Evaluation of Sample Proposal 4: Telecommunication Customer Churn Prediction

The project goal is to predict customer churn, which customers of the business may leave within the next month. The proposal also does a good job at stating the "why" for the proposal, "it is costly for a company to find new customers". This is always a useful prediction to make for a business, however, it is also helpful to know what specific problems this would solve for the business. When looking at a supervised problem, we want to look at how the model will solve, or address, the business problem. The project proposal does a good job at explaining how the problem will be addressed, but not which problem will be addressed.

The project proposal does well at defining the target variable, "churn", which will predict either "yes – customer leaves within the last month" or "no – customer does not leave within the last month". There are also many attributes that are being used to predict the target attribute, which include the attribute name, type, and description. The data source is provided and has been divided into subsets, training and prediction. The values for the target variable will be acquired by using the demographic attributes that have been obtained from customers. A couple things are missing during the data preparation process of this project proposal. Although the data is being obtained from a website, it is not stated how the attributes are being obtained from the customer and if there are any costs associated with obtaining the information. As these points pertain directly to the business budget and possibly privacy guidelines, it is important to include them.

The methods in the project proposal are stated clearly as using decision trees, logistic regression, Naïve Bayes, neural networks, lift chart, and confusion matrix. Using multiple methods could help in comparing and possibly finding the most accurate one. The proposal does well at mentioning the "missing values" will be handled during the data preparation process, as some classifications methods, such as neural networks, do not handle missing values well.

Overall, the proposal does well at defining a goal and target variable during the supervised model, but is missing information that would be important in the business understanding aspect of the problem.