Attached are two datasets; *training.dat* and *new_customers.dat*. Please provide a brief report about the structure of the data, and then use *training.dat* to predict *new_customers.dat*.

training.dat has 5 columns given by

1. **Total Sales Since 2005:** The total number of sales this particular customer has made since 2005
2. **Recent Activity Change:** A metric that aggregates data associated with customer activity on our website. Positive values indicate the customer is more active than in previous months and negative means they are less active.
3. **Days Since First Sale:** The number of days that have passed since the customer made their first purchase on our site.
4. **Customer Metadata:** Some information on the individual customer. The classes are *Child Male, Child Female, Adult Male, Adult Female, Senior Male,* and *Senior Female.*
5. **Customer Score:** The desired metric we wish to optimize. This is a proprietary metric that balances how much the customer spends on certain items.

new_customer.dat only has the first 4 columns and you should be predicting the *Customer Score* using these 4 columns.

You report should answer the following questions:

1. Is there any structure in the data?
2. Are all the features important in predicting the Customer Score? Why or Why not?
3. Does a customer need to have increased their activity to be scored high?
4. How did you handle missing data? Why?

For the prediction please indicate the following:

1. How should *training.dat* be split into testing and training data?
2. How did you evaluate your model to be sure the predictions on *new_customers.dat* are correct?
3. How confident are you in your predictions on *new_customers.dat*?

Other considerations:

1. Be clear and concise in your report.
2. Although we will review this in more detail during your interview, it should be relatively understandable to someone without further explanation.
3. Consider vagueness as open questions that will require you to list the pros/cons of various solutions.
4. You should not spend more than few hours on this.

Thank you!

Doug Sherman