Bernoulli Naive Bayes – SMS Spam Classification

# 📘 Dataset Summary

- Dataset: SMS Spam Collection  
- Task: Binary text classification (Ham vs Spam)  
- Features: Converted to binary using CountVectorizer(binary=True)  
 - 1 if a word is present in the message, 0 if absent  
- Split: 80% training, 20% test

# 📊 Performance Summary

- Accuracy: ~98.21%  
- Classification Report:  
 - Precision (Ham): 0.98  
 - Recall (Ham): 1.00  
 - F1 Score (Ham): 0.99  
 - Precision (Spam): 1.00  
 - Recall (Spam): 0.87  
 - F1 Score (Spam): 0.93

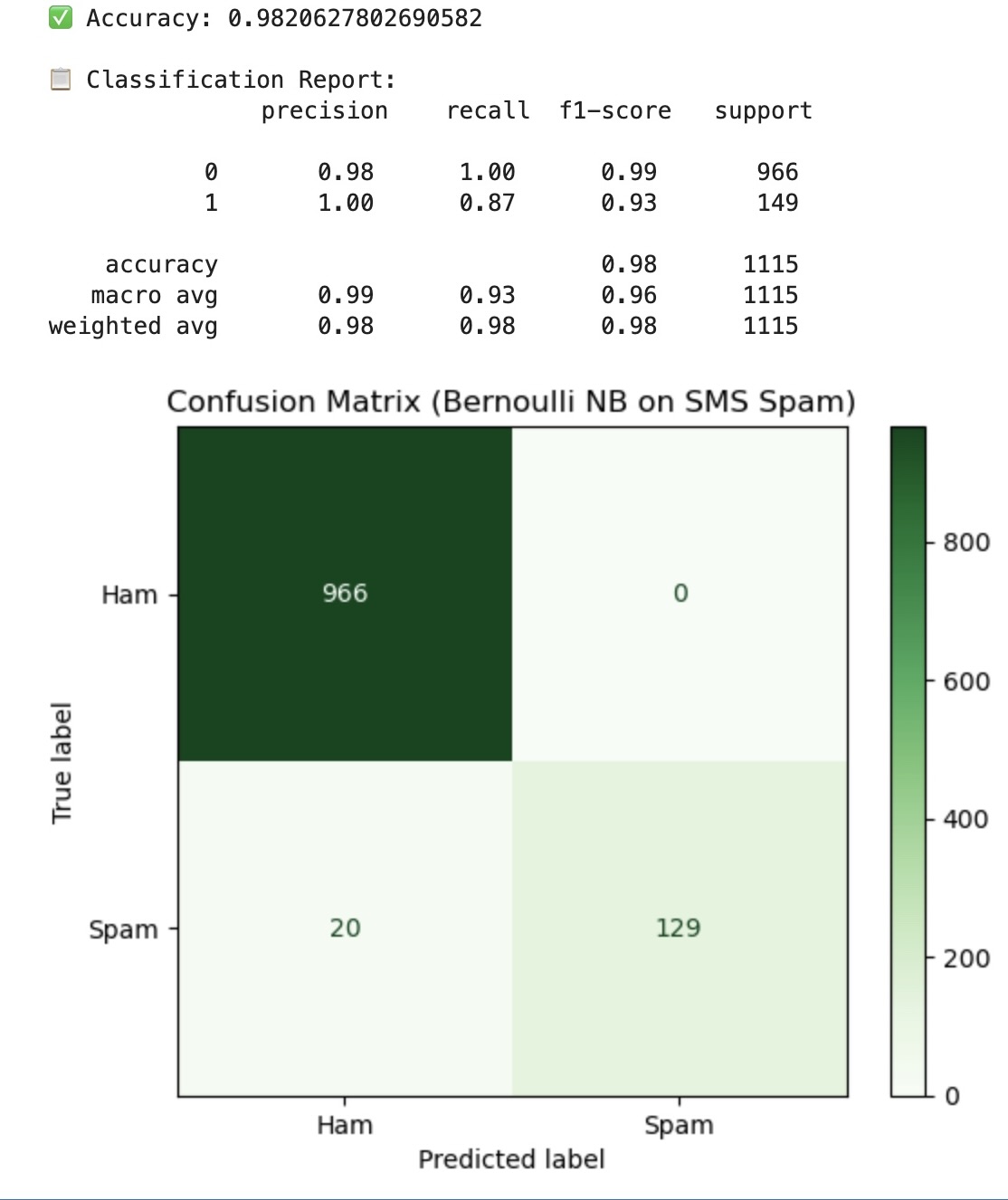
# 📉 Confusion Matrix

Predicted: Ham Predicted: Spam  
Actual Ham 966 0  
Actual Spam 20 129

# 🧠 Interpretation

- The model performed well with high precision and recall for Ham.  
- It missed 20 spam messages (false negatives), which means a few spam texts were classified as safe.  
- However, no ham messages were misclassified as spam — making the model very safe in flagging.  
- Bernoulli Naive Bayes is well-suited here due to the binary nature of the features (word presence).

# 📸 Output Visualization



Confusion Matrix for Bernoulli Naive Bayes on SMS Spam Dataset