CoIL Challenge 2000 submission

Paul Rickets
44 West 63rd Street, New York, NY 10023
paulrickets@hotmail.com

I divided the data in three sets, and build a tree for each of them. I used C4.5, and created new attributes.

The explanation for the marketing people is the following :

There are tree clusters of people :

1. People who is not a customer
2. People who is a customer and has a car policy.
3. People who is a customer and don't have a car policy.

I built a model for each of the three segments, and found this :

Cluster 3 has a very low probability of having a caravan, 3 times less than the average.In this group, people who own a mopped, have more than 60 years, have social_class_A >= 50% or social_class_D >= 24% (too rich or too poor) or have farmer >= 11% have a much more lower probability, p = 0

In cluster 2, people having a boat, or having a fire policy have a greater than average probability of buying a caravan policy. The most important attribute, was a derived one, named 'average car policy contribution' that resulted from c_car_policies/ n_car_policies. Having a large value for this attribute means a greater probability of having a caravan.

In cluster 1, married people, skilled labourers or large families (but not rural ones) have the best chances.