 A title that sums what the phenomenon is and how accurately you believe you can model  
it   
 Definition of a phenomenon.  
 An introduction of no more than 1.5 pages that talks about the phenomenon, why you  
are modeling it why it is important and possibly a short history of attempts to model it.  
  Definition of a model  
 Definition of a mathematical model  
 Description of what a model means in the context of your phenomenon  
 Description of your phenomenon  
 Description of how the prediction target is measured exactly  
 Discuss if you believe this measurement can be made accurately  
 Definition of causal drivers  
 Discussion of the phenomenon's causal drivers and how measuring them would be impossible  
 Definition of stationarity  
 Is your phenomenon and response metric stationary?  
 Definition of supervised learning  
 A discussion of how can supervised learning be used to model your phenomenon  
 Definition of independent variables (features)  
 Description of your model's features and how they are measured exactly  
 Do you believe your feature measurements are practical and can be made accurately?  
 In order to improve the model's predictive performance, do you think some of the features  
should be transformed or interacted?  
 Definition of historical data observations (training data).  
 How would you go about obtaining (sampling) a training dataset in your context? What  
would n be? Would it be possible? Expensive?  
 Definition of the three sources of error.  
 In your modeling scenario, which of these sources do you anticipate would be large and  
why? Which are small and why?  
 Definition of prediction using models.  
 How can your model be used to predict?  
 Who will be predicting using your model and for what purpose?  
 Definition of prediction error metrics.  
 Which error metric would you employ in your modeling context?  
 What is the threshold for usefulness in your context?  
 Definition of interpolation and extrapolation.  
 When your model will be used in the real world, will its users be interpolating or extrapolating?  
 What is an algorithm and candidate set? What is machine learning?  
 What are some algorithm choices in this modeling context?  
 Definition of the model section problem and how it arises during modeling.  
 How would you select a model from those set of choices?  
 Which model do you think will ultimately get selected and why?  
 For the algorithm you are considering after selection, what is the null model you seek to  
outperform?  
 Do you have enough sample in your historical data to fit this selected model? Discuss.  
 Define underfitting.  
 Could your chosen model be underfit?  
 Define overfitting.  
 Could your chosen model be overfit?  
 Define validation using concepts such as in-sample and out of sample.  
 How would you validate your chosen model?  
 Conclude: is the title of your essay correct and why? Tie the answer to what you believe  
will be your chosen model's predictive performance.  
 Throughout the essay you must use all the following notation where appropriate: