



MSBD566 – Predictive Modeling and Analytics

Assignment 2

Instructions

We will prepare a midterm project (which you can also use the same data for the final project).

For this assignment, you will need to identify a dataset that you would like to work on for the project. The data should be:

1. More than 500 samples
2. At least 4 categories/columns that we can train/analyze, with at least 1 column containing a categorical type.

Submit a description of the dataset that you have in a PDF file containing the following information:

1. Your name, course name, date, and assignment title
2. Data Title
3. Source: where you get it
4. Data description: what is it about
5. Data dictionary: all the definitions, what the columns mean, etc.
6. Any disclaimer or notes that are needed to be aware of about the data.

Homework must be submitted by **10/8/2025 11:59 pm CT**. Late submissions will be accepted up to 2 days with penalty of 5 points per day. **No submission will be accepted after 10/10/2025 11:59 pm CT without prior approval.**

AI Policy: AI can be used in terms of understanding terms and asking for examples but not directly to answer the questions.