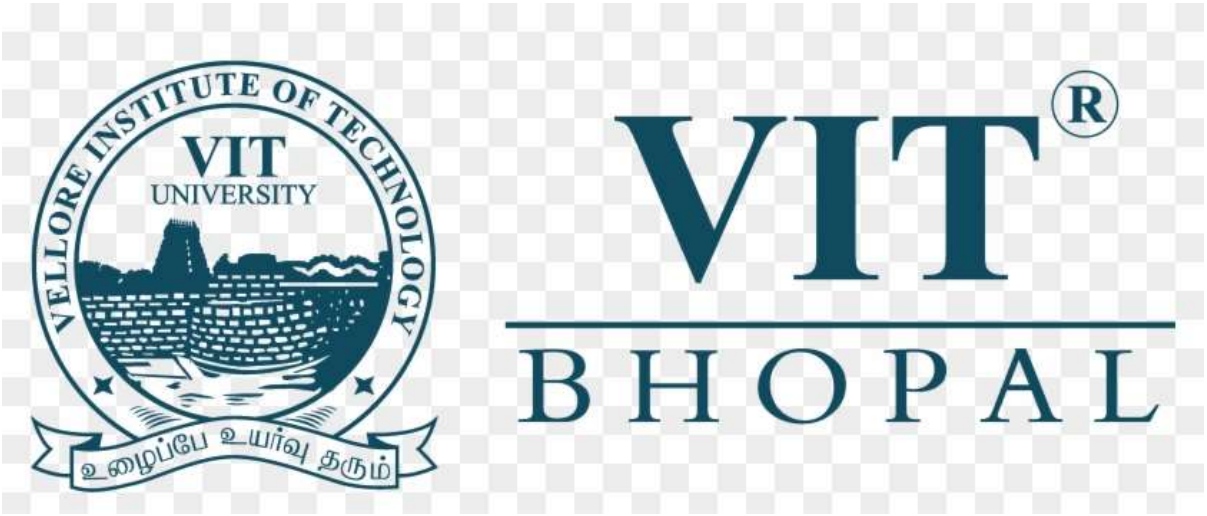


SMS Spam Detection Project Report

Name: Satya Prakash Sahu (25BAC10023)

Subject: Introduction To Problem Solving

2. Introduction

This report presents an SMS spam detection system developed using machine learning and NLP

techniques.

3. Problem Statement

The goal is to classify SMS messages as spam or ham to reduce user exposure to fraudulent or unwanted content.

4. Functional Requirements

- Load dataset

- Preprocess text

- Train ML model

- Evaluate performance

- Save trained model

5. Non-functional Requirements

- Accuracy

- Scalability

- Maintainability

- Usability

- Performance

6. System Architecture

The system includes data ingestion, preprocessing, vectorization, model training, and prediction

pipeline.

7. Design Diagrams

Use Case Diagram, Workflow Diagram, Sequence Diagram, Class Diagram, ER Diagram

(conceptual only).

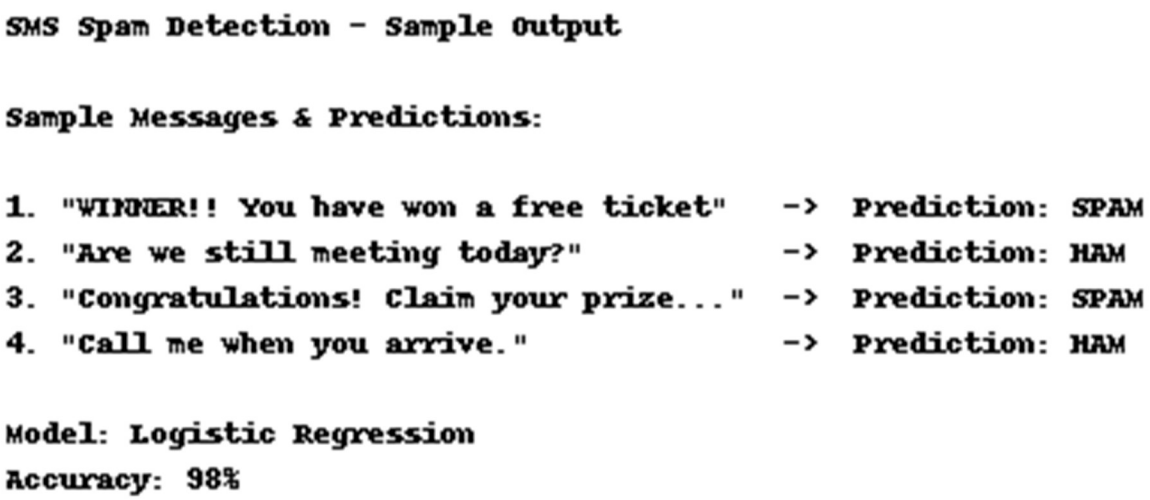
8. Design Decisions & Rationale

Logistic Regression chosen for simplicity, interpretability, and strong performance on text classification.

9. Implementation Details

Implemented in Python using pandas, scikit-learn, and NLTK. TF-IDF used for feature extraction.

10. Screenshots / Results



11. Testing Approach

Train/test split and cross-validation used to validate performance.

12. Challenges Faced

Handling text noise, selecting optimal preprocessing, tuning hyperparameters.

13. Learnings & Key Takeaways

NLP preprocessing critically impacts model performance. Simpler models can perform strongly on

text tasks.

14. Future Enhancements

Use deep learning, deploy API, integrate real-time detection, expand dataset.

15. References

UCI Machine Learning Repository, scikit-learn documentation,
NLTK documentation,Kaggle ,github