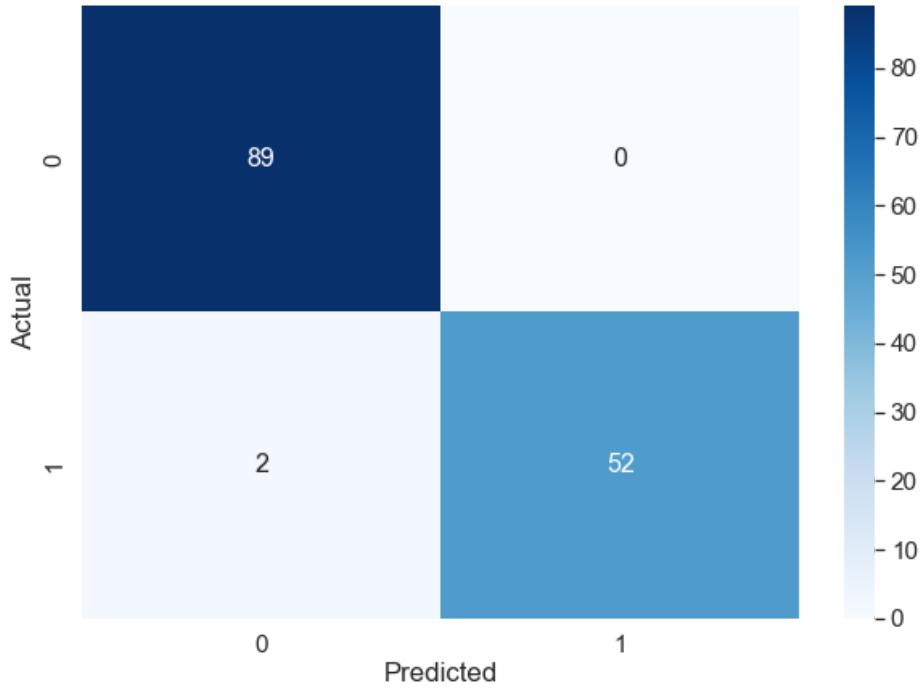
Logistic Regression Model

Pre-processing steps for each model:

- Split data into X = data & y = diagnosis
- Label encoded the y data to change from string (B for benign, M for malignant) to int (0 for benign, 1 for malignant)
- Scaled the X data using StandardScaler
- Split the datasets into training & testing using train_test_split with random_state = 42
- From sklearn.linear_model import LogisticRegression
- Fit the training data to the model
- Testing Data Score: **0.986**
- Made predictions for y values
- Created confusion matrix to compare predictions to actual
- Confusion matrix shows that the model did not overpredict malignancy, and it did well on a low number of False Negatives (2 predicted benign but actually malignant)



Benign = 0 & Malignant = 1

The above confusion matrix visually shows how our model did:

89 True Positives = True Benign Tumors

0 False Positives = Predicted to be Malignant Tumors, but actually Benign Tumors

2 False Negatives = Predicted to be Benign Tumors, but actually Malignant Tumors

52 True Negatives = True Malignant Tumors