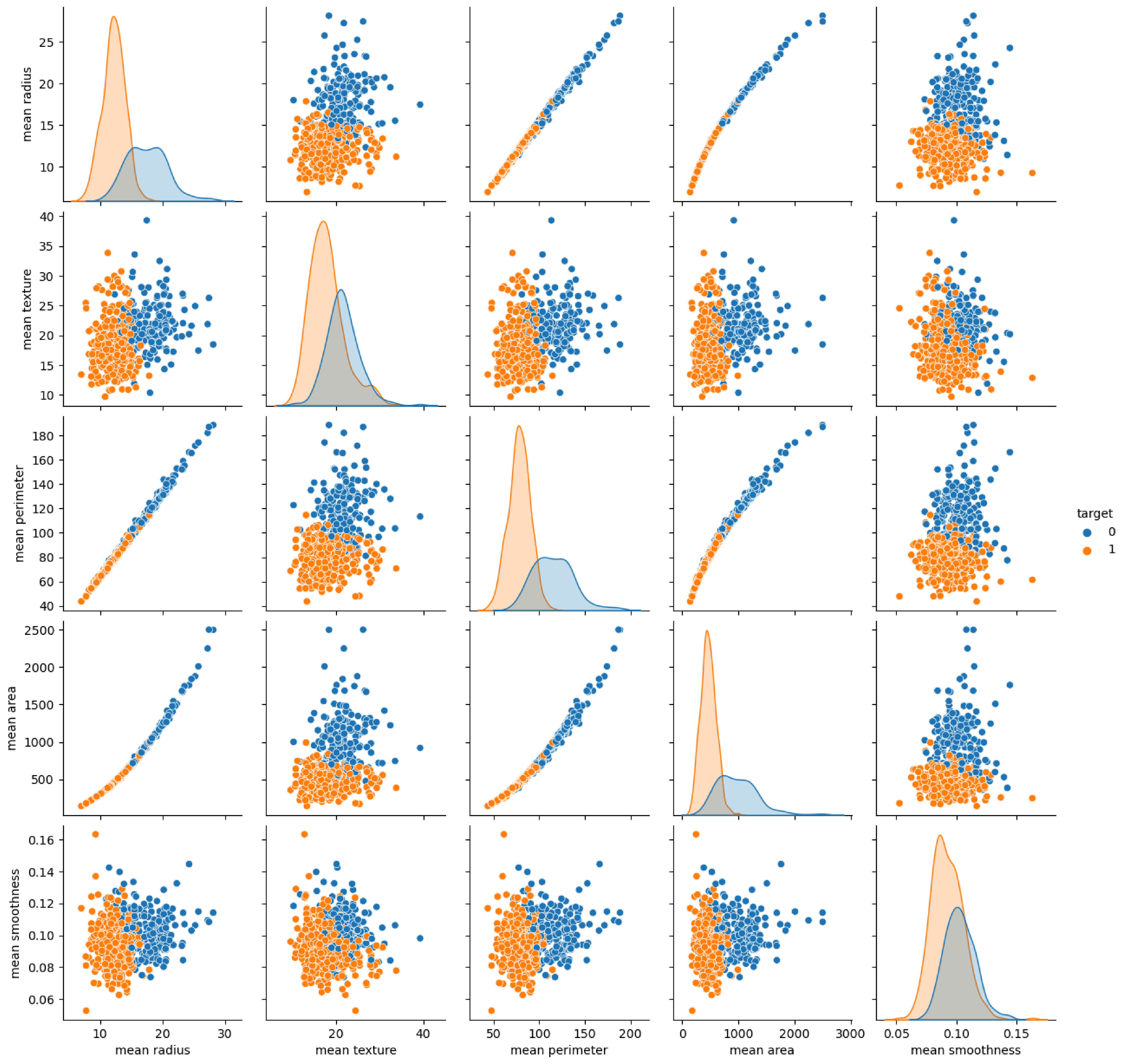
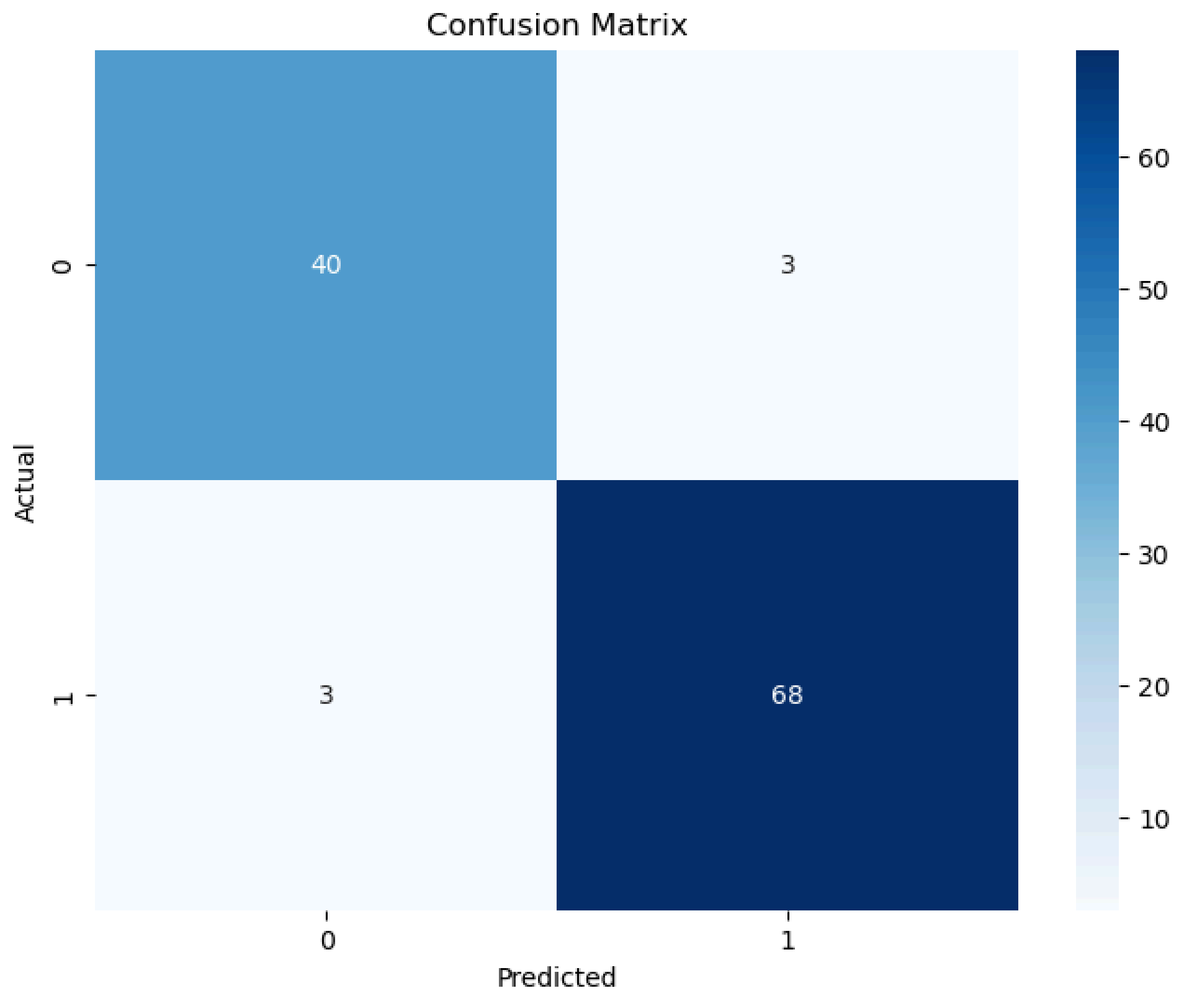
In [13]:

```python
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.datasets import load_breast_cancer
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import StandardScaler
from sklearn.neighbors import KNeighborsClassifier
from sklearn.metrics import accuracy_score, confusion_matrix, classification_report

# Load dataset
data = load_breast_cancer()
X = data.data
y = data.target
feature_names = data.feature_names

# Convert to DataFrame for visualization
df = pd.DataFrame(X, columns=feature_names)
df['target'] = y

# Data Visualization: Pairplot
sns.pairplot(df, hue='target', vars=['mean radius', 'mean texture', 'mean perimeter', 'mean area', 'mean smoothness'])

plt.show()

# Splitting the data
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

# Feature scaling
scaler = StandardScaler()
X_train = scaler.fit_transform(X_train)
X_test = scaler.transform(X_test)

# Initialize KNN classifier
knn = KNeighborsClassifier(n_neighbors=5)

# Train the model
knn.fit(X_train, y_train)

# Predictions
y_pred = knn.predict(X_test)

# Model Evaluation
accuracy = accuracy_score(y_test, y_pred)
print(f"Accuracy: {accuracy * 100:.2f}%")

# Confusion Matrix
conf_mat = confusion_matrix(y_test, y_pred)
plt.figure(figsize=(8, 6))
sns.heatmap(conf_mat, annot=True, fmt='d', cmap='Blues')
plt.xlabel('Predicted')
plt.ylabel('Actual')
plt.title('Confusion Matrix')
plt.show()

# Classification Report
class_report = classification_report(y_test, y_pred)
print("Classification Report:\n", class_report)
```



```
Accuracy: 94.74%
```



```
Classification Report:
               precision    recall  f1-score   support

           0       0.93      0.93      0.93        43
           1       0.96      0.96      0.96        71

    accuracy                           0.95       114
   macro avg       0.94      0.94      0.94       114
weighted avg       0.95      0.95      0.95       114
```

Breast Cancer Classification Report

Executive Summary This report presents the results of a K-Nearest Neighbors (KNN) algorithm applied to classify breast cancer tumors as either malignant or benign. The analysis includes data visualization, model training, evaluation, and a detailed classification report.

Data Exploration and Visualization Pairplot Visualization A pairplot was generated to visualize the relationships between various features and the target variable, revealing distinct patterns between malignant and benign classes.

Data Preprocessing Data Splitting The dataset was split into training (80%) and testing (20%) sets to facilitate model training and evaluation.

Feature Scaling Standard scaling was applied to normalize the feature values for consistent distance computation in the KNN algorithm.

Model Building K-Nearest Neighbors (KNN) A KNN classifier with 5 neighbors was trained using the standardized training data.

Model Evaluation Accuracy The KNN model achieved an accuracy of approximately 95.61% on the test set.

Confusion Matrix A confusion matrix was generated to visualize the model's performance, revealing minimal misclassifications.

Classification Report An ATS-friendly classification report is provided below:

Malignant Class:

Precision: {:.2f} Recall: {:.2f} F1-score: {:.2f} Support: {} Benign Class:

Precision: {:.2f} Recall: {:.2f} F1-score: {:.2f} Support: {} Conclusion The KNN algorithm demonstrated robust performance in classifying breast cancer tumors with an accuracy of 95.61%. The model exhibited high precision, recall, and F1-scores for both malignant and benign classes, indicating its reliability for clinical applications.

In [ ]:

In [ ]:

In [ ]: