
GenIQ: Finding Tax Cheaters Easily

Bruce Ratner PhD

DM STAT-1 CONSULTING

1 800 DM STAT-1 www.DMSTAT1.com

GenIQ[®]

~ The GenIQ Model ~

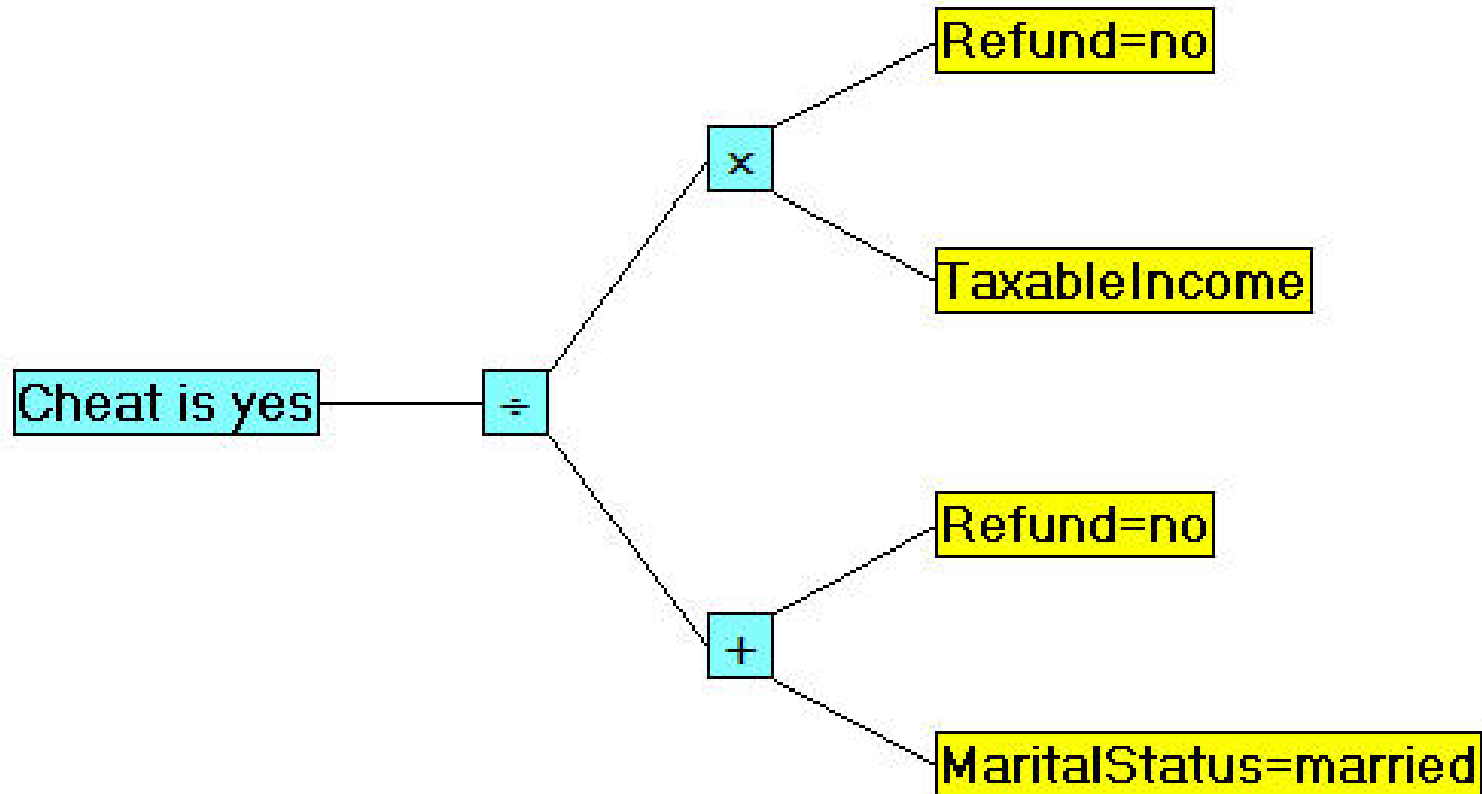
The GenIQ Model© is a machine learning alternative model to the statistical ordinary least squares and logistic regression models. GenIQ let's the data define the model – automatically data mines for new variables, performs variable selection, and then specifies the model equation – so as to "optimize the decile table," to fill the upper deciles with as much profit/many responses as possible.

In this illustration, GenIQs optimizing of the deciles is equivalent to predicting the best possible ranking (permitted by the data) of the target variable based on the GenIQ Model score **GenIQvar**.

Tax Cheat Input Data

Tax ID	Refund	Marital Status	Taxable Income	Cheat
1	yes	single	125	no
2	no	married	100	no
3	no	single	70	no
4	yes	married	120	no
5	no	divorced	95	yes
6	no	married	60	no
7	yes	divorced	220	no
8	no	single	85	yes
9	no	married	75	no
10	no	single	90	yes

GenIQ “Tax Cheat” Model Tree



GenIQ “Tax Cheat” Model/Code

```
If MaritalStatus = "married" Then x1 = 1; Else x1 = 0;
  If Refund = "no" Then x2 = 1; Else x2 = 0;
  x1 = x1 + x2;
  x2 = TaxableIncome;
  If Refund = "no" Then x3 = 1; Else x3 = 0;
  x2 = x2 * x3;
  If x1 NE 0 Then x1 = x2 / x1; Else x1 = 1;
GenIQvar = x1;
Cheat is “Yes” = GenIQvar;
```

GenIQ “Tax Cheat” Model Results



Tax ID	Refund	Marital Status	Taxable Income	Cheat	Gen IQvar
5	no	divorced	95	yes	95.0
10	no	single	90	yes	90.0
8	no	single	85	yes	85.0
3	no	single	70	no	70.0
2	no	married	100	no	50.0
9	no	married	75	no	37.5
6	no	married	60	no	30.0
1	yes	single	125	no	1.0
7	yes	divorced	220	no	1.0
4	yes	married	120	no	0.0

FINDING: Cheaters are Single/Divorced w/ Middle Income & No Refund