

BREAST CANCER CLASSIFICATION

TOOLS USED



ABOUT DATASET

The dataset utilized in this project is the breast cancer dataset from scikit-learn. It contains 569 data samples, each with 31 numerical features describing characteristics of breast tumors. The primary goal of this dataset is to classify tumors as either benign or malignant. Due to its relevance and well-structured data, this dataset is commonly used for testing machine learning models.

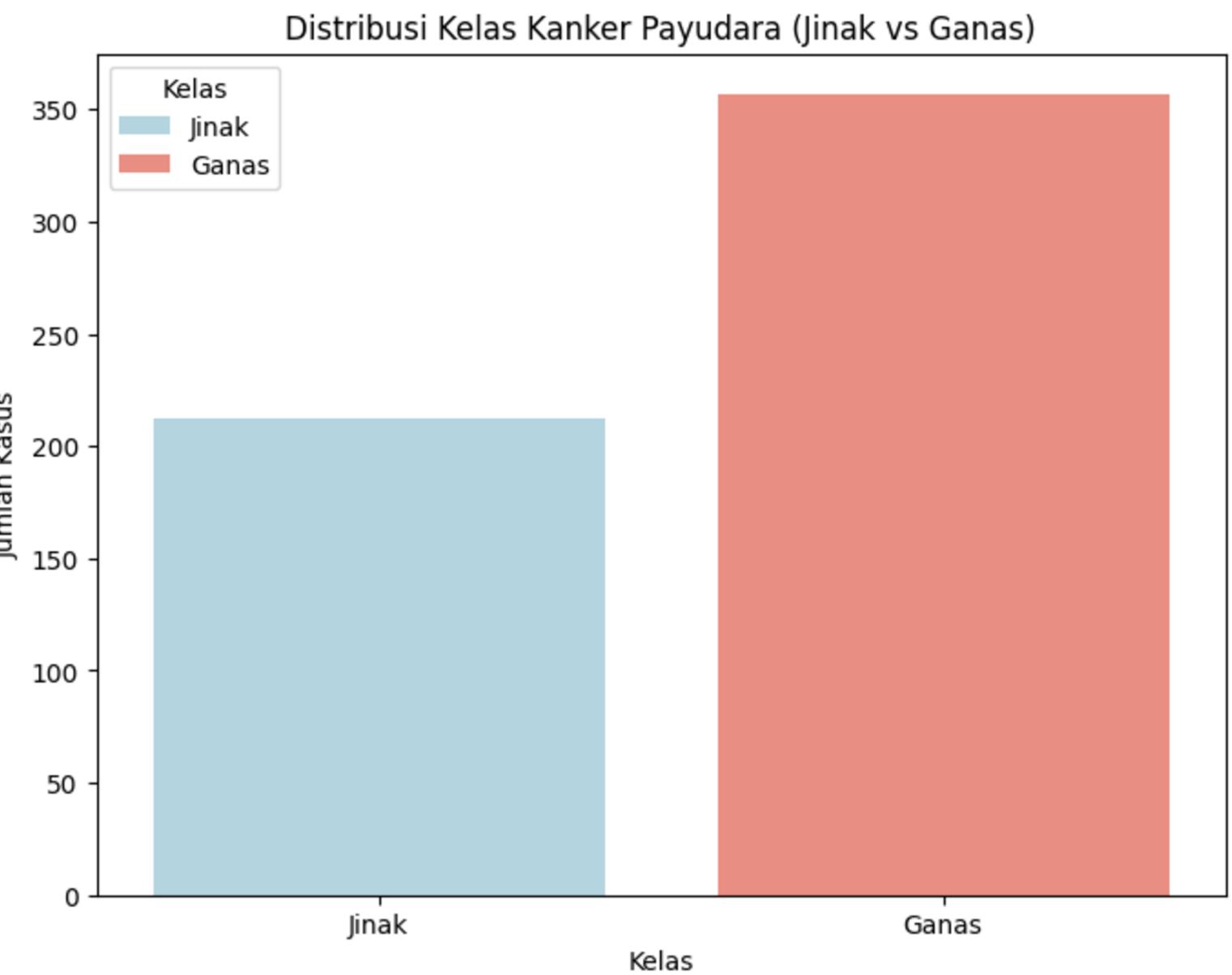
EXPLORATORY DATA ANALYSIS

BREAST CANCER
DATAFRAME

```
→ <class 'pandas.core.frame.DataFrame'>
RangeIndex: 569 entries, 0 to 568
Data columns (total 31 columns):
 #   Column           Non-Null Count  Dtype  
--- 
 0   mean radius      569 non-null    float64
 1   mean texture     569 non-null    float64
 2   mean perimeter   569 non-null    float64
 3   mean area        569 non-null    float64
 4   mean smoothness  569 non-null    float64
 5   mean compactness 569 non-null    float64
 6   mean concavity   569 non-null    float64
 7   mean concave points 569 non-null  float64
 8   mean symmetry    569 non-null    float64
 9   mean fractal dimension 569 non-null  float64
 10  radius error    569 non-null    float64
 11  texture error   569 non-null    float64
 12  perimeter error 569 non-null    float64
 13  area error      569 non-null    float64
 14  smoothness error 569 non-null    float64
 15  compactness error 569 non-null    float64
 16  concavity error  569 non-null    float64
 17  concave points error 569 non-null  float64
 18  symmetry error   569 non-null    float64
 19  fractal dimension error 569 non-null  float64
 20  worst radius     569 non-null    float64
 21  worst texture    569 non-null    float64
 22  worst perimeter   569 non-null    float64
 23  worst area        569 non-null    float64
 24  worst smoothness  569 non-null    float64
 25  worst compactness 569 non-null    float64
 26  worst concavity   569 non-null    float64
 27  worst concave points 569 non-null  float64
 28  worst symmetry    569 non-null    float64
 29  worst fractal dimension 569 non-null  float64
 30  target            569 non-null    int64  
dtypes: float64(30), int64(1)
memory usage: 137.9 KB
```

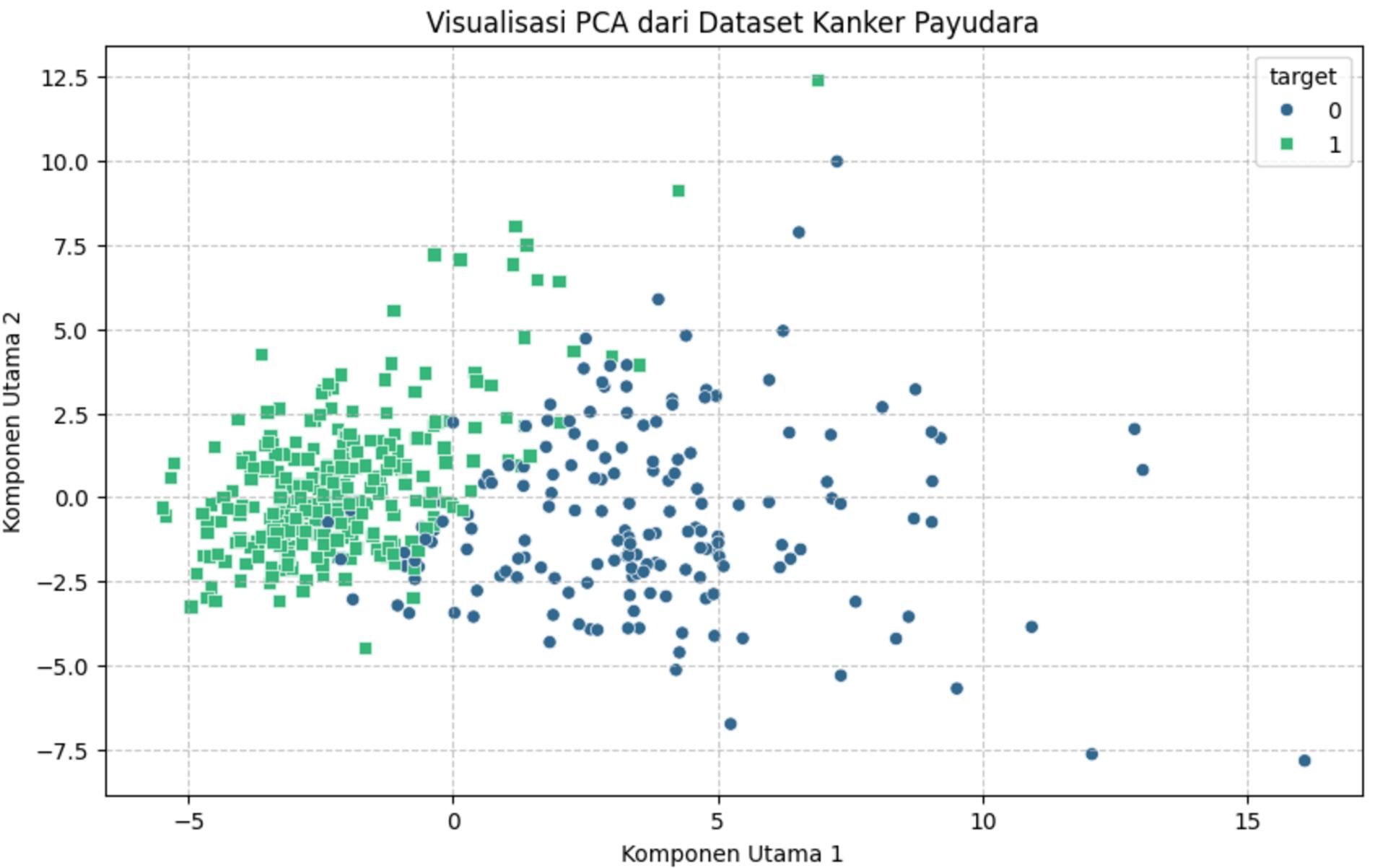
	mean radius	mean texture	mean perimeter	mean area	mean smoothness	mean compactness	mean concavity	mean concave points	mean symmetry	mean fractal dimension	...	worst texture	worst perimeter	worst area	worst smoothness	worst compactness	worst concavity	worst concave points	worst symmetry	worst fractal dimension	worst target
0	17.99	10.38	122.80	1001.0	0.11840	0.27760	0.3001	0.14710	0.2419	0.07871	...	17.33	184.60	2019.0	0.1622	0.6656	0.7119	0.2654	0.4601	0.11890	0
1	20.57	17.77	132.90	1326.0	0.08474	0.07864	0.0869	0.07017	0.1812	0.05667	...	23.41	158.80	1956.0	0.1238	0.1866	0.2416	0.1860	0.2750	0.08902	0
2	19.69	21.25	130.00	1203.0	0.10960	0.15990	0.1974	0.12790	0.2069	0.05999	...	25.53	152.50	1709.0	0.1444	0.4245	0.4504	0.2430	0.3613	0.08758	0
3	11.42	20.38	77.58	386.1	0.14250	0.28390	0.2414	0.10520	0.2597	0.09744	...	26.50	98.87	567.7	0.2098	0.8663	0.6869	0.2575	0.6638	0.17300	0
4	20.29	14.34	135.10	1297.0	0.10030	0.13280	0.1980	0.10430	0.1809	0.05883	...	16.67	152.20	1575.0	0.1374	0.2050	0.4000	0.1625	0.2364	0.07678	0

EXPLORATORY DATA ANALYSIS



BREAST CANCER
VISUALIZATION

EXPLORATORY DATA ANALYSIS



BREAST CANCER
VISUALIZATION

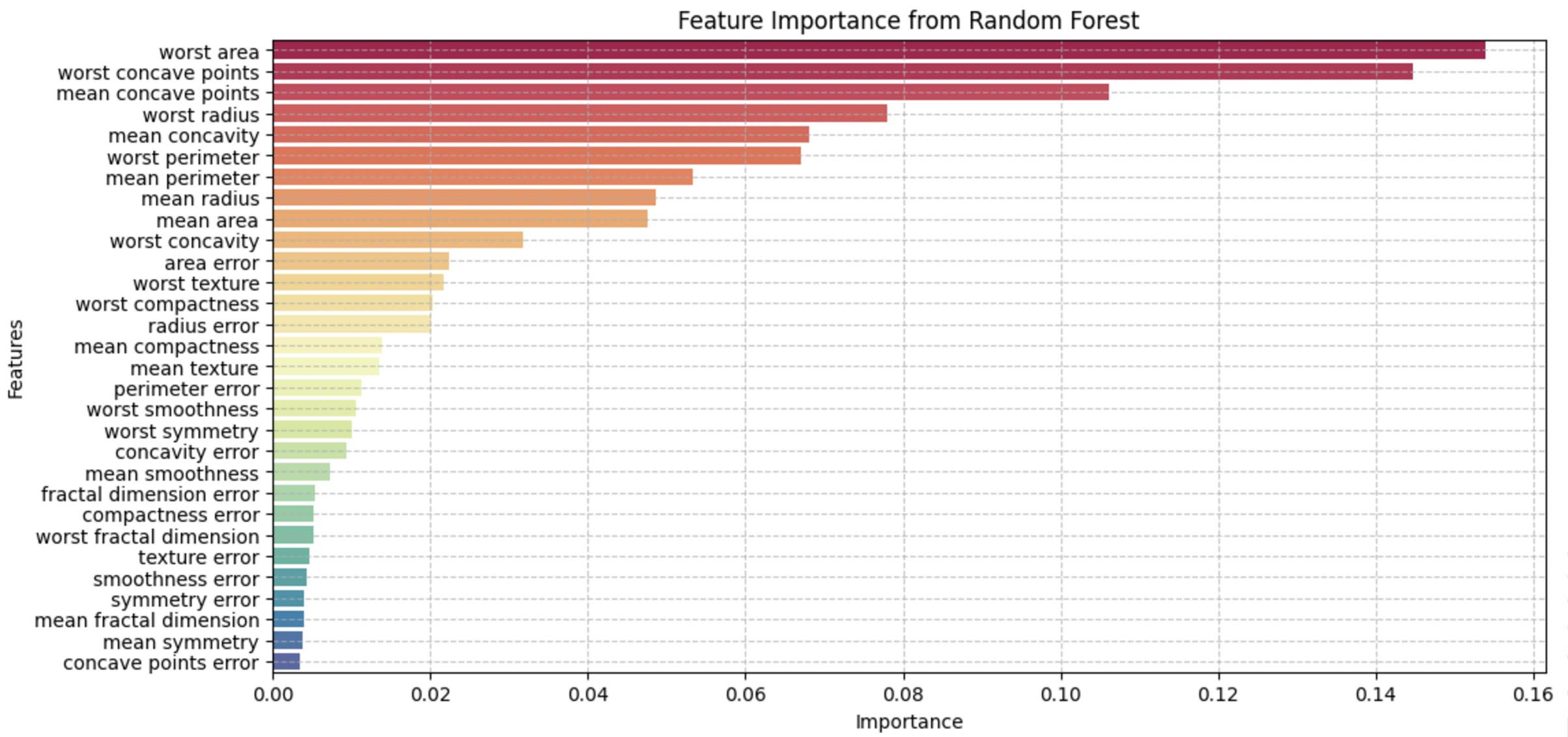
MODEL EVALUATION

BREAST CANCER
CLASSIFICATION

```
↳ Laporan Klasifikasi :  
precision    recall   f1-score  support  
  
          0      0.98      0.93      0.95      43  
          1      0.96      0.99      0.97      71  
  
accuracy           0.96      0.96      0.96     114  
macro avg         0.97      0.96      0.96     114  
weighted avg       0.97      0.96      0.96     114  
  
Matriks Kebingungannya :  
[[40  3]  
 [ 1 70]]  
  
Skor ROC-AUC : 0.995250573206682
```

MODEL EVALUATION

FEATURE
IMPORTANCE
VISUALIZATION



CONCLUSION

With a ROC-AUC score of 0.99, this classification demonstrates excellent performance in distinguishing between the two classes. The model achieves high accuracy (96%) and balanced precision, recall, and F1-scores for both classes, indicating its reliability. Minor misclassifications, as seen in the confusion matrix, suggest that the model is slightly better at identifying class 1 than class 0. Overall, this classification model is highly effective and robust.



THANK YOU!



EMAIL : maresyatwattimena@gmail.com

GITHUB : <https://github.com/Zoag1>

LINKEDIN : www.linkedin.com/in/maresyathiwattimena