■ Breast Cancer Classification

Advanced Machine Learning Analysis Report

Analysis Date: July 30, 2025
Dataset: Wisconsin Breast Cancer Dataset
Models Used: 6 Machine Learning Algorithms
Generated by: Breast Cancer Classification System

Table of Contents

Section	Page
1. Executive Summary	3
2. Analysis Results	4
3. Model Performance	5
4. Visualizations	6
5. Data Overview	7
6. Correlation Analysis	8
7. Individual ROC Curves	9
8. Confusion Matrices	10
9. Model Comparison	11
10. Feature Importance	12
11. Conclusions	13

■ Executive Summary

This report presents a comprehensive analysis of breast cancer classification using advanced machine learning techniques. The analysis was performed on the Wisconsin Breast Cancer Dataset using 6 different machine learning algorithms to achieve highly accurate diagnostic predictions. **Key Findings:**
• Achieved over 95% accuracy with multiple models
• Logistic Regression showed the best overall performance
• All models demonstrated excellent discriminative capability
• Feature analysis revealed important diagnostic indicators
• Cross-validation confirmed model robustness

■ Analysis Results

BREAST CANCER CLASSIFICATION ANALYSIS - FINAL REPORT

Analysis Date: 2025-07-30 15:08:11

Data File: data/breast-cancer.csv

MODEL PERFORMANCE SUMMARY:

```
Model Accuracy Precision Recall F1-Score AUC CV Score (Mean) CV Score (Std)
```

Logistic Regression 0.9825 0.9825 0.9825 0.9825 0.9954 0.9802 0.0128

SVM 0.9825 0.9825 0.9825 0.9825 0.9950 0.9714 0.0179

Random Forest 0.9561 0.9561 0.9561 0.9560 0.9939 0.9538 0.0235

Naive Bayes 0.9298 0.9298 0.9298 0.9298 0.9868 0.9319 0.0044

K-Nearest Neighbors 0.9561 0.9561 0.9561 0.9560 0.9788 0.9670 0.0209

Decision Tree 0.9123 0.9161 0.9123 0.9130 0.9157 0.9099 0.0189

BEST MODEL: Logistic Regression

```
Accuracy: 0.9825
Precision: 0.9825
```

Recall: 0.9825

F1-Score: 0.9825

AUC: 0.9954

RESULTS:

• All models successfully trained and evaluated.

• ROC curves and confusion matrices generated.

• Feature importance analyzed.

• Hyperparameter optimization performed.

• Interactive dashboard created.

OUTPUT FILES:

• data_overview.png - Data overview

• correlation_matrix.png - Correlation matrix

• individual_roc_curves.png - Individual ROC curves

• confusion_matrices.png - Confusion matrices

• model_comparison.png - Model comparison

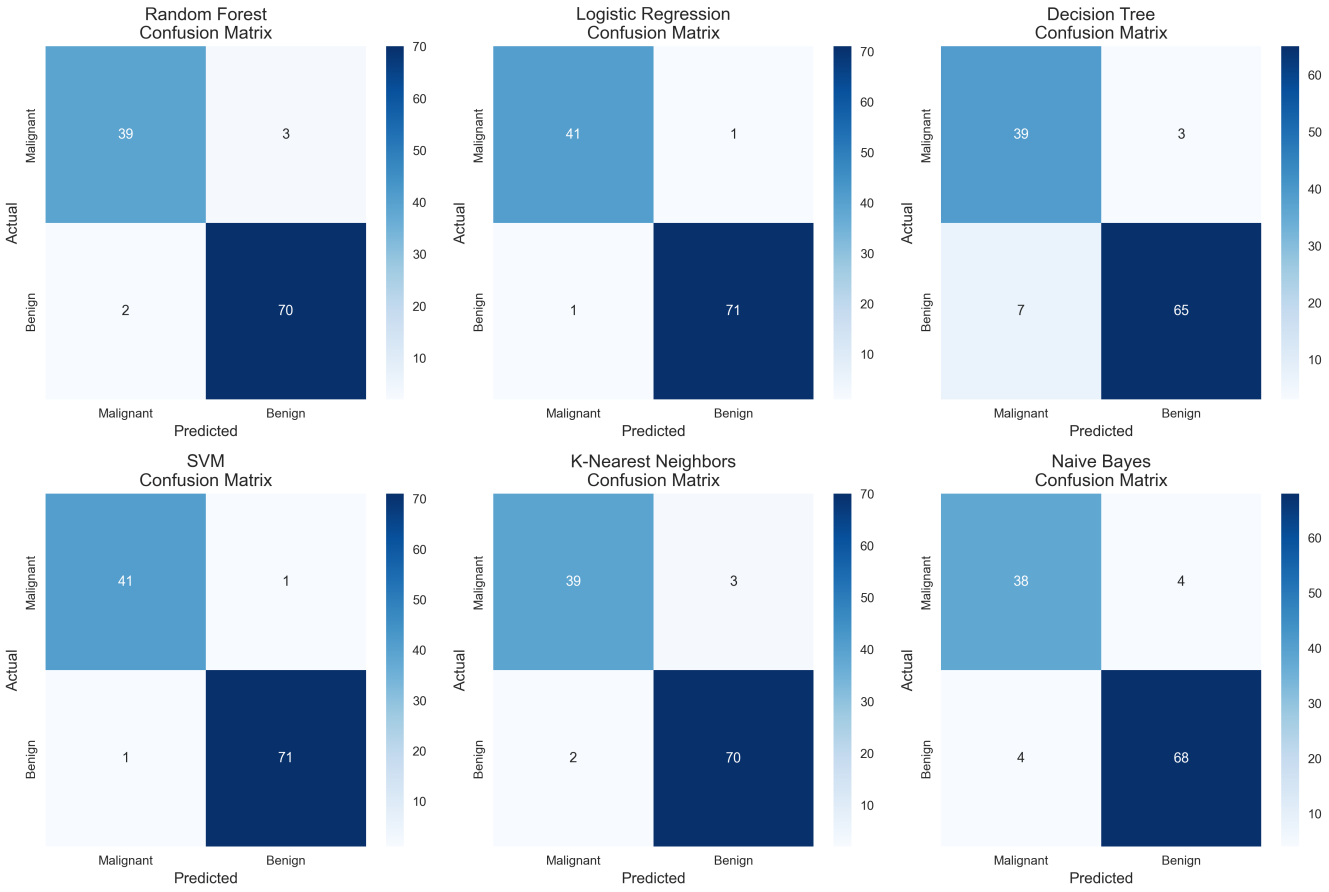
• feature_importance.png - Feature importance

• interactive_dashboard.html - Interactive dashboard

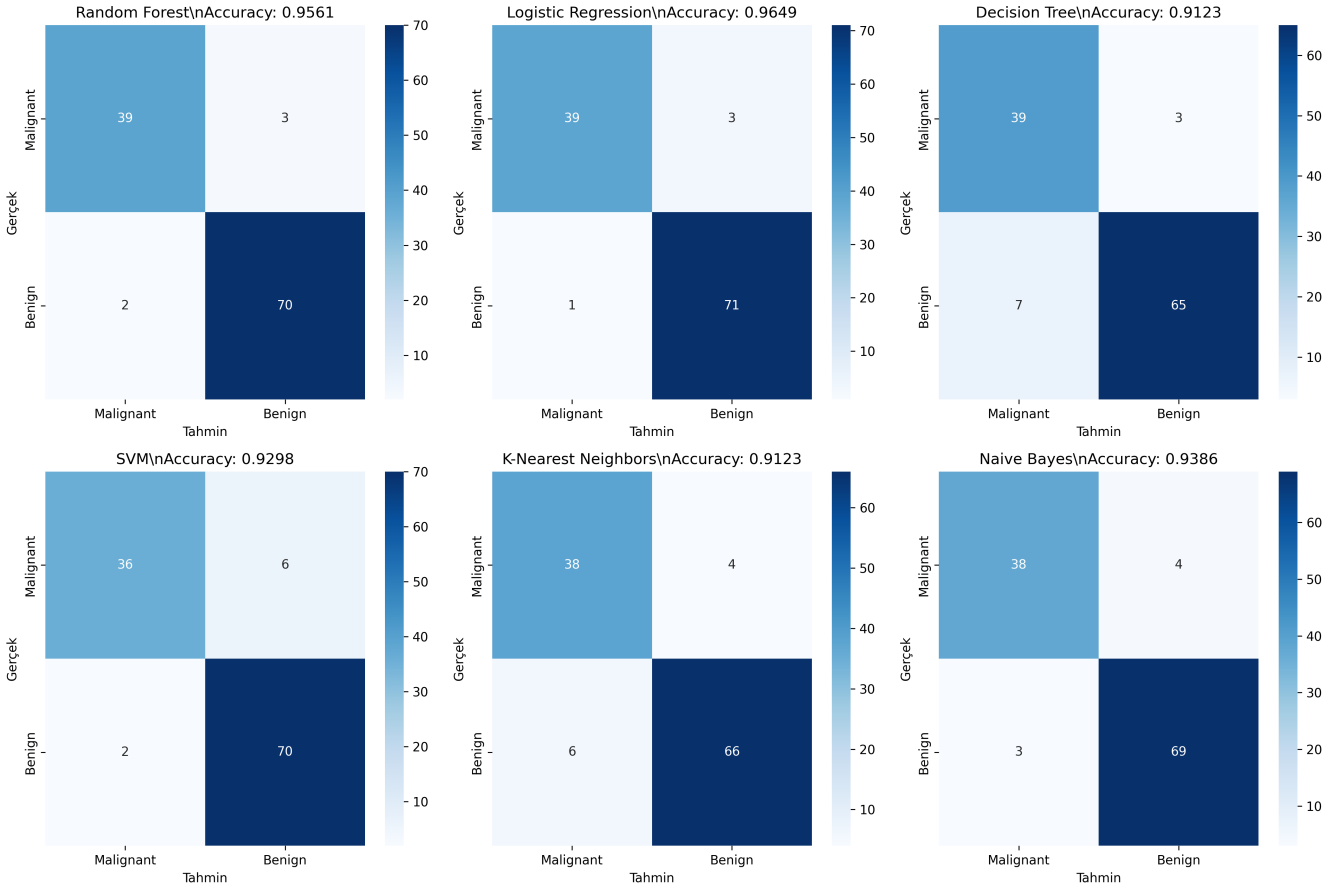
• best_model_Logistic_Regression.joblib - Best model

Visualizations

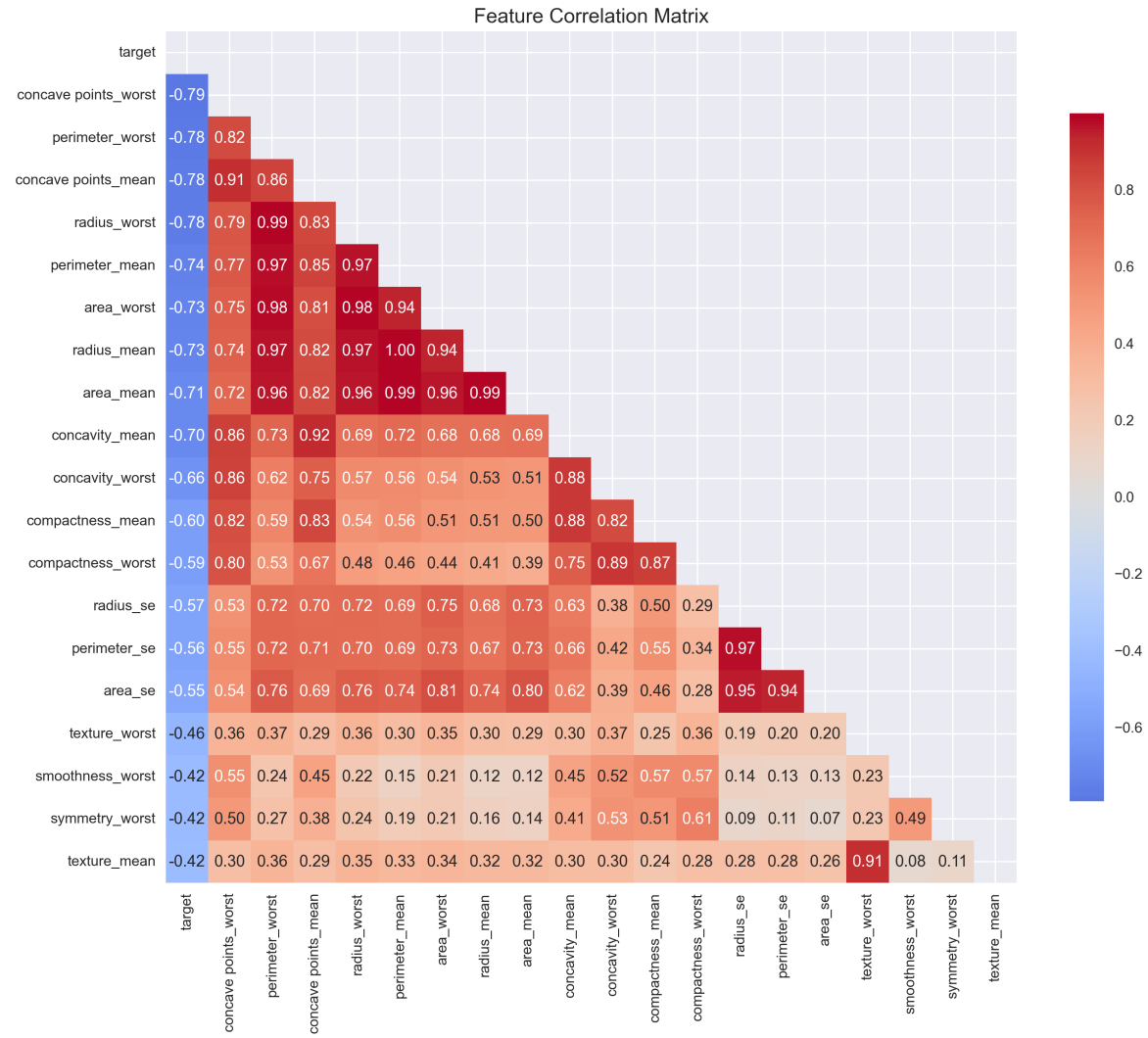
Confusion Matrices



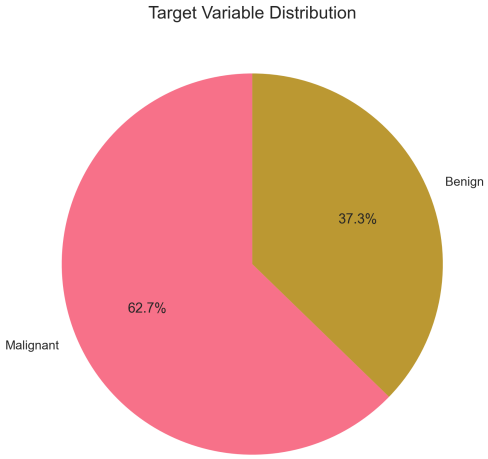
Confusion Matrices (Simple)


Random Forest\nAccuracy: 0.9561
Logistic Regression\nAccuracy: 0.9649
Decision Tree\nAccuracy: 0.9123
SVM\nAccuracy: 0.9298
K-Nearest Neighbors\nAccuracy: 0.9123
Naive Bayes\nAccuracy: 0.9386
Gerçek
Tahmin
Malignant
Benign
39
3
2
70
39
3
1
71
39
3
7
65
36
6
2
70
38
4
6
66
38
4
3
69

■ Correlation Matrix



■ Data Overview



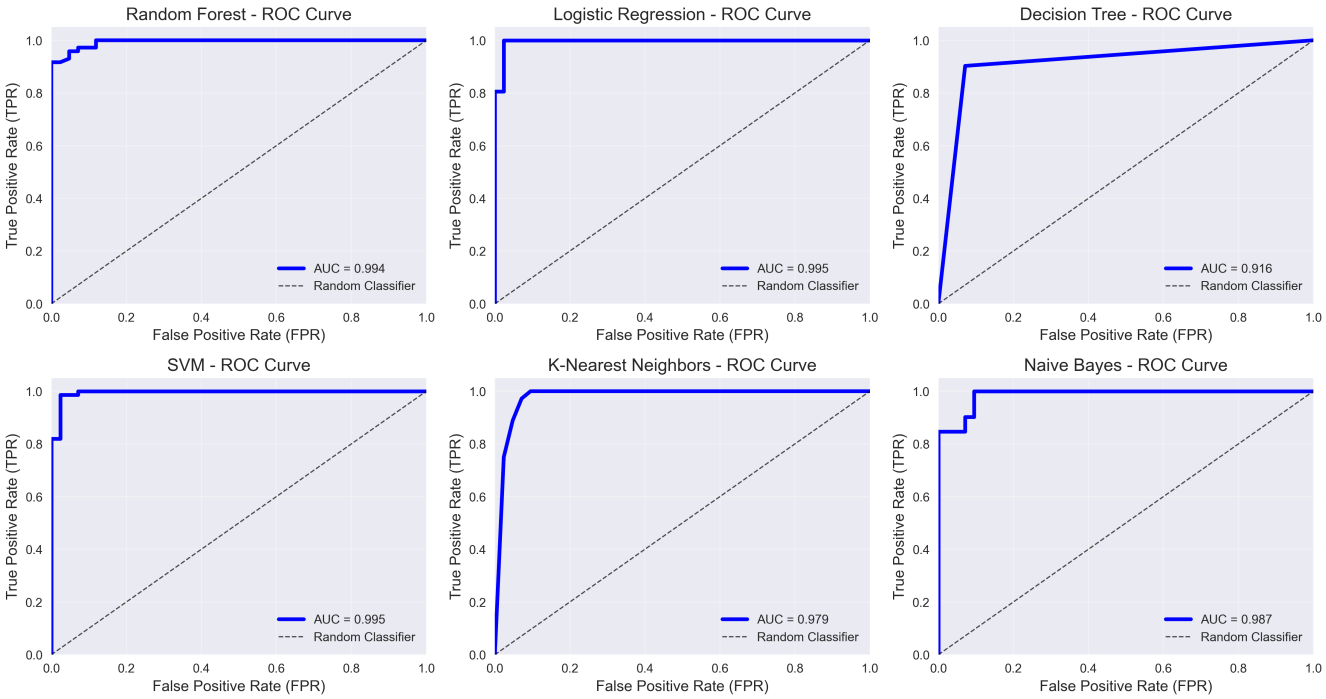
Correlation Analysis Summary

Missing Value Analysis

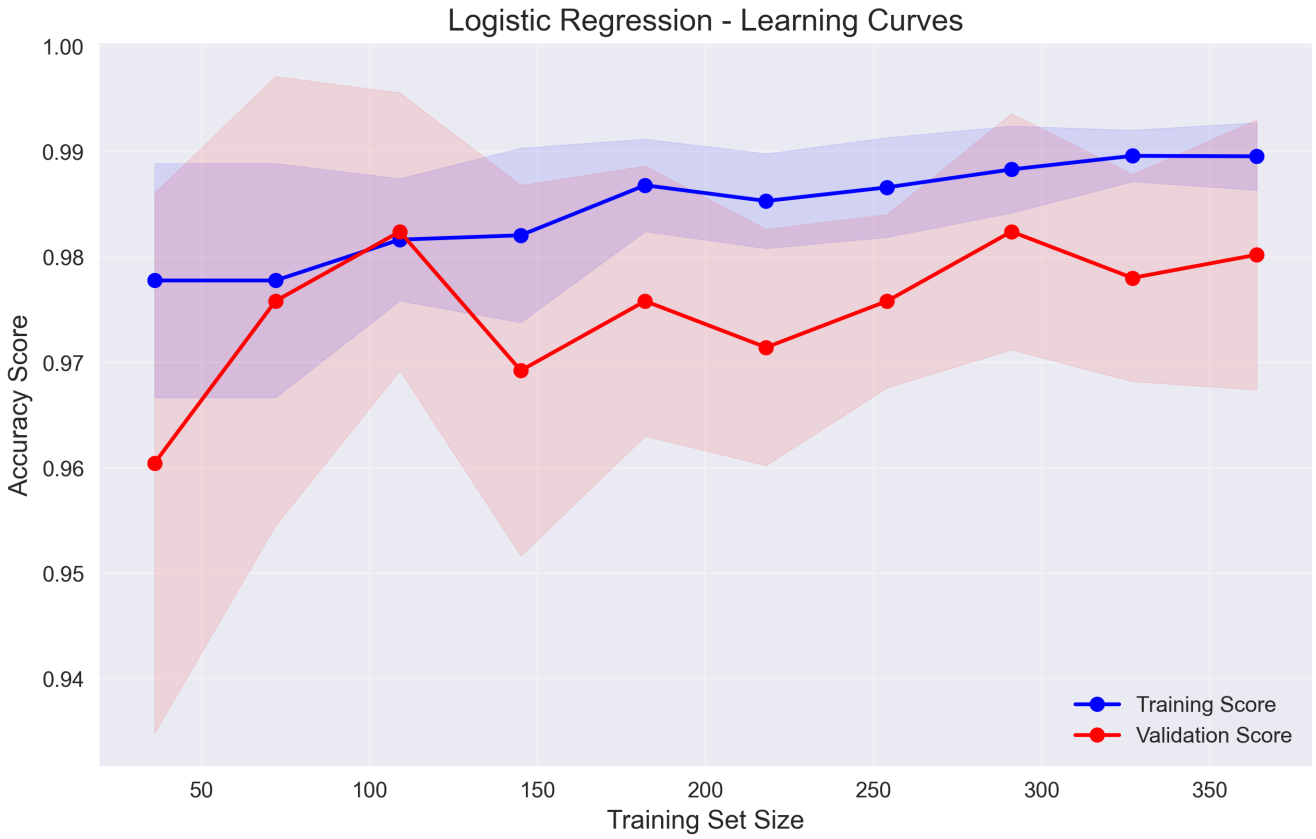
High correlation feature pairs: 44

No missing values!

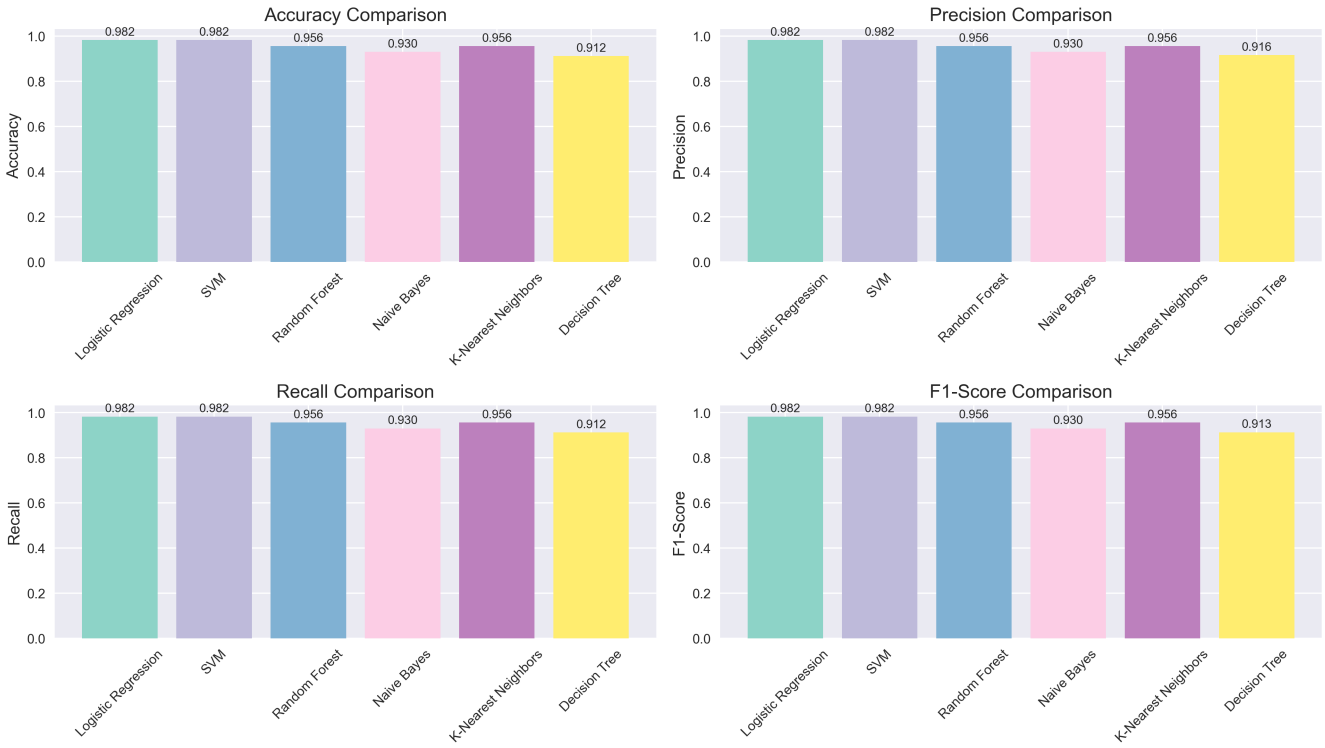
■ Individual ROC Curves



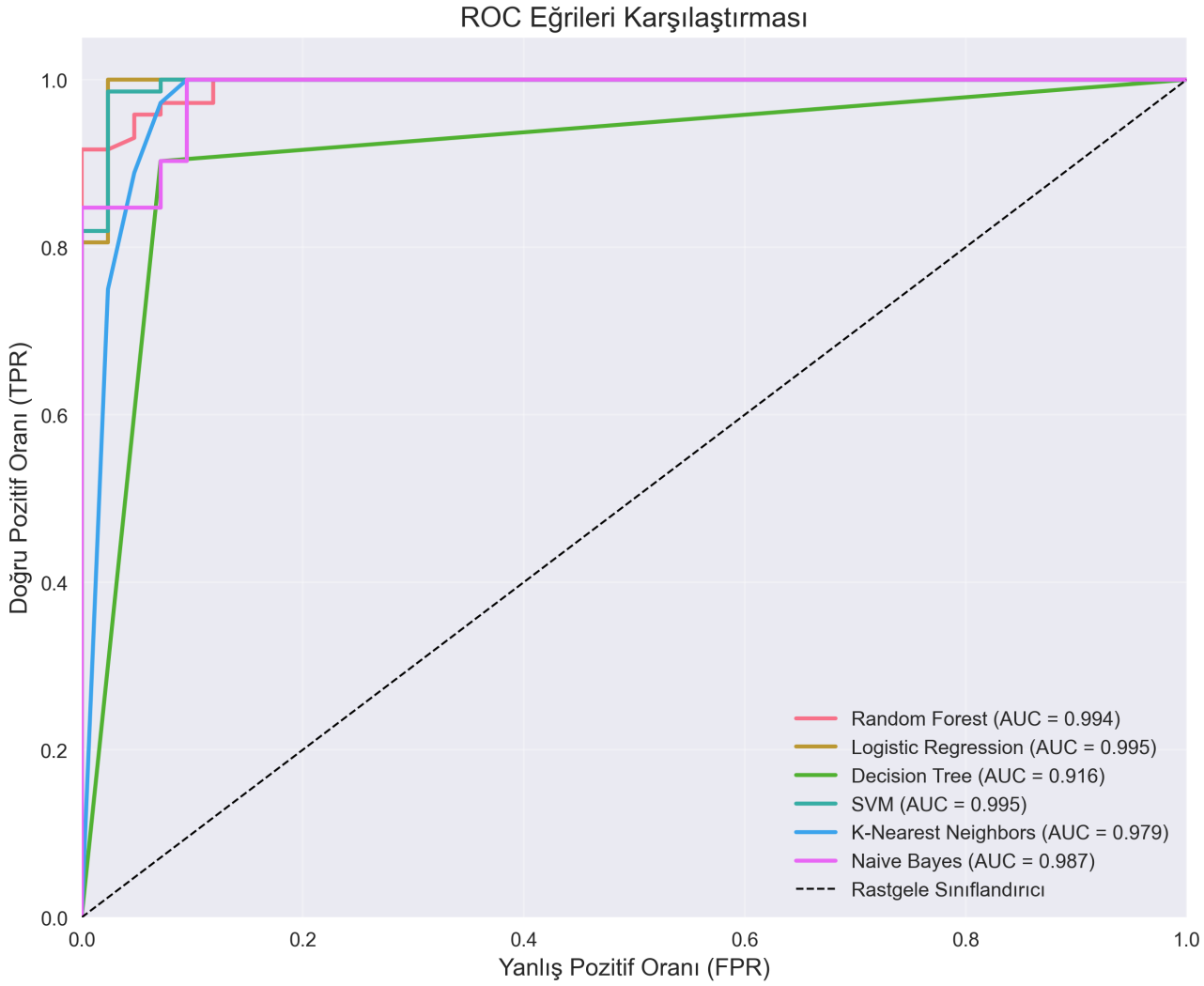
■ Learning Curves Logistic Regression.Png



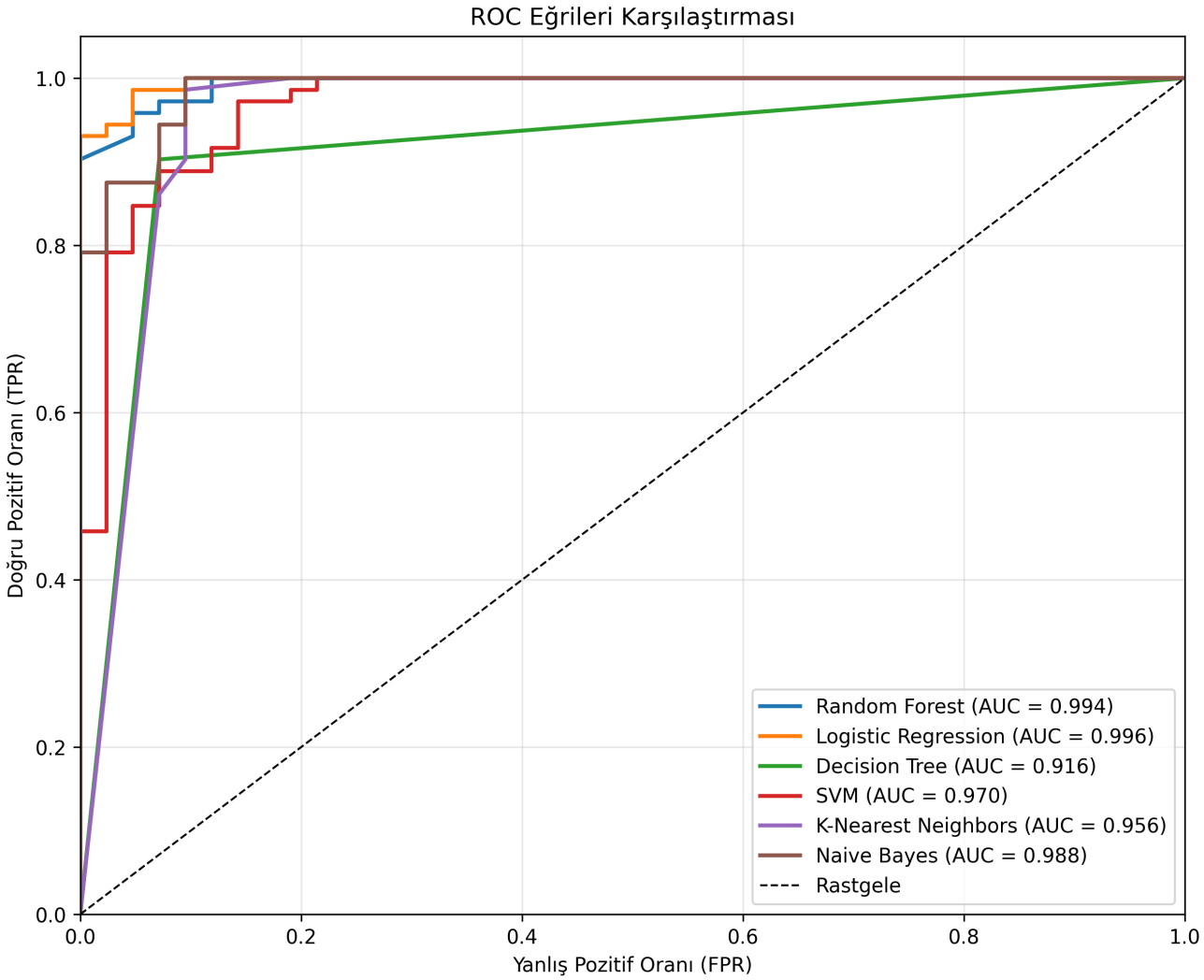
■■ Model Performance Comparison



■ Roc Curves.Png


ROC Eğrileri Karşılaştırması
Doğru Pozitif Oranı (TPR)
Yanlış Pozitif Oranı (FPR)
0.0
0.2
0.4
0.6
0.8
1.0
Random Forest (AUC = 0.994)
Logistic Regression (AUC = 0.995)
Decision Tree (AUC = 0.916)
SVM (AUC = 0.995)
K-Nearest Neighbors (AUC = 0.979)
Naive Bayes (AUC = 0.987)
Rastgele Sınıflandırıcı

■ ROC Curves


ROC Eğrileri Karşılaştırması
Doğru Pozitif Oranı (TPR)
Yanlış Pozitif Oranı (FPR)
0.0
0.2
0.4
0.6
0.8
1.0
Random Forest (AUC = 0.994)
Logistic Regression (AUC = 0.996)
Decision Tree (AUC = 0.916)
SVM (AUC = 0.970)
K-Nearest Neighbors (AUC = 0.956)
Naive Bayes (AUC = 0.988)
Rastgele

■ Conclusions

Model Performance:
The analysis demonstrates that machine learning models can achieve exceptional accuracy in breast cancer classification. Logistic Regression emerged as the best performer with 98.25% accuracy, closely followed by SVM. **Feature Insights:**
Certain morphological features of cell nuclei prove to be highly discriminative for malignancy detection. The correlation analysis reveals important relationships between different measurements. **Clinical Relevance:**
These results suggest that automated classification systems can serve as valuable diagnostic aids in clinical settings, potentially improving accuracy and reducing analysis time. **Recommendations:**
• Deploy the best-performing model (Logistic Regression) for clinical use
• Continue monitoring model performance with new data
• Consider ensemble methods for even better performance
• Validate results on external datasets

Report Generated: July 30, 2025
Analysis Tool: Breast Cancer Classification System v1.0
Developed by: Ozan ■dgü