



Machine Learning Prediction of Injuries Inpatient Length of Stay

Alessio Pitidis¹. Gianni Fondi². Carlo Mamo³. Marco Giustini². IDB-FDS Reference Group⁴

¹ B2C Innovation. Italy

² Italian National Institute of health. Italy

³ Epidemiology Unit, Local Health Unit T03, Italy

⁴ European Association for Injury Prevention and Safety Promotion (EuroSafe-IDB Network). The Netherlands

ML techniques were applied to the Full Data Set (FDS) of the European Injury DataBase (EU-IDB) which provides detailed information on the external causes and diagnoses of injury observed at the Emergency Departments (ED). The IDB-FDS provides more than 3.800.000 ED records for the period 2008-19 in 19 Countries. LASSO (Least Absolute Shrinkage and Selection Operator) cross-validated linear regression technique was used for variable selection and parameter regularization.

Countries that provided IDB-FDS data (2008-2019)



Cross validation was performed randomly assigning the records on 5 folds. On the selected variables a cross-validated linear model was performed on 5 folds which were randomly sampled assigning 80% of records to the training sample and 20% to the testing one.

Predictors of Length of Stay	Relative Explained variability	Fold 4 regression coefficient
EUROCCOST-39 categories:	86.20%	0.920 (95%CI: 0.918-0.922)
AgeGroup	8.57%	0.070 (95%CI: 0.068-0.070)
ActivityWhenInjured*	2.04%	0.219 (95%CI: 0.210-0.228)
MechanismOfInjury*	2.04%	0.287 (95%CI: 0.280-0.295)
TransportInjuryEvent	1.21%	0.971 (95%CI: 0.940-1.001)
PlaceOfOccurrence	0.92%	0.135 (95%CI: 0.126-0.144)
SexOfPatient	0.77%	-0.694 (95%CI: -0.729--0.660)
Intent	0.29%	0.607 (95%CI: 0.590-0.624)

* In combination ActivityWhenInjured & MechanismOfInjury

The ML model explains a significant part of the Length of Stay (LoS) variability and this measure is stable in the different training and testing samples used to cross validate the estimates. The main part of the relative variability is explained by the diagnoses reclassified according to the EUROCCOST disability standardization method. For instance, in the training sample of the median fold (Fold 4) the LoS ranges from 0.45 days for strain of hand/fingers up to 11.65 days for multi-trauma. The respective figures in the testing sample are 0.46 and 12.00 days.