After running one million tests of each form of prediction consisting almost entirely unique slices, I have found that on average the Tree method is the most accurate, with a 78.5% accuracy, with Random Forest being the second most accurate with a 78.4% accuracy. However, if you look at standard deviation, Random Forest appears to be better with a standard deviation of 0.0283 while Tree had a standard deviation of 0.0285. Now, seeing as we are not looking for god-tier accuracy where everything must be disputed, I will assume that the Tree method is more accurate.

For the following calculations, I ran 20,000 simulations using randomly generated slices, and got to watch my amazingly lovely loading bar in action:

1. An adult male in the 3rd class:
   1. Tree avg % chance of survival: 0.145
   2. Nearest avg % chance of survival: 3.415
   3. Forest avg % chance of survival: 0.24

It can be inferred from the above data that the character most likely would not survive, with an optimistic 3.4% chance of survival from the most inaccurate prediction method. From what I deemed to be the most accurate, there is not even a remote chance of survival, weighing in at under a percent.

1. An adult female in 1st class
   1. Tree avg % chance of survival: 99.68
   2. Nearest avg % chance of survival: 99.17
   3. Forest avg % chance of survival: 99.6

It is almost guaranteed survival for an adult female. Again, a high of 99.68% chance of living and a low of 99.17% chance. It appears that you would have to purposely sacrifice yourself to die in this position.

1. A female child in 1st class
   1. Tree avg % chance of survival: 99.49
   2. Nearest avg % chance of survival: 99.32
   3. Forest avg % chance of survival: 99.06

Surprisingly, a female child has a lower survival chance than an adult female. My speculation is that children often do stupid things, and are also less resilient in harsh weather conditions such as freezing water. That being said, female children in the first class had insanely high survival chance.

The most important feature in the dataset seems to be gender. If you are a female, your chances of survival increase significantly, by about 101% compared to a male of the same class and age group. The above number came from a comparison between a third-class adult male and female. Now, the percent difference varies between each class and age group, but the biggest disparity occurs between the genders, not class or age group. I think this may be because of the “women and children first” policy, and the fact that the entire orchestra composed of men died playing music to keep the passengers calm during evacuation.

Overall, I learned that each form of machine learning is suited to a different task, and that getting a huge variety of data with a large scale test is key to normalizing results to make a good decision about what to use. I also learned that optimization on all fronts is key to making a reliable program. I initially had to create safeguards to prevent buffer overflow due to output spam. This meant I would not be able to accurately see what my program was doing, and couldn’t interact with it efficiently. Not to mention 450+~ lines of code would not only take a long time to go through, but it would also cause issues in the IDE itself. Once I reduced the code to under 250 lines again, speed improved significantly, both in IDE and runtime. I did this by using a function to predict estimated time until completion, and by also designing a loading bar. This means that I am able to print logs on one line, and change them without adding more lines, removing the risk of crashing. Another optimization I made was to the slicing method. I originally created a new slice partition every time I wanted to make a calculation in the same for loop. This caused mass stuttering as the kernel tried to juggle massive amounts of variables and functions from the classifiers and random functions. Once I made the partitions get pre-generated into a list, runtime decreased significantly as well.

Tree avg % chance of survival: 14.705

Nearest avg % chance of survival: 42.24

Forest avg % chance of survival: 18.7