**Individual Peer Evaluation Form**

Your name:

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 topic is clear, identifying cases of credit card fraud. The questions I would consider would be what variables are more likely to be of use in identifying fraud cases. Andrew has stated that this may not be possible due to the anonymized nature of the data (the variable names are labeled V#).

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

Andrew has identified the issue of the low ratio of fraudulent vs non-fraudulent cases. He has posed the idea of removing non-fraudulent cases or introducing additional data of fraudulent cases. Andrew acknowledges that it would he tough to verify if any of his data would be duplicates, however. Upsampling of fraudulent cases would be a solution but would not solve the issue of duplicate data though.

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

I agree that adding additional fraud cases won’t be a bad idea. My only concern is that if they are from a different dataset, it’s possible that the data either will not have the exact same fields or the fields will not be in the same order as the primary dataset. I would not know how to verify that if the variable names are anonymized. This concern is magnified since the primary dataset’s 28 fields are the result of Principal Component Analysis. Recreating this without the same number of fields or the eigenvalues and eigenvectors from the original PCA to apply to additional datasets would be incredibly difficult. Without that, additional data could end up harming the model.

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