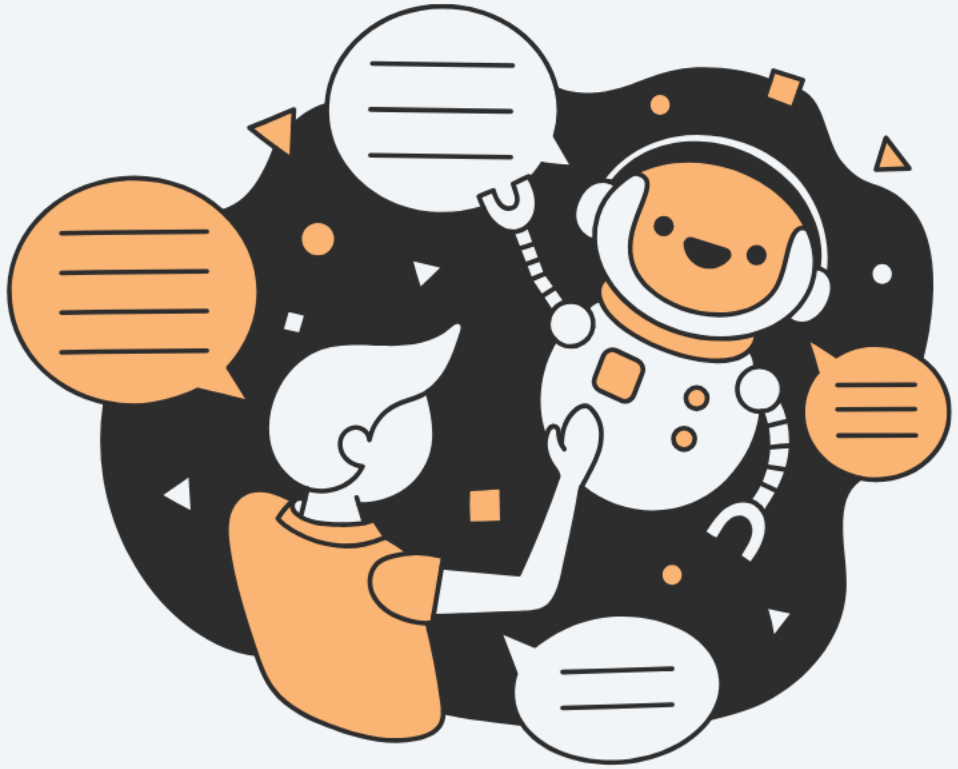


An Introduction to Naive Bayes Classifiers

How It Works, Where We See It, and Why It Matters

Spams vs. Hams

hey friend! how are things...

I HAVE A DATE ON SUNDAY WITH WILL!!

congratulations! win a price...

Eh u remember how 2 spell his name...

WINNER!! As a valued network customer...

আমি গুগল-এর একজন নিয়োগ ব্যবস্থাপক।

Thanks for your subscription to...

Did you hear about the...

Wah lucky man...

Ok lar... Joking wif u oni...

Congrats! 1 year special...

Please call our customer service...

Tell where you reached...

Spams vs. Hams

hey friend! how are things...

I HAVE A DATE ON SUNDAY WITH WILL!!

congratulations! win a price...

Eh u remember how 2 spell his name...

WINNER!! As a valued network customer...

আমি গুগল-এর একজন নিয়োগ ব্যবস্থাপক।

Thanks for your subscription to...

Did you hear about the...

Wah lucky man...

Ok lar... Joking wif u oni...

Congrats! 1 year special...

Please call our customer service...

Tell where you reached...

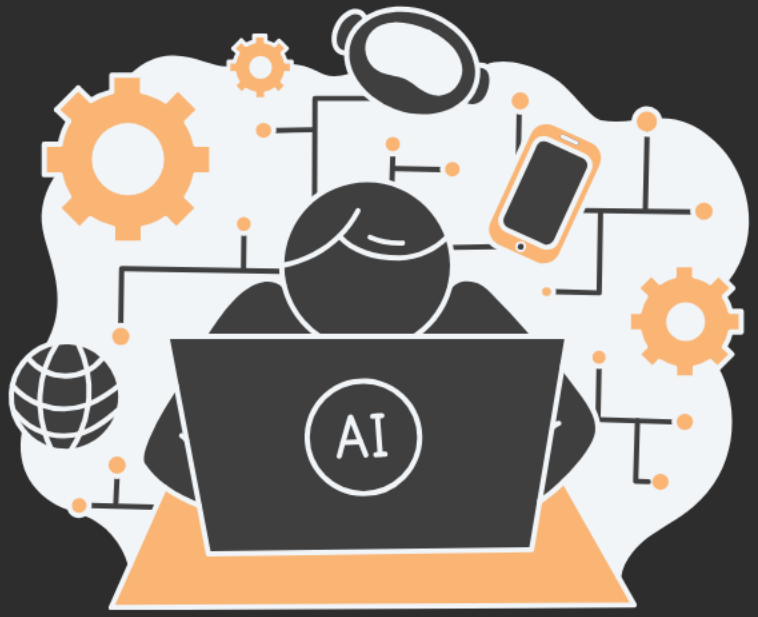
Naive Bayes Models

- **Bayes' Theorem:** What we know to guess what's true.
- **Naive Assumption:** Assumes every word is independent
- **Use cases:** *text classification*, *sentiment analysis*, *spam filtering*.

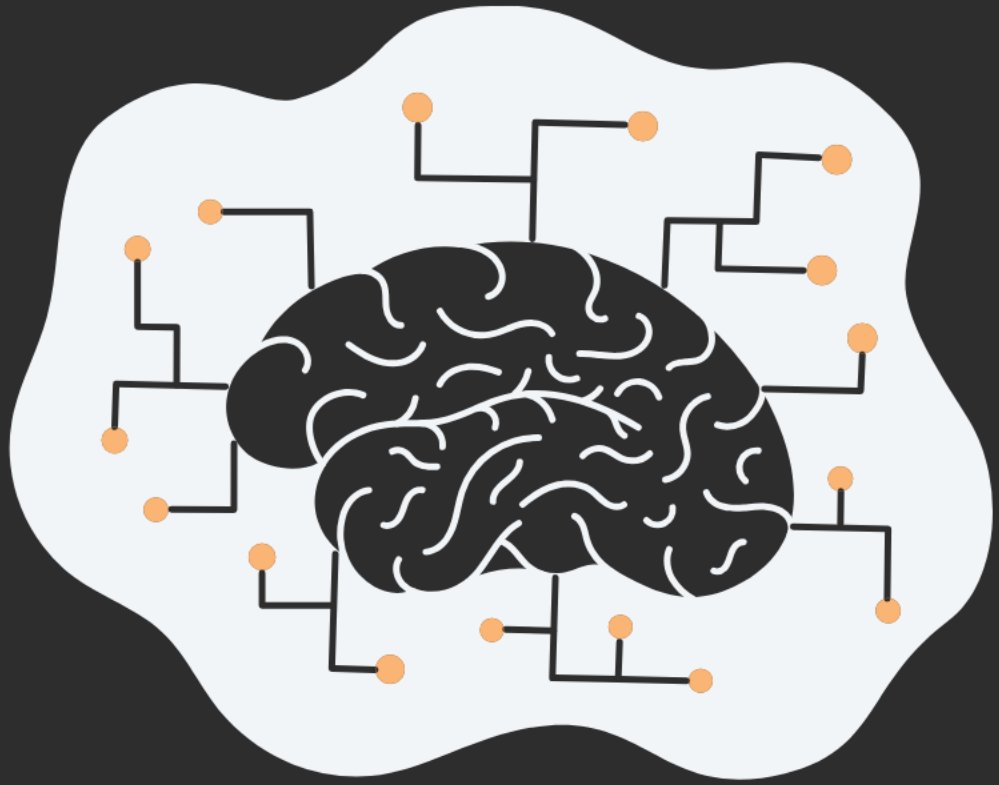
Naive Bayes Models

*A **probability-based** classifier that assumes all features work **independently**.*

Three Naive Bayes Models



- Multinomial Naive Bayes
- Bernoulli Naive Bayes
- Gaussian Naive Bayes



Multinomial Naive Bayes

Count Matters

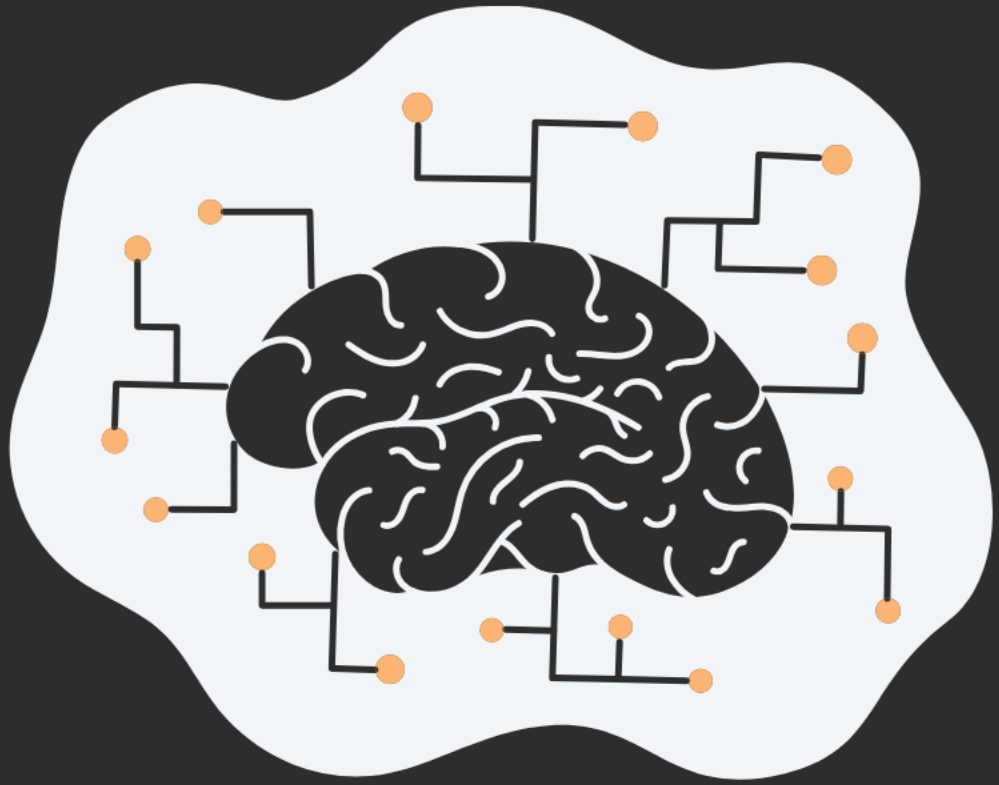
Perfect for text data where we care about **how many times a word appears**

Example

The word ***free*** shows up 5 times → **5**

Practical Example

Message	Words	Count
Free prize now	“free”, “prize”, “now”	**free:** 1 **prize:** 1 **now:** 1
Free free unlimited offer	“free”, “offer”	**free:** 2 **unlimited:** 1 **offer:** 1
Hello friend, hello	“hello”, “friend”	**hello: 2** **friend:** 1



Bernoulli Naive Bayes

Presence Matters

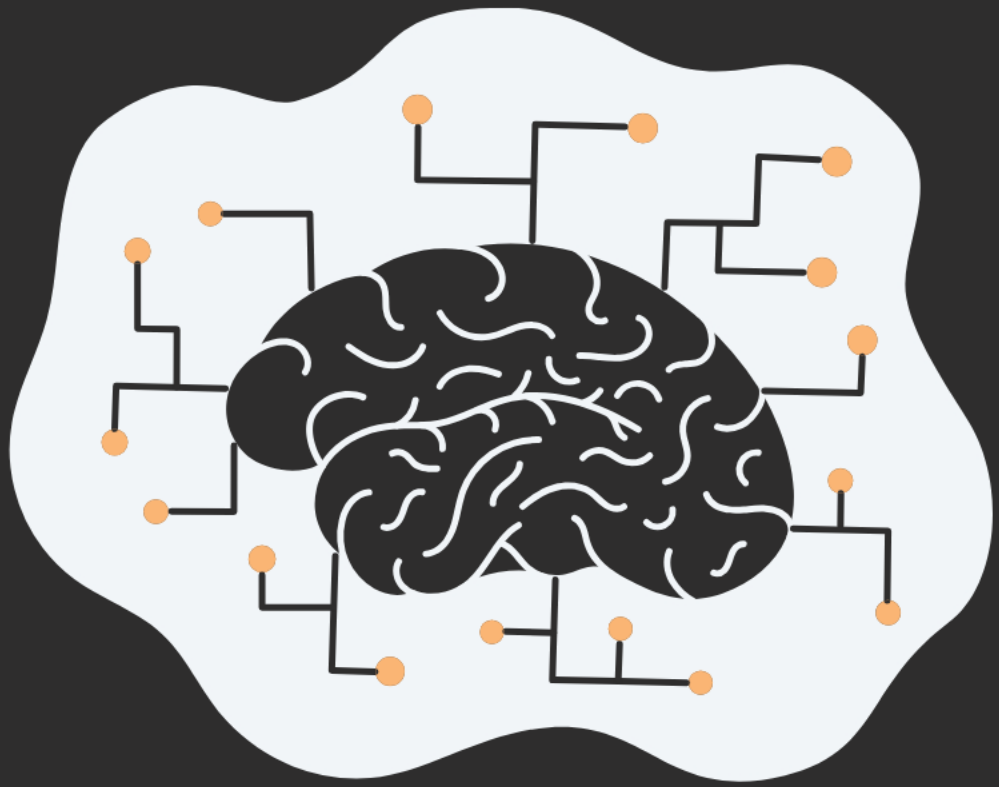
Doesn't care how many times a word appears; **just whether it's there or not**

Example

The word ***free*** shows up 5 times → **1** (Present)

Practical Example

Message	Words	Presence
Free prize now	“free”, “prize”, “now”	**free:** 1 **prize:** 1 **now:** 1
Free free unlimited offer	“free”, “offer”	**free:** 1 **unlimited:** 1 **offer:** 1
Hello friend, hello	“hello”, “friend”	**hello:** 1 **friend:** 1



Gaussian Naive Bayes

Normal Distribution

Numerical data that follows a **normal distribution**

Use Cases

Continuous numbers, sensor readings, pixel intensities

Practical Example

Message	Words	TF-IDF Scores
Free prize now	“free”, “prize”, “now”	**free:** 0.82 **prize:** 0.54 **now:** 0.36
Free free unlimited offer	“free”, “offer”	**free:** 0.91 **unlimited:** 0.62 **offer:** 0.48
Hello friend, hello	“hello”, “friend”	**hello:** 0.73 **friend:** 0.41

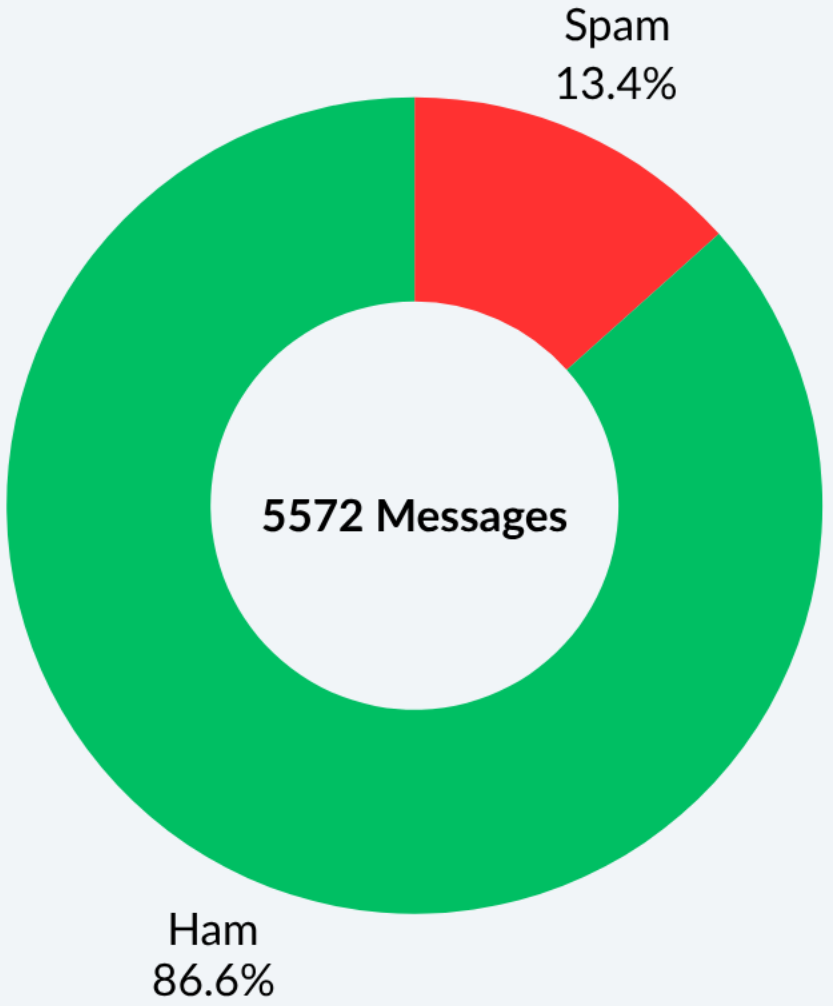
Final Recap

Model	Input Type	What it Looks for
MultinomialNB	Word Counts	**Frequency of each word**
BernoulliNB	Binary Values	**Whether a word appears or not**
GaussianNB	Continuous Values	**Statistical distribution of features**

SMS Spam Dataset

Sample Dataset

- **Ham:** U dun say so early hor... U c already then say...
- **Ham:** Ok lar... Joking wif u oni...
- **Spam:** Free entry in 2 a wkly comp to win FA Cup final tkts 21st May 2005.
- **Spam:** WINNER!! As a valued network customer you have been selected to receivea £900 prize



Report for MultinomialNB

Accuracy Score: 98.68%				
	precision	recall	f1-score	support
0	0.99	1.00	0.99	1448
1	**0.97**	**0.93**	**0.95**	224
accuracy			0.99	1672
macro avg	0.98	0.96	0.97	1672
weighted avg	0.99	0.99	0.90	1672

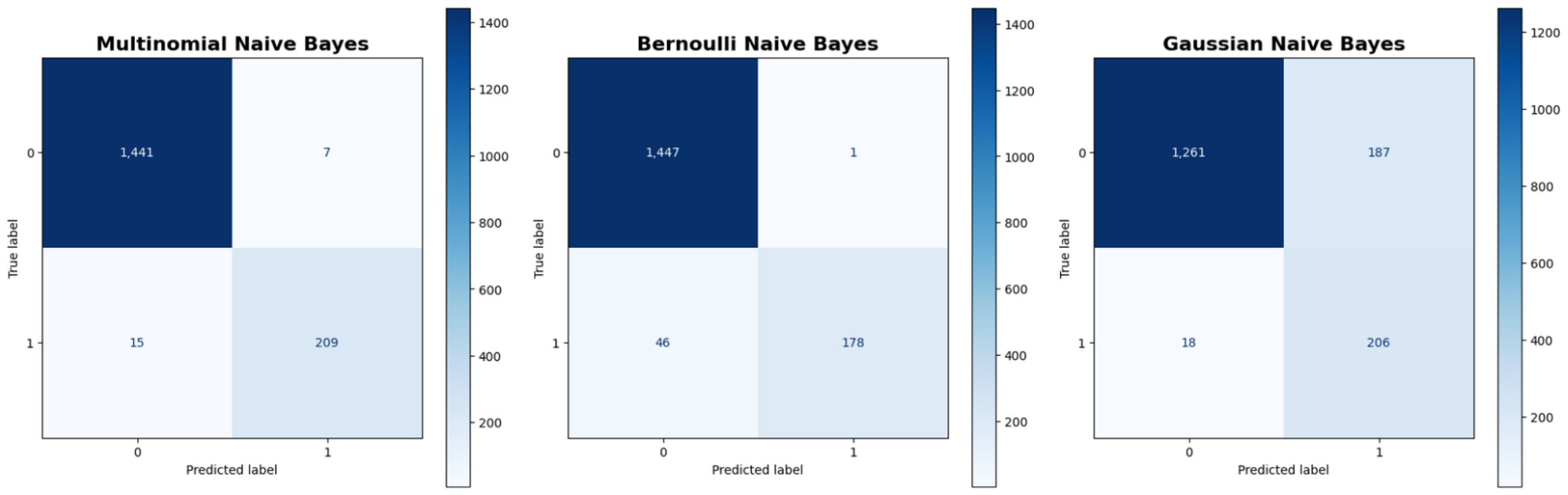
Report for BernoulliNB

Accuracy Score: 97.19%				
	precision	recall	f1-score	support
0	0.97	1.00	0.98	1448
1	**0.99**	**0.79**	**0.88**	224
accuracy			0.97	1672
macro avg	0.98	0.9	0.93	1672
weighted avg	0.97	0.97	0.97	1672

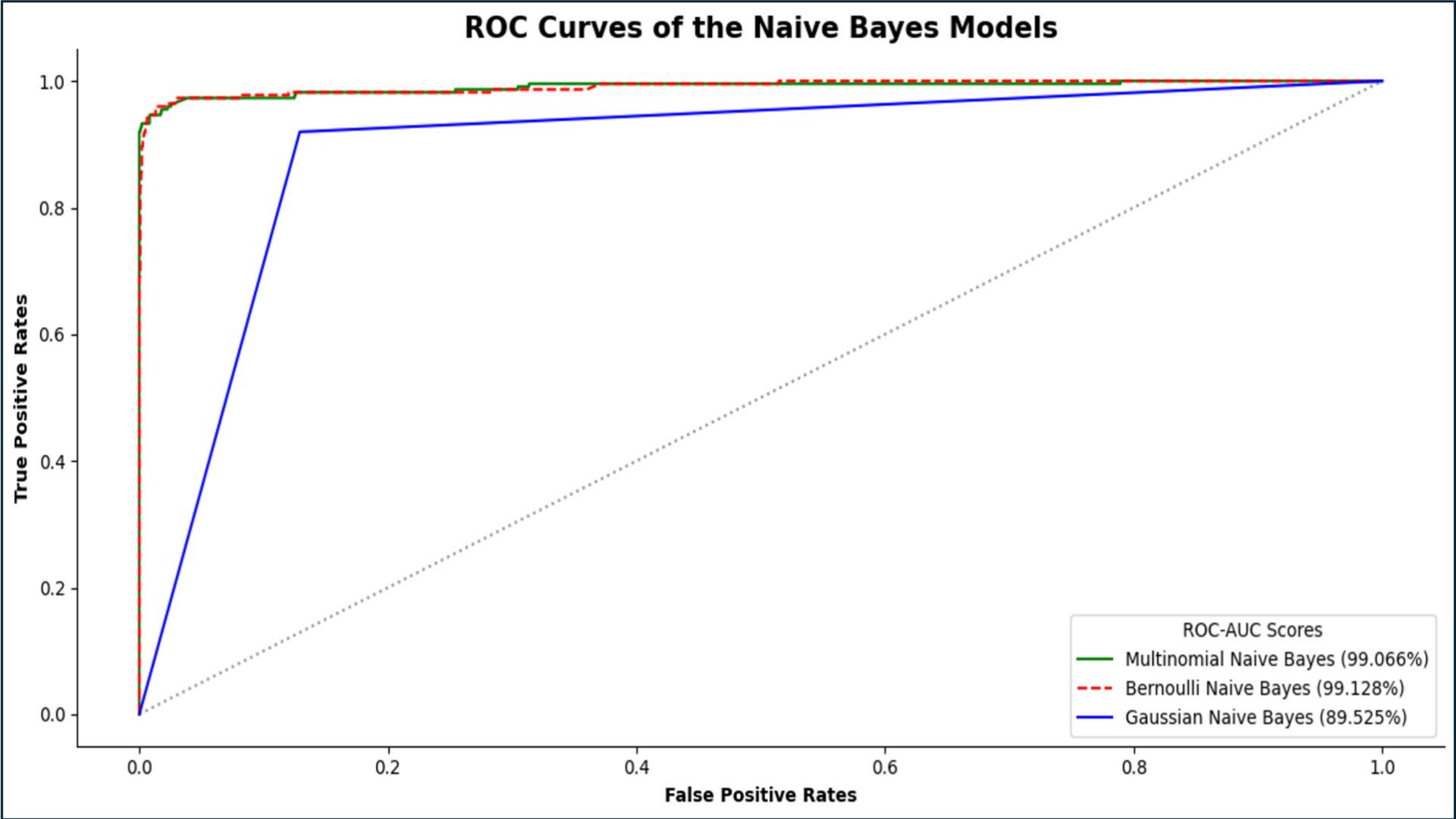
Report for GaussianNB

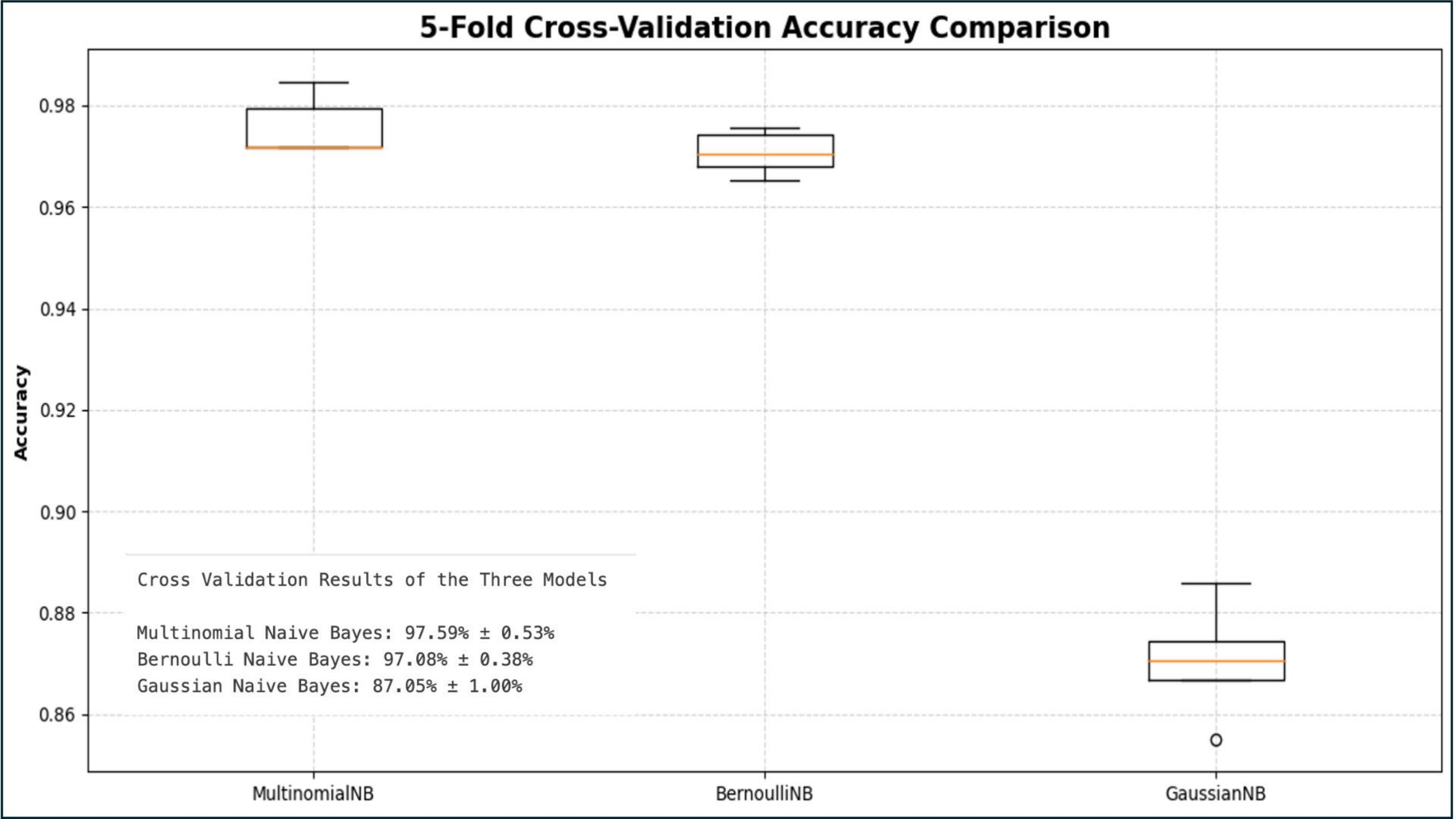
Accuracy Score: 87.74%				
	precision	recall	f1-score	support
0	0.99	9.87	0.92	1448
1	**0.52**	**0.92**	**0.67**	224
accuracy			0.88	1672
macro avg	0.76	0.90	0.8	1672
weighted avg	0.52	0.98	0.89	1672

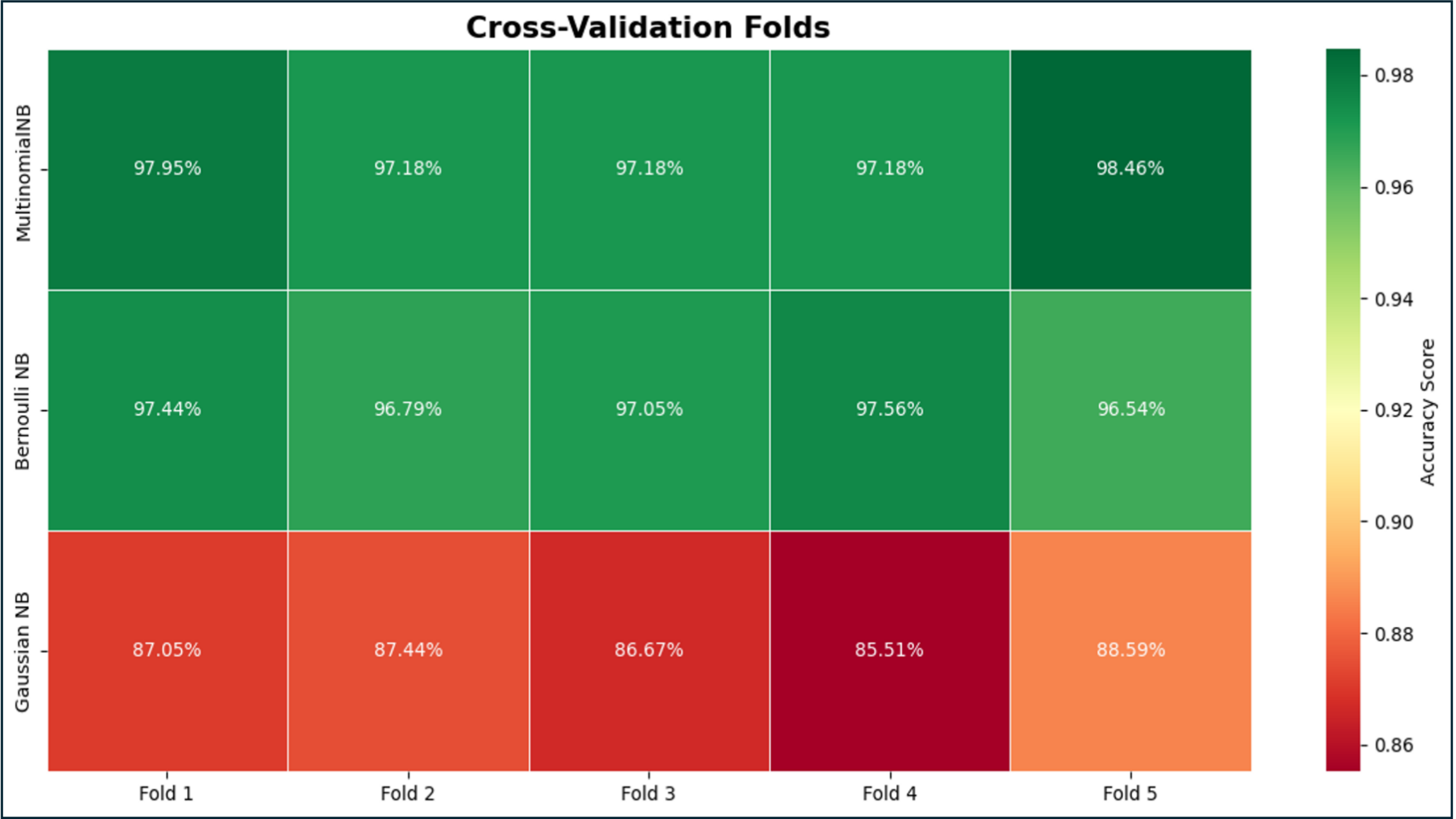
Confusion Matrices

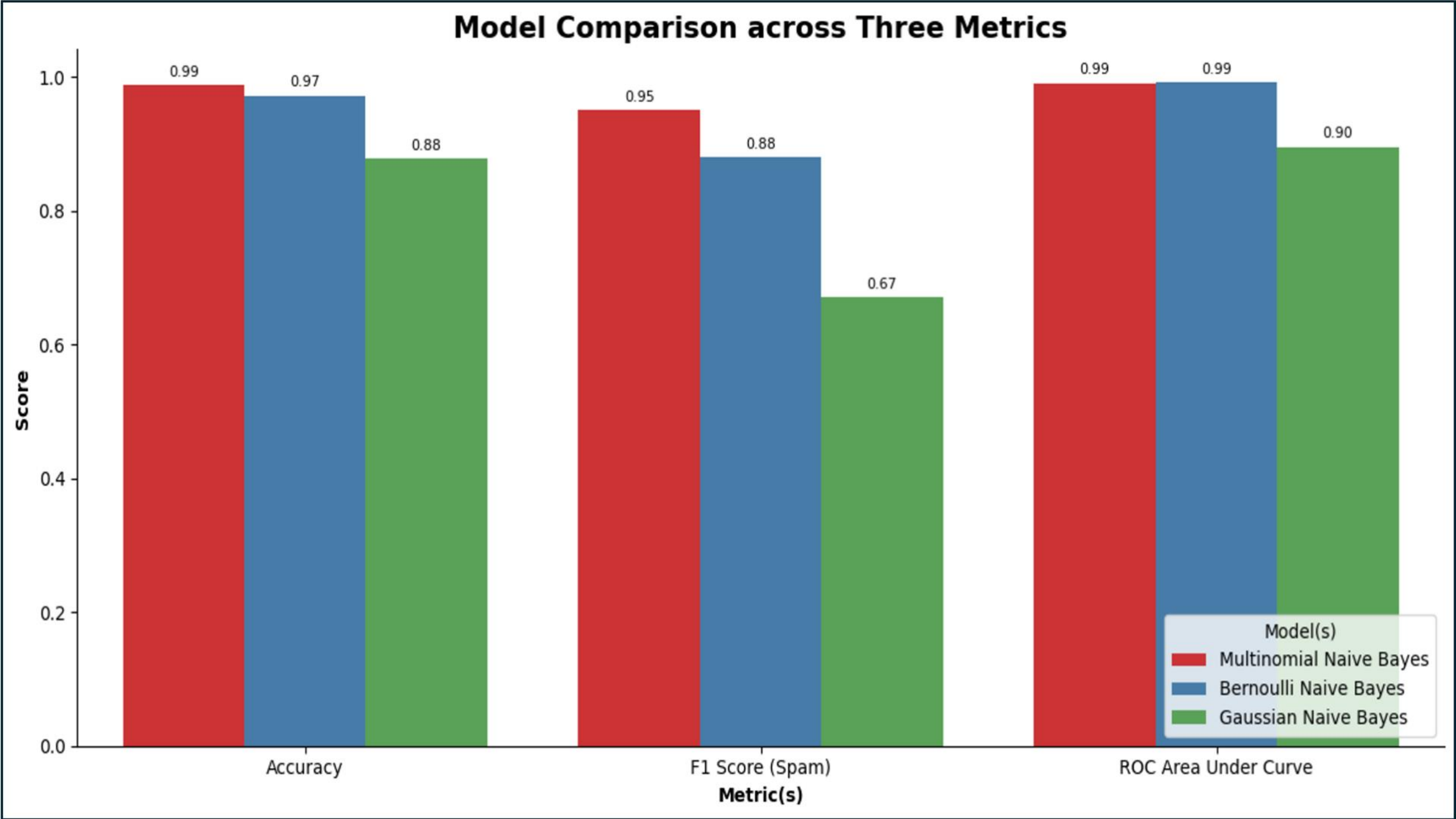


- **Minimize False Positives** → Bernoulli Naive Bayes
- **Minimize False Negatives**→ Multinomial Naive Bayes
- **Overall Best Model** → Multinomial Naive Bayes
- **Overall Worst Model** → Gaussian Naive Bayes


ROC Curves of the Naive Bayes Models
True Positive Rates
False Positive Rates
0.0
0.2
0.4
0.6
0.8
1.0
ROC-AUC Scores
Multinomial Naive Bayes (99.066%)
Bernoulli Naive Bayes (99.128%)
Gaussian Naive Bayes (89.525%)


5-Fold Cross-Validation Accuracy Comparison
Accuracy
0.98
0.96
0.94
0.92
0.90
0.88
0.86
Cross Validation Results of the Three Models
Multinomial Naive Bayes: 97.59% ± 0.53%
Bernoulli Naive Bayes: 97.08% ± 0.38%
Gaussian Naive Bayes: 87.05% ± 1.00%
MultinomialNB
BernoulliNB
GaussianNB


Cross-Validation Folds
MultinomialNB
Bernoulli NB
Gaussian NB
97.95%
97.18%
97.18%
97.18%
98.46%
97.44%
96.79%
97.05%
97.56%
96.54%
87.05%
87.44%
86.67%
85.51%
88.59%
Fold 1
Fold 2
Fold 3
Fold 4
Fold 5
Accuracy Score
0.98
0.96
0.94
0.92
0.90
0.88
0.86

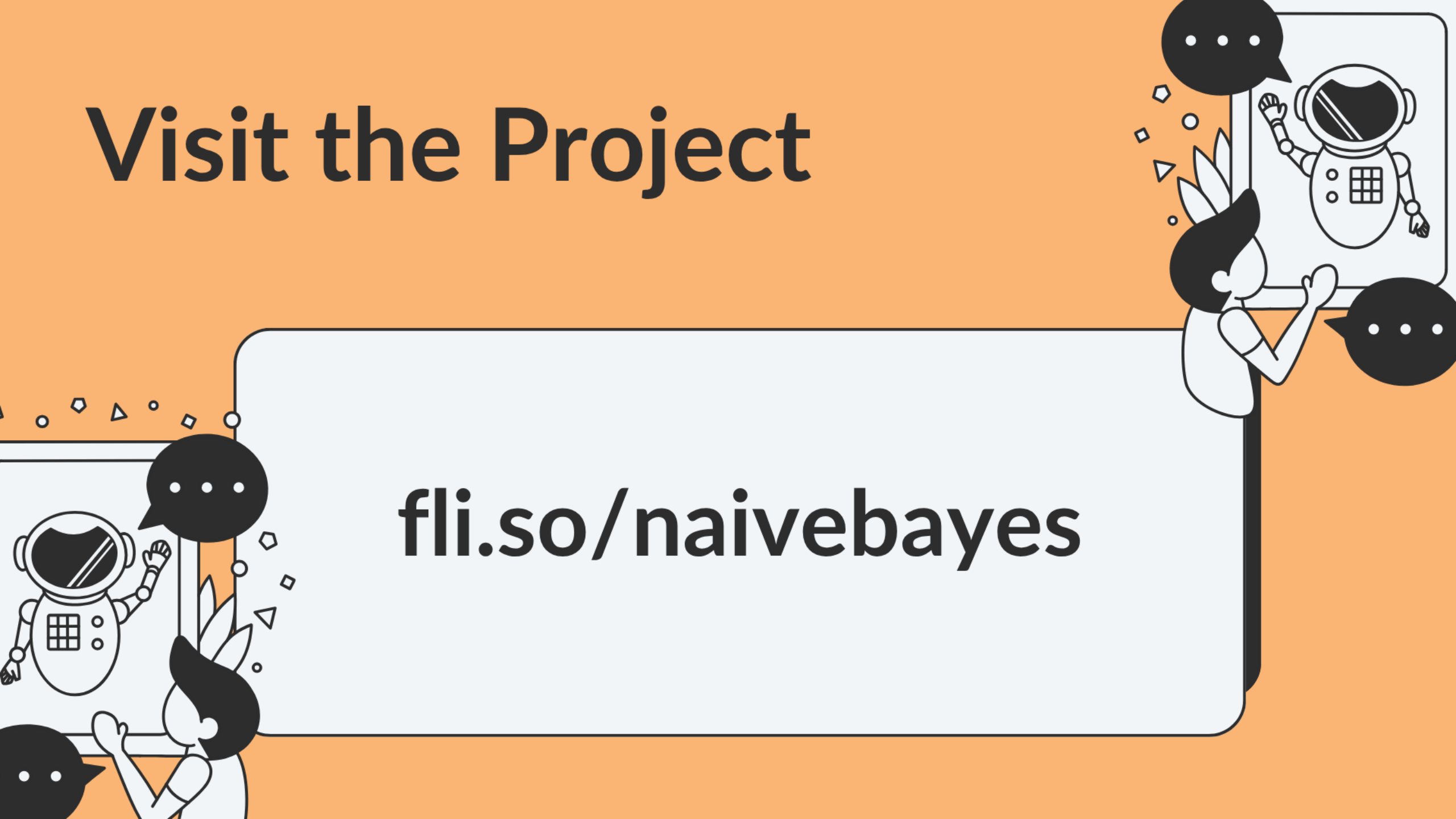

Model Comparison across Three Metrics
Score
1.0
0.8
0.6
0.4
0.2
0.0
0.99
0.97
0.88
0.95
0.88
0.67
0.99
0.99
0.90
Accuracy
F1 Score (Spam)
ROC Area Under Curve
Metric(s)
Model(s)
Multinomial Naive Bayes
Bernoulli Naive Bayes
Gaussian Naive Bayes

Best Model: Multinomial NB

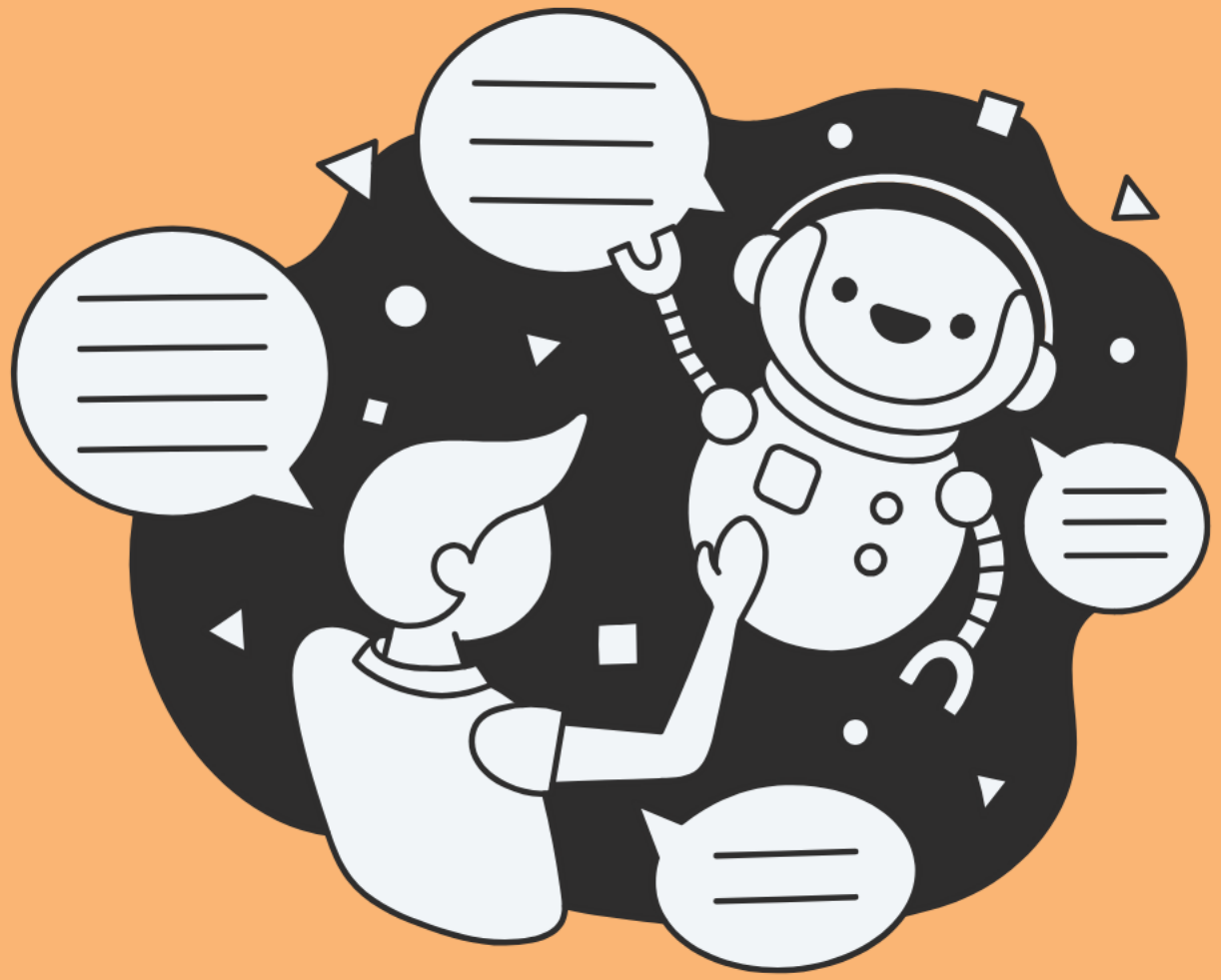
Accuracy	Precision	Recall	F1-Score
98.68%	97.00%	93.00%	95.00%

ROC-AUC	FPR Rate	FNR Rate
99.07%	0.48%	6.70%

Visit the Project



fli.so/naivebayes



The End

Thank you for listening!