Misha Suresh

1. In the correlation heatmap, a value of 1 represents a strong positive correlation, whereas a value of -1 represents a strong negative correlation. When looking at the heatmap for correlation between different features, the correlation between income and age, capital gain, and hours per week had a substantially higher positive correlation value compared to the other features. The columns that stuck out to me were capital gain and capital loss, in correlation with each other, which had a value of -0.031.
2. When observing the histograms for ‘hours.per.week’ and ‘education’ I noticed that the for salaries which were ≤ 50k, the most popular columns were (in order) for education: HS-grad, some-college, and school and for hours per week: 40, 50, and 20. On the other hand, most popular columns were (in order) for education for salaries < 50k were: Bachelors, HS-grad, and some-college, and for hours per week: 40, 50, and 45/60 being tied.
3. For the accuracy performance, we can observe that the model is consistently accurate, showcasing its effectiveness. For the log-likelihood graph, we can see that it is increasing, which means that it is an effective model.
4. The ROC curve for the training and testing data is quite similar. We can see that they have similar characterstics to each other.