Analysing Harsh Events with Spatiotemporal Machine Learning Techniques Using Mobile Data

Armira Kontaxi[1], Haris Sideris [2], Dimitris Oikonomopoulos [2] and George Yannis [1]

[1] National Technical University of Athens, Athens, Greece

[2]OSeven Single Member Private Company, Athens, Greece

Introduction

Driver behaviour monitoring has gained prominence in transportation research, but collecting accurate, real-time data at scale remains difficult and costly. The ubiquity of smartphones offers a practical alternative: app-based sensing enables low-cost data collection, trip analytics, and real-time feedback that can reduce crashes and injuries.

Smartphones, beyond passively collecting driving data, **enable driver feedback and gamified incentives** that can strengthen safety culture. Evidence on feedback/gamification is largely positive, with effectiveness contingent on feedback modality, timing, and incentive design.

Methodologically, studies employ statistical, machine learning, and deep learning approaches tailored to the outcome of interest. **Harsh events are commonly used as surrogate safety measures**, with prior work applying generalized mixed-effects models and spatial ML to relate harsh events to driving-behavior indicators and exposure. However, integrated spatiotemporal modeling that explicitly examines feedback/ gamification effects remains limited.

Objectives

Addressing this gap, the present study aims to:

- **Use large-scale naturalistic driving data** from smartphone sensors
- Evaluate the **impact of feedback/gamification** on harsh events
- **Employ XGBoost models** to assess whether, and to what extent, gamification delivered via a smartphone application reduces the incidence of harsh events during driving

Experiment design

To achieve the research objective, an innovative smartphone application developed by OSeven (www.oseven.io) for the purpose of the "O7Insurance" (implemented under the National Recovery and Resilience Plan Greece 2.0) research project was exploited **aiming to record, analyse and improve driver behavior.**

The **naturalistic driving experiment spanned five months** and comprised a baseline and a 30-day competition. Two modes of feedback were used:

- **Phase A (baseline)** provided personalized, non-competitive feedback via the DrivingStar app, trip lists, scorecards (0–100), maps, and highlights to pinpoint unsafe behaviours (speeding, phone use, harsh braking, harsh acceleration), delivered after each trip.

- **Phase B (competition)** introduced social gamification and incentives for safe driving over 30 days; competition points equalled the sum across trips of distance travelled × a driving-behaviour factor, re-warding safer performance.

Smartphone application

In order to achieve the research objective, an innovative smartphone application developed by **OSeven (www.oseven.io)** for the purpose of the "O7Insurance" (implemented under the National Recovery and Resilience Plan Greece 2.0) research project was exploited aiming to record, analyse and improve driver behavior. "O7Insurance" introduces a **novel approach to vehicle insurance management** by enabling drivers to handle all aspects of their coverage through a mobile application, supported by OSeven's telematics technology.

The app utilizes the **smartphone's sensors and APIs to capture detailed trip data**, which is stored locally and then uploaded to a secure cloud database for processing. Using advanced signal processing, machine learning algorithms, data fusion, and big data techniques. Figure 1 depicts and summarizes the OSeven data flow system.

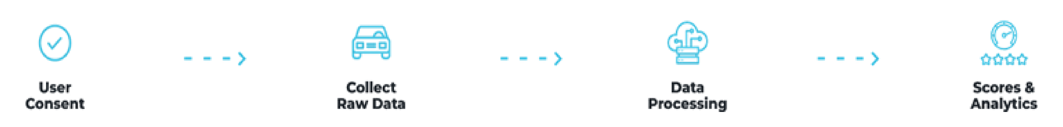

Figure 1. OSeven data flow

These indicators cover **both exposure variables** (e.g., trip distance, driving duration, road type, rush hour conditions) **and behavioral variables** (e.g., speeding episodes, severity of harsh braking or acceleration, and mobile phone distraction). Figure 2 illustrates examples of the app's feedback features across the different experimental phases.

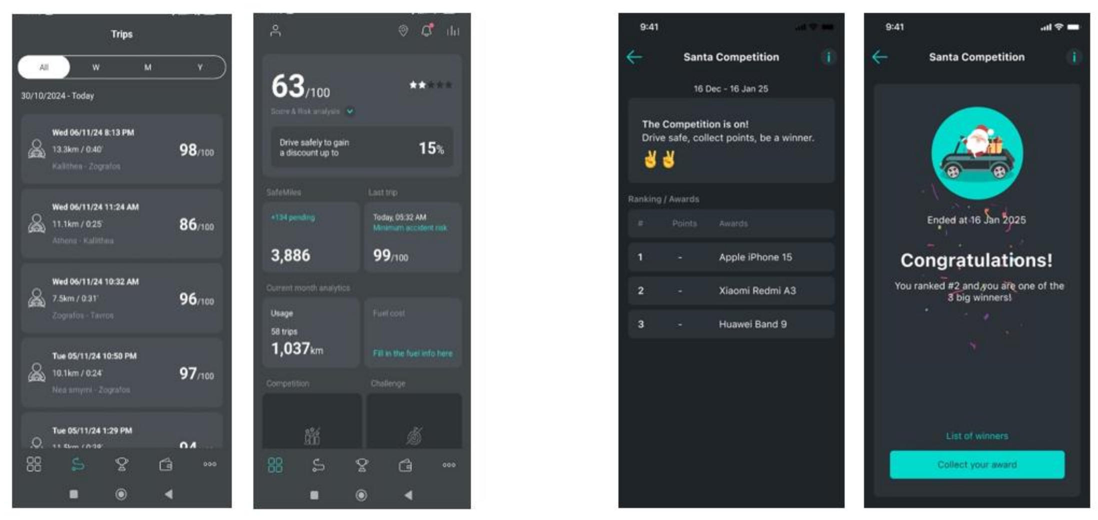

Figure 2: Example screenshots from the application features in Phase A – Baseline (left) and Phase B – Competition (right)

Theoretical background

Extreme Gradient Boosting (XGBoost) classifier is a scalable, regularized gradient-boosted tree method that builds an additive ensemble of CART trees to minimize a penalized objective, delivering speed and accuracy.

Let f_k denote the k-th regression tree; the model is:

$$\hat{y} = \varphi(x_i) = \sum_{k=1}^{K} f(x_i) \quad (1)$$

with a differentiable convex training loss $l(\hat{y}_i, y_i)$ (e.g., mean squared error),

$$l(\varphi_i) = \sum_{i=1}^{l} (\hat{y}_i - y_i)^2 \quad (2)$$

and a complexity penalty per tree,

$$\Omega(f) = \gamma T + \frac{1}{2}\lambda\|c\|^2 \quad (3)$$

where T is the number of leaves and ccc are leaf weights. The overall objective combines fit and regularization:

$$L(\varphi_i) = \sum_{i=1}^{l} l(\hat{y}_i, y_i) + \sum_{k=1}^{K} \Omega(f) \quad (4)$$

Regularization (e.g., λ, γ) helps control overfitting, and tree structures are relatively robust to multicollinearity. **Feature importance** is summarized by Gain (loss reduction), Cover (proportion of samples affected), and Frequency (usage count). **Key hyperparameters**—learning rate (eta), max depth, gamma, min_child_weight, subsample/ colsample, and ℓ1/ℓ2 regularization (α, λ), are typically tuned via cross-validation (often with early stopping).

Spatial distribution of harsh events

As **a preliminary spatial exploration of safety-critical behaviour** indicators, using per-second GPS telemetry, records with harsh-event intensity > 0 were mapped on an OpenStreetMap basemap in Leaflet. Clustered point markers, coloured by intensity (Low/Medium/High), were overlaid with a severity-weighted kernel heatmap (weights 1/2/3) to highlight co-location of frequency and severity.

The visualization (Figure 3) indicates pronounced **hotspots in the urban core and along major arterial corridors**, with secondary clusters at junctions and approach roads; peripheral areas exhibit sparser, lower-intensity activity.

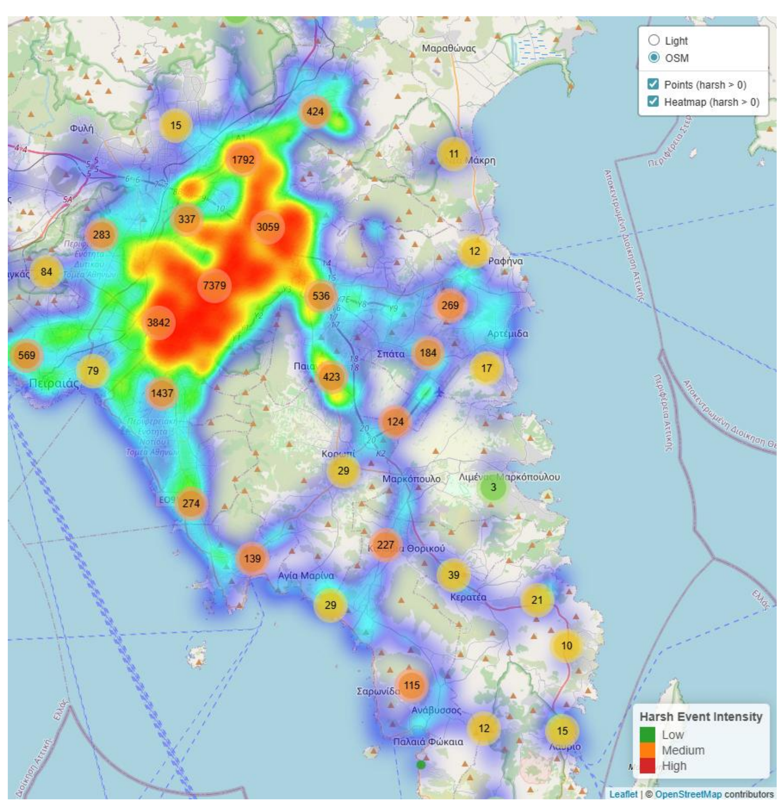

Figure 3: Severity-weighted heatmap and clustered counts of harsh events (intensity > 0) on OSM

Descriptive Statistics

Prior to model fitting, **descriptive rates of harsh events per 100 km** were examined across experiment phases (Table 1). This monotonic downward trend across phases indicates that the competition period was associated with fewer harsh maneuvers, with the most pronounced improvement evident after the intervention. A possible explanation for this finding is that drivers who already improved their behavior during the competition, continued to aim for safe driving.

Table 1. Descriptive rates of harsh events per 100 km across experiment phases

Experiment phase	Average of harsh_acc_per100km	Average of harsh_brk_per100km	Count of trips
pre	14.265	15.815	5627
during	12.321	13.671	6528
after	9.939	9.169	4404

Competition phase and harsh events via XGBoost

XGBoost classifiers were trained on data split 80/20 into training (n = 13,247 trips) and test (n = 3,312) using stratified sampling. Because the binary harsh-event label was imbalanced, the training set was upsampled (random resampling with replacement via caret) to a 1:1 class ratio; the test set was retained unchanged. Features included speed behavior, time of day, speeding, experiment phase, and mobile usage. At the chosen operating point (threshold 0.80), the model achieves high recall (0.82) with moderate precision (0.56), indicating most true harsh events are captured while more than half of flagged cases are correct. **The ROC–AUC of 0.87** suggests strong ranking ability across thresholds. However, the relatively low balanced accuracy (0.52) suggests some limitations in distinguishing between classes under imbalance conditions.

Table 2. Feature importance (XGBoost)

Rank	Feature	Gain	Cover	Frequency
1	Total_distance	0.5254	0.6537	0.8265
2	Hour_of_day	0.0528	0.0551	0.1034
3	Speeding	0.0428	0.0384	0.0758
4	Experiment_phase	0.0208	0.0190	0.0542
5	Mobile_usage	0.0060	0.0132	0.0091

Feature importance (Table 2) is dominated by trip distance, with additional signal from hour of day and speeding; experiment phase contributes modestly, and mobile usage minimally. More precisely, **total trip distance is the most informative variable** which contribute to the prediction of harsh events during a trip. As for **experiment phase**, this variable contributes a **smaller but measurable effect**: during a competition/intervention phase, behaviour may shift (e.g., temporary caution or attention), while post-phase periods can show rebound effects; the net direction depends on how incentives interacted with routes and driver habits.

Conclusions

The findings suggest that **smartphone-based feedback and gamification**, such as those provided by the telematics app of OSeven, **promote safer driving**, while they can also be operationalized within a risk-modelling workflow to prioritize locations, times, and drivers for targeted interventions. Moreover, the combination of smartphone telematics and explainable ML can support near-real-time safety management by flagging high-risk corridors and hours, guiding targeted enforcement and feedback.

Future research could incorporate **richer contextual variables** such as road classification, congestion, and weather. Furthermore, future modelling efforts should also consider probability calibration, precision–recall–AUC as a complement to ROC metrics, and advanced spatiotemporal approaches.

Acknowledgements

This paper was carried out within the framework of OSeven's project "O7Insurance", which is implemented under the National Recovery and Resilience Plan Greece 2.0, funded by the European Union – NextGenerationEU, under the call "Innovation funding – Horizon 2020 (Implementation body: MIA RI) .

Greece 2.0 NATIONAL RECOVERY AND RESILIENCE PLAN

Funded by the European Union NextGenerationEU

Contact Information

Armira Kontaxi

Address

Tel: +30.210.7721575

Email: akontaxi@mail.ntua.gr

Web: https://www.nrso.ntua.gr/p/akontaxi/