**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:  Sameer Nepal |
| 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. | 4 |
| TOTALS | 24 |

Feedback on Individual’s project topic:

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

Sameer Nepal’s is planning to investigate the effect of smoking to the birth of the baby. His project topic is very clear. He has explained very well why he has chosen this project, what he aims to analyze and find out in this project. He has also described who will be benefitted from his model. His model not just helps a particular group of people, but also to everyone as everyone would want to know if smoking affects the birth and birth weight of the baby. The topic he has chosen doesn’t raise any questions for me.

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

All the variables in the dataset have been detailed well. He has considered all the factors which would influence the birth weight of the baby. He has a solid plan on how the missing variables will be handled and why he has chosen that approach. He got a clear idea of what variables he is going to choose for the model and how he is going to choose those variables. Model performs better with numerical variables so he has converted all the categorical variables using label encoder. He did his exploratory data analysis very well by all possible means. Finally, he has deployed several models on the cleaned data and compared the accuracy of all the models to determine the best model. I would say he has taken almost everything into consideration to develop the best model which would gain all set of people.

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

I suggest he could use cross validation technique like GridSearchCV to search through the best parameter values from the given set of the grid of parameters.  Grid Search assesses the performance for each possible combination of the hyperparameters and their values, chooses the combination with the best performance.

Also, there are chances of imbalanced data distribution when observations in one of the classes is much higher or lower than the other classes.  If he has imbalanced data distribution in his dataset, then his model would become more prone to the case when minority class has negligible or very lesser recall. SMOTE is one of the most commonly used oversampling methods to solve the imbalance problem. My suggestion is to check if there is oversampling in the data and handle it using SMOTE if the data is oversampled.

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