Spam Detection with Naive Bayes

Aachal Kushwaha 202410116100001

Anshika Srivastava 202410116100032

Anshu Nishad 202410116100033

Disha Seth 201410116100064

Introduction to Spam Email Classification

Spam email classification aims to differentiate spam from legitimate emails.

This project uses machine learning techniques for accurate filtering.

Our focus is on the Naive Bayes algorithm for its simplicity and efficiency.

Understanding the Problem

What is Spam Email?

Unsolicited bulk email, often unwanted or harmful.

Spam vs. Ham

Spam is unwanted mail; ham is legitimate email.

Importance of Detection

Protects from phishing, malware, and inbox clutter.

Exploring the Dataset

Dataset Source:
Loaded from mail_data.csv, containing labeled SMS/email messages.

Columns Present:

Category (label: spam or ham)

Message (actual email/text content)

Missing Values Handled:

Dropped rows with missing Category or Message using dropna().

Class Distribution:

Checked using value_counts()

Imbalanced: More "ham" (non-spam) than "spam".

Text Type:

Messages are raw text, requiring preprocessing for machine learning.

Dataset Details

- **5,574 SMS messages**
- **Labels: spam or ham**
- **80% training, 20% testing split**

Naive Bayes: The Algorithm

Naive Bayes altdstton fo forflection aligrilation

1	Lable	Date	date	date
2	Lg.bit	505	595	506
3	Lable	705	565	508
3	La.Pic	369	597	309
4	Lg.bic	361	967	968
5	La.ble	456	305	524
5	Lable	944	905	308
6	Labic	305	975	955
7	Lable	255	565	366
9	La.bic	308	995	504
10	La.ble	389	599	500

→ P (=() = B

→ **P((a/ = A)**

P (/A)) = =|B

P(a)(B/=(A/ ≠ = |B)

Probabilistic Classifier

Uses Bayes' Theorem to predict classes.

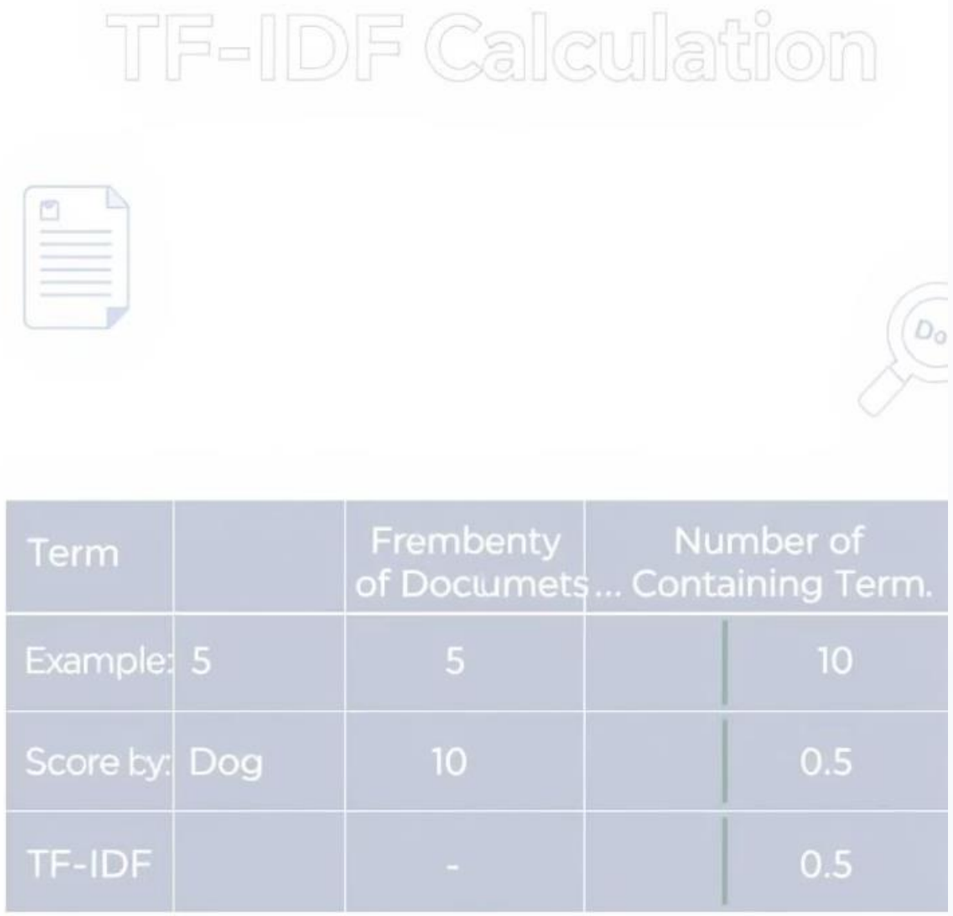
Feature Independence

Assumes features are independent within classes.

Simple & Effective

Widely used for text classification tasks.

Feature Extraction: TF-IDF



Term Frequency (TF)

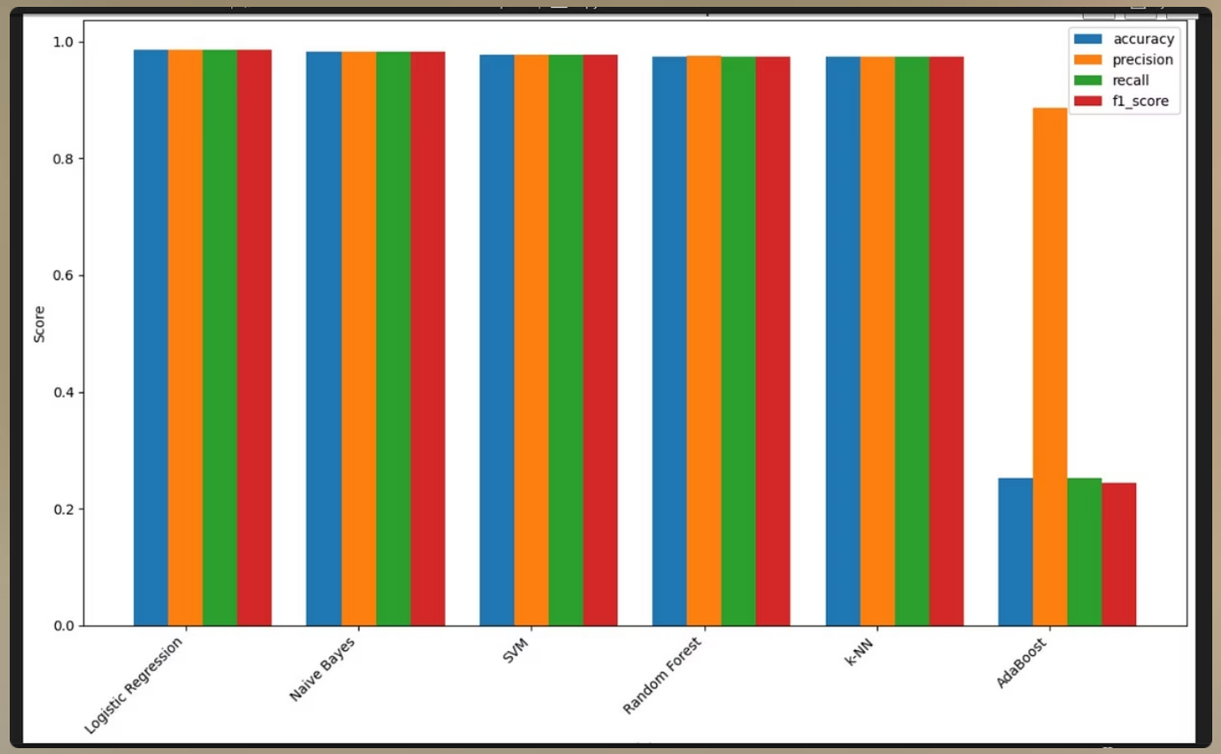
Measures how often a term appears in a document.

Inverse Document Frequency (IDF)

Downweights common words across documents.

TF-IDF Value

Highlights important but rare terms in text.



Results and Evaluation

Model	Accuracy	F1 Score	Time (s)
SVM	0.9865	0.9864	12.90
Logistic Regression	0.9821	0.9820	0.06
Naive Bayes	0.9776	0.9778	0.02
Random Forest	0.9749	0.9747	0.64
AdaBoost	0.9740	0.9733	3.94
k-NN	0.2449	0.2449	0.35

Conclusion and Next Steps

Powerful & Fast

Naive Bayes delivers simple, accurate spam detection.

Enhancements

Feature engineering and ensembles can boost results.

Future Work

Deploy as web service and integrate with email clients.