Team Members:

**Ayesha Maryam, Athul Madhusudhanan, Mujahid Iqbal, Edward Vaughan, Daniel Daniel**

Topic:

**Predict which employees are likely to leave a company based on their work history, demographics, and performance metrics.**

Dataset (link it if possible):

**https://www.kaggle.com/datasets/pavansubhasht/ibm-hr-analytics-attrition-dataset**

1. Is this a regression or classification task?

**Classification**

1. What are your inputs? (features/columns/etc. that your model will train on)

**All columns except for “Attrition”**

1. What is your target? (the output that your model will predict)

**“Attrition”**

1. What machine learning model(s) are you planning on using?  
   **Logistic regression OR Neural Net / Deep Learning**
2. One of the criteria for the project is for you to optimize your model. What are some parameters or other things you could change about your model as you test its performance to optimize it?

**Increasing/decreasing the number of epochs**

**Changing the number of hidden layers**

**Changing the number of neurons in a layer**

**Changing the number of input columns**

**Removing outliers and erroneous data**

**Placing data into “bins”**

**Use a larger training dataset**

**Removing columns which don’t impact the output as much (by weight)**

1. What steps will you need to take to clean and prepare your data?  
   **Identify outliers**

**Drop empty records**

**Identify features which have the most impact on attrition**

1. Do you have any questions or concerns you think you will need help with?

**Is there a method, if any, to find out which features have the most impact on the target(s)?, like feature\_importances for Random Forests?**

Link to GitHub Repository:

<https://github.com/ddaniel2024/Project-4.git>