Bonus Exercise: Car Insurance

\*For convenient use, press the links placed on the right side of each sheet ([Car Insurance.xlsx](Car%20Insurance.xlsx))

1. For adjusting the car insurance price, we have a lot of attributes we can take in mind, from Wobi's 6 attributes, I chose 3 attributes and guessed if these attributes are considered as sensitive or less sensitive (sensitive: the impact on the price). The attributes I chose is:
   * **Claims in past 3 years:** In my opinion this attribute should be the one of the most influencing on the car insurance price, from this attribute we can understand the most important information about our driver, its directly related to our goal target– whether the car driver will do another car accident or not and how many times he was involved with lawsuits related to car accidents, therefore I would score this attribute as P=20% (sensitivity).
   * **Car model and details:** important attribute as well, its obvious that the car insurance price will be affected - the expensive the vehicle is the expensive the car insurance will be (considering the cost of repairing it). Moreover, if the car has new vehicle warning systems the probability of being involved in car accident are reducing therefore the price of insurance should decreased in accordance. I would give this attribute P=15% (sensitivity).
   * **Car use purpose:** by this attribute the insurance company can measure the level of attachment of driver to his vehicle and thus assess the extent of his concern about the vehicle, if it’s a leasing car that used mostly for work purposes the driver won't care about the car and the probability to active the car insurance will be higher. In addition, by checking the amount of this car usage (with questions - does your partner has a car? Is it a leasing car?) the insurance company can estimate the price as will therefore the price will be sensitive to this attribute as well, I would score this attribute as P=10% (sensitivity).

3. The sensitivity measuring achieved by ID3 algorithm, with that algorithm I computed the probabilities for success (Bought Car Insurance = Yes). The algorithm is checking which attribute is the most influential in purchasing the car insurance. At branches that has even probabilities the algorithm chose randomly (between Yes/No). By measuring which branch has the highest probability for success (Bought Car Insurance = Yes) I can tell which category has the highest demand, therefore, I can raise its price, under consideration that customers will still purchase the insurance due to the high demand.  
In our case, the leading attribute is **Car Type** (Contains most of the information), that attribute divided into 4 categories: Compact, Executive, Family & SUV. Let (Bought Car Insurance = "Yes") considered as success.

* Compact: achieved 7 successes out of 9 branches, its sensitivity score will be: = 0.7777
* Executive: achieved 4 successes out of 6 branches, its sensitivity score will be: = = 0.6666
* Family: achieved 10 successes out of 16 branches, its sensitivity score will be: = 0.625
* SUV: achieved 8 successes out of 12 branches, its sensitivity score will be: = = 0.6666

We can infer from those results that Compact attribute should be the most sensitive (if our customer has Compact car his car insurance price should be the highest). Moreover, at branches with small number of nodes (2-3) and a lot of customers who bought insurance eventually (e.g. Compact>2008-2011 or Compact>2015-201 {presented at the excel file}) price should rise even more. Attributes with smaller sensitivity score will have smaller increase in their price.

Three additional attributes for car insurance (even though some of them aren't that ethical):

* Medical history of the costumer: e.g. if one has epilepsy, he/she has higher chances of activating the insurance.
* Cars that are more likely to be stolen (from statistic results)
* Drugs using behaviors