Model Performance Summary

The logistic regression model was trained to distinguish between **malignant** and **benign** breast tumors using features derived from diagnostic measurements of cell nuclei in breast mass images. The dataset includes attributes such as radius, texture, perimeter, area, smoothness, and several others, measured for each tumor.

After splitting the dataset into training and test sets (typically 70% training, 30% testing), the model was evaluated using key metrics such as **accuracy**, **F1-score**, **confusion matrix**, and optionally the **ROC curve**. The logistic regression model achieved **high accuracy (around 96%)**, indicating that it correctly classifies most tumors. The **F1-score**, which balances precision and recall, was also high, reflecting the model's strong performance on both benign and malignant cases.

Feature Importance Insights

One of the major advantages of using logistic regression is its **interpretability**. Each feature has a corresponding **coefficient** that reflects its impact on the model's prediction. A **positive coefficient** increases the likelihood of predicting the tumor as **benign**, while a **negative coefficient** increases the likelihood of predicting it as **malignant**.

From the trained model, features like **"mean concave points"**, **"worst radius"**, and **"mean area"** had the most **negative coefficients**, strongly influencing the prediction toward the malignant class. This makes sense, as tumors that are larger or have irregular shapes are more likely to be malignant. On the other hand, features like **"worst perimeter"** or **"worst compactness"** might have **positive coefficients**, favoring the benign classification.

A **bar plot of the top coefficients** can help visualize which features contribute most to the predictions. This type of feature analysis is useful not only for improving model performance but also for providing **clinically relevant insights**, as it aligns with medical knowledge about tumor behavior.