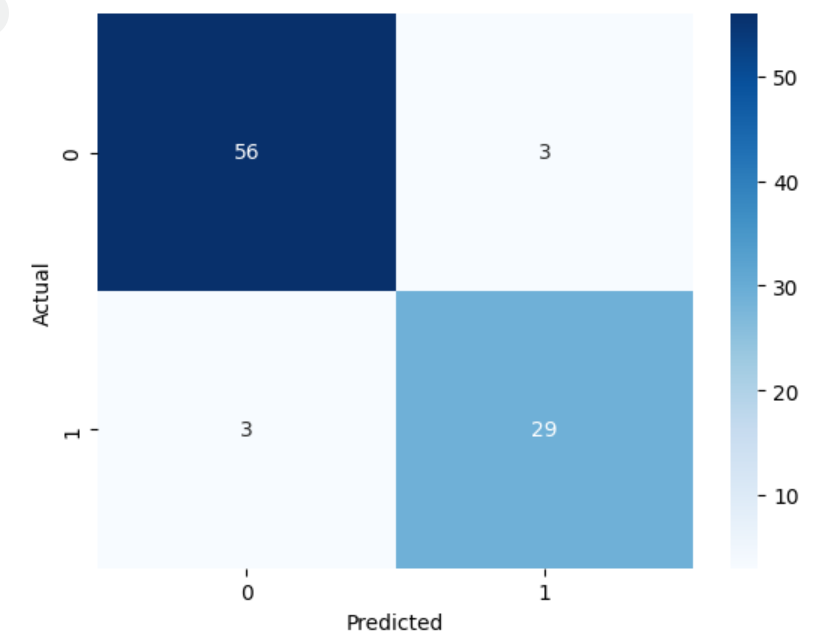
**Objective of the project:**

To distinguish a tumor between it being benign or malignant breast cancer. We want to remove unnecessary data and sift through the useful data to organize the data into these categories.

**Methods of data preprocessing:**

We downloaded all of the modules. We opened the breast cancer data set and reformatted it into a new file to make the B’s and M’s into 0’s and 1’s (binary formatting). We found the target processing. We standardized the features and split them into training tests. We train them used a RandomForestClassifier and make predictions on test set. We used a formula for accuracy and we created a confusion matrix to show how many we identified accurately. Our accuracy rate was 0.93.

* **Model Selection**
  + We tested RandomForestClassifier
  + We chose this because it handles a lot of data and it provides nonlinear relationships
* **Results & Performance Evaluation**
  + Accuracy: 0.93
  + Precision:0.95 for 0’s and 0.91 for 1’s
  + Recall and F1 scores: same as precision
  + Show a confusion matrix and discuss errors/limitations in the model:

Confusion Matrix:

Some limitations are if we had data outliers they wouldn't be shown in the model.

* Conclusion
  + We preprocessed the data to reformat the benign and malignant assignments as binary and used a RandomForestClassifier to handle the high dimensional data and train the model that was then used as a tester.
  + Limitations: the classifier is generic compared to other models so it might have overlooked details/underlying patterns.

Citations for code: ChatGPT