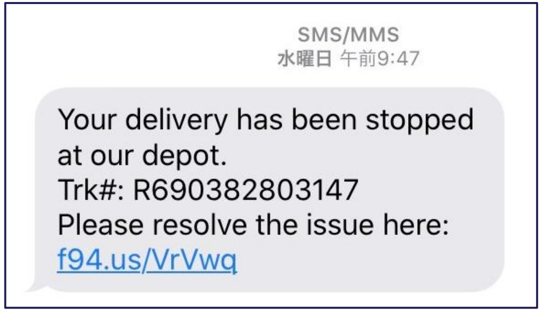
Spam Message Classification

Leo Siu-Yin

METIS

Spam and Scam Messages



- Spam: Messages sent to a large group of recipients without their prior consent.
- Usually advertise for goods/services.
- Scam messages form a high percentage of spam messages.
- Typically trick people into giving away money or personal details by offering an attractive/false deal.
- This year Jan–June: scammed amount increased by >S$8 million.

Motivation

SECURITY

Spam Message Classification:

A step towards building a tool for scam message identification and early scam detection.

Contents

Dataset

- Spam Message Collection Dataset from Kaggle.
- 5572 rows of messages.
- All messages are classified into either spam or ham.
- 13.4% spam, 86.6% ham

Tools Used

python

NumPy

scikit learn

pandas

matplotlib

NLTK
Natural Language Toolkit

seaborn

Methodology

Data Pre-processing

1. Word tokenize
2. Convert to lower case
3. Remove punctuation except '!'
4. Remove stopwords
5. Remove words containing digits
6. Exploratory Data Analysis

Model Training
Model Comparison

1. Train Test Split data
2. CountVectorizer
3. GridSearchCV across 10 folds
4. Fit data to models
5. Comparison of model results

Model Evaluation

1. Result Summary
2. Confusion Matrix
3. Precision-Recall Curve

Data Pre-processing

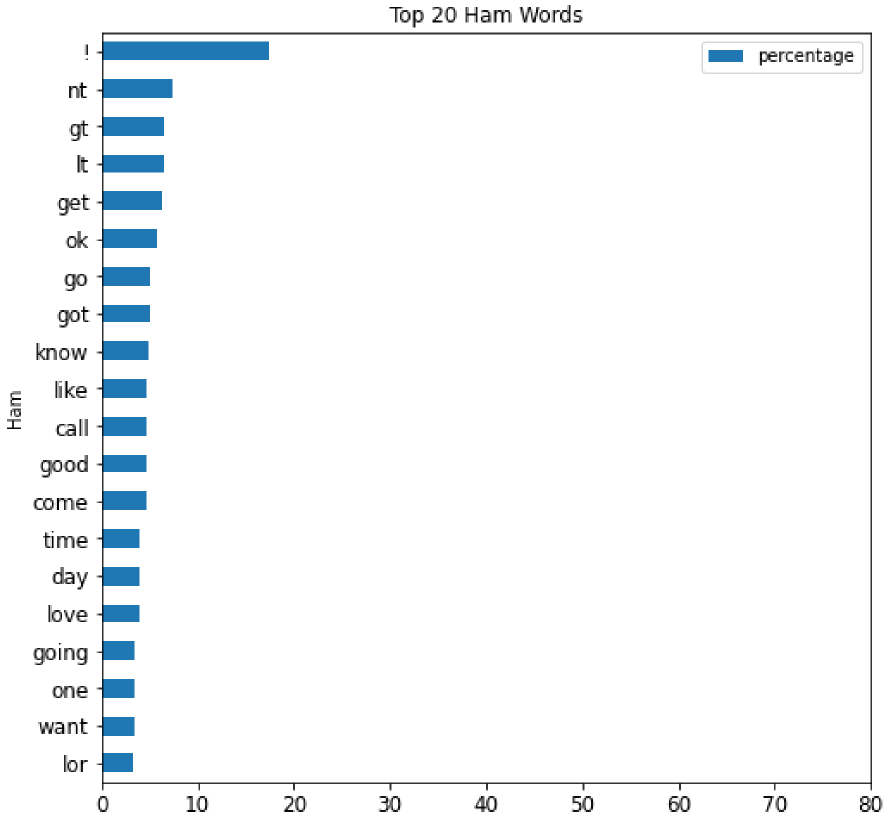
1. After Word Tokenize

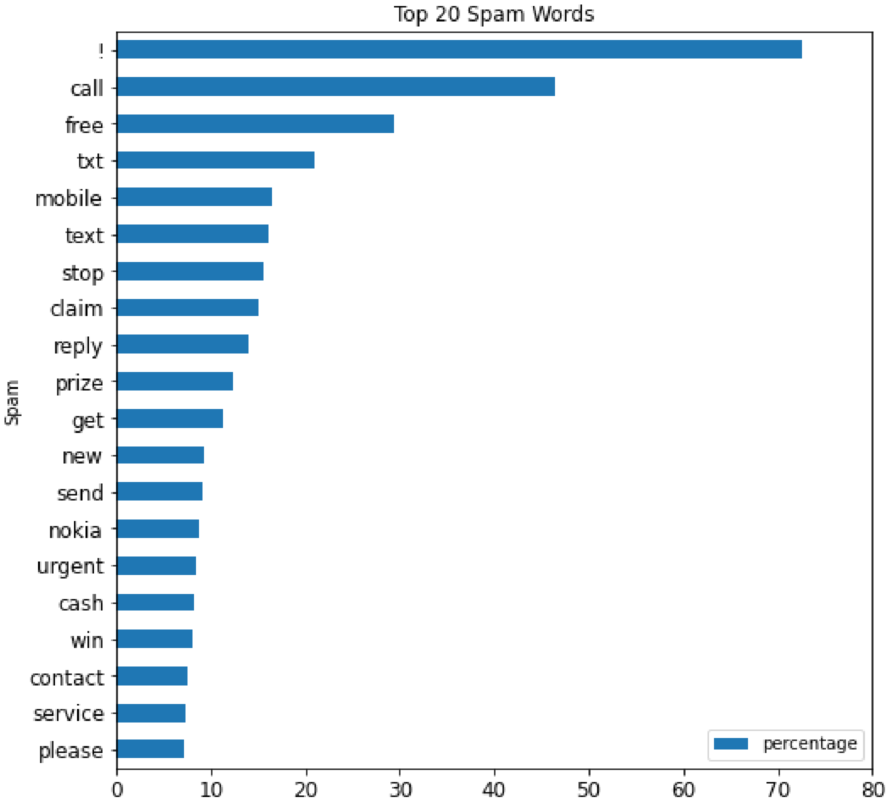
['Hello', '!', 'How', "'s", 'you', 'and', 'how', 'did', 'saturday', 'go', '?', 'I', 'was', 'just', 'texting', 'to', 'see', 'if', 'you', "'d", 'decided', 'to', 'do', 'anything', 'tomo', '.', 'Not', 'that', 'i', "'m", 'trying', 'to', 'invite', 'myself', 'or', 'anything', '!']

2. Conversion to lower case
3. Punctuation removed except '!'
4. Stopwords and words with digits removed

['hello', '!', 'saturday', 'go', 'texting', 'see', 'decided', 'anything', 'tomo', 'trying', 'invite', 'anything', '!']

Exploratory Data Analysis





Exploratory Data Analysis: Topic Modeling

Topic #0 (Ham):

nt ok like got go come good get know time love day going home sorry lor still see want da

Topic #1 (Spam):

call gt lt free txt text get mobile stop reply new claim send please number prize week message phone win

Models Used

Logistic Regression

CountVectoriser

Tfidf Vectoriser

CountVectoriser with LSA

CountVectoriser with LDA

K- Nearest Neighbours

CountVectoriser

Random Forest

CountVectoriser

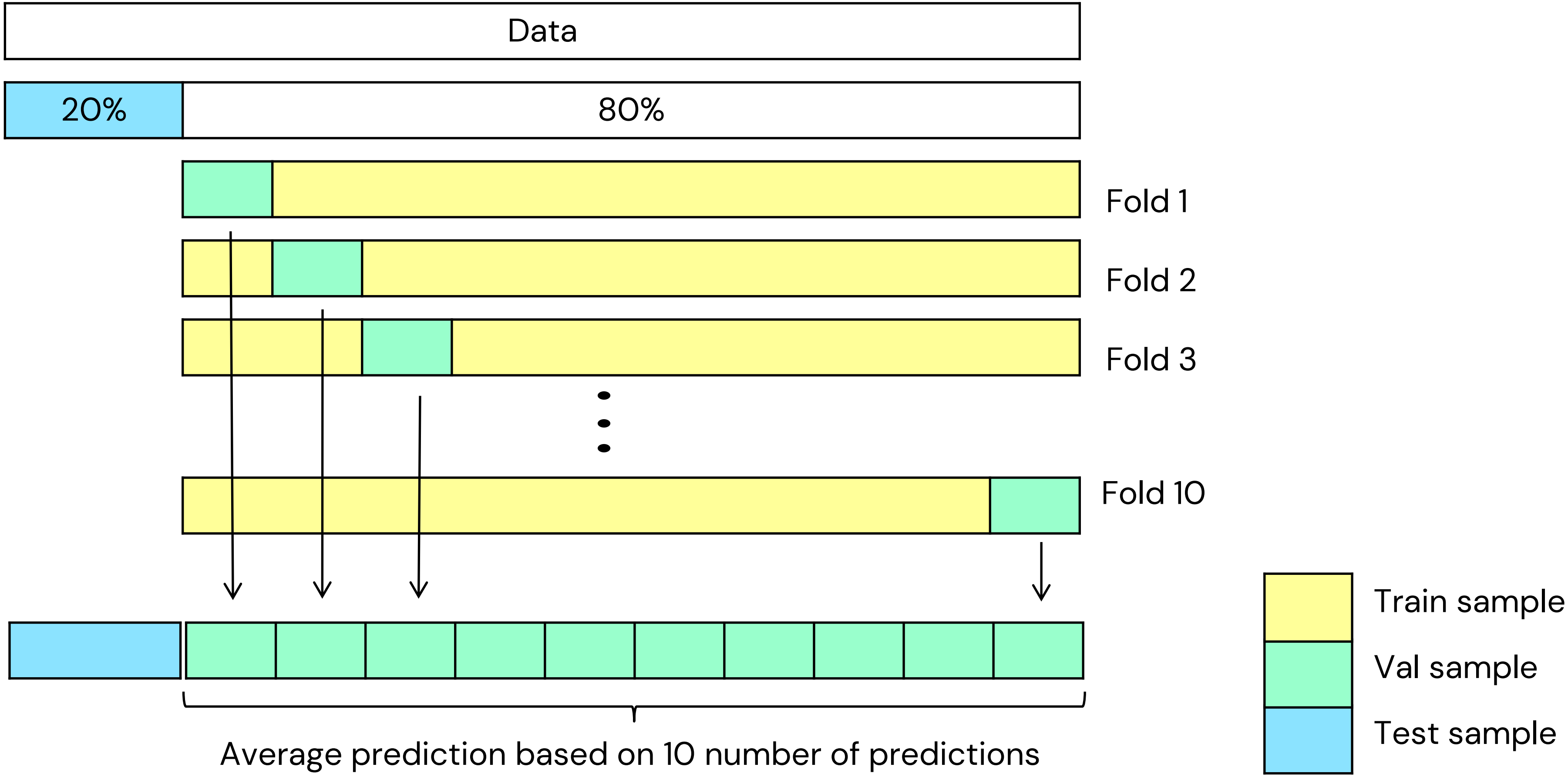
Bernoulli Naïve Bayes

CountVectoriser

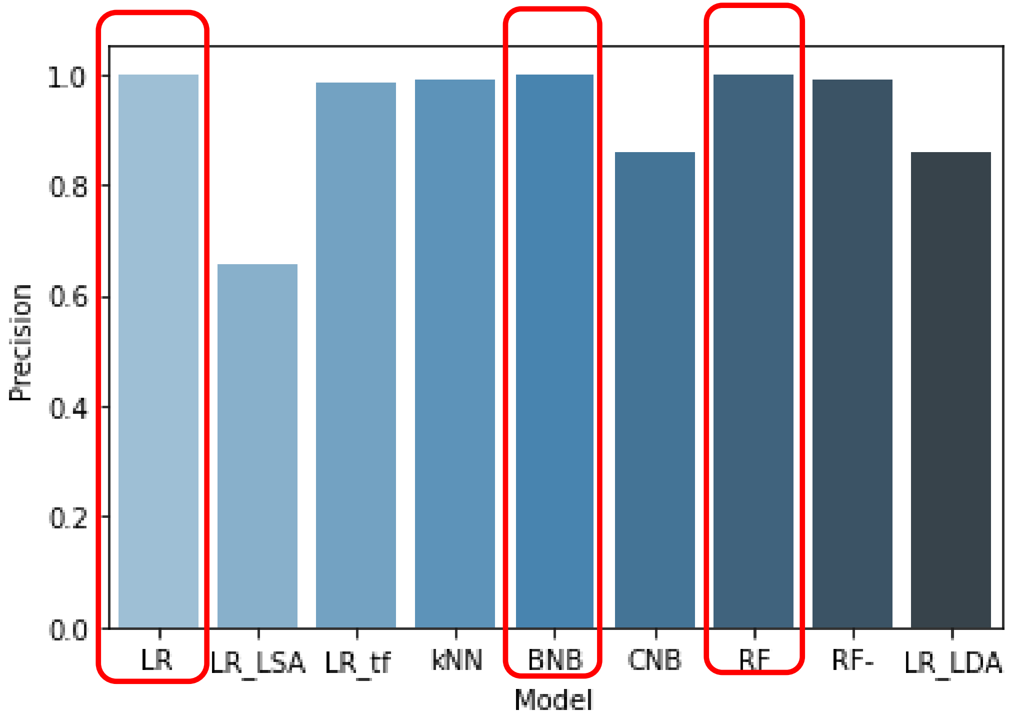
Complement Naïve Bayes

CountVectoriser

GridSearchCV

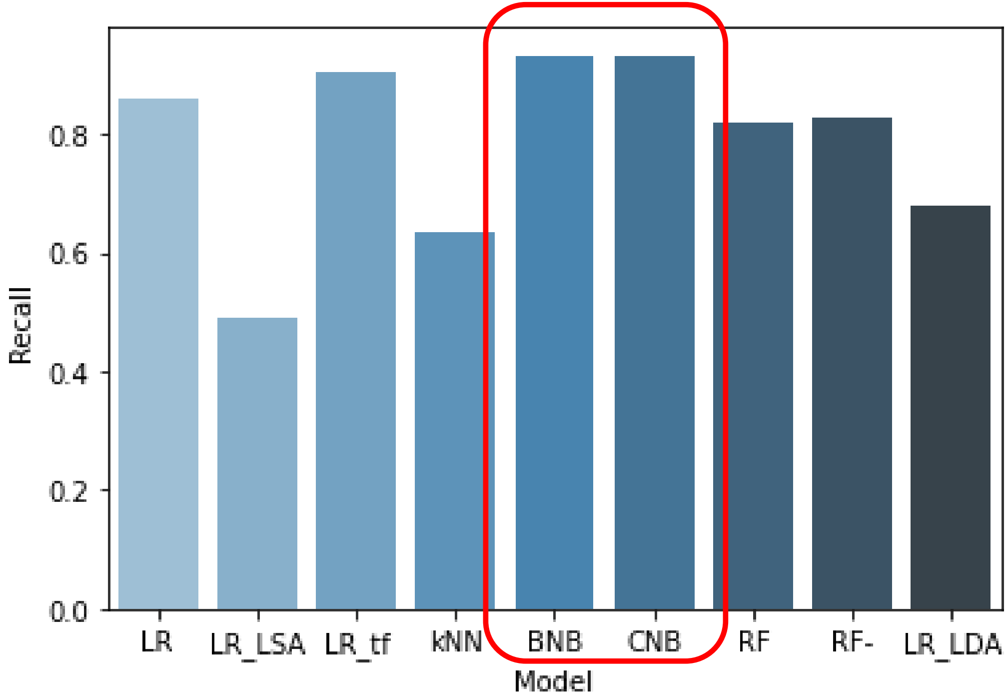


Modeling Results and Comparison: Precision



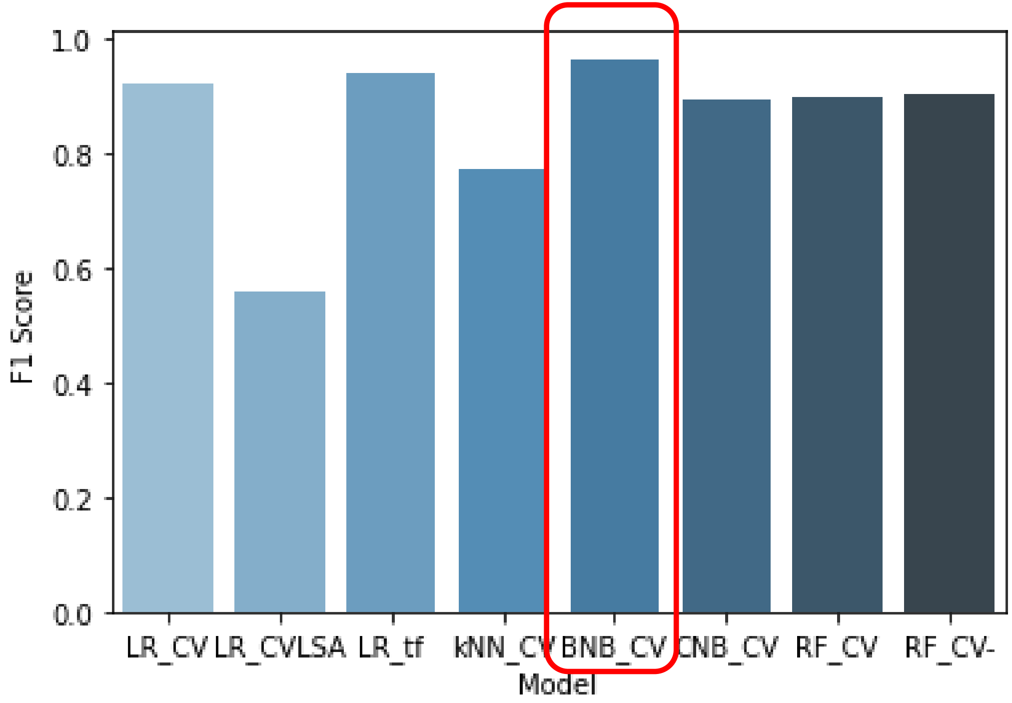
Model	LR	LR_LSA	LR_tf	kNN	BNB	CNB	RF	RF-	LR_LDA
Precision	1	0.66	0.98	0.99	1	0.86	1	0.99	0.86

Modeling Results and Comparison: Recall



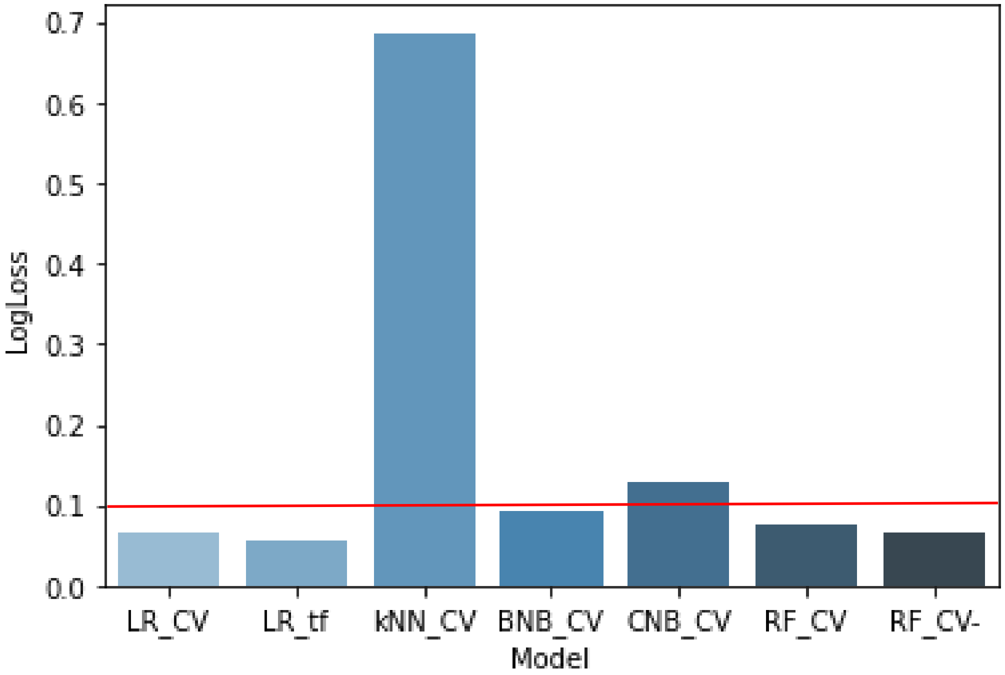
Model	LR	LR_LSA	LR_tf	kNN	BNB	CNB	RF	RF-	LR_LDA
Recall	0.86	0.49	0.90	0.63	0.93	0.93	0.82	0.83	0.68

Modeling Results and Comparison: F1 Score



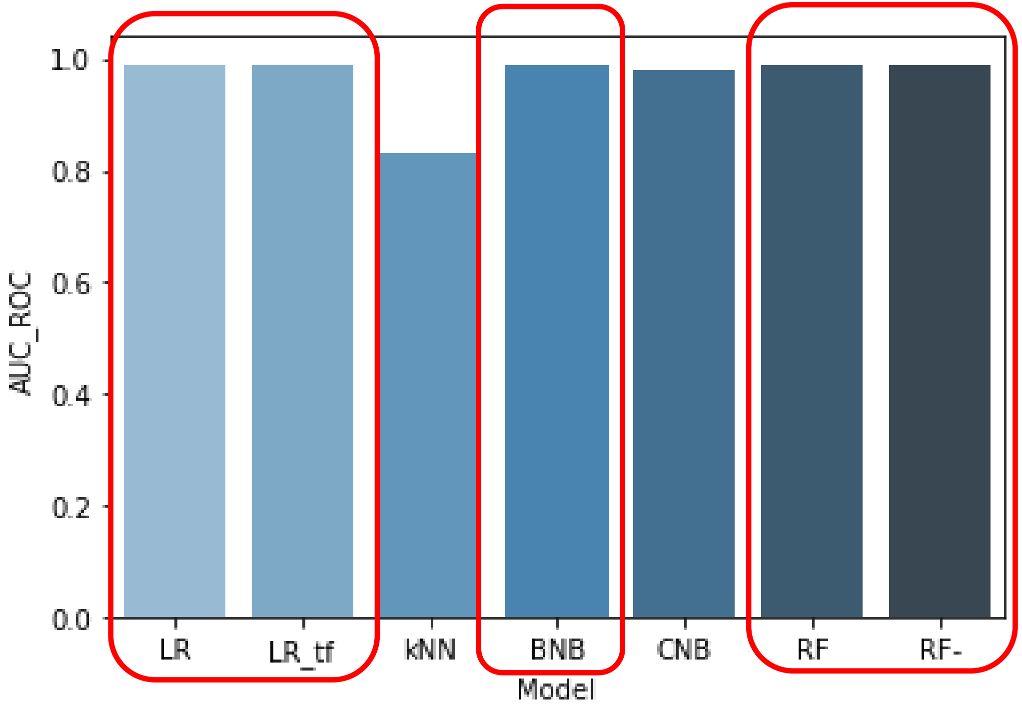
Model	LR	LR_LSA	LR_tf	kNN	BNB	CNB	RF	RF-	LR_LDA
F1 Score	0.92	0.56	0.94	0.77	0.97	0.90	0.90	0.90	0.76

Modeling Results and Comparison: Log Loss



Model	LR	LR_tf	kNN	BNB	CNB	RF	RF-
LogLoss	0.07	0.06	0.69	0.09	0.13	0.07	0.07

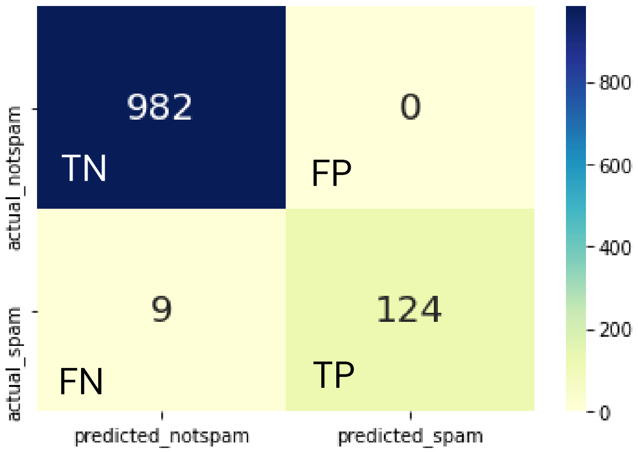
Modeling Results and Comparison: AUC ROC

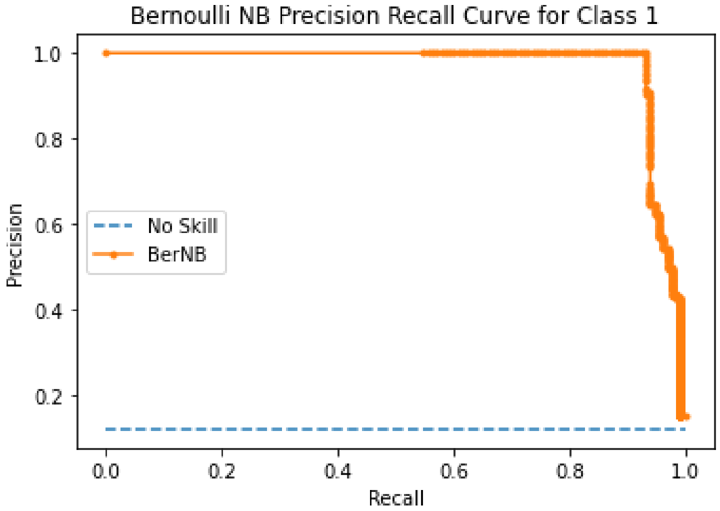


Model	LR	LR_tf	kNN	BNB	CNB	RF	RF_CV-
AUC_ROC	0.99	0.99	0.83	0.99	0.98	0.99	0.99

Result Summary

Model	LR	LR_LSA	LR_tf	kNN	BNB	CNB	RF	RF-	LR_LDA
Precision	1	0.66	0.98	0.99	1	0.86	1	0.99	0.86
Recall	0.86	0.49	0.90	0.63	0.93	0.93	0.82	0.83	0.68
F1 Score	0.92	0.56	0.94	0.77	0.97	0.90	0.90	0.90	0.76
LogLoss	0.07	N.A.	0.06	0.69	0.09	0.13	0.07	0.07	N.A.
AUC ROC	0.99	N.A.	0.99	0.83	0.99	0.98	0.99	0.99	N.A.





Conclusion

- A more customised pre-processing step is important for precision.

- Logistic Regression and Naive Bayes models have performed better than other models.

- A model with 100% precision has been built.

Future Work

1. Investigate performance of models based on Tfidf vectoriser.

2. Use word embeddings trained with neural-network for classification.

THANKS!

Do you have any questions?
syleo22@gmail.com
https://github.com/syleo22/SiuYin_Projects
https://www.linkedin.com/in/syleo/

CREDITS: This presentation template was created by **Slidesgo**, including icons by **Flaticon**, infographics & images by **Freepik** and illustrations by Stories

Please keep this slide for attribution.