Approach:

I decided to use Adaline model implementation found in the Python Machine Learning book, however, I needed to make some changes in order for the model to work with the Titanic data. Specifically, I had to update the activation function to utilize a Sigmoid function rather than a linear activation function. This option was outlined in the textbook as an option for logistic regression, and I found it necessary to use this function because I needed to deal with a probability rate in terms of either 0 or 1 rather than a continuous value. When I tested the model with just a linear activation function the predictions.csv file simply outputted 1 (survived) for all passengers, which I understood intuitively and historically could not be correct. I trained the Adaline model on the following features: passenger class (Pclass), gender (Sex), age (Age), and fare (Fare).

Data Preprocessing:

In order to work with the training and test data it was necessary to do some preprocessing of the data in order for the Adaline model to work correctly. I mapped gender to numerical values (0 = male, 1 = female) and filled missing values in age and fare the average ages and fare values. When I tried running the model as is I found I was having runtime overflow errors and nonsensical outputs in my predictions.csv. In chapter 2 the section "Improving gradient descent through feature scaling" outlined a method for scaling that seemed like a way to optimize the performance of the algorithm. The chapter describes feature scaling as simply subtracting the sample mean for each and example and dividing by its standard deviation which I make use of in the standardize function. I then tuned the learning rate to .001 and number of iterations to 100 because I was still receiving runtime overflow errors. I then decided to save the predictions generated by the Adaline model to a predictions.csv and merge the predictions with the test data into a new csv named merged_data.csv. Utilizing the merged_data.csv I used the pandas library along with some print statements to find the survival rates for overall passengers, Pclass, gender, age and fare.

Results:

Overall:

Overall Survival Rate: 0.36

Pclass:

- **Pclass 1: 0.551402**
- **Pclass 2: 0.322581**
- **Pclass 3: 0.288991**

This makes sense since historically we know that first-class passengers had a higher survival rate

Sex:

- **Male: 0.033835**
- **Female: 0.940789**

This rate does not make as much sense because historically we know that females were more likely to survive than males.

Age:

- **0-10: 0.500000**
- **10-20: 0.510638**
- **20-30: 0.389313**
- **30-40: 0.276596**
- **40-50: 0.260870**
- **50-60: 0.550000**
- **60-70: 0.300000**
- **70-80: 1.000000**

This shows high survival rates for the youngest and oldest age groups, which makes sense to me logically.

Fare:

- **0-10: 0.235294**
- **10-20: 0.280488**
- **20-30: 0.422535**
- **30-40: 0.454545**
- **40-50: 0.375000**

- **50-60: 0.611111**
- **60-70: 0.636364**
- **70-80: 0.300000**
- **80-90: 0.750000**
- **90-100: 0.500000**
- **100-200: 0.692308**
- **200-300: 0.705882**

These results coincide with the Pclass results where higher fare passengers had a higher survival rate than lower fare passengers.