This PDF should describe how you improved the application compared to the presented in the lecture.

I made a boxplot for the features fare and age since I assume those two will be correlated to the survival rate the most

After I found out there are quite a few outliers, I removed them … and got the same accuracy. so I figured we probably dropped too much data.

Instead of dropna which drops more than 50% of the data I drop the column cabin which have the most row of NA and doesn't correlate much to the survival rate.

Instead of using
df['TicketNumber'] = split_value.str[-1]
Which will cause the TicketNumber to be "LINE"" if the ticket is "LINE" which will cause error since the TicketNumber column is supposed to be Numeric
I changed it to
df['TicketNumber'] = pd.to_numeric(split_value.str[-1], errors='coerce')

Old model results:
Accuracy: 0.784

	precision	recall	f1-score	support
0	0.70	0.58	0.64	12
1	0.81	0.88	0.85	25
accuracy			0.78	37
macro avg	0.76	0.73	0.74	37
weighted avg	0.78	0.78	0.78	37

New model results:
Accuracy: 0.810

	precision	recall	f1-score	support
0	0.81	0.90	0.85	110
1	0.81	0.67	0.73	69
accuracy			0.81	179
macro avg	0.81	0.78	0.79	179
weighted avg	0.81	0.81	0.81	179