**Individual Peer Evaluation Form**

Brian Reppeto:

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:  Sauda Haywood |
| 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?

* The project topic is clearly defined, focusing on predicting the likelihood of an individual developing obesity based on their lifestyle, demographic data, and health metrics. The explanation provides a solid understanding of the objectives, which includes building predictive models using features such as age, weight, height, physical activity, and eating habits.

A few questions come to mind regarding the scope and methodology:

* Data Balance: Is there any imbalance in the target variable (obesity vs. non-obesity) that might affect model performance, and how will this be addressed?
* Granularity of Lifestyle Factors: How granular are the lifestyle factors? For example, are specific eating habits like meal timing and quality included, and how could these influence the model’s predictions?

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

* Bias in the Data: Where is the dataset sourced from? Are there cultural biases in the model. The model may not generalize well to other populations due to differences in lifestyle, diet, and physical activity. Bias mitigation techniques should be considered, such as re-sampling or adjusting model weights.
* Data Quality: Lifestyle data can be subjective (e.g., self-reported physical activity levels or eating habits). Such data might have inconsistencies or inaccuracies that could impact the model’s reliability. Proper data preprocessing and verification methods should be implemented to minimize this risk.

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

* Feature Engineering: Consider performing feature engineering to create more informative variables. Additionally, creating interaction terms between features (combining physical activity and diet quality) could help capture more complex relationships.
* Cross-Validation and Hyperparameter Tuning: To ensure model robustness, use cross-validation techniques to evaluate model performance. Hyperparameter tuning for the Random Forest model, such as adjusting the number of trees or maximum depth, could lead to significant performance improvements.
* Bias Mitigation Strategies: To address potential biases, implementing re-sampling techniques (e.g., SMOTE) if there is an imbalance in the target classes to ensure the model does not disproportionately affect any group.

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