Animal Health Classification Report

- ## Milestone 1 : Problem Understanding

Problem Statement:
In veterinary healthcare, early identification of potentially dangerous health conditions in animals can lead to timely intervention and better outcomes. This project aims to build a machine learning model to classify whether an animal's condition is dangerous based on a set of symptoms.

Objective:
To classify animals as either in a dangerous or non-dangerous condition using symptoms data. This model will support veterinary professionals in making quick assessments .

- ## Milestone 2:Data Collection and Preparation

Data Used:
File: animal_health.csv
871 rows , 7 columns

Steps Performed:

1. *Uploaded the CSV to Google Colab*
2. *Checked for null values*
3. *Dropped irrelevant or incomplete rows*
4. *Encoded categorical columns using Label Encoding*
5. *Split data into X (features) and y (target)*

- **Milestone 3:Exploratory Data Analysis(EDA)**

What We Did:

Used df.info() and df.describe() to understand data
Checked unique value counts

Visualized:
Class distribution of "Dangerous"
Most common symptoms
Found most animals are labeled "Yes" under Dangerous
Confirmed good diversity in symptoms data

- **Milestone 4:Model Building**

Model Used:
RandomForestClassifier from sklearn

Why Random Forest?

1. *Handles categorical and numeric data well*
2. *Works efficiently for binary classification*
3. *Less prone to overfitting than single decision trees*

Training Done:

1. *Split data: 80% training, 20% testing*
2. *Trained model with model.fit(X_train, y_train)*
3. *Model was successfully trained*

- **Milestone 5 : Performance Testing**

Evaluation Metrics Used:

1. *Accuracy: 98.85%*

2. *Classification Report: Precision, Recall, F1-score*
3. *Feature Importance plotted*
4. *Pie chart for class distribution*

Why the Accuracy is High ?

1. ***Strong Signal**: The symptoms in the dataset have a clear connection to whether the animal is in a dangerous condition or not.*
2. ***Balanced Dataset:** The dataset might be well-balanced (not too many more "Yes" than "No").*
3. ***Random Forest Strength:** This model works well with categorical or encoded data and handles complexity without overfitting too quickly.*

Conclusion:

Model performs very well
Very high precision and recall
Imbalanced data slightly favors "Yes" class, but results are still valid

- ## Milestone 6: Model Deployment

Basic Deployment Insight:

Since deployment isn't mandatory, we simulated predictions using test data.
model.predict() was used to generate predictions.
In real deployment: You could create a web app using Flask or Streamlit.

animal-health-app
├── animal_health_app.py # Streamlit web app
├── best_model.pkl # Trained SVM model

├── README.md # Project documentation

- **Conclusion:**

This project successfully demonstrates how machine learning can be used to aid in automated animal health classification. With further data and enhancements, such systems can be used in veterinary diagnostics, animal welfare programs, and research labs.