**Accident Severity Prediction**

Shahid Shakil

September 17, 2020

**1. Introduction**

**1.1 Background**

In recent years, road traffic accidents have become one of the largest national health issues in the world and it is leading cause for deaths. The burden of road accident casualties and damage

is much higher in developing countries than in developed countries. Many factors (driver, environment,

vehicle, etc.) are related to traffic accidents, some of those factors are more important in determining the

accident severity than others. The analytical data mining solutions can significantly be employed to determine

and predict such influential factors among human, vehicle and environmental factors. In this research, the

classification technique i.e., Logistic Regression algorithm is used to identify relevant patterns and for classifying the

type of accident severity of various traffic accidents with the help of influential environmental features of accidents

that can be used to build the prediction model. This technique was tested using a real dataset. A decision system

has been built using the model generated by Logistic Regression that will help decision makers to

enhance the decision making process by predicting the severity of the accident.

**1.2 Interest**

The Governing bodies and Traffic Departments would be interested in this analysis and will benefit from the prescribed suggestions.

**2. Data acquisition and cleaning**

**2.1 Data sources**

The data on which analysis has to be performed can be found in this Kaggle dataset <https://www.kaggle.com/tsiaras/uk-road-safety-accidents-and-vehicles>. The UK government collects and publishes (usually on an annual basis) detailed information about traffic accidents across the country. This information includes, but is not limited to, geographical locations, weather conditions, type of vehicles, number of casualties and vehicle maneuvers, making this a very interesting and comprehensive dataset for analysis and research.

* It covers a wider date range of events.
* Most of the coded data variables have been transformed to textual strings using relevant lookup tables, enabling more efficient and "human-readable" analysis.
* It features detailed information about the vehicles involved in the accidents.