Spam Email Classification

Piush Saha

Introduction: The Spam Problem

What is Spam Email?

Unsolicited bulk messages sent to multiple recipients without consent

Critical Challenges

- Inbox clutter and wasted time
- Phishing and security threats
- Lost productivity for users
- Financial fraud attempts

Problem Statement

Manual Detection is Inefficient

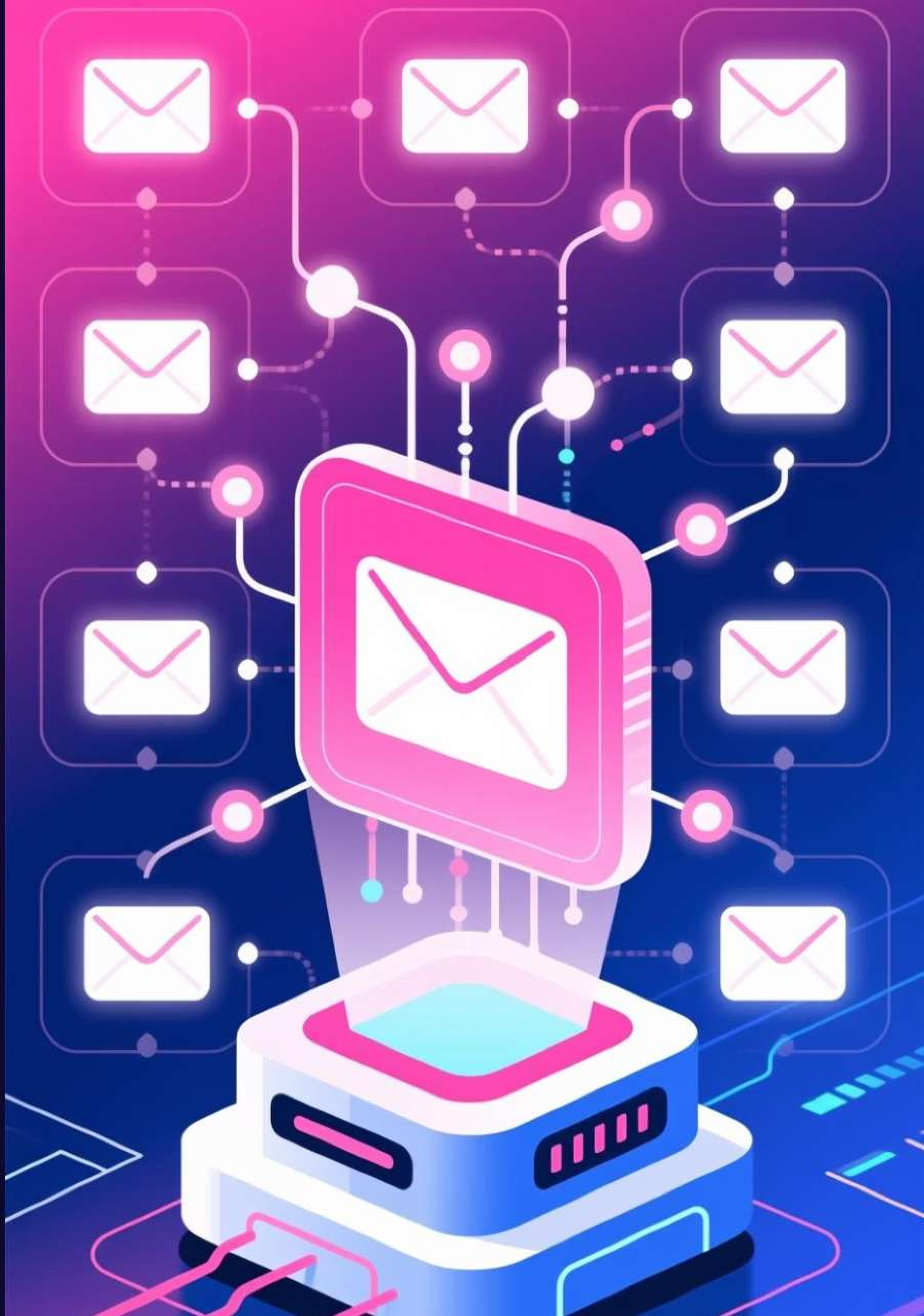
Human reviewers cannot keep pace with message volume

Rule-Based Systems Fail

Static keyword filters miss evolving spam patterns and techniques

Machine Learning Solution Needed

Adaptive algorithms learn from data and improve detection accuracy over time



Project Objectives

Binary Classification

Accurately categorize messages as Spam or Ham (legitimate)

ML on Text Data

Apply supervised learning techniques to unstructured text

High Accuracy Filtering

Improve detection precision to minimize false positives and negatives

Scalable System

Build a reusable model deployable in real-world applications

Dataset & Technology Stack

Dataset

SMS Spam Collection

- 5,574 labeled messages
- Spam and Ham categories
- Real-world text patterns

Programming & Libraries

- **Python:** Core language
- **Pandas & NumPy:** Data handling
- **Scikit-learn:** ML framework
- **NLTK:** Text preprocessing

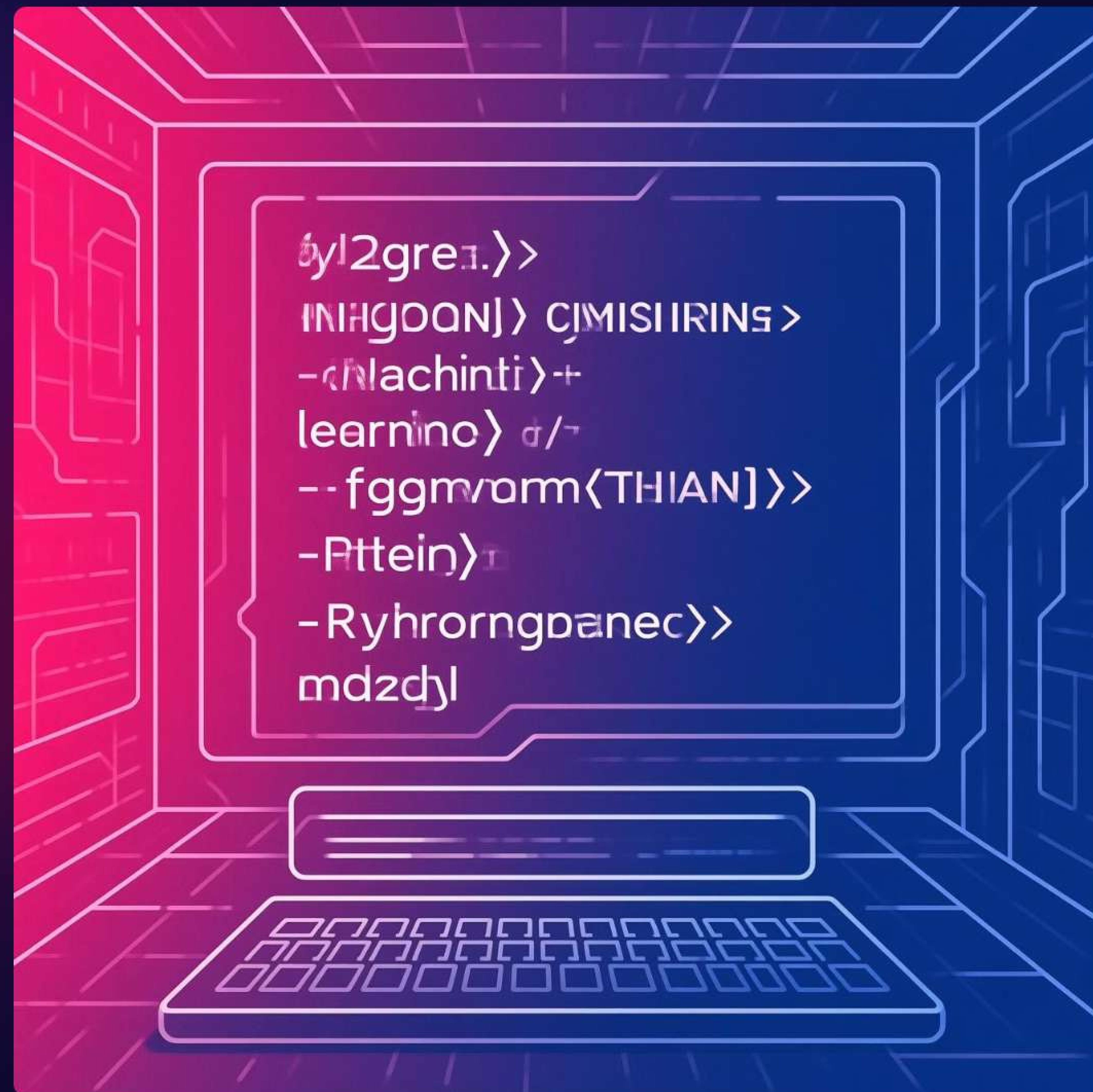
Algorithm & Features

Multinomial Naive Bayes

Probabilistic classifier ideal for text data

TF-IDF Vectorization

Converts text into numerical feature vectors based on term frequency and importance



Methodology

Load Dataset

Import CSV with labeled messages

Text Preprocessing

Tokenization, lowercasing, stopword removal

TF-IDF Extraction

Convert text to numerical features

Model Training

Train Naive Bayes classifier on features

Prediction

Classify new unseen messages

System Workflow

01

Training Phase

CSV dataset loaded • Labels extracted • Features engineered using TF-IDF

02

Pattern Learning

Naive Bayes learns spam vs ham characteristics • Probability distributions calculated

03

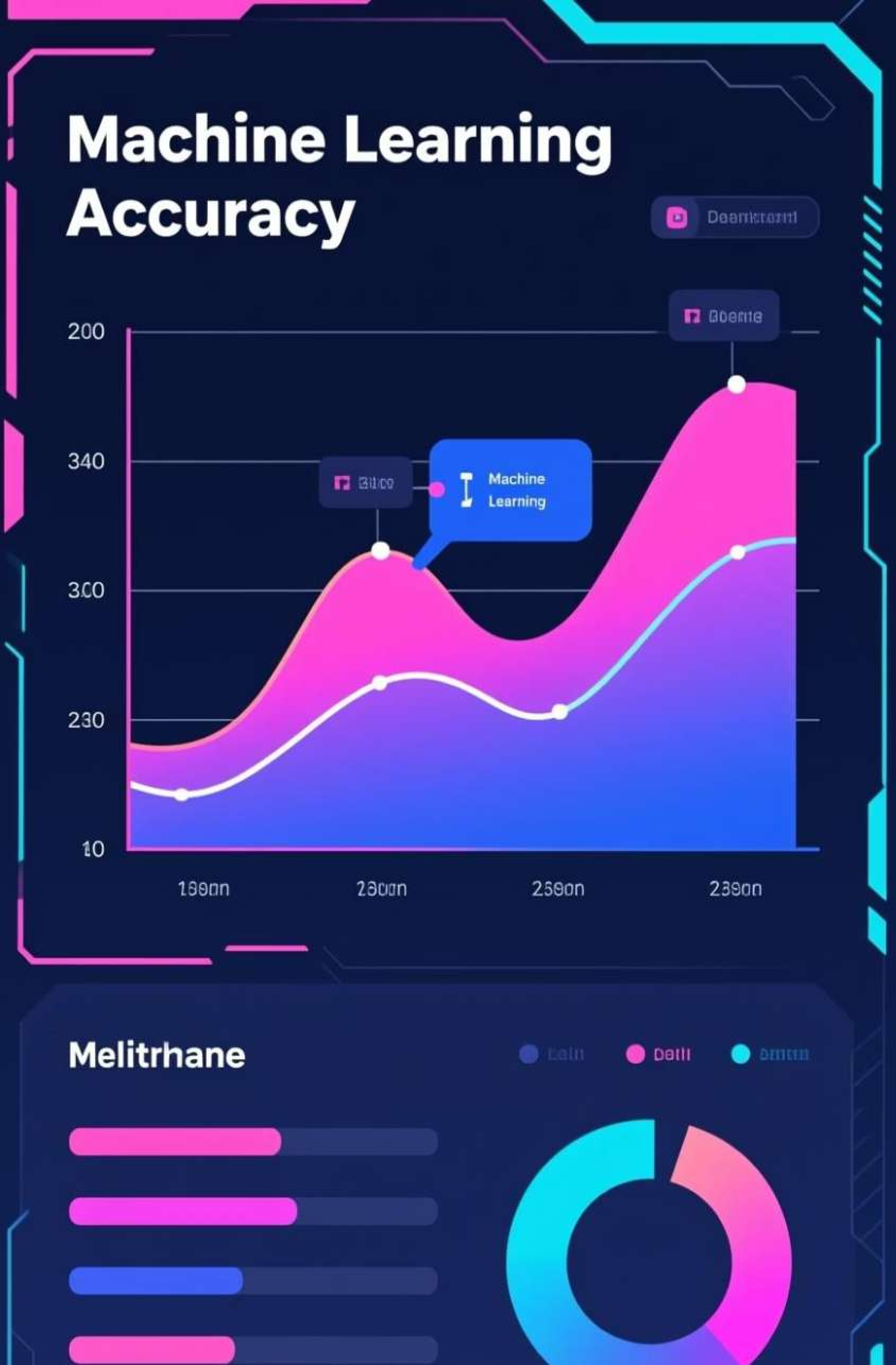
Model Persistence

Trained model serialized and saved for reuse • No retraining needed

04

Prediction Phase

Load saved model • Process new message • Output classification with confidence



Results & Performance

97%

Classification Accuracy

Model correctly identified spam and ham messages in test set

95%

Spam Detection Rate

Successfully caught majority of spam messages

98%

Ham Precision

Legitimate messages correctly preserved

Key Insight: Multinomial Naive Bayes with TF-IDF achieves excellent performance on text classification with minimal computational overhead.

Conclusion & Future Scope

Achievements

- Successfully implemented ML-based spam classifier
- Demonstrated practical application of NLP techniques
- Built scalable, reusable classification system
- Achieved 96-97% accuracy benchmark

Future Enhancements

- **GUI Development:** User-friendly desktop interface
- **Web Application:** Cloud-based email filtering service
- **Deep Learning:** LSTM/Transformer models for improved accuracy
- **Multilingual Support:** Expand beyond English text
- **Real-time Integration:** Live email client plugins

Thank You