**Algorithms for Predicting Traffic Accidents Severity in France**

**Mogondo M. Edwin**

**B.E Chemical Engineer**

**IBM Data Science Program, Coursera E-Learning**

**Houston, TX**

**Email:** [emogondo@gmail.com](mailto:emogondo@gmail.com)

1. **Data source**

**The data used for the purpose of this study was acquired from Kaggle.com open source for data distribution center. The dataset contains information about road crashes for France from 2005 through 2016. The data contains 839,985 of recorded crashes.**

**Accident severity data was categorical as follows: 1: unscathed, 2: Killed 3: Serious injury and 4: Light injury as shown in Figure 2.**

![Chart

Description automatically generated]()

**Figure 2: Distribution of Severity**

**Over 50,000 people lost their lives during the over 10 years period accounting to about 2.70 % of mortality from crashes in France.**

1. **Data Pre-processing**

**Preprocessing of data was performed before model training and testing. These processes included cleaning (data wrangling), normalization, feature selection(extraction), and transformation. The dataset used integer values for entire attributes. Transformation was done on categorical data to contain 0s and 1s. Missing values were accounted for by mean (average) and frequently occurring integers. Factors that affected severity of accidents were chosen based on importance as follows in Table 1.**

![Table

Description automatically generated]()

**Table 1. Feature Selection Table**