**MESSAGE SPAM DETECTION ABHAY PRABHAKAR**

**1. Introduction**

Spam SMS detection is critical for user security. This project develops a classifier to distinguish spam from ham using the SMSSpamCollection dataset (5,572 messages: 4,825 ham, 747 spam). The dataset’s imbalance (86.6% ham vs. 13.4% spam) mirrors real-world challenges where spam is rare but high-risk.

**2. Exploratory Data Analysis**

* Class Distribution:
  + Ham: 4,825 messages (86.6%).
  + Spam: 747 messages (13.4%).
* Message Length:
  + Ham: Avg. 71 characters (short, conversational).
  + Spam: Avg. 138 characters (longer, promotional).
* Visualizations:
  + Count plots highlighted class imbalance.
  + Histograms showed spam messages are consistently longer.

**3. Methodology**

**Preprocessing:**

1. Lowercasing, punctuation removal, stopword filtering, and lemmatization.
2. TF-IDF converted text to numerical features.

**Models:**

1. **Naive Bayes:**
   * Optimized with alpha=1.0 (Grid Search CV).
   * Strengths: Zero false positives.
2. **Logistic Regression:**
   * Optimized with C=100 and class\_weight='balanced'.
   * Strengths: Handles imbalance effectively.

**4. Model Performance**

| **Metric** | **Naive Bayes** | **Logistic Regression** |
| --- | --- | --- |
| **Accuracy** | 97.40% | 98.03% |
| **Spam Recall** | 83% | 89% |
| **F1-Score** | 90% | 92% |
| **False Positives** | 0 | 5 |
| **False Negatives** | 25 | 17 |

**Key Insights:**

* Logistic Regression outperformed Naive Bayes in accuracy (98.03%) and spam recall (89%).
* Trade-offs:
  + Naive Bayes: Zero false positives but missed 25 spam messages.
  + Logistic Regression: 5 false positives but missed only 17 spam messages**.**

**5. Challenges & Solutions**

* Class Imbalance:
  + Solution: Stratified sampling and class weights in Logistic Regression.
* Text Noise:
  + Solution: Regex cleaning and lemmatization.
* Hyperparameter Tuning:
  + Solution: Grid Search CV for automated optimization**.**

**6. Conclusion**

Logistic Regression is ideal for deployment, balancing accuracy (98.03%) and spam recall (89%). While it misclassifies 5 ham messages, it minimizes missed spam (17 FN), aligning with real-world priorities. Future work could explore SMOTE to reduce false negatives.