Bayesian Networks

Butterfly Effects

By

Bharat Nagaraju (A0178258N)

Edwin Tam (A0178396J)

Vignesram Andiappan Selvaraj (A0178215A)

Contents

[Domain & Approaches 2](#_Toc5526470)

[Preparing the Data 3](#_Toc5526471)

[Our Network Models 4](#_Toc5526472)

[How Good Are Our Network Models? 5](#_Toc5526473)

# Domain & Approaches

## **Just what are the factors impact vehicle safety?**

We are given a car crash dataset from Bayesia Website to clean, review, and infer from using 2 Bayesian Networks.

The dataset contains 19 usable features, of which OA\_MAIS is the target variable (it measures injury severity). Before we describe the preparation, network generation, and results etc – let’s have a looksee into the dataset. We’ll use a mixture of Genie & Excel to get the work done.

**Dataset Shape:** 21 columns X 20,247 observations

**Continuous Variables**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Variable** | **Mean** | **Variance** | **StdDev** | **Min** | **Max** | **Count** | **Missing %** | **Remarks** |
| GV\_CURBWGT | 1617.3 | **154900.5** | 393.6 | 670 | 4310 | 20204 | 0.2% |  |
| GV\_DVLAT | 0.0 | 169.5 | 13.0 | -114 | 118 | 14049 | 30.6% |  |
| GV\_DVLONG | -14.8 | 311.8 | 17.7 | -145 | 84 | 14049 | 30.6% |  |
| GV\_ENERGY | 505.2 | **416986.1** | 645.7 | 4 | 9852 | 14049 | 30.6% |  |
| GV\_LANES | 3.3 | 1.8 | 1.4 | 1 | 7 | 20244 | 0.0% |  |
| GV\_MODELYR | 2003.6 | 7.6 | 2.8 | 2000 | 2012 | 20247 | 0.0% |  |
| GV\_OTVEHWGT | 1630.2 | **169212.7** | 411.4 | 640 | 4540 | 18147 | 10.4% |  |
| GV\_SPLIMIT | 40.7 | 126.4 | 11.2 | 0 | 75 | 20016 | 1.1% |  |
| OA\_AGE | 40.2 | 301.8 | 17.4 | 0 | 97 | 20190 | 0.3% |  |
| OA\_HEIGHT | 170.8 | 115.5 | 10.7 | 59 | 216 | 17508 | 13.5% |  |
| OA\_MAIS | 0.9 | 1.1 | 1.0 | 0 | 6 | 19203 | 5.2% |  |
| OA\_MANUSE | 0.9 | 0.1 | 0.3 | 0 | 1 | 19774 | 2.3% |  |
| OA\_WEIGHT | 78.7 | 385.9 | 19.6 | 28 | 150 | 17599 | 13.1% |  |
| VE\_ORIGAVTW | 154.7 | 58.6 | 7.7 | 105 | 185 | 20014 | 1.2% | Not Used |
| VE\_WHEELBAS | 281.0 | 824.7 | 28.7 | 141 | 481 | 20238 | 0.0% | Not Used |
| VE\_PDOF\_TR | 152.6 | 4557.2 | 67.5 | 5 | 355 | 18298 | 9.6% |  |
| GV\_FOOTPRINT | 4.4 | 0.4 | 0.6 | 2.4684 | 7.7952 | 20010 | 1.2% |  |

**Categorical Variables**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **GV\_WGTCDTR** | **Count** | **Percentage** |  | **OA\_BAGDEPLY** | **Count** | **Percentage** |
| Passenger Car | 12474 | 62% | Deployed | 9593 | 47% |
| Truck (<=10000 lbs.) | 2542 | 13% | Not Deployed | 10654 | 53% |
| Truck (<=6000 lbs.) | 5231 | 26% |  |  |  |
|  | | | | | | |
| **OA\_SEX** | **Count** | **Percentage** |  | **VE\_GAD1** | **Count** | **Percentage** |
| (missing) | 234 | 1% | (missing) | 789 | 4% |
| Female | 9938 | 49% | Front | 11793 | 58% |
| Male | 10075 | 50% | Left | 3145 | 16% |
|  |  |  | Rear | 1741 | 9% |
|  |  |  | Right | 2779 | 14% |

# Preparing the Data

# Our Network Models

# How Good Are Our Network Models?