

# SMS SPAM DETECTION USING NLP and ML

By: Prashant Rijal

# ABOUT THE PROJECT

## Project Overview

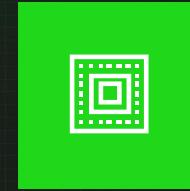
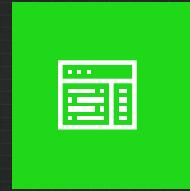
- Text classification project for SMS spam detection
- Classifies messages as **Spam** or **Ham**
- Uses Machine Learning and NLP techniques

## Why This Topic Matters

- Increase in SMS phishing (Smishing)
- Protects users from fraud and scams
- Improves trust in mobile communication



# Machine Learning Models



## Multinomial Naive Bayes

- Probabilistic classifier
- Assumes feature independence
- Very efficient for text data

## Logistic Regression

- Linear classification model
- Uses sigmoid function
- Outputs probability of spam

## KNN

- Distance-based classifier
- Instance-based learning

... . . .

# Research Evidence

**Dataset:** SMS Spam Collection Dataset (5,574 messages)

**Why chosen:**

- Publicly available and widely used benchmark dataset
- Contains both spam and ham messages for supervised learning

**Model Selection Rationale:**

- **Multinomial Naive Bayes:** Performs well with word frequency features, lightweight, accurate for text classification
- **Logistic Regression:** Predicts probability of spam, interpretable, robust for binary classification
- **K-Nearest Neighbors (KNN):** Simple, distance-based classifier, provides comparison with probabilistic models

# Used Architecture

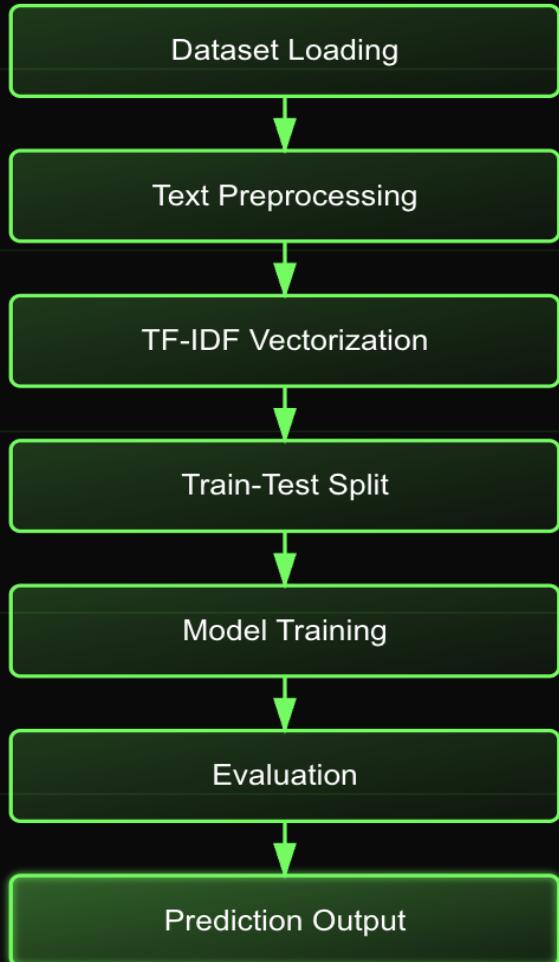
## Steps in the solution:

- Data Cleaning (remove punctuation, lowercase, remove stop words)
- Feature Extraction (TF-IDF vectorizer)
- Train ML models on labelled SMS data
- Evaluate the model
- Predict incoming messages as Spam or Ham

## Application Function:

- User inputs a message
- System predicts Spam or Ham

## System Workflow



# Solution

## **Steps in the solution:**

- Data Cleaning (remove punctuation, lowercase, remove stop words)
- Feature Extraction (TF-IDF vectorizer)
- Train ML models on labelled SMS data
- Evaluate the model
- Predict incoming messages as Spam or Ham

## **Application Function:**

- User inputs a message
- System predicts Spam or Ham

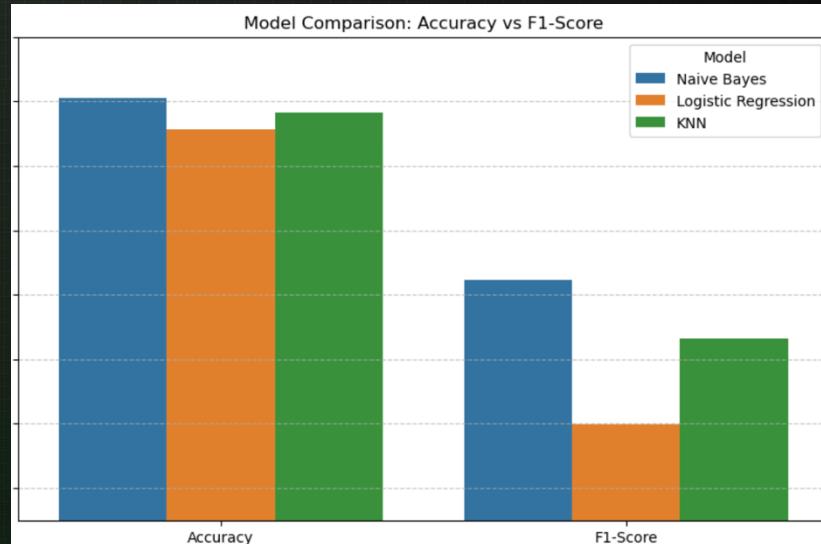
# Results

For Ham

Model	Precision	Recall	F1 score
Multinomial Naïve Bayes	0.95	0.78	0.86
Logistic Regression	0.97	0.73	0.84
K-Nearest Neighbor	0.98	0.75	0.85

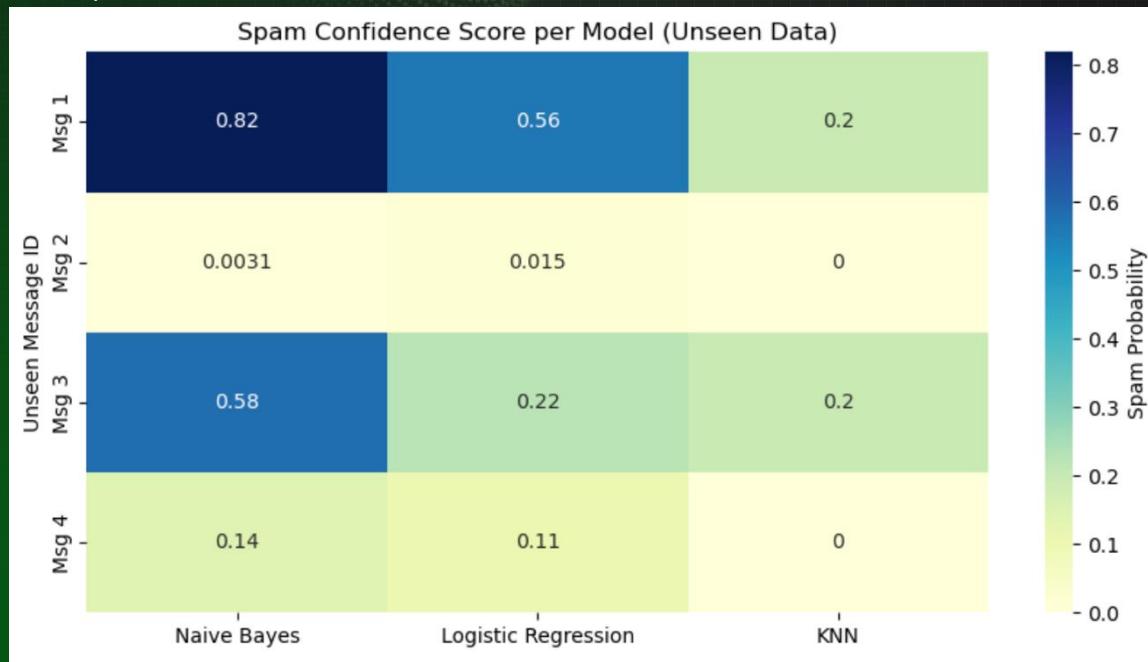
For Spam

Model	Precision	Recall	F1 score
Multinomial Naïve Bayes	0.97	0.99	0.98
Logistic Regression	0.96	1.0	0.98
K-Nearest Neighbor	0.96	1.0	0.98



The models perform excellently in the labelled data but when unseen data is given to them not all the models perform well.

- **Best Model:** Naive Bayes (due to its efficiency and accuracy on text)
- **How results help:**
  - Detect spam messages automatically
  - Reduce user exposure to malicious content



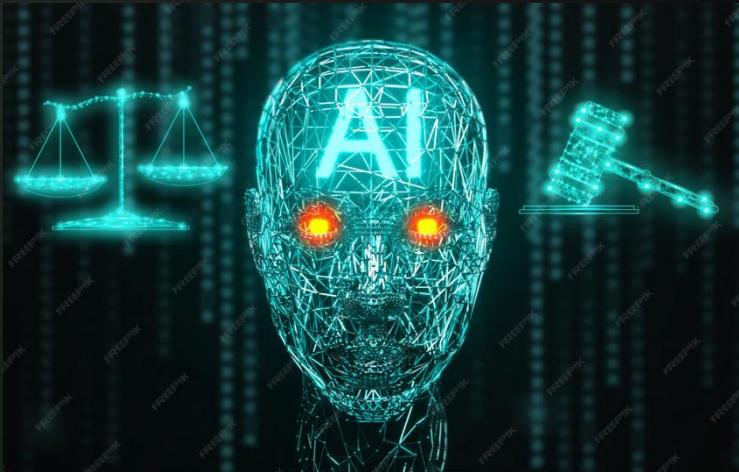
# Ethical Considerations

## Ethical & Social Concerns

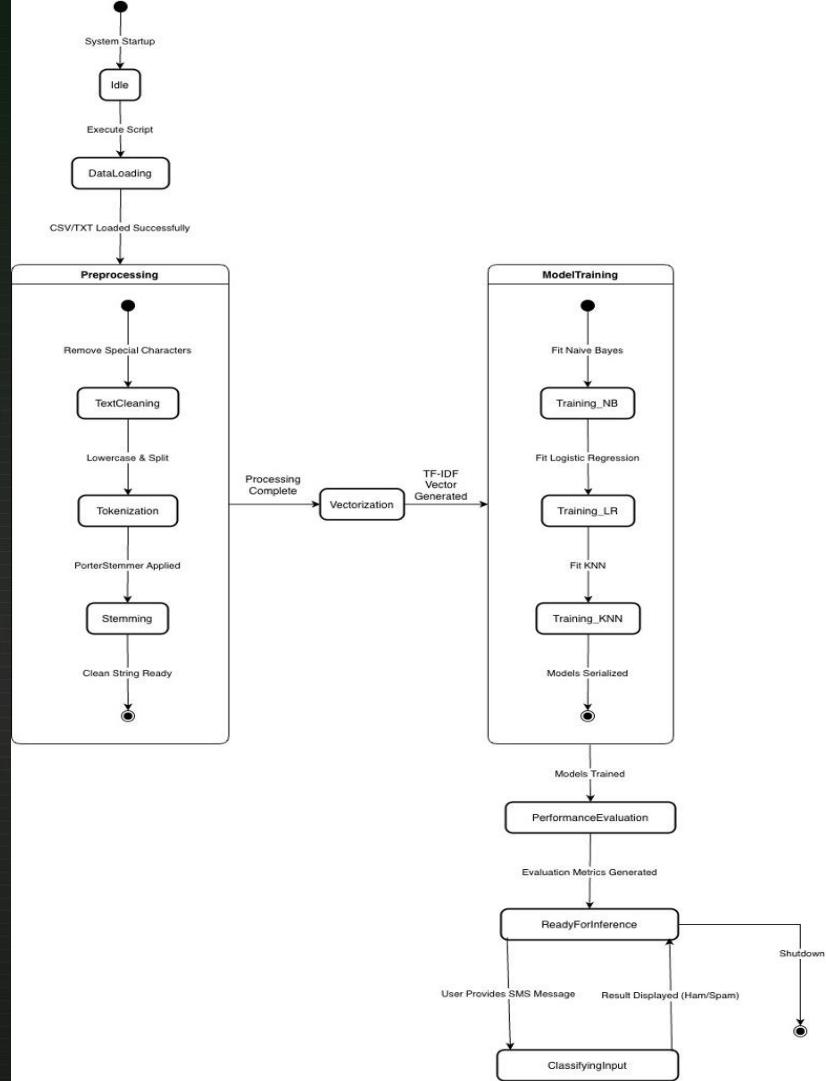
- Privacy of SMS content
- Risk of false positives
- Bias due to class imbalance

## Responsible AI

- Secure data handling
- Transparent decision-making



# State Transition Diagram



THANK You For Listening!!