Provide your written answers to the following prompts in the spaces provided. Use whole sentences and clear grammar – answers that are difficult to interpret will lose credit. Please use a text color other than black for your answers. When you have completed the exam, save the document as a PDF, submit it on the course D2L website and email it to [dhowell@oru.edu](mailto:dhowell@oru.edu). Due May 1, 2020.

Name: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

|  |  |
| --- | --- |
| 1. What is Data Science? |  |
| 1. Why is Data Science an important field, given the current state of technology? |  |
| 1. What are the two main types of data covered in this class? Give an example of each. |  |
| 1. Explain the difference between a regression algorithm and a classification algorithm. What does the output of each algorithm look like? |  |
| 1. Explain the difference between supervised and unsupervised machine learning. Which type of machine learning works best with a data set which has no clear dependent variable? |  |
| 1. What does it mean for an algorithm to be computationally expensive? When does computational cost matter and when does it not? |  |
| 1. Choose three of the following types of algorithms and explain how they work. You do not need to include the equations used by each algorithm; all that is needed is a simple, succinct, and accurate description of the mechanics of the algorithm.  * Linear Regression * Logistic Regression * K-means Clustering * Random Forest * Naïve Bayes * Support Vector Machines |  |