11. APPENDIX

Source Code

```
# Import necessary libraries
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.feature_extraction.text import CountVectorizer
from sklearn.model_selection import train_test_split
from sklearn.naive_bayes import MultinomialNB
from sklearn.metrics import classification_report, confusion_matrix, accuracy_score

# Load the dataset
data = pd.read_csv('/kaggle/input/emails/emails.csv')

# Display a sample of the dataset
print(data.head())

# Explore the distribution of the labels
sns.countplot(x='label', data=data)
plt.title('Distribution of Phishing and Legitimate Emails')
plt.show()
```