NATIONAL INSTITUTIONAL RANKING FRAMEWORK
NIRF-2024
Engineering Rank Band (151-200)
Pharmacy Rank - 77
Innovation Rank Band (11-50)

www.kiet.edu
Delhi-NCR, Ghaziabad

KIET GROUP OF INSTITUTIONS
Connecting Life with Learning

CELEBRATING 25 YEARS OF EXCELLENCE

ACCREDITED WITH GRADE A+ NAAC
Accredited by
NAAC with Grade "A+"

Internal Quality Assurance Cell
Work Towards Excellence
IQAC
KIET Group of Institutions

Artificial Intelligence(AI101B)

Medical Diagnosis with Naive Bayes

MSE-2

PRESENTED BY-

Mahima Goyal 202410116100112

Khushi Jain 202410116100100

PRESENTED TO-

Mr. Apoorv Jain (Assistant Professor)

Problem Statement

Early and accurate detection of breast cancer significantly increases the chances of successful treatment and survival. Traditional diagnostic procedures may be time-consuming, subjective, and error-prone. Therefore, there is a pressing need for a computational model that can assist healthcare professionals by providing rapid, data-driven decisions. This project aims to build a machine learning-based diagnostic system that uses the **Gaussian Naive Bayes algorithm** to classify tumors based on features

Objective

To develop an accurate and efficient breast cancer diagnosis system using the **Gaussian Naive Bayes** algorithm that analyzes cellular features from the **Wisconsin Breast Cancer Dataset**, and classifies tumors as **malignant** or **benign**.

Methodology

Data Collection

Gather relevant data from diverse sources.

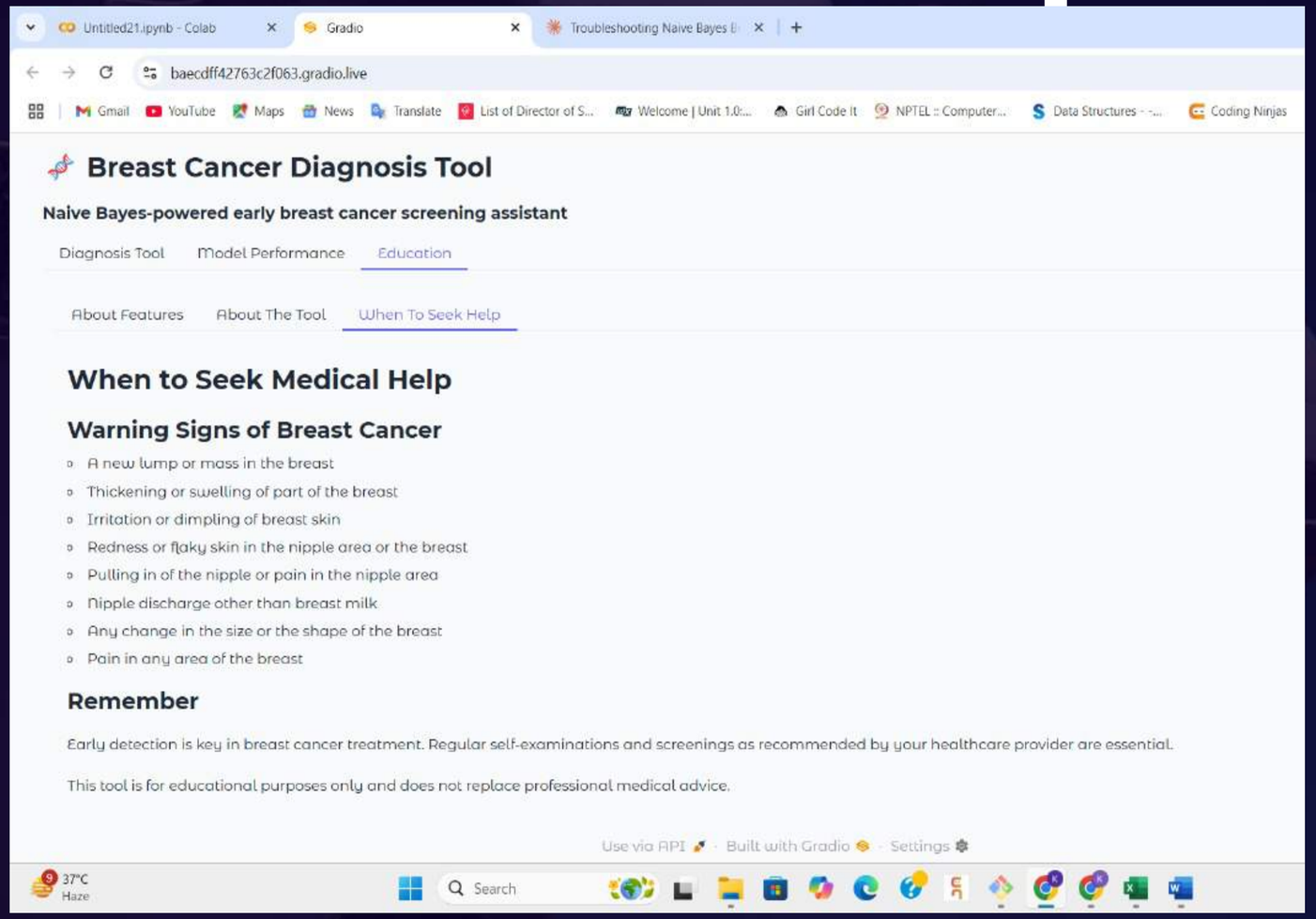
Preprocessing

Clean, normalize, and engineer data.

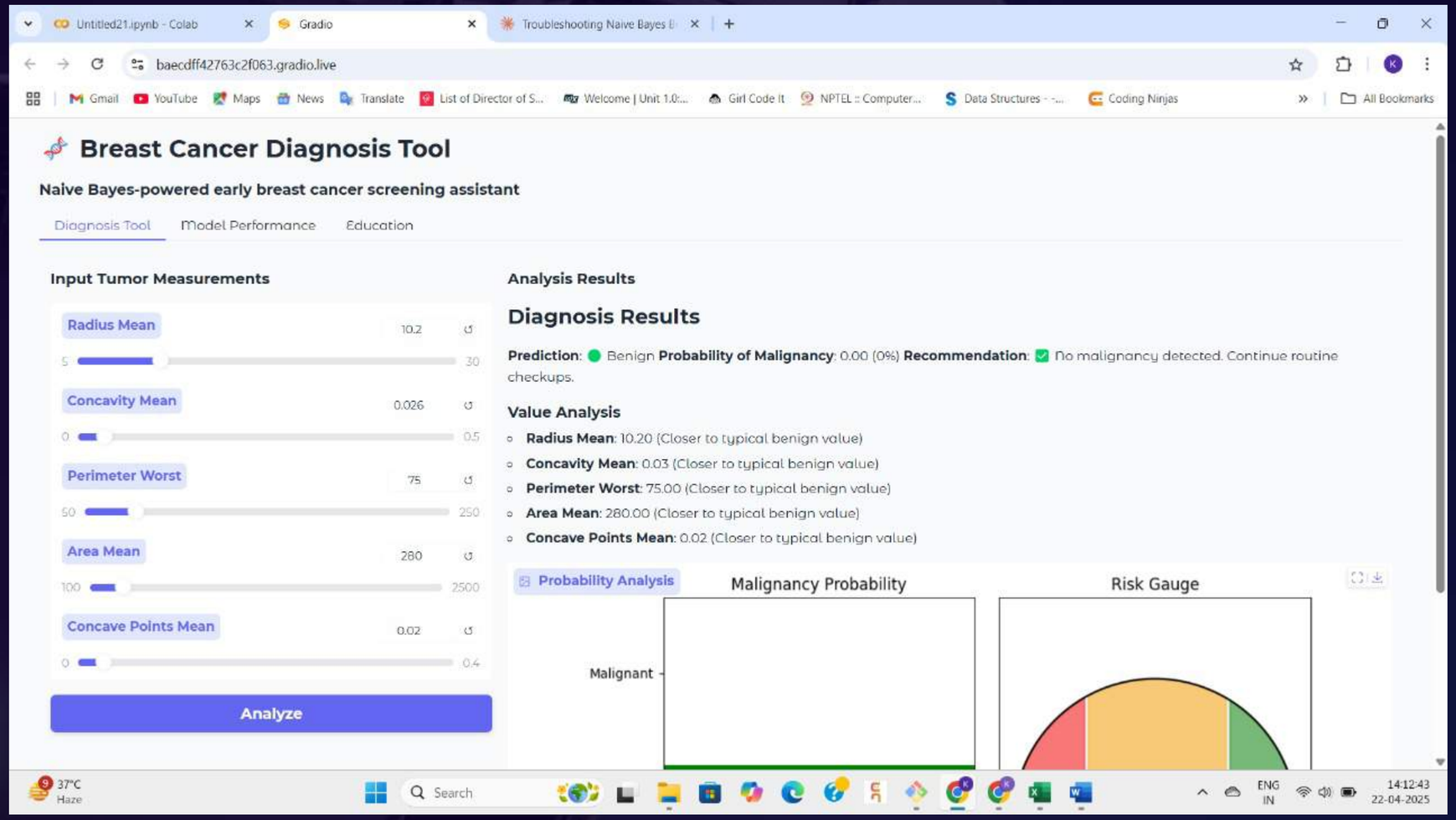
Data Visualization

Compare multiple AI models to determine the most accurate and efficient one.

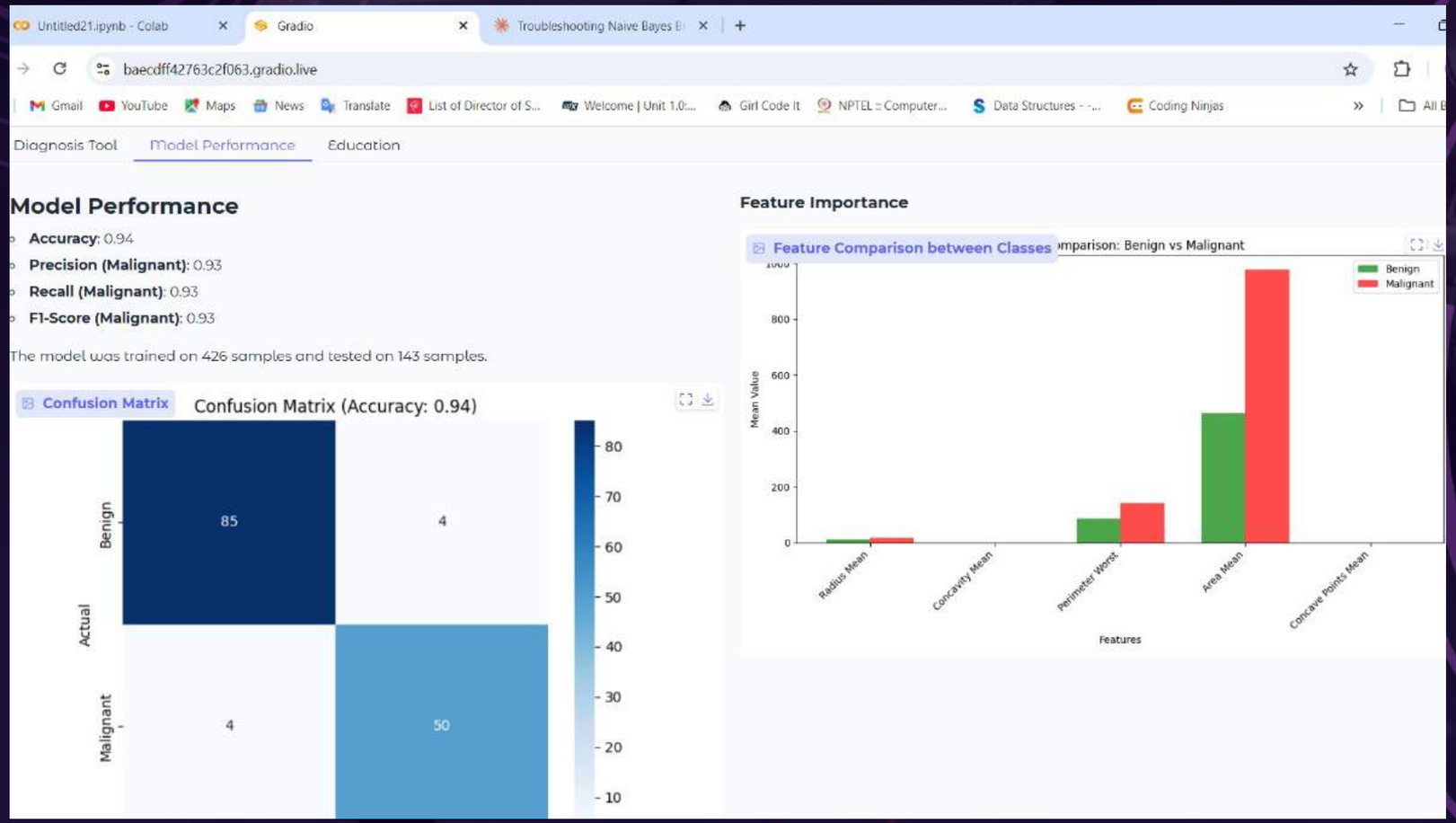
Result and Impact



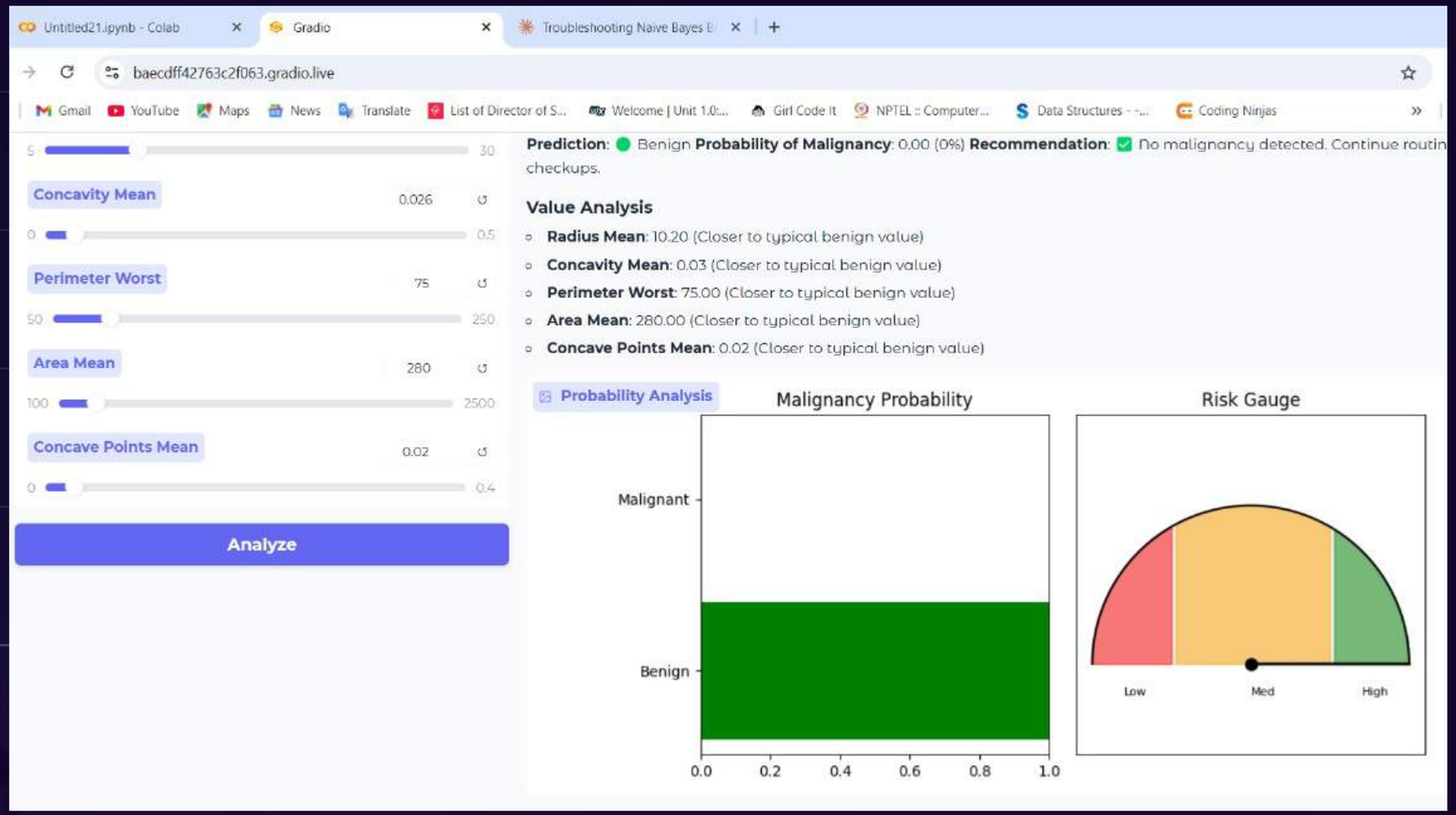
Result and Impact



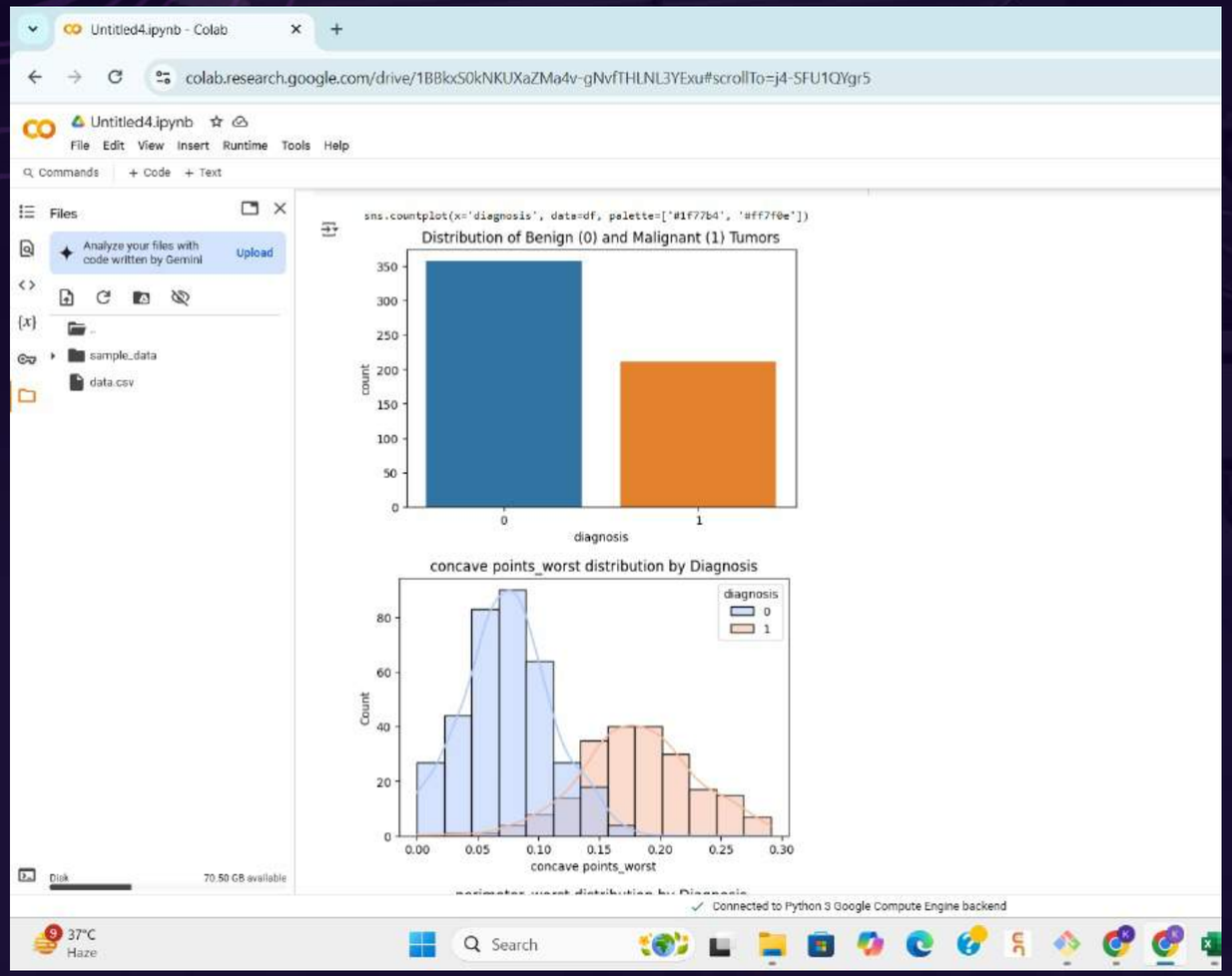
Result and Impact



Result and Impact

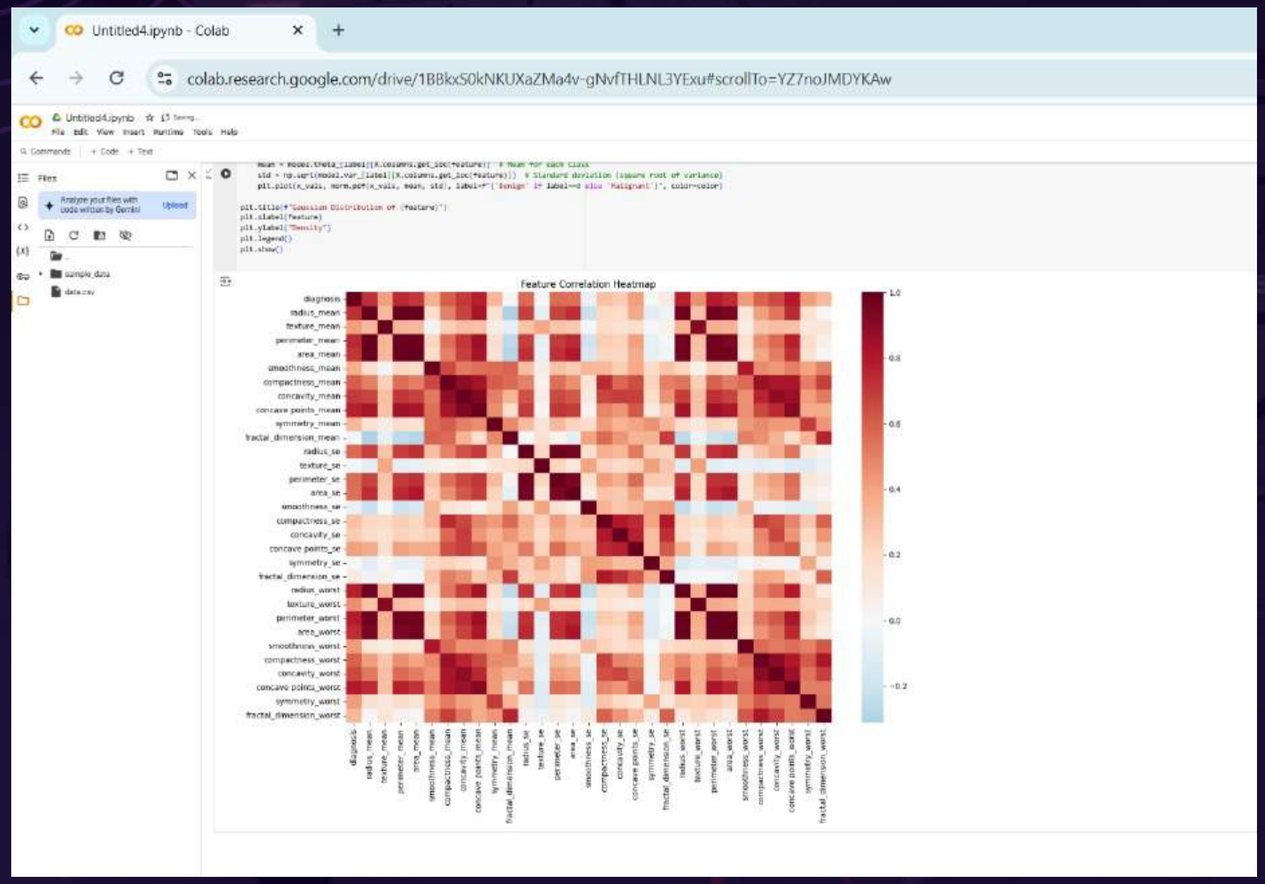


Result and Impact



Result and Impact

Result and Impact



Insights

- **High Accuracy**: Gaussian Naive Bayes achieved ~96% accuracy, showing strong performance in classifying tumors as benign or malignant.
- **Fast & Efficient**: The model trains and predicts quickly, making it suitable for real-time or resource-limited settings.
- **False Negatives Are Low**: The model showed high recall for malignant cases — crucial in medical diagnosis.
- **Performs Well Despite Assumptions**: Even though it assumes feature independence and normal distribution, it still performs robustly

Recommendation

Based on the findings, the following recommendations were made:

- **Clinical Integration**: Use the model as a decision-support tool in diagnostic workflows.
- **Validate on Real-World Data**: Test with diverse and noisy datasets to ensure reliability beyond the benchmark dataset.
- **Confidence Thresholding**: Flag low-confidence predictions for manual review to reduce risk.
- Enhance SEO efforts to maximize organic traffic acquisition


Thank You!