|  |  |
| --- | --- |
| Laura Dooley  July 15, 2024  Purpose is to work with the Titanic train dataset and create a training model. | |
| Q.1 | 2. The survivor, age, sibling/spouse, parent/child, and the fare histograms were skewed to the right. The passenger class was skewed to the left. 3. The boat, body, and home.dest are features that seem would decrease the training model’s predictions. There are large amounts of missing data between the deleted features. When comparing these deleted features with the kept features the data could be considered redundant. There are kept features that are more complete that present similar information. |
| Q. 2 | a) A feature that would allow a better prediction would be to try and create a column that would classify a passenger as a parent or a sibling. I have a suspicion the passengers with children, or a sibling would have a higher chance of survival. I would write a code that would take passengers over a certain age that have a value in the sibling/spouse column and a value in the parent/child column would be classified as parent. The others that are under the age value with a value in the sibling column will be considered sibling.  b) A screenshot of a computer  Description automatically generated  In my opinion the 2 most important features are sex, and survivor. The reason I believe this is there is a higher probability that a survivor is woman, and considering the models are trying to predict a survivor the survivor feature is important. |
| Q. 3 | a) A screenshot of a computer  Description automatically generated  b) A screenshot of a computer program  Description automatically generated |
| Q.3 | c) A screenshot of a computer program  Description automatically generated   1. The training model outperformed the testing model. This suggests the model is overfitting. |
|  |  |
|  |  |
|  |  |