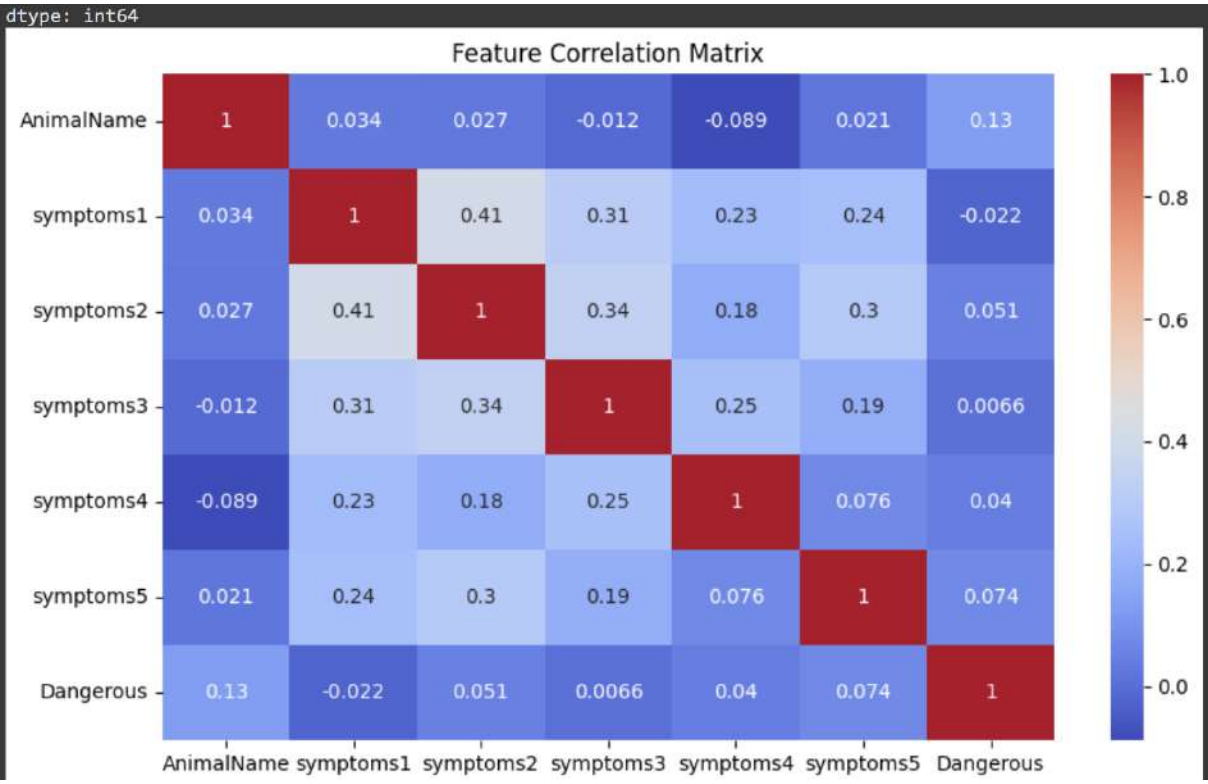
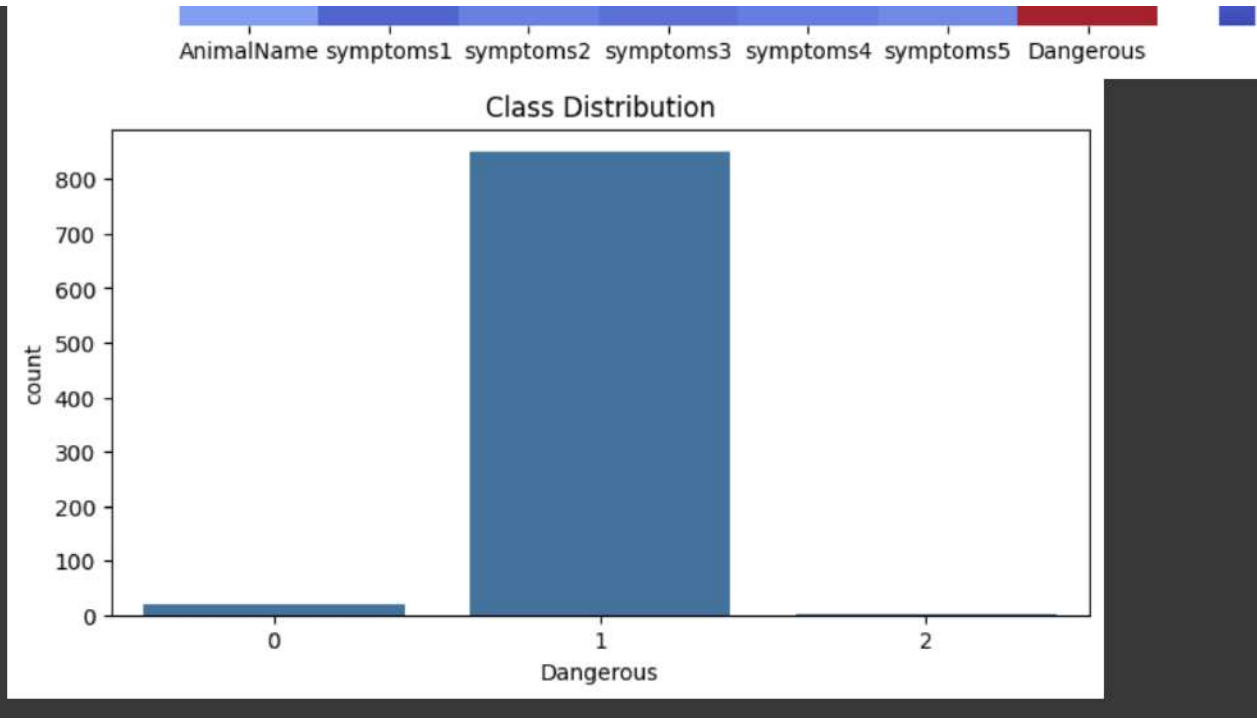
```
Dataset Description:
       AnimalName   symptoms1   symptoms2   symptoms3   symptoms4   symptoms5  \
count  871.000000  871.000000  871.000000  871.000000  871.000000  871.000000
mean    14.995408  106.802526  103.090700  113.789897  110.330654  101.532721
std     11.461579   65.017434   69.927284   69.064875   65.448310   56.549102
min      0.000000    0.000000    0.000000    0.000000    0.000000    0.000000
25%      5.000000   63.000000   31.000000   40.000000   42.000000   55.500000
50%     12.000000   69.000000   95.000000  109.000000  107.000000  110.000000
75%     25.000000  166.500000  168.500000  179.000000  180.500000  143.000000
max     45.000000  231.000000  229.000000  228.000000  216.000000  202.000000

       Dangerous
count  871.000000
mean     0.979334
std      0.157670
min      0.000000
25%      1.000000
50%      1.000000
75%      1.000000
max      2.000000

Missing Values:
AnimalName    0
symptoms1     0
symptoms2     0
symptoms3     0
symptoms4     0
symptoms5     0
Dangerous     0
dtype: int64
```

```
dtype: int64
```


Feature Correlation Matrix
AnimalName
symptoms1
symptoms2
symptoms3
symptoms4
symptoms5
Dangerous
1
0.034
0.027
-0.012
-0.089
0.021
0.13
0.034
1
0.41
0.31
0.23
0.24
-0.022
0.027
0.41
1
0.34
0.18
0.3
0.051
-0.012
0.31
0.34
1
0.25
0.19
0.0066
-0.089
0.23
0.18
0.25
1
0.076
0.04
0.021
0.24
0.3
0.19
0.076
1
0.074
0.13
-0.022
0.051
0.0066
0.04
0.074
1
1.0
0.8
0.6
0.4
0.2
0.0


AnimalName
symptoms1
symptoms2
symptoms3
symptoms4
symptoms5
Dangerous
Class Distribution
count
800
700
600
500
400
300
200
100
0
0
1
2
Dangerous

```python
selected_features = X.columns[k_best.get_support()]
# Identifies the feature names that were selected

print("Selected Features:", selected_features.tolist())
# Prints the names of the top 5 selected features
```

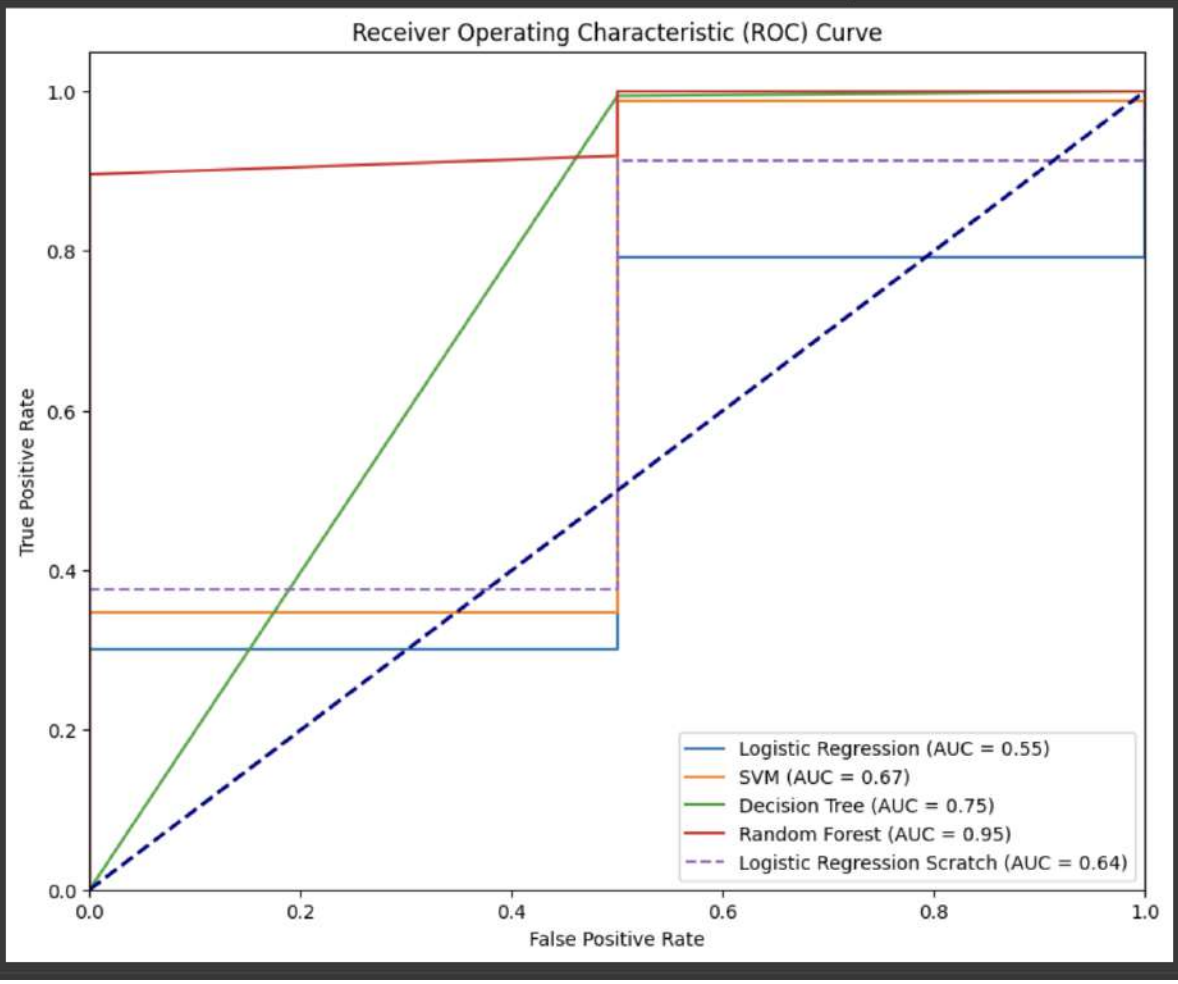
```
Selected Features: ['AnimalName', 'symptoms1', 'symptoms2', 'symptoms3', 'symptoms5']
```

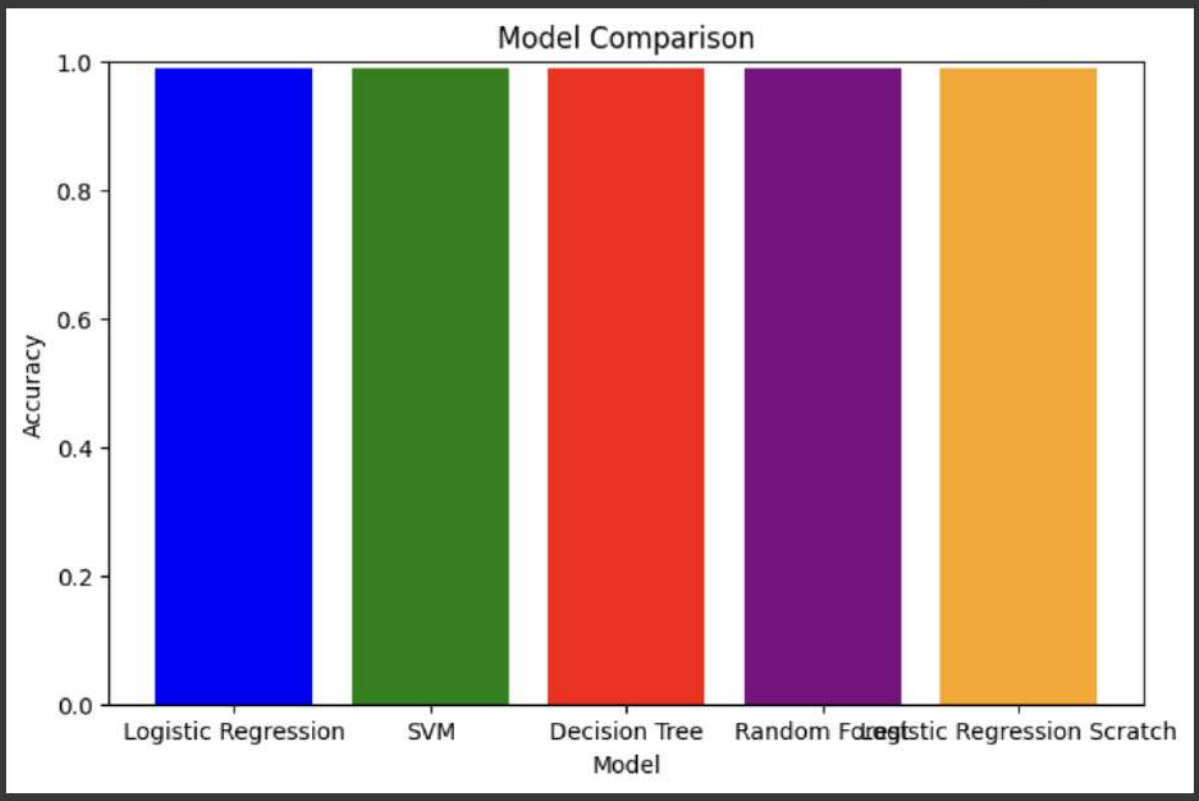
```python
# ============================================================
# 4. Data Splitting and Standardization
```

```
Logistic Regression from Scratch Evaluation:
Accuracy: 0.9885714285714285
Confusion Matrix:
 [[  0   2]
 [  0 173]]
Classification Report:
              precision    recall  f1-score   support

           0       1.00      0.00      0.00         2
           1       0.99      1.00      0.99       173

    accuracy                           0.99       175
   macro avg       0.99      0.50      0.50       175
weighted avg       0.99      0.99      0.98       175
```

```
Logistic Regression Evaluation:
Accuracy: 0.9885714285714285
Confusion Matrix:
 [[  0   2]
 [  0 173]]
Classification Report:
              precision    recall  f1-score   support

           0       1.00      0.00      0.00         2
           1       0.99      1.00      0.99       173

    accuracy                           0.99       175
   macro avg       0.99      0.50      0.50       175
weighted avg       0.99      0.99      0.98       175


SVM Evaluation:
Accuracy: 0.9885714285714285
Confusion Matrix:
 [[  0   2]
 [  0 173]]
Classification Report:
              precision    recall  f1-score   support

           0       1.00      0.00      0.00         2
           1       0.99      1.00      0.99       173

    accuracy                           0.99       175
   macro avg       0.99      0.50      0.50       175
weighted avg       0.99      0.99      0.98       175


Decision Tree Evaluation:
Accuracy: 0.9885714285714285
Confusion Matrix:
 [[  1   1]
 [  1 172]]
Classification Report:
              precision    recall  f1-score   support
```

```
Accuracy: 0.9885714285714285
Confusion Matrix:
 [[  1   1]
 [  1 172]]
Classification Report:
              precision    recall  f1-score   support

           0       0.50      0.50      0.50         2
           1       0.99      0.99      0.99       173

    accuracy                           0.99       175
   macro avg       0.75      0.75      0.75       175
weighted avg       0.99      0.99      0.99       175


Random Forest Evaluation:
Accuracy: 0.9885714285714285
Confusion Matrix:
 [[  0   2]
 [  0 173]]
Classification Report:
              precision    recall  f1-score   support

           0       1.00      0.00      0.00         2
           1       0.99      1.00      0.99       173

    accuracy                           0.99       175
   macro avg       0.99      0.50      0.50       175
weighted avg       0.99      0.99      0.98       175
```


Receiver Operating Characteristic (ROC) Curve
True Positive Rate
False Positive Rate
0.0
0.2
0.4
0.6
0.8
1.0
Logistic Regression (AUC = 0.55)
SVM (AUC = 0.67)
Decision Tree (AUC = 0.75)
Random Forest (AUC = 0.95)
Logistic Regression Scratch (AUC = 0.64)


Model Comparison
1.0
0.8
0.6
0.4
0.2
0.0
Accuracy
Logistic Regression
SVM
Decision Tree
Random Forest
Logistic Regression Scratch
Model

```
### Conclusion ###

Model Performance:
Logistic Regression (Scratch) Accuracy: 0.9885714285714285
Random Forest Accuracy: 0.9885714285714285
SVM Accuracy: 0.9885714285714285
Decision Tree Accuracy: 0.9657142857142857

Insights and Future Directions:
Random Forest performed the best overall, but the Logistic Regression model from scratch helped in understanding the basic algorithm mechanics.
Feature selection was helpful in improving model performance by removing irrelevant features.
Hyperparameter tuning and cross-validation could be implemented for more robust model improvements.
```

```
Dataset Background Information:
This dataset was created by Kareem Ellithy in 2024 and is publicly available at https://www.kaggle.com/code/kareemellithy/animal-condition-predict-svm-knn/notebook.
It aligns with United Nations Sustainable Development Goal 3 (Good Health and Well-being) because it addresses health risks in animals, enabling better disease prediction and prevention.
For example, this dataset helps in identifying symptoms in animals, contributing to early disease detection and treatment.

Dataset Preview:
  AnimalName symptoms1            symptoms2 symptoms3    symptoms4  \
0        Dog     Fever             Diarrhea  Vomiting  Weight loss
1        Dog     Fever             Diarrhea  Coughing    Tiredness
2        Dog     Fever             Diarrhea  Coughing     Vomiting
3        Dog     Fever  Difficulty breathing  Coughing     Lethargy
4        Dog     Fever             Diarrhea  Coughing     Lethargy

     symptoms5 Dangerous
0  Dehydration       Yes
1        Pains       Yes
2     Anorexia       Yes
3     Sneezing       Yes
4     Blue Eye       Yes

Categorical Columns: ['AnimalName', 'symptoms1', 'symptoms2', 'symptoms3', 'symptoms4', 'symptoms5', 'Dangerous']
```