Reviewer’s report

Title: EFS: An Ensemble Feature Selection Tool implemented as R-package and Web-Application

Version: 0 Date: 21 Apr 2017

Reviewer: Benjamin Reading

Reviewer’s report:

Overview:

The authors provide a well written manuscript describing a new software package "EFS" that is designed to evaluate feature importance. The EFS uses several feature selection methods and an example if its use is also provided along with example output figures. There are a small number of minor comments below.

General Comments:

Page 1, Abstract, Conclusion: The authors may consider revising the first sentence to "The EFS identifies relevant features while…".

In the Introduction, the authors may consider adding a very brief (1-2 sentence long) general description of why feature selection is important and how it is typically used (e.g., a sensitivity analysis to reduce data dimensionality by ranking a list of features by importance and then eliminating those features that are less relevant to the data mining/machine learning question. This improves the overall performance of the model because it addresses the problem of overfitting.). This will demonstrate the outright importance of using the described methods.

Page 2, under "Median": "As importance measure the p-values were used." is an awkward sentence. Do the authors mean to state "The resulting p-values are used as a measure of feature importance."?

Page 2, under "Random Forest": the authors should spell out the abbreviation of area under the ROC curve (AUC) when it is first used. Also, ROC should be defined here as well.

Page 2, "Ensemble learning". Words in all of the other subheadings start with capitalized letters (should this be "Ensemble Learning"?)

Page 3, last paragraph under "R-Package": A period (.) is used where a comma (,) should be: Consider revising to "Due to the high computational costs of the RFs, the default selection is set to"
Page 4 under "Performance evaluation by logistic regression": Receiver operating characteristic (ROC) curve should be spelled out on page 2 when the abbreviation is first used (See comment above).

Page 4, under "Permutation of class variable": The authors should consider revising the last sentence to "...the AUC from the original LR model using a Student's t-Test." or some similar wording.

Page 5, last paragraph under "Example": "Figure 2" is stated with a capital "F" however "figure 3" is not. Should this be "Figure 3"?

Page 5, last paragraph under "Example": The sentence "Low values of variances imply that the importances are stable and reliable" is awkward. The authors may consider revising this sentence to "A low value of variance implies that the importance of a feature is stable and reliable" or some similar wording.

**Level of interest**
Please indicate how interesting you found the manuscript:

An article of importance in its field

**Quality of written English**
Please indicate the quality of language in the manuscript:

Acceptable

**Declaration of competing interests**
Please complete a declaration of competing interests, considering the following questions:

1. Have you in the past five years received reimbursements, fees, funding, or salary from an organisation that may in any way gain or lose financially from the publication of this manuscript, either now or in the future?

2. Do you hold any stocks or shares in an organisation that may in any way gain or lose financially from the publication of this manuscript, either now or in the future?

3. Do you hold or are you currently applying for any patents relating to the content of the manuscript?

4. Have you received reimbursements, fees, funding, or salary from an organization that holds or has applied for patents relating to the content of the manuscript?

5. Do you have any other financial competing interests?

6. Do you have any non-financial competing interests in relation to this paper?
If you can answer no to all of the above, write 'I declare that I have no competing interests' below. If your reply is yes to any, please give details below.

I declare that I have no competing interests.

I agree to the open peer review policy of the journal. I understand that my name will be included on my report to the authors and, if the manuscript is accepted for publication, my named report including any attachments I upload will be posted on the website along with the authors' responses. I agree for my report to be made available under an Open Access Creative Commons CC-BY license (http://creativecommons.org/licenses/by/4.0/). I understand that any comments which I do not wish to be included in my named report can be included as confidential comments to the editors, which will not be published.

I agree to the open peer review policy of the journal

Do you want to get recognition for reviewing this manuscript? Add a record of this review to Publons to track and showcase your reviewing expertise across the world’s journals. Signing up is quick, easy and free!

Yes