**Individual Peer Evaluation Form**

Your name: David Culhane

Write the name of your classmate you are preparing this review for in the designated column. Using a scale of 1-4 (1=strongly disagree; 2=disagree; 3=agree; 4=strongly agree) answer each question. If you aren’t able to answer the question based on what is posted in the discussion board, reach out to your classmate for more information via the discussion board. Total the numbers in each column. **Make sure to answer the questions on the 2nd page.**

|  |  |
| --- | --- |
| Evaluation Criteria | Peer Name:  Andrew Pfeifer |
| Has plan in place to complete course project. | 4 |
| Has found datasets/data sources to support project idea. | 4 |
| Has solidified project idea. | 4 |
| Has identified resources for project. | 4 |
| Topic is related to data science and demonstrates topics learned to date through program. | 4 |
| Risks and potential issues have been identified. | 3 |
| TOTALS | 23 |

Feedback on Individual’s project topic:

1. How clear is the classmate’s project topic? What questions does their topic make you consider?

Andrew’s project topic is clear and clearly executed. Credit card fraud is a topic that is highly relevant to data science and he takes multiple approaches when attempting to create models that can do a good job at predicting credit card fraud. Logistic Regression and Random Forest Classification are classic models to use for this classification task.

The only thing I would consider given his code is what the effect of up-sampling the fraud cases would do for the models themselves. On the one hand, it could end up reinforcing fraud trends. It would be something to look into by re-running the models after up-sampling to see if it also helped identify cases that were fraudulent but not predicted as such (the false negative cases).

1. What risks or issues should your classmate consider while working on their project?

Andrew has already detailed wishes to go back through his models and add standardization to the fields of his data.

Andrew also detailed steps involving feature reduction. In a previous peer-review I mentioned how the data being used for modeling is already the result of Principal Component Analysis (PCA). The section with regard to feature removal has outputs suppressed by commenting so I’m unsure how many columns were removed on the basis of having high correlation (corr > 0.9) with another column that itself was the result of PCA. I would also be interested in models created without removing these columns since each column that is the result of PCA is already a blend of data from each of the original fields that were fed into PCA. Removing post-PCA columns may end up removing other aspects of the data that are no longer being captured by the models.

1. Additional suggestions/comments that might be beneficial to your peer?

I do not have any additional suggestions or comments at this time.

Adapted from a peer evaluation form developed at Johns Hopkins University (October, 2006)