Model selection challenge

1. Predict the running times of prospective Olympic sprinters using data from the last 20 Olympics.
   1. This is a regression problem. Linear regression seems like it would work just fine. I’m sure there is an ensemble method that could be added to make it more powerful. A knn regresor might be interesting here as well.
2. You have more features(columns) than rows in your dataset.
   1. While I am not sure what type of outcome you are looking for(classification or regression), the fact that there are so many columns indicate that you could probably drop some without too much loss of variance. If it is a regression problem then Lasso regression would help get rid of a lot of the useless features.
3. Identify the most important characteristic predicting the likelihood of being jailed before age 20.
   1. The feature importance method would easily solve this problem but I am not sure if that is specific to one type of model.
4. Implement a filter to highlight e-mails that might be important to the recipient.
   1. This is a classification problem and sounds like it could be solved similarly to the Spam prediction model. A naïve bayes model would predict it well and could be made stronger by using boosting.
5. You have 1000+ features
   1. Lasso or Ridge would both work here assuming it is a regression issue.
6. Predict whether someone who adds items to their cart on a website will purchase the items.
   1. An ensemble method like random forest would probably work well. As far as which weak learner to use, I might go with KNN.
7. Your dataset dimensions are 982400 x 500
   1. Ridge regression would help deal with this massive amount of data
8. Identify faces in an image
   1. This is a classification issue. I think KNN would work here and then run through an ensemble method like boosting.
9. Predict which of three flavors of ice cream will be most popular with boys vs girls.
   1. This is a question best answered by logistic regression(classification)