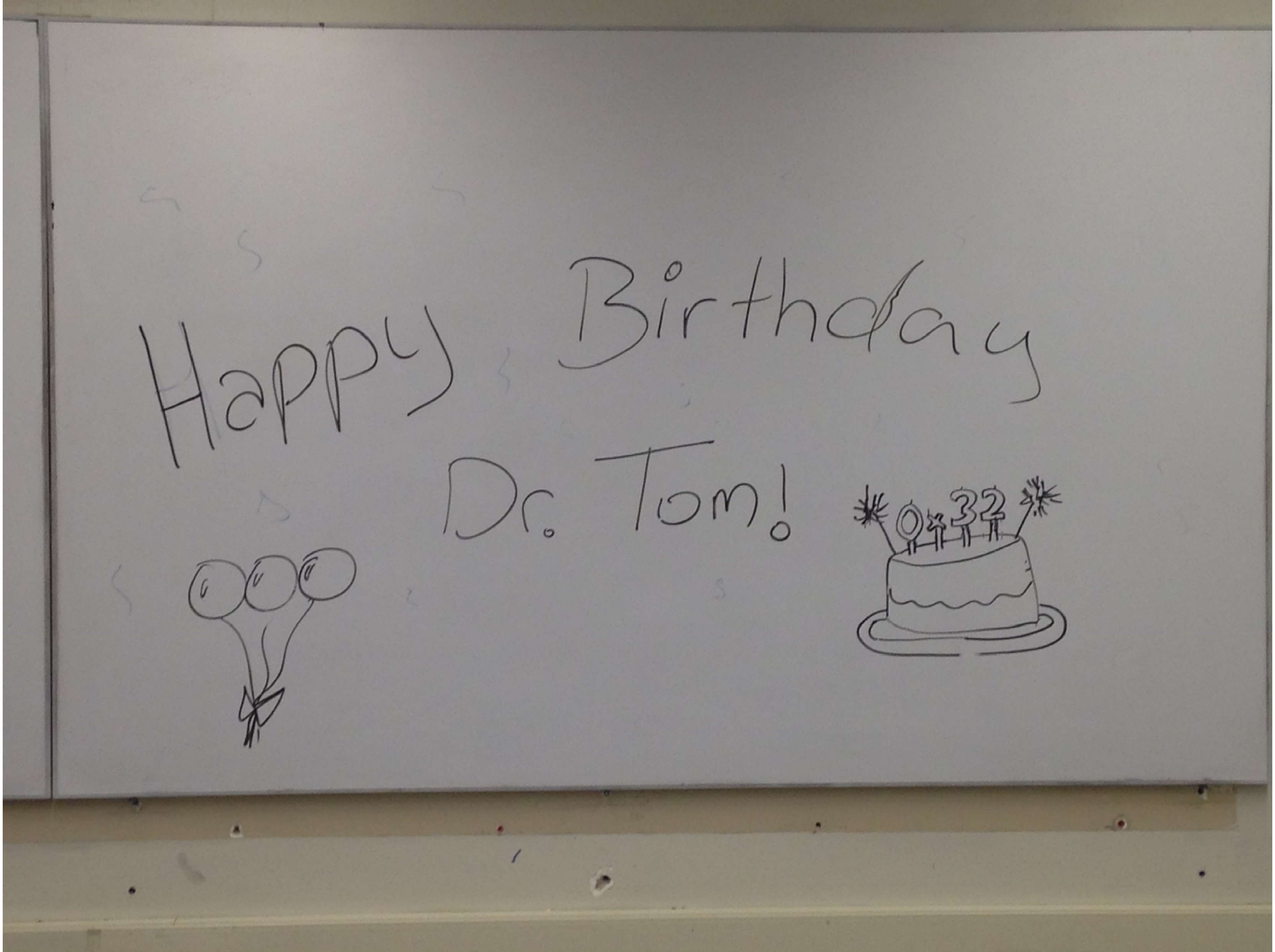
A Hybrid Approach to Email Spam Detection-Random Forest and Sentiment Analysis

Shuyan Liu
sl10158@nyu.edu

Eleftheria Pissadaki, PhD

ep3041@nyu.edu

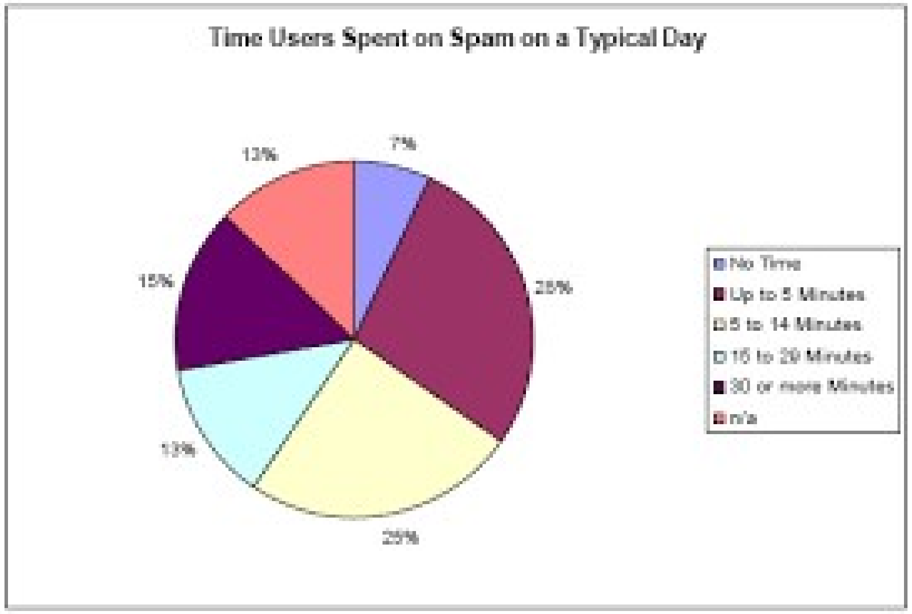
Thomas M. Schmidt, D. Sc.
tms493@nyu.edu

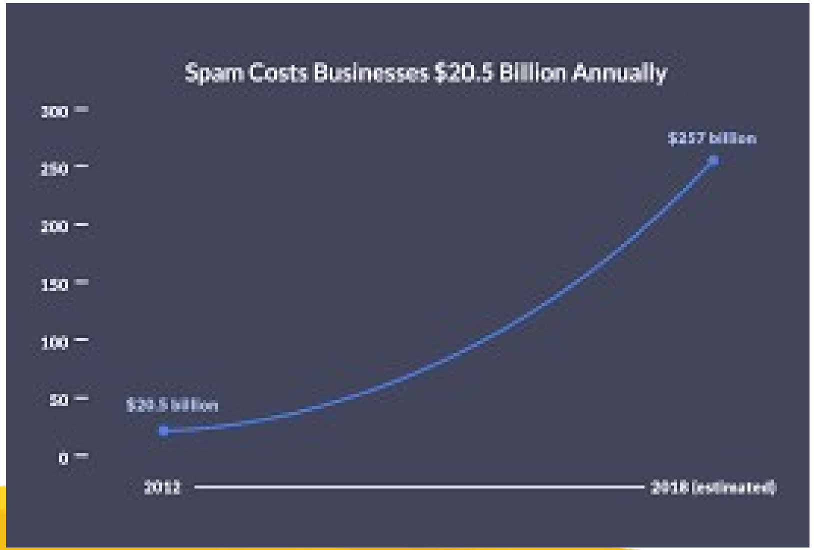

Happy Birthday
Dr. Tom!
0x32

Agenda

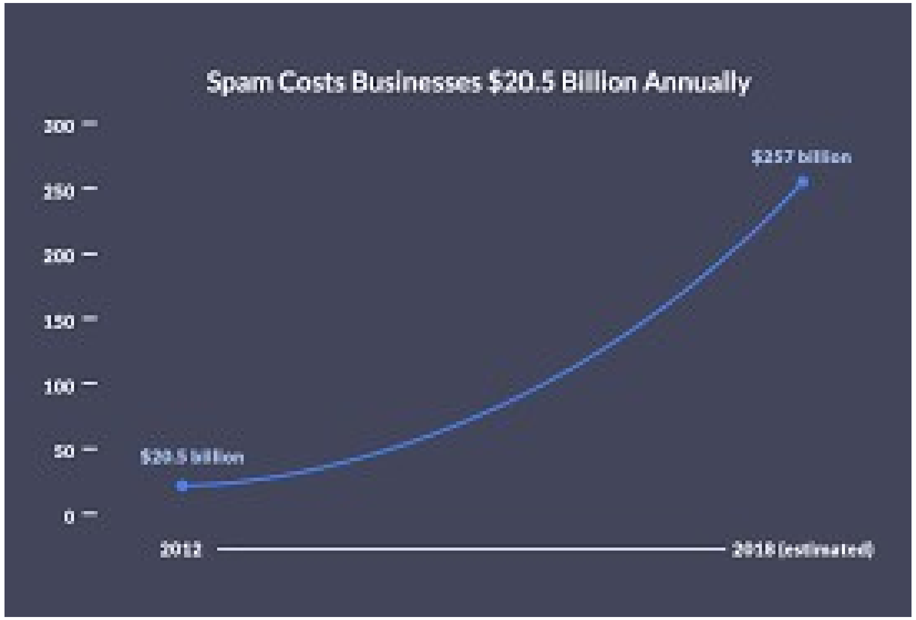
- Spam Detection
- Types of Errors
- Pattern Recognition
- Classifiers
- Sentiment Analysis
- Building the model
- Results
- Future Directions

The Spam Problem



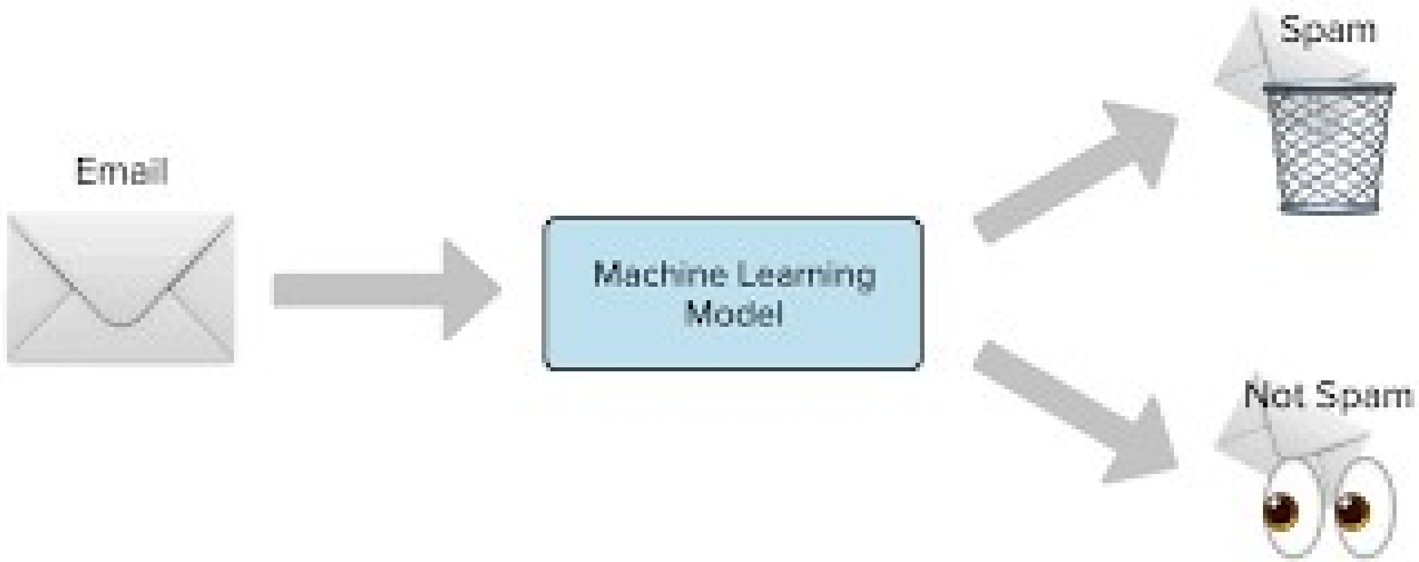


The Spam Problem

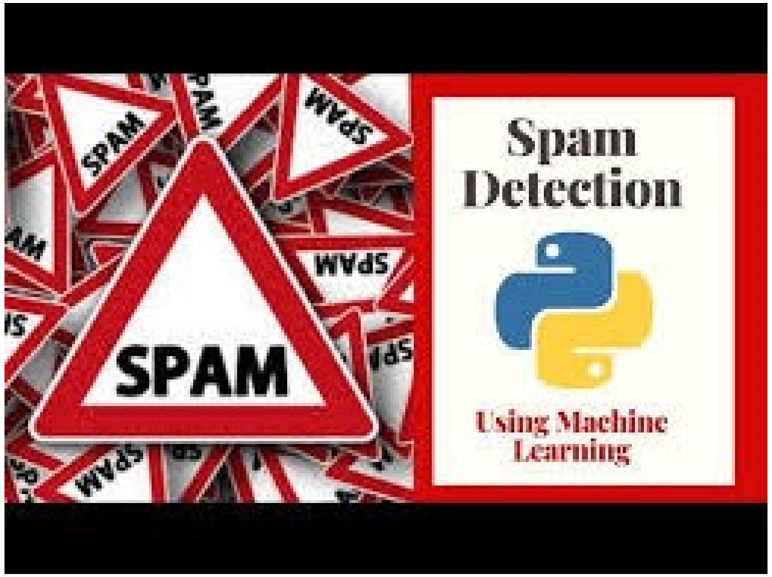


Proof of Identity

Email Spam Detection



Email Spam Detection: ML problem



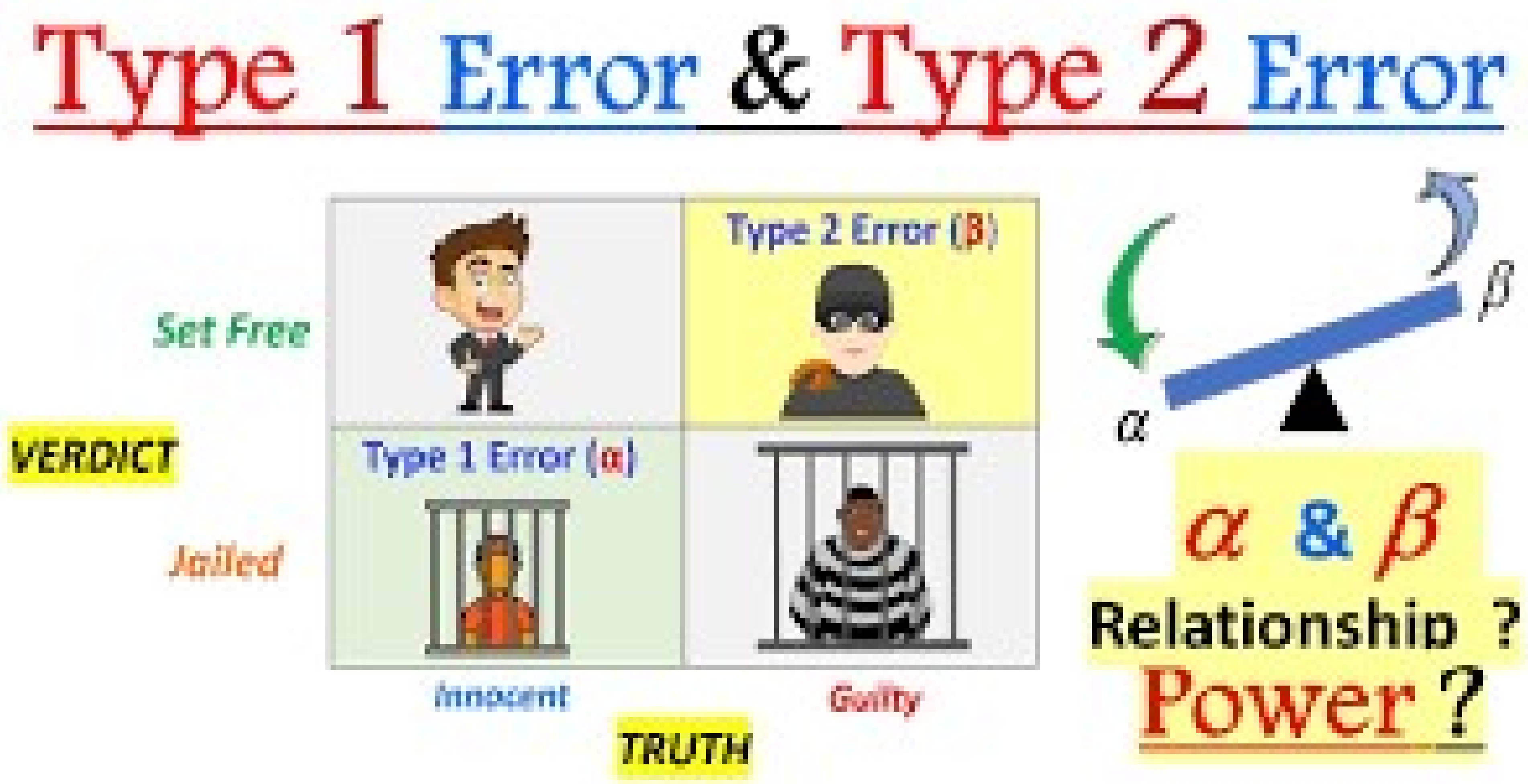
PGP


HOW TO USE PGP TO VERIFY THAT AN EMAIL IS AUTHENTIC:
LOOK FOR THIS TEXT AT THE TOP:
REPLY
-----BEGIN PGP SIGNED MESSAGE-----
HASH: SHA256
HEY,
IF IT'S THERE, THE EMAIL IS PROBABLY FINE.

Type 1 and Type 2 Errors

	Null Hypothesis is TRUE	Null Hypothesis is FALSE
Reject null hypothesis	Type I Error (False positive)	Correct Outcome! (True positive)
Fail to reject null hypothesis	Correct Outcome! (True negative)	Type II Error (False negative)

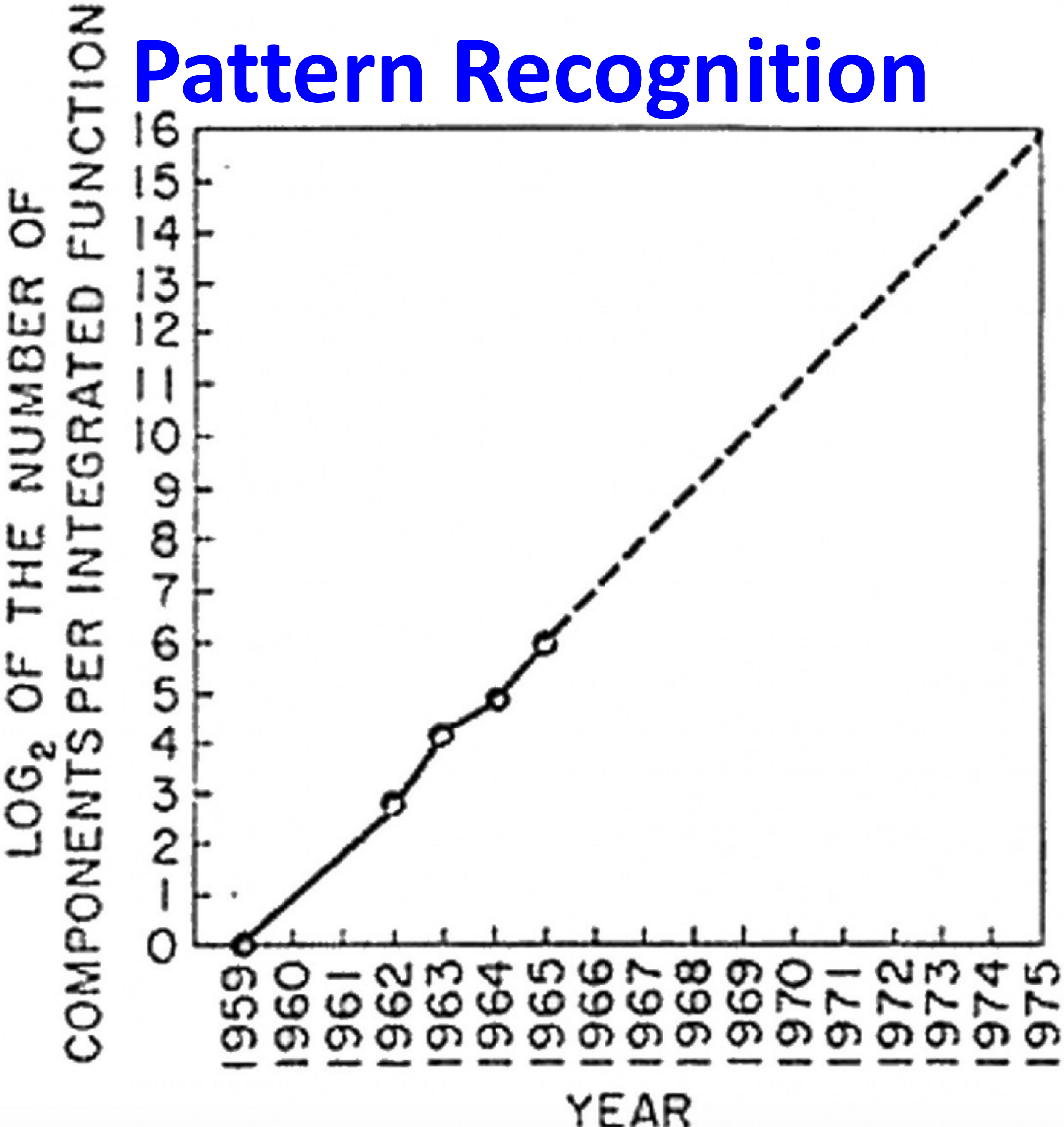
Type 1 and Type 2 Errors



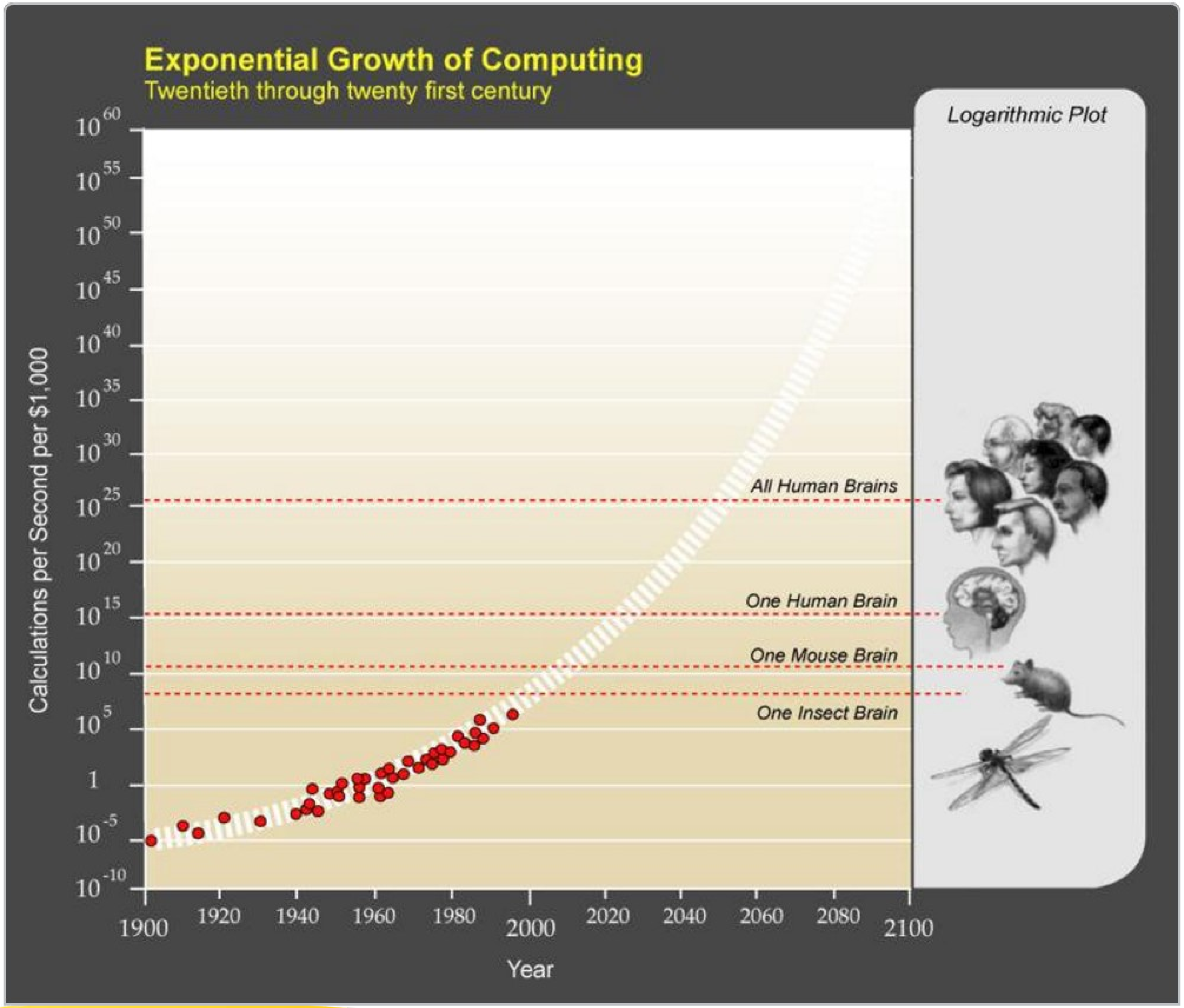
Type 1 and Type 2 Errors, Spam Detection

	Null Hypothesis (H_0)	
Machine Learning Classifier	Actual Spam Email	Actual Not-Spam Email
Predicted Spam (Action: Delete)	222 (True Negative)	32 (False Negative) TYPE II ERROR
Predicted Not-Spam (Action: Keep)	22 (False Positive) TYPE I ERROR	39 (True Positive)
Sum	244	71

Pattern Recognition



Is the Pattern Clearer now?



Naïve classification?

Naïve Bayes

$$P(c \mid x) = \frac{P(x \mid c)P(c)}{P(x)} \qquad (1)$$

Naïve Bayes

$$P(c \mid x) = \frac{P(x \mid c)P(c)}{P(x)} \qquad (1)$$

Naïve Bayes

$$P(c \mid x) = \frac{P(x \mid c)P(c)}{P(x)} \qquad (1)$$

Naïve Bayes

$$P(c \mid x) = \frac{P(x \mid c)P(c)}{P(x)} \qquad (1)$$

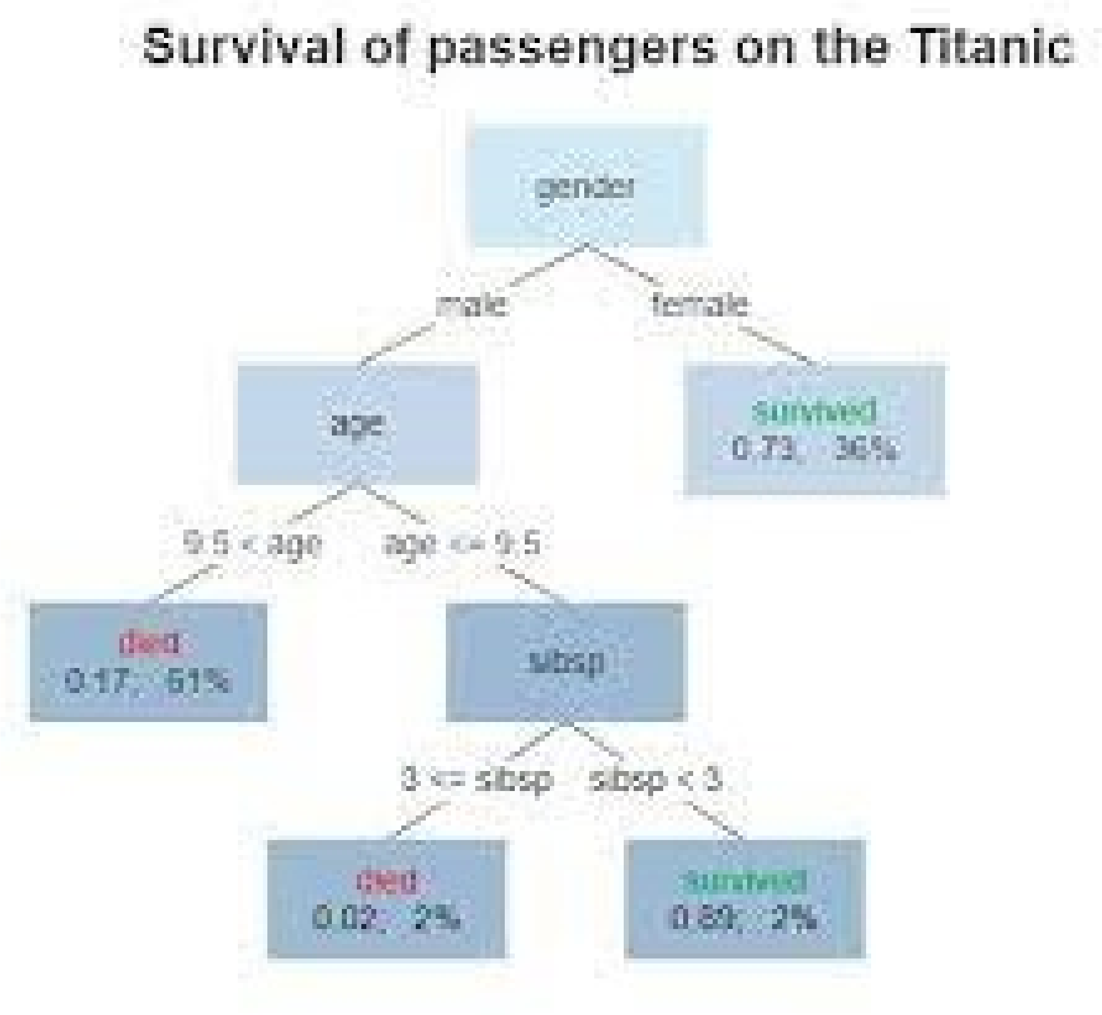
Naïve Bayes

$$P(c \mid x) = \frac{P(x \mid c)P(c)}{P(x)} \qquad (1)$$

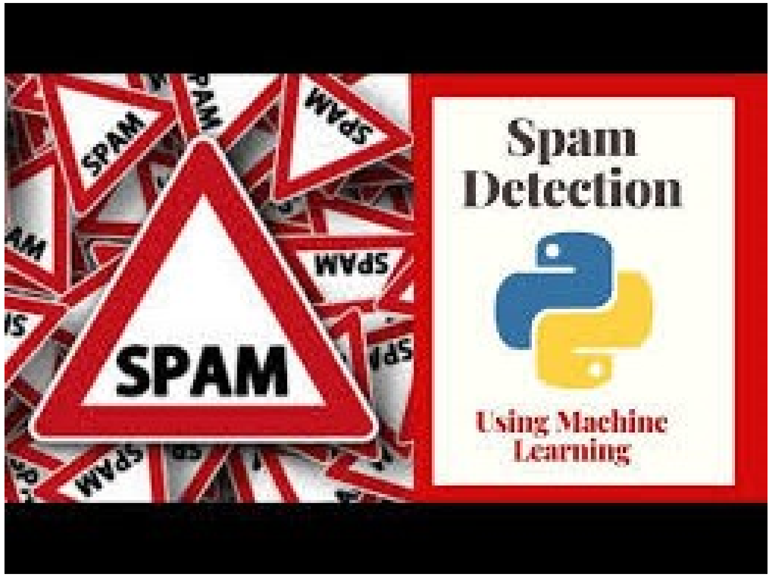
Naïve Bayes

$$P(c \mid x) = \frac{P(x \mid c)P(c)}{P(x)} \qquad (1)$$

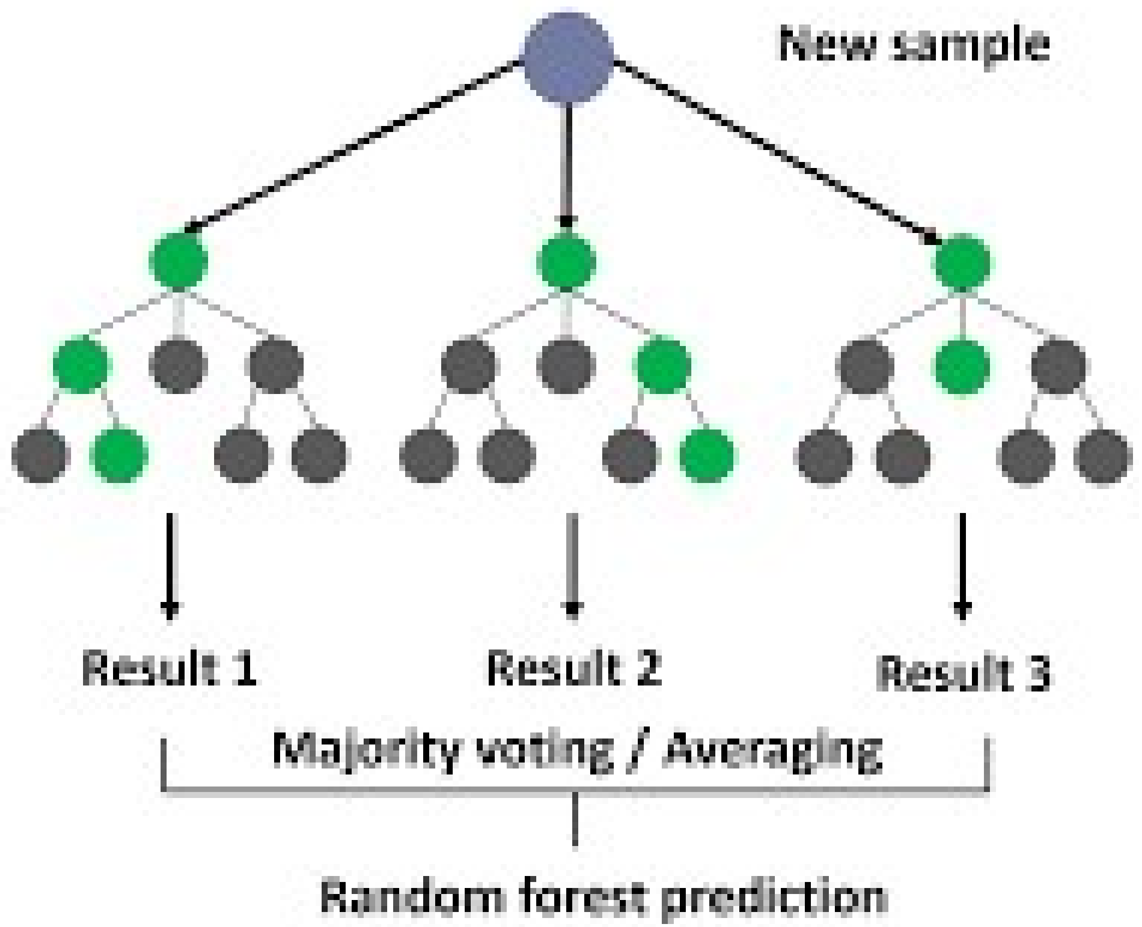
Titanic Decision Tree



Email Spam Detection: ML problem

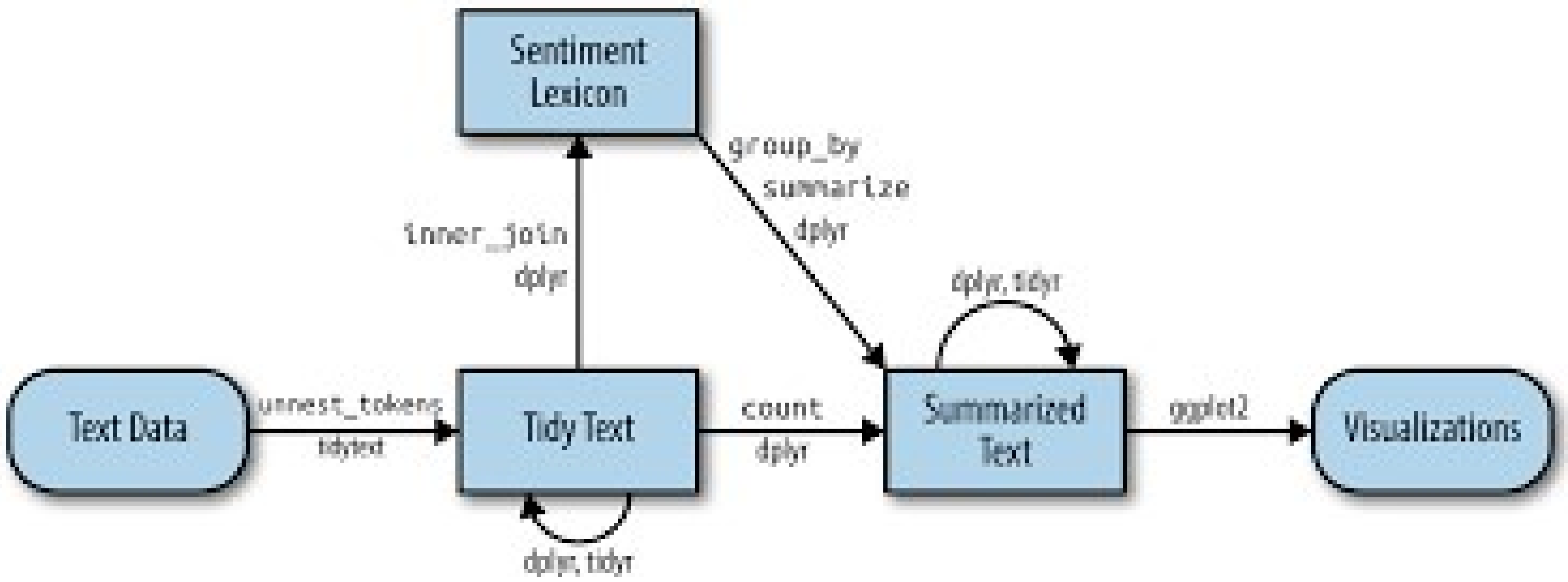


Random Forest as ML?



Sentiment Analysis

Sentiment Analyzer



Model Build Flow Chart

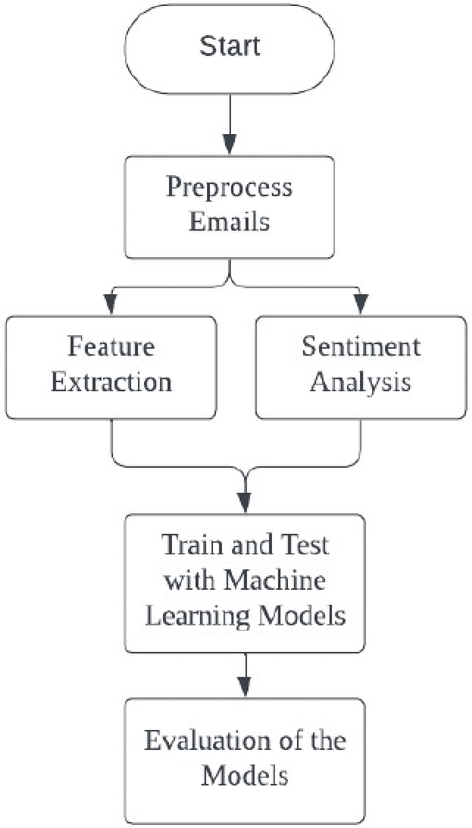


Fig. 1. Model Build Flow Chart

Model Build Flow Chart

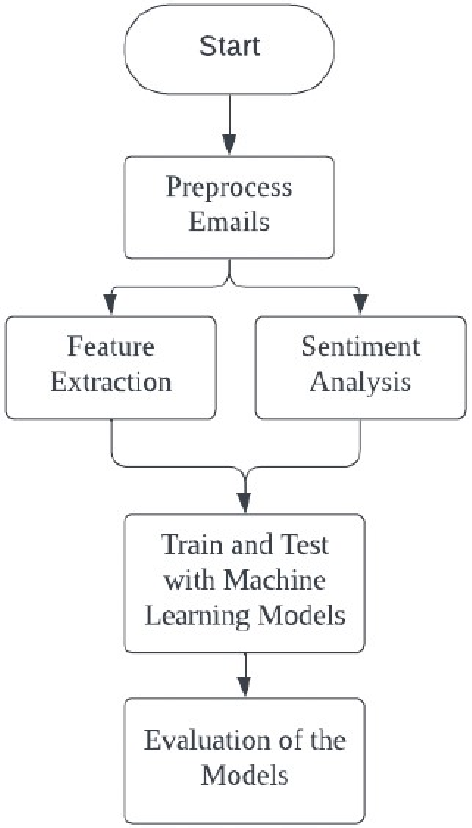


Fig. 1. Model Build Flow Chart

Model Build Flow Chart

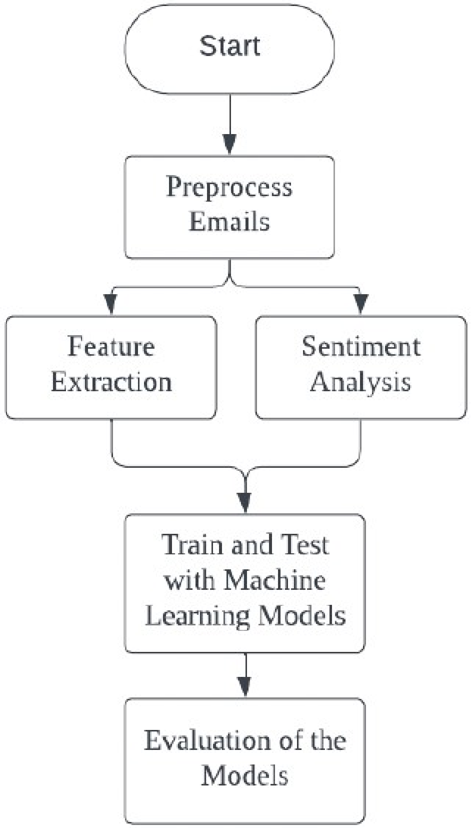


Fig. 1. Model Build Flow Chart

Model Build Flow Chart

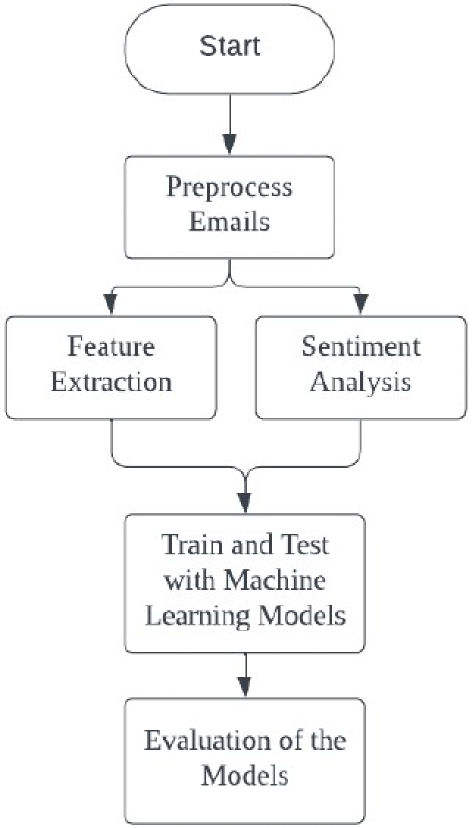


Fig. 1. Model Build Flow Chart

Model Build Flow Chart

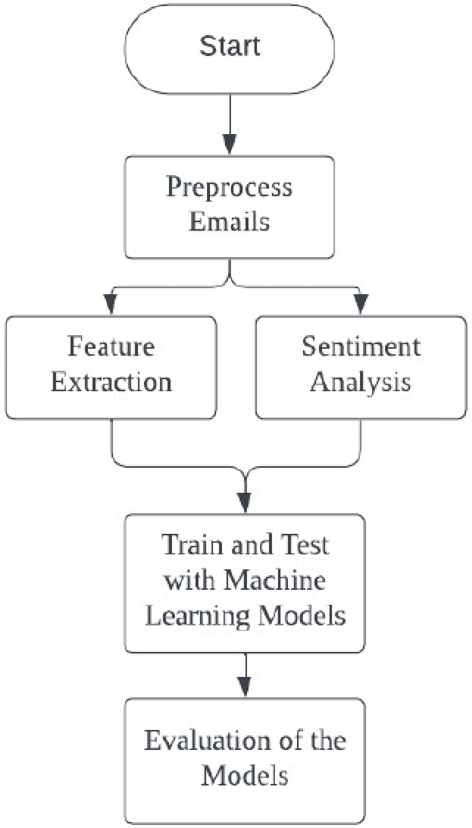


Fig. 1. Model Build Flow Chart

Model Build Flow Chart

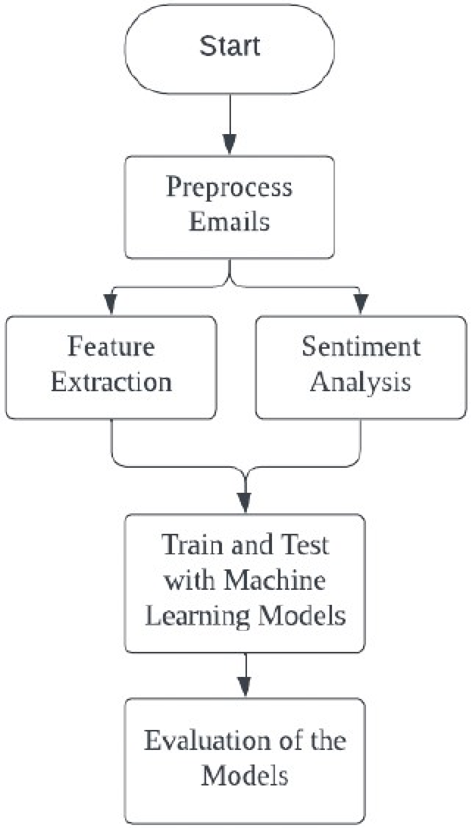


Fig. 1. Model Build Flow Chart

Throw that all into our ML Blender

Classification Results for Proposed Models

TABLE I
CLASSFICATION RESULT FOR PROPOSED MODELS

Model	FP-Rate	FN-Rate	Accuracy	Precision	Recall
NB	26.25%	38.91%	66.67%	68.10%	66.67%
RF	15.23%	10.67%	87.32%	87.31%	87.32%
NB-SA	23.55%	21.24%	77.80%	78.10%	77.80%
RF-SA	7.59%	3.69%	94.54%	94.56%	94.54%

Classification Results for Proposed Models

TABLE I
CLASSFICATION RESULT FOR PROPOSED MODELS

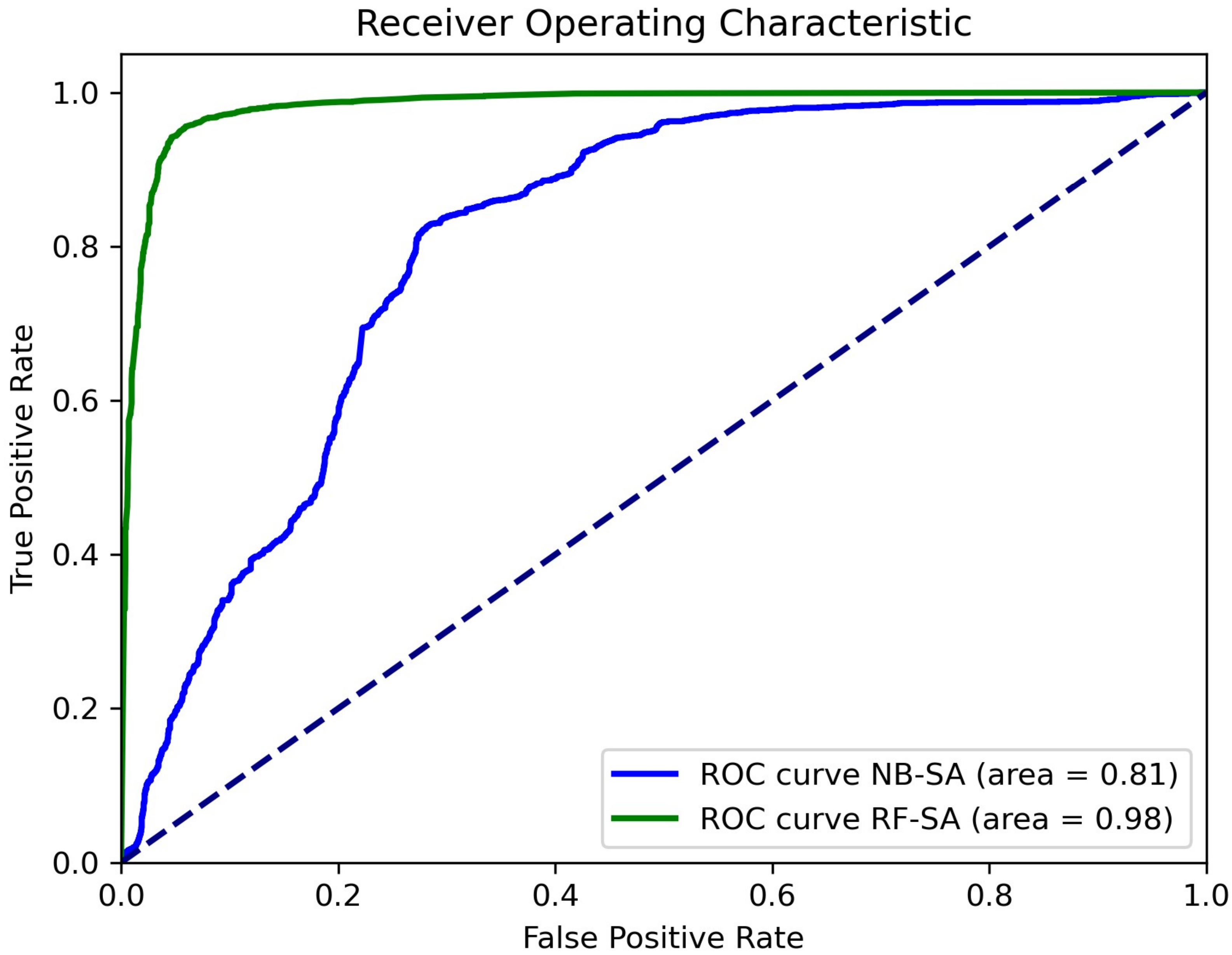
Model	FP-Rate	FN-Rate	Accuracy	Precision	Recall
NB	26.25%	38.91%	66.67%	68.10%	66.67%
RF	15.23%	10.67%	87.32%	87.31%	87.32%
NB-SA	23.55%	21.24%	77.80%	78.10%	77.80%
RF-SA	7.59%	3.69%	94.54%	94.56%	94.54%

Classification Results for Proposed Models

TABLE I
CLASSFICATION RESULT FOR PROPOSED MODELS

Model	FP-Rate	FN-Rate	Accuracy	Precision	Recall
NB	26.25%	38.91%	66.67%	68.10%	66.67%
RF	15.23%	10.67%	87.32%	87.31%	87.32%
NB-SA	23.55%	21.24%	77.80%	78.10%	77.80%
RF-SA	7.59%	3.69%	94.54%	94.56%	94.54%

ROC Curves for NB-SA and RF-SA Models



Future Directions

Future Directions

Thanks To

Chinese American Scholars Association

Questions