

### 4.4 Exercise do it yourself:

*Download Source Data Set from GitHub link* : <https://github.com/Smartbrain2024/Mastering_AI_2.git>

**Chapters/Chp\_04/4.4/DIY\_Exercise/Automobile\_data.csv**

#### About Dataset

##### Context

This dataset consists of data from 1985 Ward's Automotive Yearbook. Here are the sources.

Sources:

1. 1985 Model Import Car and Truck Specifications, 1985 Ward's Automotive Yearbook.  
   2) Personal Auto Manuals, Insurance Services Office, 160 Water Street, New York, NY 10038  
   3) Insurance Collision Report, Insurance Institute for Highway Safety, Watergate 600, Washington, DC 20037

##### Content

This data set consists of three types of entities: (a) the specification of an auto in terms of various characteristics, (b) it assigned insurance risk rating, (c) its normalized losses in use as compared to other cars. The second rating corresponds to the degree to which the auto is riskier than its price indicates. Cars are initially assigned a risk factor symbol associated with its price. Then, if it is riskier (or less), this symbol is adjusted by moving it up (or down) the scale. Actuaries call this process "symbolling". A value of +3 indicates that the auto is risky, -3 that it is safe.

The third factor is the relative average loss payment per insured vehicle year. This value is normalized for all autos within a particular size classification (two-door small, station wagons, sports/specialty, etc. .), and represents the average loss per car per year.

Note: Several of the attributes in the database could be used as a "class" attribute.

Inspiration

Please bring it on whatever inferences you can get it. through them, you can predict each data point’s likelihood of error.