SMS Spam Classifier: Detailed Project Report

1. Project Title

SMS Spam Classification using Machine Learning

2. Objective

Automatically classify SMS messages as 'spam' or 'ham' (not spam) using machine learning to reduce manual effort and protect users from fraudulent and unwanted messages.

3. Dataset Description

- **File:** spam.csv
- **Source:** UCI SMS Spam Collection
- **Description:** 5,574 SMS messages tagged as ham or spam.
- **Columns:**
    - label: ham or spam
    - message: SMS text content

4. Project Workflow

Data Collection

- Loaded using pandas
- Checked nulls and duplicates
- Converted ham to 0 and spam to 1

Data Preprocessing

- Lowercased text
- Removed punctuation and special characters
- Removed stopwords
- Lemmatized words to their base form

Feature Extraction

- Used TF-IDF Vectorizer to convert text into numerical vectors
- Limited maximum features for control

Train-Test Split

- 80% training, 20% testing using stratified split

Model Selection and Training

- Tried Multinomial Naive Bayes, SVM, Logistic Regression, Random Forest
- Chose **Multinomial Naive Bayes** for its text classification efficiency

Model Evaluation

- Evaluated using Accuracy, Precision, Recall, F1-Score, Confusion Matrix, ROC-AUC

5. Results

- **Accuracy:** ~98%
- **Precision:** High (low false positives)
- **Recall:** High (catches spam effectively)
- **F1-Score:** High, balanced
- **Confusion Matrix Example:**
- [[965, 1],
- [18, 131]]
- Spam keywords detected: "free", "win", "claim", "call"
- Ham keywords: "ok", "come", "home"

6. Visualizations

- Word clouds for spam and ham
- Confusion matrix plot for clarity

7. Conclusion

- Successfully built a high-accuracy SMS spam classifier
- Protects users from unwanted messages automatically
- Ready for integration into messaging systems