Question 1 of 13: Build a decision tree (using operator W-J48 from the Weka Extension Pack) on the entire data set. What is the non-cross-validated kappa?

0.346

Question 2 of 13: That's the correct answer, but let's think about it. The kappa value you just obtained is artificially high – the model is over-fitting to which student it is. What is the non-cross-validated kappa, if you build the model (using the same operator), excluding student?

0.159

Question 3 of 13: Some other features in the data set may make your model overly specific to the current data set. Which data features would not apply outside of the population sampled in the current data set? Select all that apply.

- [ ] A) School
- [ ] B) Class
- [ ] C) Grade
- [ ] D) Coder
- [ ] E) Gender
- [ ] F) Obsnum

Submit

Question 4 of 13: What is the non-cross-validated kappa, if you build the W-J48 decision tree model (using the same operator), excluding student and the variables from Question 3? Recall that you decided to eliminate School, Class, and Coder, as well as STUDENTID.

0.171

Question 5 of 13: What is the non-cross-validated kappa, for the same set of variables you used for question 4, if you use Naïve Bayes?

-0.007

Question 6 of 13: What is the non-cross-validated kappa, for the same set of variables you used for question 4, if you use W-JRip?

0.100

Question 7 of 13: What is the non-cross-validated kappa, for the same set of variables you used for question 4, if you use Logistic Regression? (Hint: You will need to transform some variables to make this work; RapidMiner will tell you what to do)

0.00

Question 8 of 13: Wow, that was a lot of waiting for nothing. What is the non-cross-validated kappa, for the same set of variables you used for question 4, if you use Step Regression (called Linear Regression)?

0.001

Question 9 of 13: What is the non-cross-validated kappa, for the same set of variables you used for question 4, if you use k-NN instead of W-J48? (We'll discuss the results of this test later).

1.000

Question 10 of 13: What is the kappa, if you delete School, Class, Coder, and STUDENTID, use W-J48, and conduct 10-fold stratified-sample cross-validation?

0.106 +/- 0.019

Question 11 of 13: Why is the kappa lower for question 11 (cross-validation) than question 4 (no cross-validation?)

- A) K-fold-cross-validation is not as good as leave-out-one-cross-validatio
- B) You shouldn't use kappa with k-fold-cross-validation
- C) Using cross-validation overfits to the training sample
- D) Not using cross-validation overfits to the training sample