

Introduction.

- The number of deaths on the Irish roads has been reduced year after year since 2005, but the decreasing trend has led to a minimum from 2012 to 2019 that remains approximately invariant (Figure 1).
- This minimum in the fatalities is proving to be hard to overcome and in order to take the right measures, analysing the data available is indispensable.
- In this project, we will quantify the risk that drivers experience while they are on the road based on historical and live data.

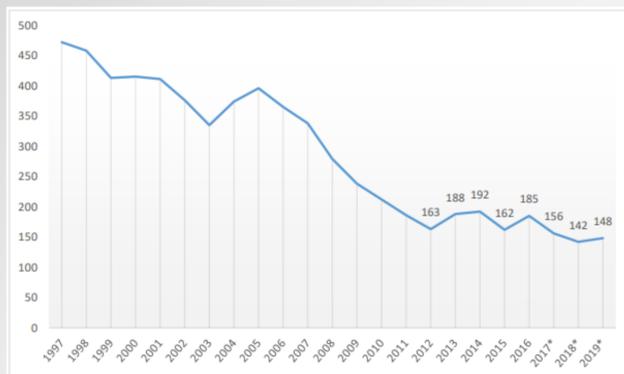


Figure 1. Fatalities on the Irish roads from 1997 to 2019. (Road Safety Authority, 2020)

Historical Data.

- **Road Safety Authority Dataset.**
 - This dataset records the traffic accidents on the Irish roads from 2011 to 2016 and its characteristics. (Figure 8)
 - This dataset contains 27.743 observations.
- **Met Éireann Dataset.** Met Éireann provides open historical weather data. (Irish Weather hourly data, 2020)
 - This dataset contains historical data from 1989 to 2018.
 - This dataset contains data from 25 stations across 15 counties in Ireland. (Figure 9)
- **Transport Infrastructure Ireland Dataset.**
 - TII offers traffic volume data on different roads along the country from 2013 until today. (Transport Infrastructure Ireland, 2020) (Figure 10)

Live Data.

- **Met Éireann.** Hourly data of different stations can be downloaded from the website.
- **Transport Infrastructure Ireland Dataset.** TII offers daily traffic volume data on the website.

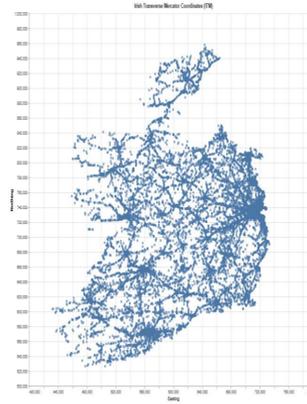


Figure 8. RSA Dataset Observations.



Figure 9. Met Éireann Stations.



Figure 10. TTI radar stations.

Methodology.

- There are two methodologies that have been used to face this problematic:
 - 1) **Regression.** The researcher tries to predict the frequency of accidents in a certain area in a definite time interval. This method is used to find correlations between accidents and different attributes.
 - 2) **Binary Classification.** The researcher aims to predict if there will be an accident or not in a certain area in a definite time interval.
- Our goal is to do binary classification. Given certain conditions, we aim to predict whether there will be an accident or not. This approach generally works better than regression.
- **Negative Sampling.** To build a classification model we need to create non-accident observations, negative samples. To do so we will follow the process carried out by (Yuan, Zhou, Yang and Tamerius, 2020) where they change one property (coordinates, hour or day) and leave the rest the same, if this new observation is not in the positive pool we will add it.
- **Normalisation.** Continuous variables will be normalised and categorical variables will be encoded.
- **Live implementation.** We will daily scrape weather and traffic volume data using Python from Met Éireann and TII websites and create an app that will find black spots based on live data.

Conclusion.

- The aim of this project is creating a base for future sophisticated live accident prediction models.
- Our model performance will be compared to previous work.
- In the future, these models will be of vital importance if the current value of accidents per year is to be decreased.

Technologies.



References.

- Kaggle.com, 2020. *Irish Weather (Hourly Data)*. [online] Available at: <https://www.kaggle.com/connorrot/irish-weather-hourly-data> [Accessed 3 April 2020].
- Road Safety Authority, 2020. *RSA.ie - Provisional Reviews*. [online] Available at: <https://www.rsa.ie/RSA/Road-Safety/RSA-Statistics/Collision-Statistics/Provisional-Reviews/> [Accessed 3 April 2020].
- Transport Infrastructure Ireland, 2020. *05 Live*. [online] Available at: <https://www.nratrafficdata.ie/c2/gmapbasic.asp?sgid=ZvyVmXU8jBt9PIE5c7UXt6> [Accessed 3 April 2020].
- Yuan, Z., Zhou, X., Yang, T. and Tamerius, J., 2020. Predicting Traffic Accidents Through Heterogeneous Urban Data : A Case Study. [online] Semanticscholar.org. Available at: <https://www.semanticscholar.org/paper/Predicting-Traffic-Accidents-Through-Heterogeneous-Yuan-Zhou/fdeaa1a8d25518b0424a66d98c04cf7bde4982ba> [Accessed 3 April 2020].

Research Objective.

- The objective of this project is investigating if an accurate classification model can be built to predict accidents on the Irish roads in real-time.
- Testing different models and compare their performance.

Rationale.

- Creating good models for accident prediction has many beneficial aspects for society.
 - 1) Optimising public transportation.
 - 2) Enabling safer routes.
 - 3) Improving infrastructures cost-effectively.
 - 4) Urban planning.
 - 5) Traffic control.
 - 6) Personalized insurance.
 - 7) Faster action by law enforcement and ambulances.
 - 8) Safe route recommendation.

State of the Art.

- Many studies have been carried out in recent years but there is not a general working method and the appearance of new models has improved the results. Usually, these studies lack comprehensive post-evaluation because of its complexity.
- The effectiveness of the models is still low and the new studies seek to outperform the previous ones, even if only a little.
- Algorithms that have been used for this problematic comprise Support Vector Machines, Decision Trees, Random Forests and Deep Neural Networks from lower to higher performance. SVMs have an accuracy from 60 to 70%, on the other hand, DNNs' accuracy is around 90%.
- There are different problems that make predicting accidents a difficult task:
 - 1) **Lack of Data.** Only in the last years, data has been started to be collected in developed countries. Furthermore, it is known that many variables that may influence accidents are very difficult to collect.
 - 2) **Spatial Heterogeneity.** The conditions in which accidents occur vary a lot from place to place.
 - 3) **Class Imbalance.** Data is only collected when accidents occur.
- Data that has been included in these models is historical weather data, road-network properties, traffic volume, road geometry, satellite images and demography.



Figures 2-7. Accident attributes in the period 2011-2016.