Assignment 7

Learning Objectives:

1. Explore a new machine learning algorithm, namely JRIP
2. Practice troubleshooting skills – figuring out what your model is doing

Description:
The data set you will be working with for this assignment was created using LightSIDE. It is a data set where the model is predicting whether a person will be given a job or not based on their resume. The job they are applying for is Wikipedia RFA (which is an administrative position you can read about on Wikipedia if you are interested). All of the basic text features have been extracted for you.

Step-by-Step Guide:

1. Complete the readings through week 10.

2. Use the experimenter to determine whether you get significantly different performance from J48 (tree based learning) and JRIP (rule based learning) when you use a feature selection wrapper that selections the top 50 features on each fold. Give screen shots and explain your results.

3. Troubleshoot your results so that you understand why performance was or was not different between tree and rule based learning. The point here is not to identify where the feature space is weak but to investigate why the algorithms did or didn't perform differently. This can be thought of as a more advanced version of what you did with the Titanic dataset earlier in the semester. It should be doable since you will only be considering around 50 features. You may need to be creative. Now explain your results using what you understand about tree and rule based learning and what you found in your error analysis.

4. Are the results surprising given the discussion about tree and rule based learning in the book? Why or why not?

Deliverables: Write up of your exploration process that includes your write-up from 2-4 above.

Submission Mode:

Submit the assignment to blackboard.