SMS Spam Classification Project

Task Objective

The objective of this project is to develop a robust and accurate machine learning model that classifies SMS messages as either spam or ham (legitimate). The classification is performed using a stacking ensemble method, combining multiple machine learning models to enhance prediction accuracy.

Approach

1. Dataset

The dataset used for this project is the SMS Spam Collection Dataset, which contains:

5,574 SMS messages tagged as either ham or spam.
The dataset was preprocessed to remove unnecessary columns and encode the labels for classification.

2. Methodology

Data Preprocessing:
Text Vectorization: The raw SMS text was converted into numerical representations using the TF-IDF (Term Frequency-Inverse Document Frequency) vectorizer, limited to the top 5000 features.
Label Encoding: The labels (ham and spam) were encoded into numerical values (0 for ham, 1 for spam).
Stacking Ensemble Technique:
To maximize classification performance, a stacking ensemble approach was employed:

Base Models:
Logistic Regression
Random Forest Classifier
Gradient Boosting Classifier
Support Vector Machine (SVM)
Meta Model:
Logistic Regression: Combines predictions from base models to make final classifications.
Model Training:
The dataset was split into 80% training and 20% testing sets using stratified sampling to preserve class distribution.
The stacking classifier was trained on the training data, leveraging cross-validation to ensure robustness.
Evaluation:
The model was evaluated on the test set using metrics such as:

Precision
Recall
F1-Score
Accuracy
A confusion matrix was generated to visualize classification performance.

3. Results

The model achieved the following metrics:

Metric	Ham (Legitimate)	Spam	Overall
Precision	0.99	0.97	0.98 (Macro Avg)
Recall	1.00	0.92	0.96 (Macro Avg)
F1-Score	0.99	0.94	0.97 (Macro Avg)
Accuracy	-	-	0.99

Confusion Matrix:

	Predicted Ham	Predicted Spam
Actual Ham	966	0
Actual Spam	12	137

4. Challenges Faced

Handling Class Imbalance: The dataset contained a significantly higher number of legitimate messages (ham) compared to spam, which required stratified sampling and robust evaluation to ensure fair performance.

Feature Engineering: Determining the optimal number of features for the TF-IDF vectorizer required experimentation to balance model complexity and performance.
Model Integration: Combining diverse base models using a stacking ensemble approach required careful parameter tuning and cross-validation.
5. Results Achieved
The stacking ensemble model successfully achieved:

99% accuracy on the test dataset.
High precision, recall, and F1-scores for both spam and ham classes.
A clear and interpretable confusion matrix highlighting the model's ability to correctly classify legitimate and spam messages.