217.
- [5] J. Beirlant, V. Derveaux, A.M. Meyer, M.J. Goovaerts, E. Labie, E. Labie, Statistical risk evaluation applied to car insurance, *Insurance: Mathematics and Economics* 10 (1991) 289–302.
- [6] S. Bhattacharyya, S. Jha, K. Tharakunnel, J.C. Westland, Data mining for credit card fraud: a comparative study, *Decision Support Systems* 50 (3) (2011) 602–613.
- [7] J.W. Bolderdijk, J. Knockaert, E.M. Steg, E.T. Verhoef, Effects of pay-as-you-drive vehicle insurance on young drivers' speed choice: results of a Dutch field experiment, *Accident Analysis and Prevention* 43 (3) (2011) 1181–1186.
- [8] J. Busquets, J. Rodon, J. Wareham, Adaptability in smart business networks: an exploratory case in the insurance industry, *Decision Support Systems* 47 (4) (2009) 287–296.
- [9] W.K. Chiang, D. Zhang, L. Zhou, Predicting and explaining patronage behavior toward web and traditional stores using neural networks: a comparative analysis with logistic regression, *Decision Support Systems* 41 (2) (2006) 514–531.
- [10] P.-A. Chiappori, B. Jullien, B. Salanié, F. Salanié, Asymmetric information in insurance: general testable implications, *The RAND Journal of Economics* 37 (4) (2006) 783–798.
- [11] M.L. Chipman, C.G. MacGregor, A.M. Smiley, M. Lee-Gosselin, The role of exposure in comparisons of crash risk among different drivers and driving environments, *Accident Analysis and Prevention* 25 (2) (1993) 207–211.
- [12] R. Christensen, *Log-linear models and logistic regression*, Springer, New York, 1997.
- [13] D.D. Clark, C. Partridge, R.T. Braden, et al., Making the world (of communications) a different place, *ACM SIGCOMM Computer Communications Review* 35 (3) (2005) 91–96.
- [14] E. Demidenko, Sample size determination for logistic regression revisited, *Statistics in Medicine* 26 (2007) 3385–3397.
- [15] J. Demsar, Statistical comparisons of classifiers over multiple data sets, *Journal of Machine Learning Research* 7 (2006) 1–30.
- [16] M. Denuit, M. Xavier, S. Pitrebois, J.F. Walhin, *Actuarial modelling of claim counts*, 1st ed. John Wiley & Sons Ltd., Chichester, England, 2007.
- [17] P. Desyllas, M. Sako, Profiting from business model innovation: evidence from Pay-As-You-Drive auto insurance, *Research Policy* (2012).
- [18] T.A. Dingus, S.G. Klauer, V.L. Neale, et al., The 100-car naturalistic driving study (phase II), *National Highway Traffic Safety Administration*, 2006.
- [19] D. Dreiseitl, L. Ohno-Machado, Logistic regression and artificial neural network classification models: a methodology review, *Journal of Biomedical Informatics* 35 (5–6) (2002) 352–359.
- [20] T. Fawcett, An introduction to ROC analysis, *Pattern Recognition Letters* 27 (8) (2006) 861–874.
- [21] G. Ferguson, Have your objects call my objects, *Harvard Business Review* 80 (6) (2002) 138–144.
- [22] E. Fieller, H. Hartley, E. Pearson, Tests for rank correlation coefficients, *Biometrika* 44 (3) (1957) 470–481.
- [23] P.A. Flach, J. Hernández-Orallo, C.F. Ramirez, A coherent interpretation of AUC as a measure of aggregated classification performance, *ICML* (2011) 657–664.
- [24] E.W. Frees, Hierarchical Insurance Claims Modeling, *Journal of the American Statistical Association* 103 (484) (2008) 1457–1469.
- [25] E.W. Frees, *Regression Modeling with Actuarial and Financial Applications*, Cambridge University Press, New York, USA, 2010.
- [26] L.A. Gardner, D.C. Marlett, The state of personal auto insurance rate regulation, *Journal of Insurance Regulation* 26 (2) (2008) 39–69.
- [27] N. Gershenfeld, *When things start to think*, Henry Holt, New York, NY, 1999.
- [28] D.J. Hand, Measuring classifier performance: a coherent alternative to the area under the ROC curve, *Machine Learning* 77 (1) (2009) 103–123.
- [29] L. Heinrich, *RFID and Beyond: Growing Your Business Through Real World Awareness*, John Wiley & Sons Ltd., Hoboken, NJ, 2005.
- [30] N. Johnson, F. Garwood, The analysis of the claim records of a motor insurance company, *Journal of the Institute of Actuaries* 83 (1957) 277–294.
- [31] J. Jun, J. Ogle, R. Guensler, Relationships between crash involvement and temporal-spatial driving behavior activity patterns: use of data for vehicles with global positioning systems, *Transportation Research Record* 2019, -1, 2007, pp. 246–255.
- [32] J. Jun, R. Guensler, J. Ogle, Differences in observed speed patterns between crash-involved and crash-not-involved drivers: application of in-vehicle monitoring technology, *Transportation Research Part C: Emerging Technologies* 19 (4) (2010) 569–578.
- [33] M.Y. Kiang, A comparative assessment of classification methods, *Decision Support Systems* 35 (4) (2003) 441–454.
- [34] Y. Kim, Toward a successful CRM: variable selection, sampling, and ensemble, *Decision Support Systems* 41 (2) (2006) 542–553.
- [35] G. King, L. Zeng, Logistic regression in rare events data, *Political Analysis* 9 (2) (2001) 137–163.
- [36] T. Litman, Distance-based vehicle insurance as a TDM strategy, *Transportation Quarterly* 51 (3) (1997) 119–138.
- [37] D. Lord, F. Mannering, The statistical analysis of crash-frequency data: a review and assessment of methodological alternatives, *Transportation Research Part A: Policy and Practice* 44 (5) (2010) 291–305.
- [38] D. Martens, J. Vanthienen, W. Verbeke, B. Baesens, Performance of classification models from a user perspective, *Decision Support Systems* 51 (4) (2011) 782–793.
- [39] R. Mayou, B. Bryant, R. Duthie, Psychiatric consequences of road traffic accidents, *British Medical Journal* 307 (6905) (1993) 647.
- [40] J. Paefgen, F. Michahelles, T. Staake, GPS trajectory feature extraction for driver risk profiling, international workshop on Trajectory data mining and analysis, 13th Conference on Ubiquitous Computing, Beijing, China, 2011.
- [41] X. Qin, J.N. Ivan, N. Ravishanker, J. Liu, D. Tepas, Bayesian estimation of hourly exposure functions by crash type and time of day, *Accident Analysis and Prevention* 38 (6) (2006) 1071–1080.
- [42] B. Rai, N. Singh, Customer-rush near warranty expiration limit, and nonparametric hazard rate estimation from known mileage accumulation rates, *IEEE Transactions on Reliability* 55 (3) (2006) 480–489.
- [43] S. Rogers, P. Langley, C. Wilson, Mining GPS data to augment road models, *Proceedings of the fifth ACM SIGKDD international conference on knowledge discovery and data mining*, 1999, pp. 104–113.
- [44] D.W. Ruck, S.K. Rogers, M. Kabrinsky, Feature selection using a multilayer perceptron, *Journal of Neural Network Computing* 2 (2) (1990) 40–48.
- [45] D. Samson, H. Thomas, Linear models as aids in insurance decision making: the estimation of automobile insurance claims, *Journal of Business Research* 15 (3) (1987) 247–256.
- [46] L. Santos, J. Coutinho-Rodrigues, C.H. Antunes, A web spatial decision support system for vehicle routing using Google Maps, *Decision Support Systems* 51 (1) (2011) 1–9.
- [47] M. Satyanarayanan, Pervasive computing: vision and challenges, *IEEE Personal Communications* 8 (4) (2001) 10–17.
- [48] J. Shreve, H. Schneider, O. Soysal, A methodology for comparing classification methods through the assessment of model stability and validity in variable selection, *Decision Support Systems* 52 (1) (2011) 247–257.
- [49] A.P. Sinha, H. Zhao, Incorporating domain knowledge into data mining classifiers: an application in indirect lending, *Decision Support Systems* 46 (1) (2008) 287–299.
- [50] Y. Suzuki, A decision support system of dynamic vehicle refueling, *Decision Support Systems* 46 (2) (2009) 522–531.
- [51] B. Swiderski, J. Kurek, S. Osowski, Multistage classification by using logistic regression and neural networks for assessment of financial condition of company, *Decision Support Systems* 52 (2) (2012) 539–547.
- [52] F. Temming, Judgment of the European Court of Justice (Grand Chamber) of 1 March 2010: ECJ finally paves the way for unisex premiums and benefits in insurance and related financial service contracts, *German Law Journal* 13 (1) (2011) 105–122.
- [53] T. Toledo, O. Musicant, T. Lotan, In-vehicle data recorders for monitoring and feedback on drivers' behavior, *Transportation Research Part C: Emerging Technologies* 16 (3) (2008) 320–331.
- [54] M. Weiser, The computer for the twenty-first century, *Scientific American* 265 (3) (1991) 94–104.
- [55] S.B. White, On the use of annual vehicle miles of travel estimates from vehicle owners, *Accident Analysis and Prevention* 8 (4) (1976) 257–261.
- [56] R. Wilson, R. Sharda, Bankruptcy prediction using neural networks, *Decision Support Systems* 11 (1994) 545–557.
- [57] A.C. Wolfe, The concept of exposure to the risk of a road traffic accident and an overview of exposure data collection methods, *Accident Analysis and Prevention* 14 (5) (1982) 337–340.
- [58] J.-Y. Yeh, T.-H. Wu, C.-W. Tsao, Using data mining techniques to predict hospitalization of hemodialysis patients, *Decision Support Systems* 50 (2) (2011) 439–448.
- [59] K. Yip, K. Yau, On modeling claim frequency data in general insurance with extra zeros, *Insurance: Mathematics and Economics* 36 (2) (2005) 153–163.
- [60] B. Zadrozny, Learning and evaluating classifiers under sample selection bias, *Proceedings of the 21st International Conference on Machine Learning*, Banff, Canada, 2004.

Johannes Paefgen is a doctoral student at the University of St. Gallen, Switzerland with the Chair of Operations Management. He holds Bachelor's and Master's degrees in engineering from the Swiss Federal Institute of Technology (ETH) Zurich, where he graduated in 2009. His research interests are in data mining and statistical analysis, with a focus on vehicular sensor data from which travel patterns and driver profiles can be obtained. He investigates the value of such information in an insurance business context. He is associated with the Insurance Lab at ETH Zurich and the University of St. Gallen. Currently, Johannes is a visiting researcher at the Massachusetts Institute of Technology with the group of Prof. Sanjay Sarma.

Thorsten Staake is lecturer at ETH Zurich and director of the Bits to Energy Lab, a joint research initiative of ETH and the University of St. Gallen. His work is dedicated to bringing together IS research with insights from behavioral economics in order to build systems that motivate a sustainable user behavior. Prior to joining ETH, Thorsten worked at the Auto-ID Labs at the Massachusetts Institute of Technology, Infineon Technologies, and Clariant. He holds a PhD in Business Administration from the University of St. Gallen and a diploma in Electrical Engineering and Information Technology from TU Darmstadt.

Frdéric Thiesse is Professor of Information Systems Engineering at the University of Wuerzburg. He holds a diploma degree in Management Information Systems from the University of Mannheim, a doctoral degree and a 'habilitation' (postdoctoral qualification) in Business Administration from the University of St. Gallen. His research interests include the application of RFID and sensor technologies in the supply chain, mobile commerce, enterprise integration, business analytics, and green IT. His articles have been published, among others, in *MIS Quarterly*, the *European Journal of Information Systems*, the *Journal of Strategic Information Systems*, the *Journal of Business Research*, and *Decision Support Systems*. He has worked as a consultant for several firms, including SAP, METRO Group, Infineon Technologies, Deutsche Telekom, and others.