Best PhD Dissertation Award – Finalist Presentation

Econometric methods and machine learning algorithms to investigate factors contributing to pedestrian crash severity

Maria Rella Riccardi
P.Eng., Ph.D. in Civil System Engineering
University of Naples Federico II
Department of Civil, Architectural and Environmental Engineering
maria.rellariccardi@unina.it

DECADE OF ACTION FOR ROAD SAFETY 2021 - 2030

Problem statement

The UN General Assembly proclaimed the "**2nd Decade of Action for Road Safety**"

Among the 2030 Agenda for Sustainable Development goals:

- **Goal 3.6 (health)**
  → **Halve** the number of global **deaths and serious injuries** from road traffic crashes
- **Goal 11.2 (cities)**
  → Improving road safety with special **attention to vulnerable road users**

The risk of fatality for pedestrian users is 9 times greater than that for occupants of 4-wheeler vehicles

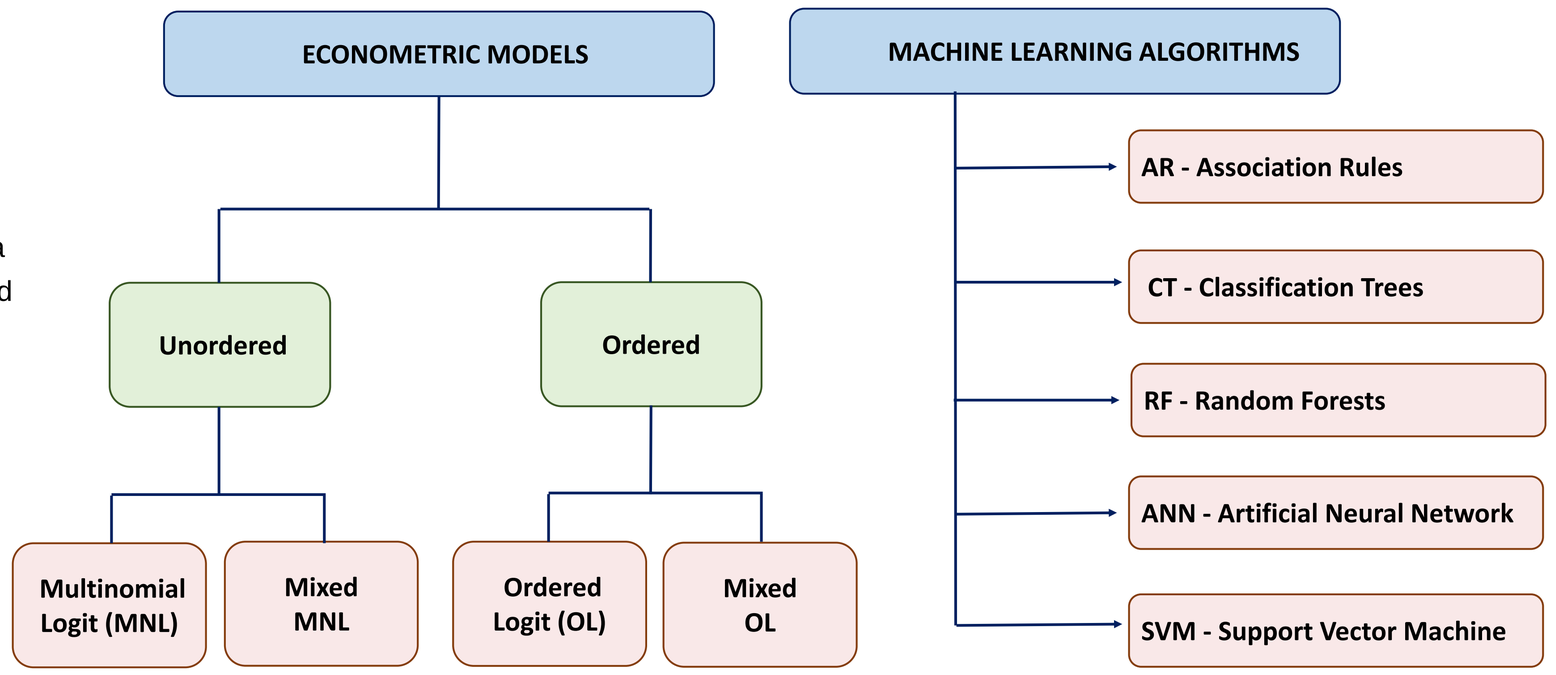
Research purposes

- To **provide support** for the choice of the appropriate prediction method
- To **compare econometric models and machine learning tools**:
  - by their capability in identifying significant explanatory variables affecting crash severity (Qualitative evaluation)
  - by their model performances (Quantitative evaluation)
- To **investigate factors associated with the most severe pedestrian crashes** considering the coexistence of pedestrian, driver, vehicle, roadway, crash, and environmental factors
- To contribute to meet the long-term goal to move close to zero deaths on roads by 2050

Methodological approach

- Three national databases were used:
  - Great Britain, 2016 – 2018
  - Sweden, 2015 – 2019
  - Italy, 2014 – 2018
- Crash severity was the dependent variable
- A weighted approach was adopted to deal with the imbalanced data
- 4 econometric models and 5 machine learning tools were developed

	Great Britain		Sweden		Italy	
Crash severity	**N**	**%**	**N**	**%**	**N**	**%**
Slight Injury	49,631	73.7	8,788	93.2	98,063	97.1
Serious Injury	16,359	24.3	426	4.5		
Fatal	1,366	2.0	212	2.3	2,969	2.9
Total	67,356	100.0	9,426	100.0	101,032	100.0



Main results

The performance metrics used to compare the models, based on the confusion matrix, were F-measure, G-mean, and Area Under the Curve.

Great Britain

	Econometric models				Machine learning				
	MNL	Mixed MNL	OL	Mixed OL	AR	CT	RF	ANN	SVM
Fatal									
F-measure	0.28	0.53	0.00	0.16	0.05	0.16	0.57	0.18	0.95
G-mean	0.50	0.65	0.04	0.33	0.36	0.72	0.77	0.66	0.95
AUC	0.87	0.94	0.85	0.85	0.79	0.82	0.88	0.78	0.88
Serious									
F-measure	0.21	0.41	0.41	0.40	0.39	0.29	0.90	0.26	0.95
G-mean	0.36	0.58	0.43	0.58	0.54	0.46	0.92	0.43	0.96
AUC	0.62	0.68	0.61	0.62	0.58	0.47	0.71	0.76	0.76

Sweden

	Econometric models				Machine learning				
	MNL	Mixed MNL	OL	Mixed OL	AR	CT	RF	ANN	SVM
Fatal									
F-measure	0.15	0.14	0.20		0.06	0.19	0.46	0.48	0.79
G-mean	0.64	0.81	0.75	*	0.18	0.61	0.76	0.82	0.86
AUC	0.85	0.86	0.89		0.80	0.71	0.95	0.91	0.86
Serious									
F-measure	0.10	0.23	0.10		0.14	0.14	0.17	0.10	0.73
G-mean	0.47	0.63	0.50	*	0.27	0.62	0.58	0.53	0.85
AUC	0.60	0.49	0.59		0.39	0.63	0.79	0.67	0.54

** The mixed OL did not arrive at convergence*

Italy

	Econometric models		Machine learning				
	Logit	Mixed Logit	AR	CT	RF	ANN	SVM
Fatal							
F-measure	0.15	0.16	0.09	0.13	0.31	0.15	0.65
G-mean	0.76	0.76	0.62	0.73	0.75	0.75	0.78
AUC	0.84	0.84	0.77	0.79	0.79	0.83	0.70

Variables identified	Fatal	Serious
Econometric and machine learning models	19	18
Only econometric models	1	1
Only machine learning	6	7

Factors affecting:

Fatal crashes

Rural area
Speed limit ≥ 50 mph
Darkness
Weekend
Wet or damp pavement
Truck involvement
Vehicle age ≥ 15
Involvement of articulated vehicles
Heavy oil vehicles
Engine capacity (CC) ≥ 3,000
Pedestrian aged over 75
Male pedestrian
Frontal vehicle impact
Male driver
Driver aged under 25
Commuting to/from work

Serious crashes

Darkness
Wet or damp pavement
Number of vehicles ≥ 2
Vehicle skidding and overturning
Pedestrian aged over 75
Driver aged under 25

Variables identified	Fatal	Serious
Econometric and machine learning models	13	12
Only econometric models	0	0
Only machine learning	5	6

Factors affecting:

Fatal crashes

Rural area
Speed limit ≥ 60 km/h
Not at intersection
Darkness
Weekday
Slippery pavement
Truck involvement
Vehicle with trailers
Pedestrians aged over 75
Male pedestrian
Pedestrian alcohol/drug use
Male driver
Driver aged 25-34 years

Serious crashes

Speed limit = 40-50 km/h
Darkness
Weekday
Vehicle with trailers
Pedestrians aged over 75
Pedestrian alcohol/drug use
Male pedestrian
Driver alcohol/drug use
Driver aged over 75
Female driver

Variables identified	Fatal
Econometric and machine learning models	13
Only econometric models	0
Only machine learning	5

Factors affecting:

Fatal crashes

Rural area
Higher speed roads
Tangent
Darkness
Weekend
Wet pavement
Truck involvement
Vehicles with deficiencies
Pedestrian aged over 75
Pedestrian alcohol use
Driver inappropriate behaviour
(disobeying stop sign, distraction, speeding, and tailgating)
Driver disobeying pedestrian crossing facilities
Driving exceeding the prescribed driving period
Driver illness
Driving under drug influence or sleeping
Driver aged under 25

Roadway factor | Environmental factor | Vehicle factor | Road user factor

Bold denotes factors significant in all databases

Conclusions

Methodological considerations

- Econometric models:
  - In small sample sizes the OL performed reasonably well
  - Unordered models outperformed ordered models when the sample size increases
  - The random parameter models provided evidence of the **existence of heterogeneity among data**
- Machine learning methods :
  - **SVM and ANN produced better prediction performances** but difficult interpretability of results
  - **AR, CT, and RF** exhibited lower performances but **produced comprehensible scenarios**
  - Uncovered **more hidden correlations among data** than the econometric models
- The **joint use of econometric methods and machine learning algorithms** may provide more insights on factors contributing to fatal and serious pedestrian crashes

Engineering considerations and safety countermeasures

- Establish a **road hierarchy giving the highest priority to pedestrians** and then to the other road users.
- Plan **specific routes for trucks**, avoiding their transit in places highly frequented by pedestrians
- Introduce **safety education** in school programs and support **safety campaigns** to reduce inappropriate driving
- **Improve lighting** in the proximity of pedestrian crossings
- In areas with relevant pedestrian activities especially with elderly pedestrians, **introduce traffic calming measures**