

## Task 3: Predictive Analytics for Resource Allocation

**Dataset:** Kaggle Breast Cancer Wisconsin Dataset

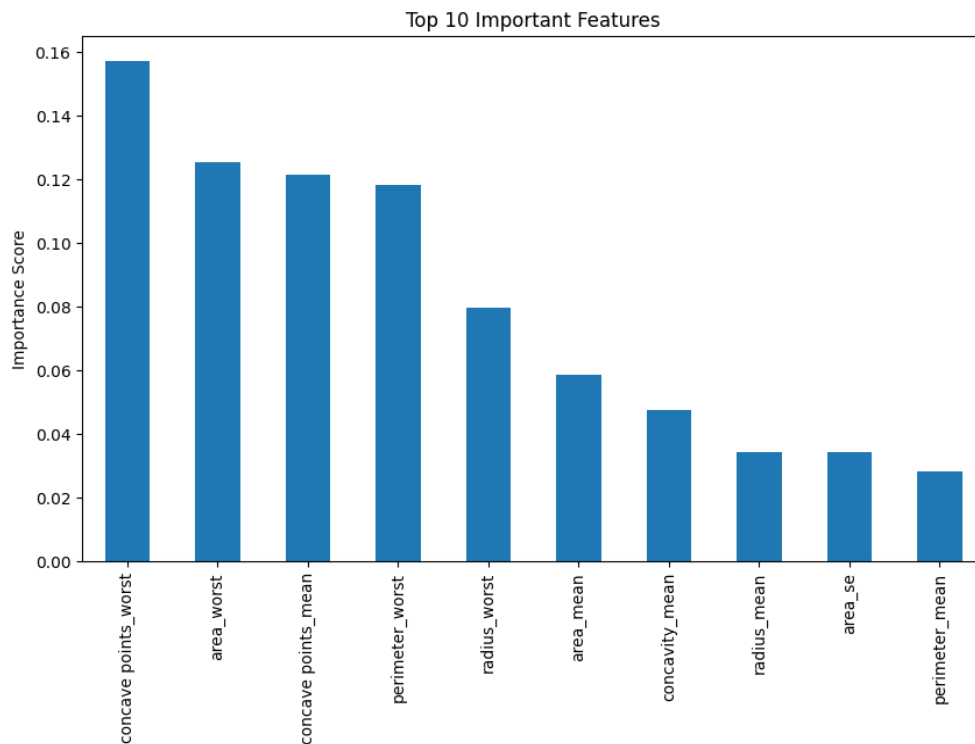
**Goal:** Train a Random Forest model to predict tumor diagnosis (malignant/benign) using preprocessed data.

**Steps:**

1. Data preprocessing (cleaning, labeling, and splitting).
2. Model training using *RandomForestClassifier*.
3. Evaluation using Accuracy and F1-score.
4. Feature importance visualization.

**Results:**

- Accuracy: 96.5%
- F1-score: 0.96
- Top predictive features: concave points\_worst, area\_worst, concave points\_mean.



### Part 3: Ethical Reflection

When deploying the predictive breast cancer model, ethical considerations become essential. The dataset used may contain biases — for example, if it overrepresents patients from certain age groups, ethnic backgrounds, or healthcare institutions, the model may perform better on those groups and worse on underrepresented ones. This could lead to unfair or inaccurate predictions, especially in diverse real-world populations. Furthermore, the historical data may embed diagnostic biases from human experts, such as misclassification of rare cases.

To address these issues, fairness tools like IBM AI Fairness 360 (AIF360) can be integrated during model evaluation. AIF360 provides metrics to detect bias (such as disparate impact or statistical parity) and algorithms to mitigate it, including reweighting or resampling methods. By applying such

tools, developers can ensure that predictions remain equitable and transparent, enhancing trust and accountability in AI-driven healthcare systems.