

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



Shuyan Liu
sl10158@nyu.edu





Eleftheria Pissadaki, PhD

ep3041@nyu.edu



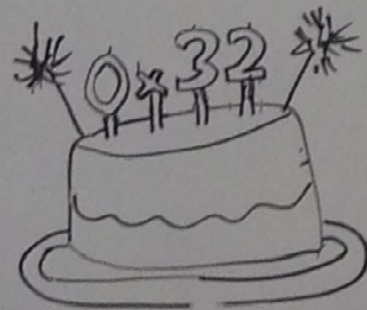
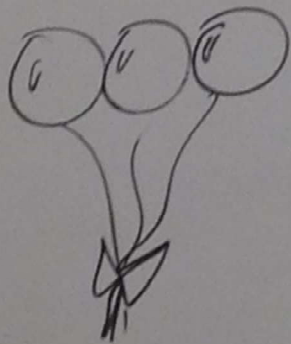
Thomas M. Schmidt, D. Sc.
tms493@nyu.edu






Happy Birthday

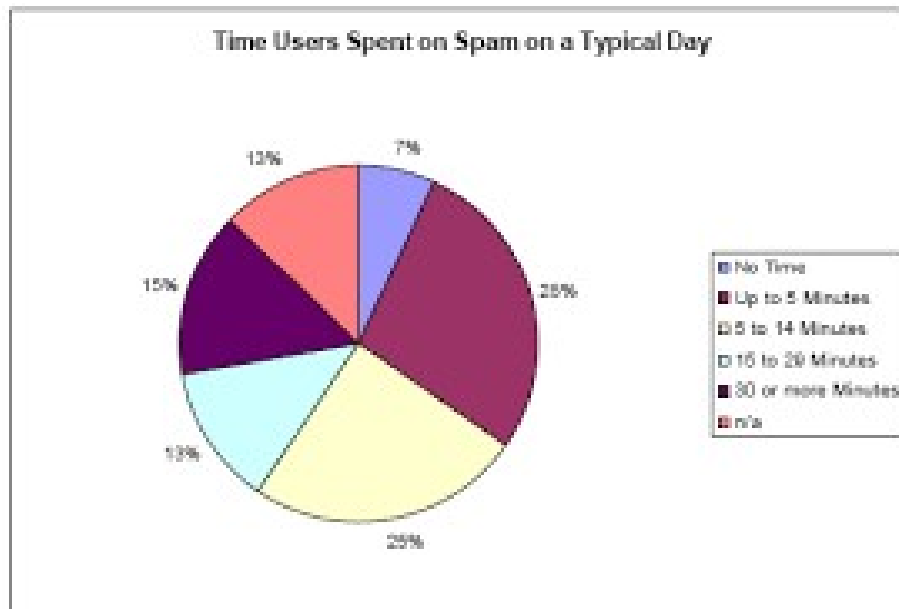
Dr. Tom!



Agenda

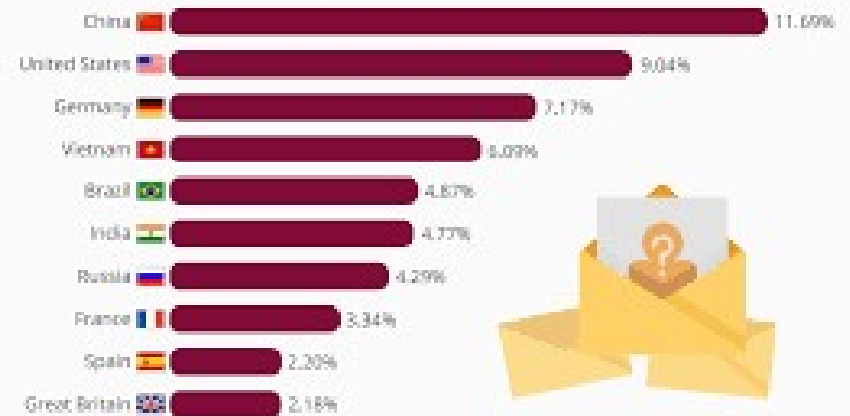
- Spam Detection
 - Types of Errors
 - Pattern Recognition
 - Classifiers
 - Sentiment Analysis
 - Building the model
 - Results
 - Future Directions
- 

The Spam Problem



Where Spam Comes From

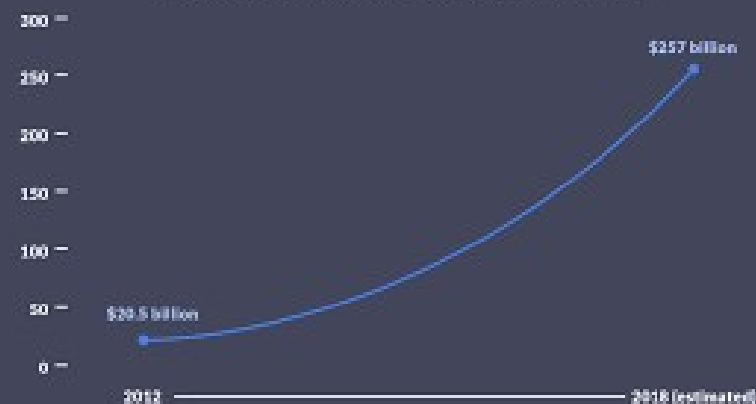
Countries from which the most spam mails originated in 2018



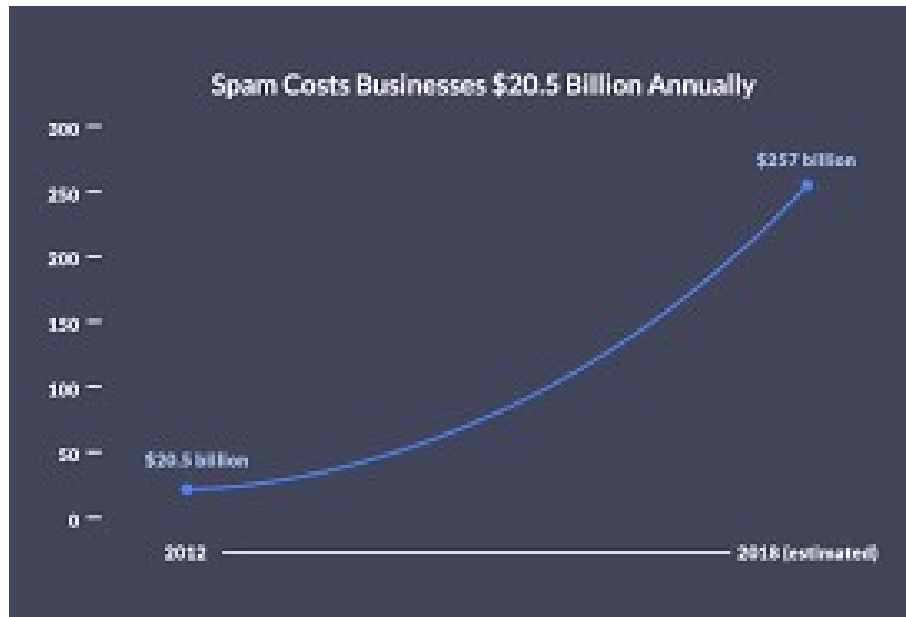
© Statista 2019 Source: statista.com

statista

Spam Costs Businesses \$20.5 Billion Annually



The Spam Problem



Rate of Spam Mails is Dropping

Share of spam mails as percentage of total e-mail traffic:

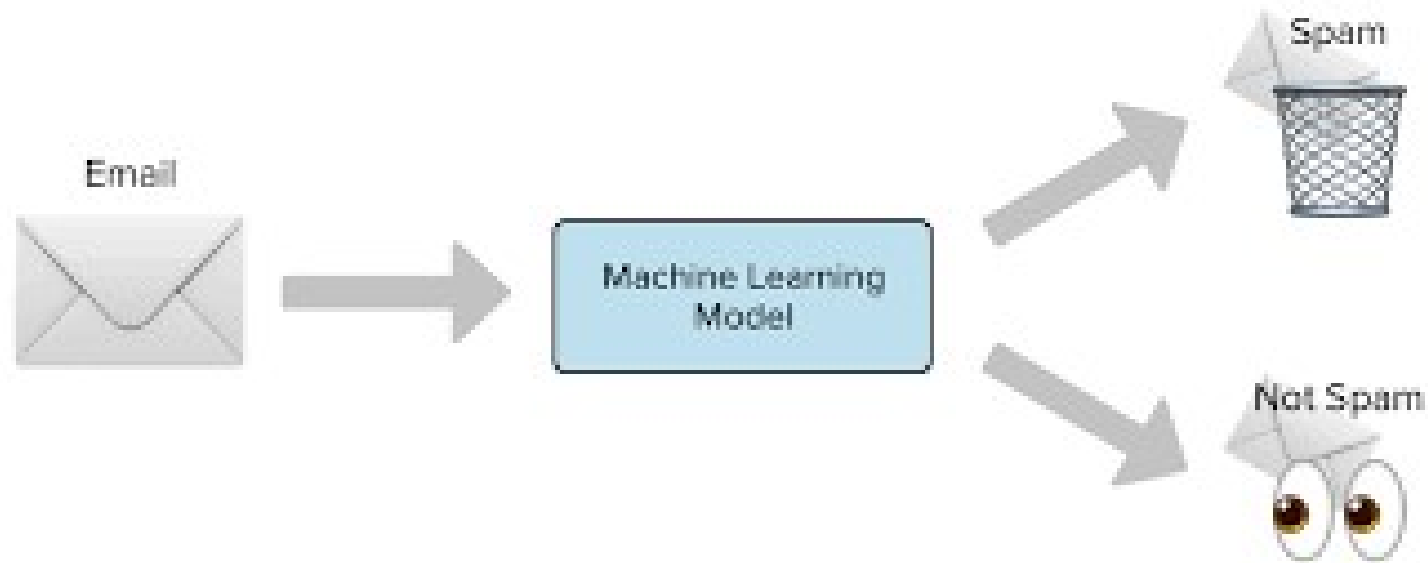


© Statista 2018. All rights reserved. Source: Statista

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.



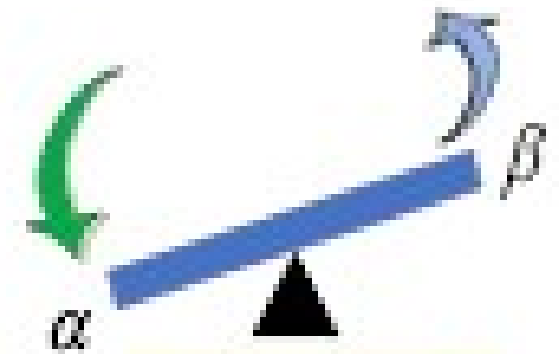
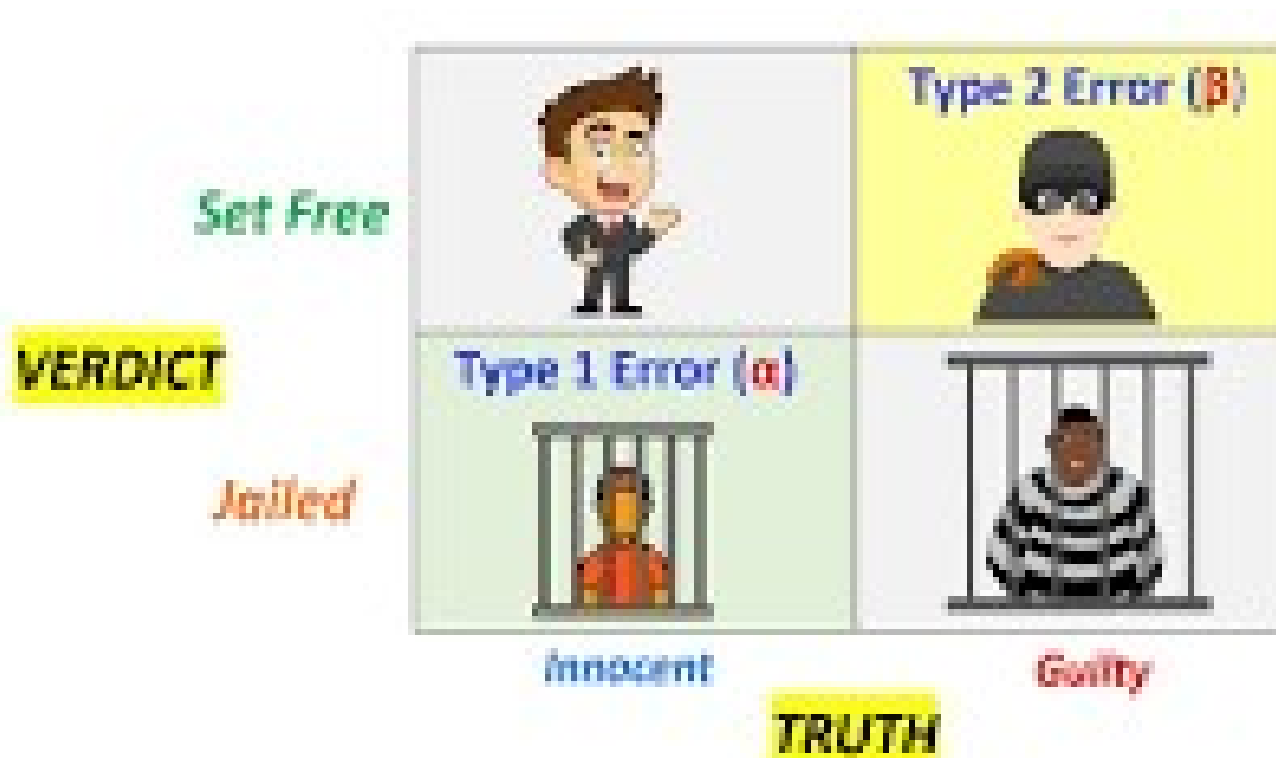
IF IT'S THERE, THE EMAIL IS PROBABLY FINE.

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 Error & Type 2 Error

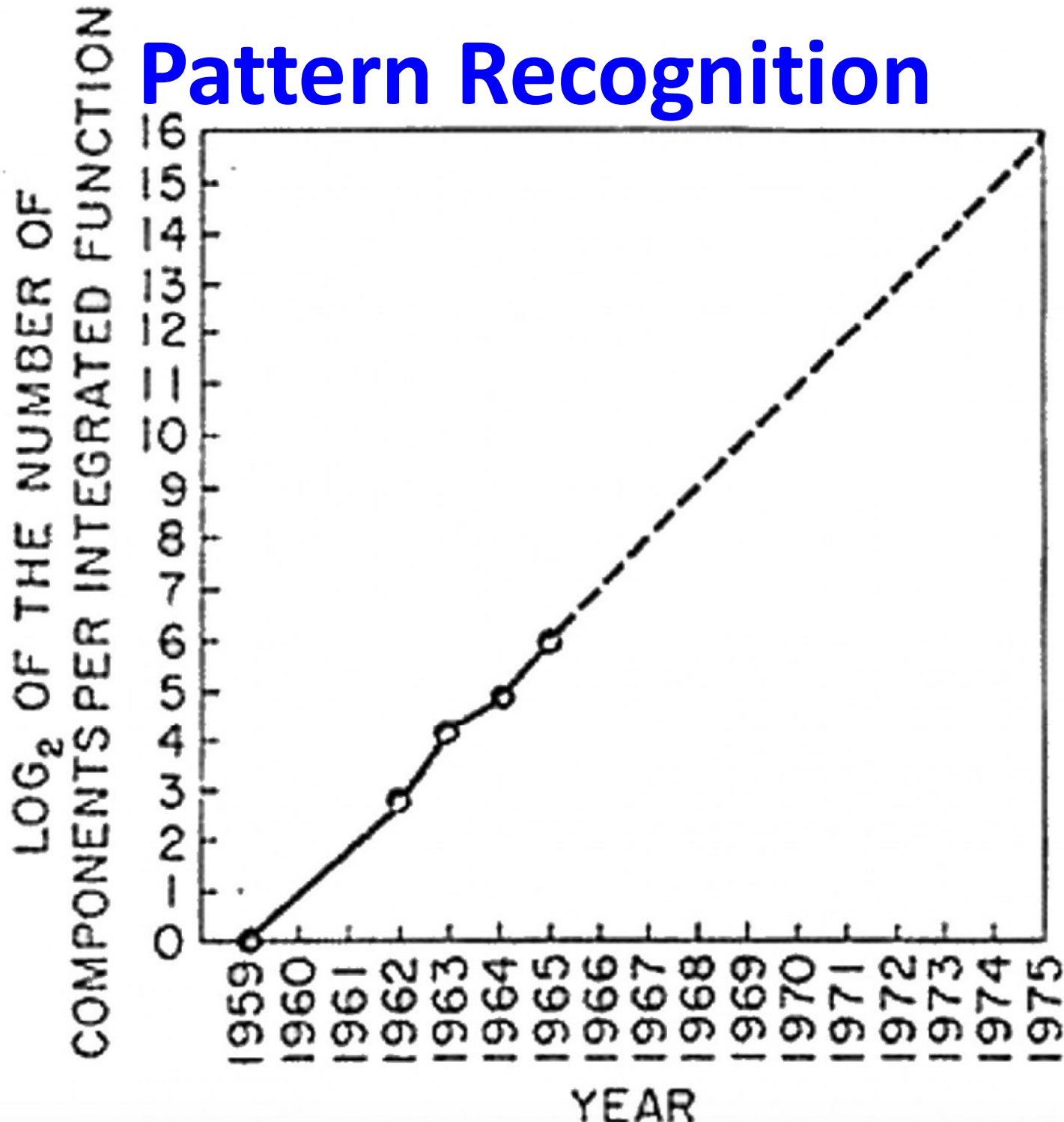


α & β
Relationship ?
Power ?

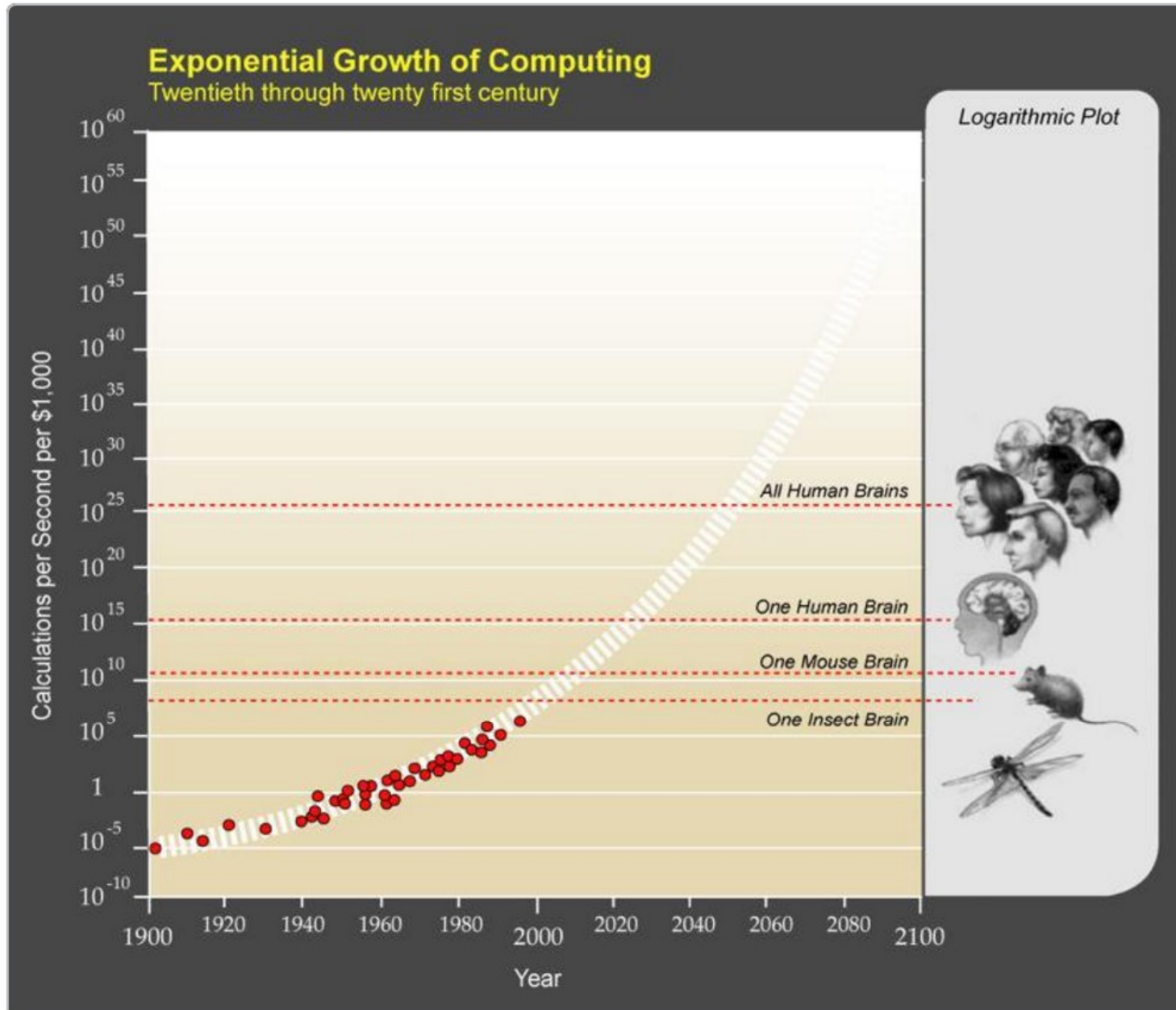
Type 1 and Type 2 Errors, Spam Detection

	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

$$P(c | x) = \frac{P(x | c)P(c)}{P(x)} \quad (1)$$

Naïve Bayes

$$P(c | x) = \frac{P(x | c)P(c)}{P(x)} \quad (1)$$

Naïve Bayes

$$P(c | x) = \frac{P(x | c)P(c)}{P(x)} \quad (1)$$

Naïve Bayes

$$P(c | x) = \frac{P(x | c)P(c)}{P(x)} \quad (1)$$

Naïve Bayes

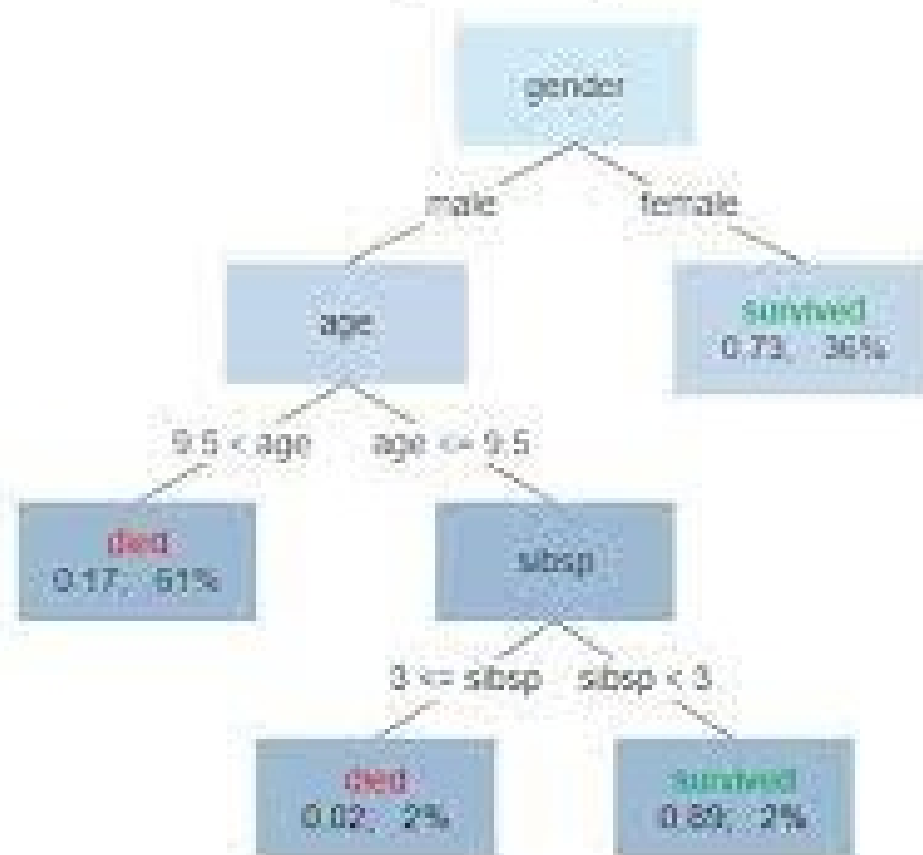
$$P(c | x) = \frac{P(x | c)P(c)}{P(x)} \quad (1)$$

Naïve Bayes

$$P(c | x) = \frac{P(x | c)P(c)}{P(x)} \quad (1)$$

Titanic Decision Tree

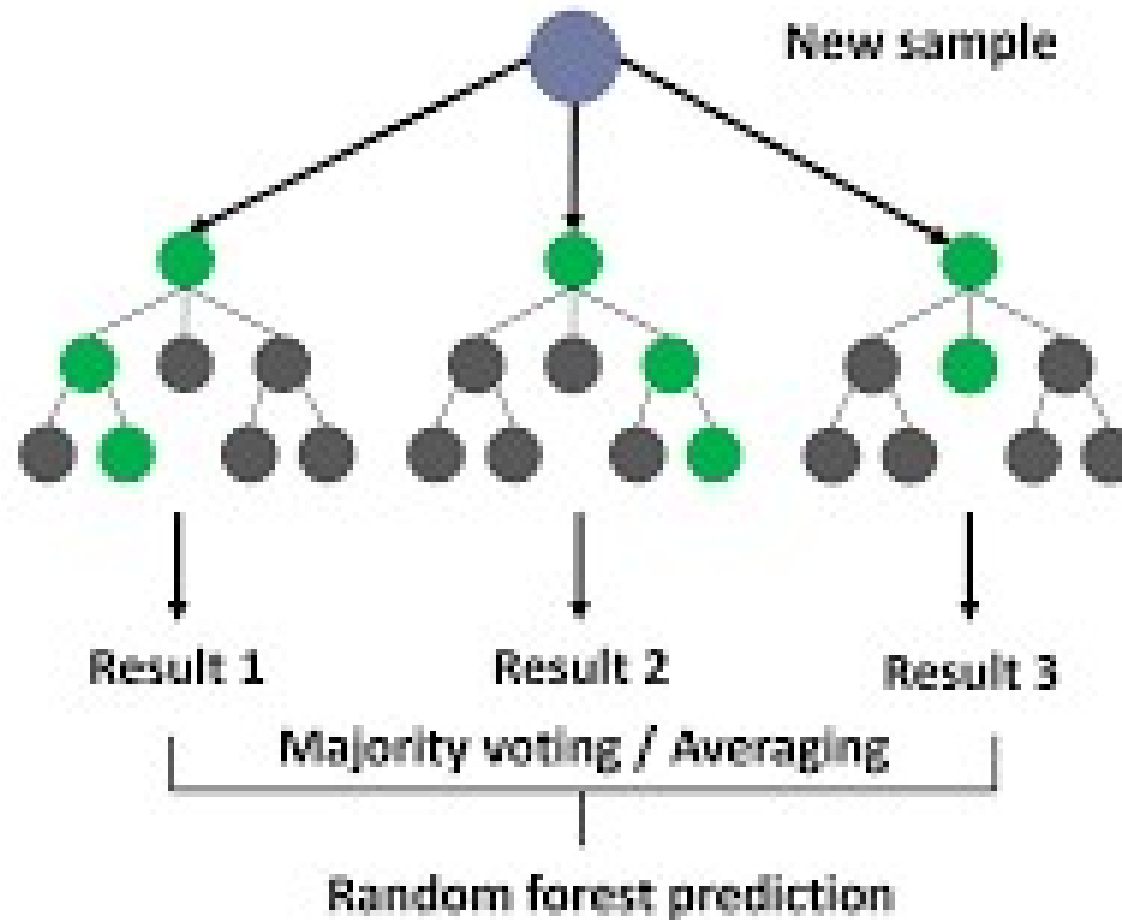
Survival of passengers on the Titanic



Email Spam Detection: ML problem



Random Forest as ML?



Sentiment Analysis

Sentiment Analysis

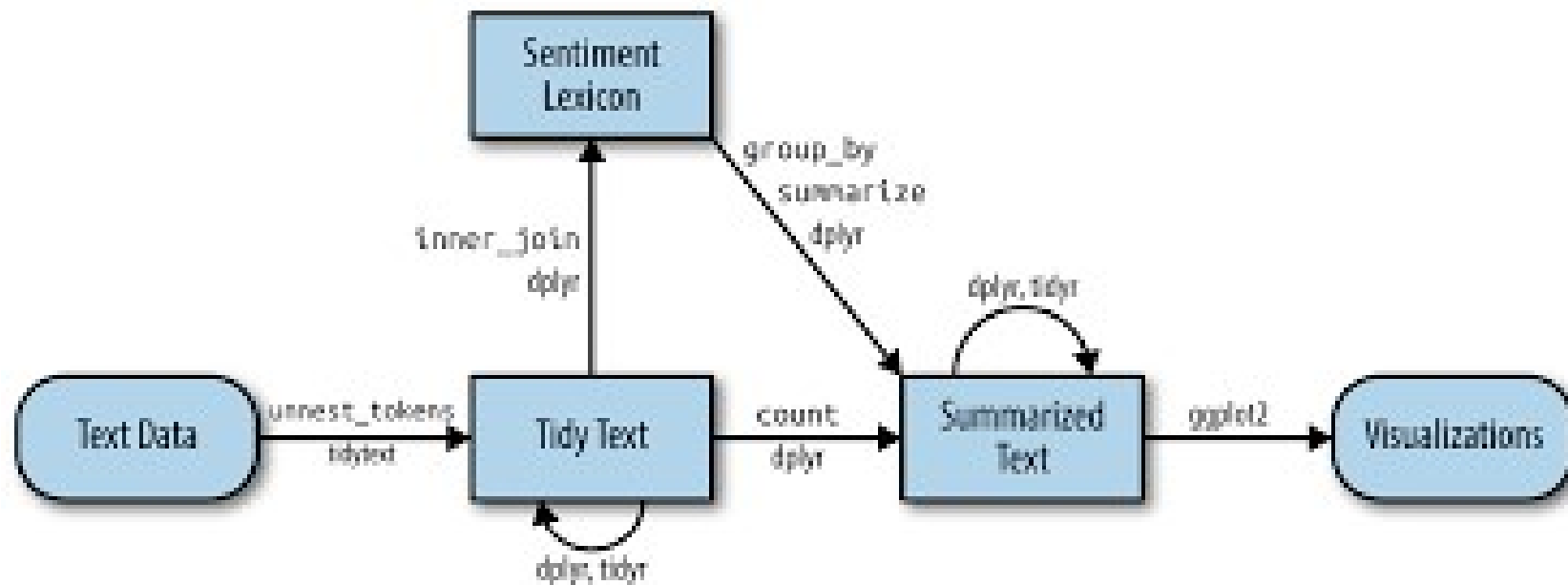


The image displays three vertical panels illustrating sentiment analysis. Each panel features an emoji at the top, a text snippet in the middle, and a sentiment label at the bottom. The first panel shows a smiling face emoji, the text 'My experience so far has been fantastic!', and a green 'POSITIVE' label. The second panel shows a neutral face emoji, the text 'The product is ok! guess', and a yellow 'NEUTRAL' label. The third panel shows an angry face emoji, the text 'Your support team is useless.', and a red 'NEGATIVE' label.

Emoji	Text	Sentiment
😊	My experience so far has been fantastic!	POSITIVE
😐	The product is ok! guess	NEUTRAL
😡	Your support team is useless.	NEGATIVE

MonkeyLearn

Sentiment Analyzer



Model Build Flow Chart

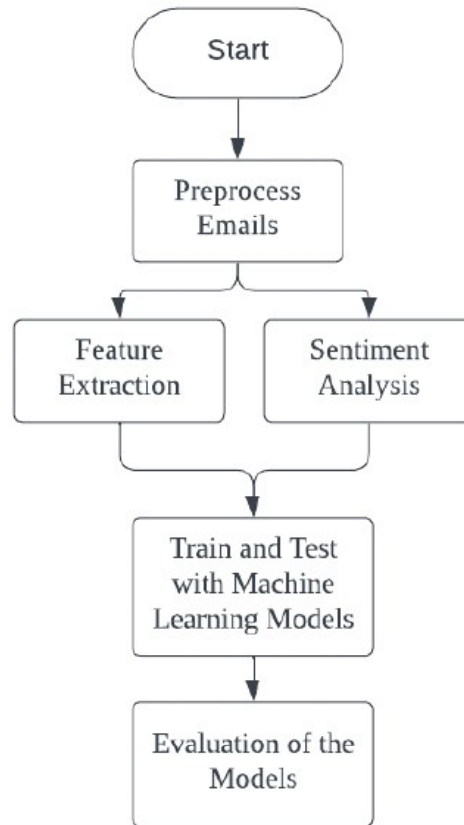


Fig. 1. Model Build Flow Chart

Model Build Flow Chart

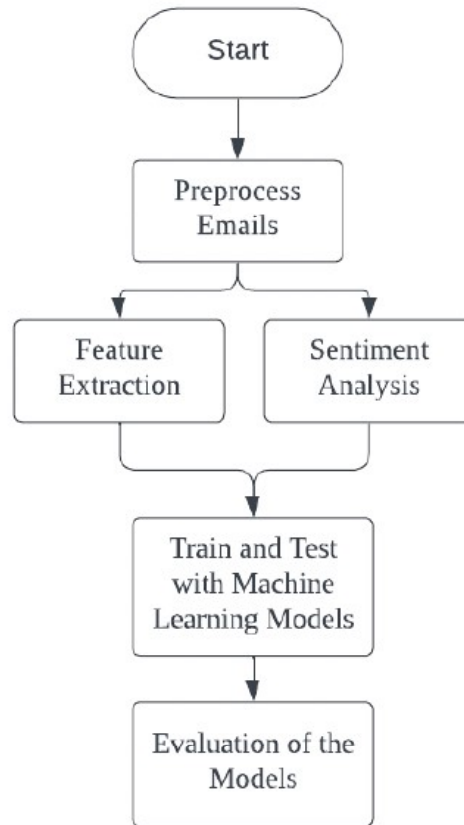


Fig. 1. Model Build Flow Chart

Model Build Flow Chart

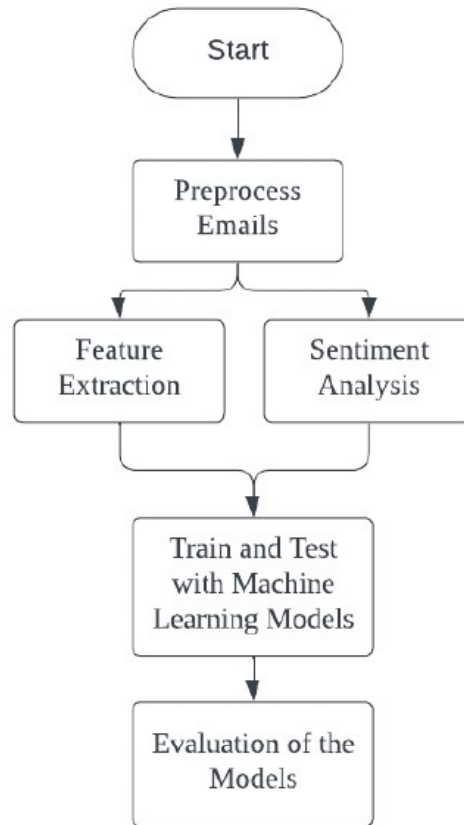


Fig. 1. Model Build Flow Chart

Model Build Flow Chart

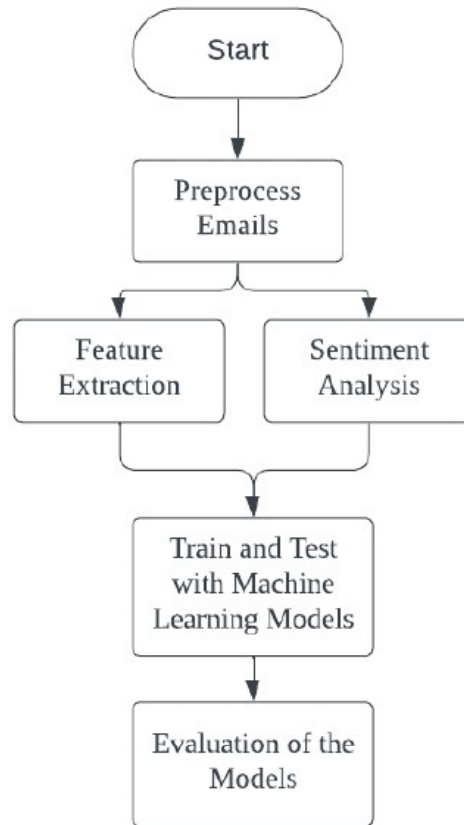


Fig. 1. Model Build Flow Chart

Model Build Flow Chart

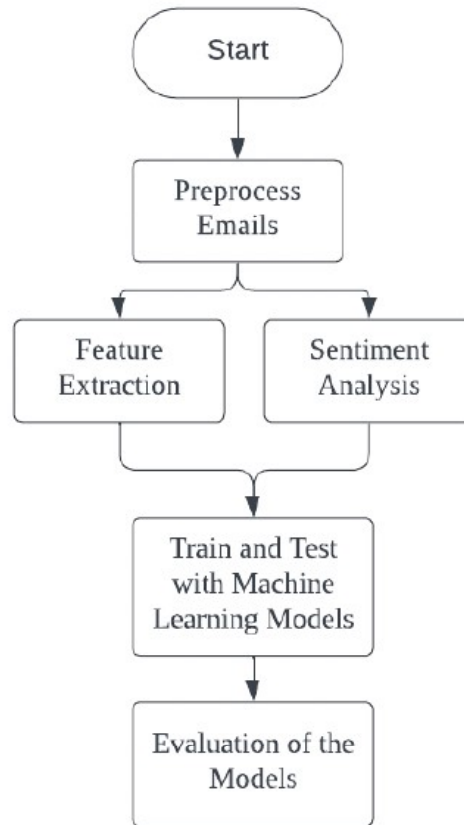


Fig. 1. Model Build Flow Chart

Model Build Flow Chart

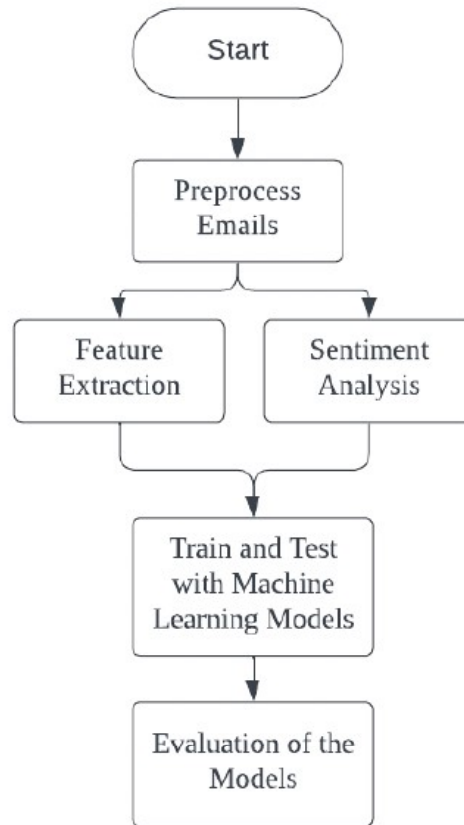


Fig. 1. Model Build Flow Chart

Throw that all into our ML Blender



Classification Results for Proposed Models

TABLE I
CLASSIFICATION 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
CLASSIFICATION 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%

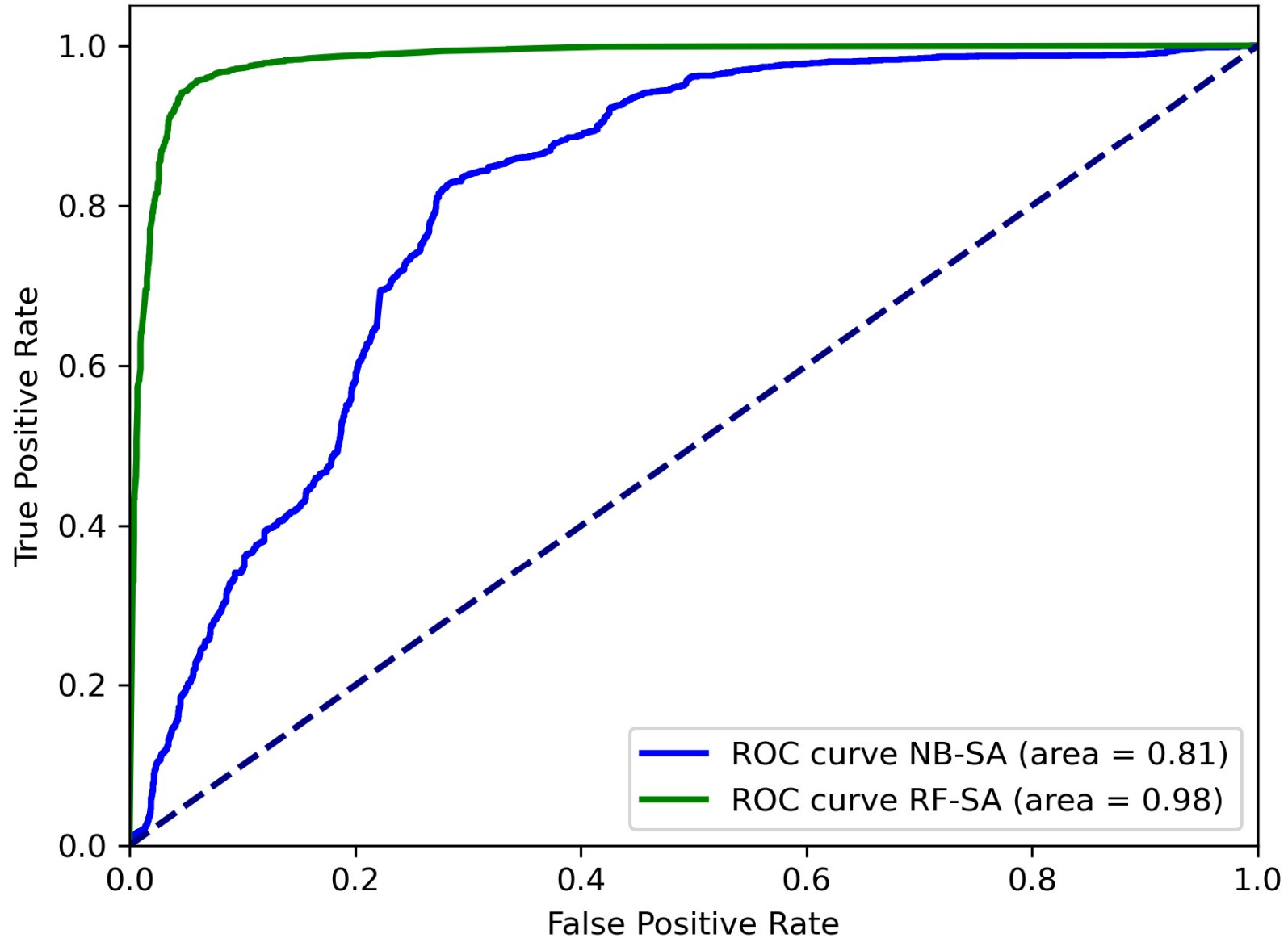
Classification Results for Proposed Models

TABLE I
CLASSIFICATION 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

Receiver Operating Characteristic



Future Directions



Future Directions



Thanks To



Chinese American
Scholars Association



Questions

