Bayesian Machine Learning and Email Spam Classification

K. V. Huynh, T. B. Khondowe, S. Yang

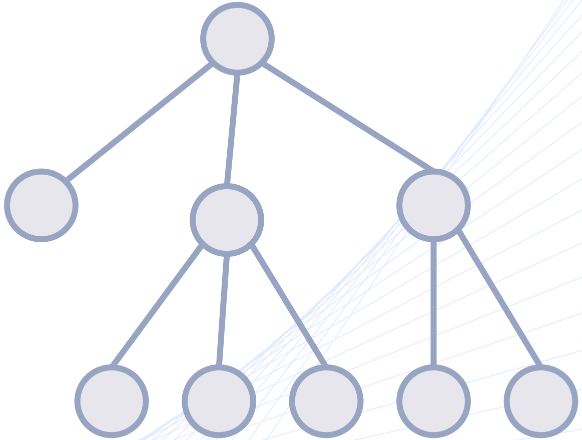
The Spam Problem

- Spam means unsolicited email message(s)
- Costs businesses and individuals
  - Time and money
- Spam as a vehicle for malicious software
  - The Ransomware Tsunami
- Artificial intelligence can be used for spam detection
- We focused on two Bayesian classifiers
  - Bernoulli Multivariate Model
  - Multinomial Event Model + Thresholding

*

Related Work

- There are multiple spam email classifiers
  - Bayesian Classifiers
  - Random Forests (DT)
  - Support Vector Machines
  - K-Nearest Neighbor
- Classifiers can be used as ensemble
- Ensembles and Random Forests perform better
- Multinomial is known to be better



Bayesian Machine Learning - Naive Bayesian

- Bayesian Machine Learning
  - A probabilistic machine learning approach
  - Learning from prior probability distributions

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

$$P(\theta|D,m) = \frac{P(D|\theta,m)P(\theta|m)}{P(D|m)}$$

- Naive Bayesian
  - Subfield of Bayesian Machine Learning
  - Assumption:
    - Features are mutually independent
    - Distribution of the data

Bernoulli Multivariate Event Model

- A Naive-Bayesian approach
  - Email is presented as a vector
  - Value for each words will be 0 (not presented) or 1 (presented)
- Assumption:
  - Classification of instances are mutually independent
  - The appearance of words are mutually independent

$$P[t][c] = \frac{NumberOfDocs_{ct} + 1}{NumberOfDocs_{c} + 2}$$

Term t in the document: *PredictScore[class] = PredictScore x P[t][c]*

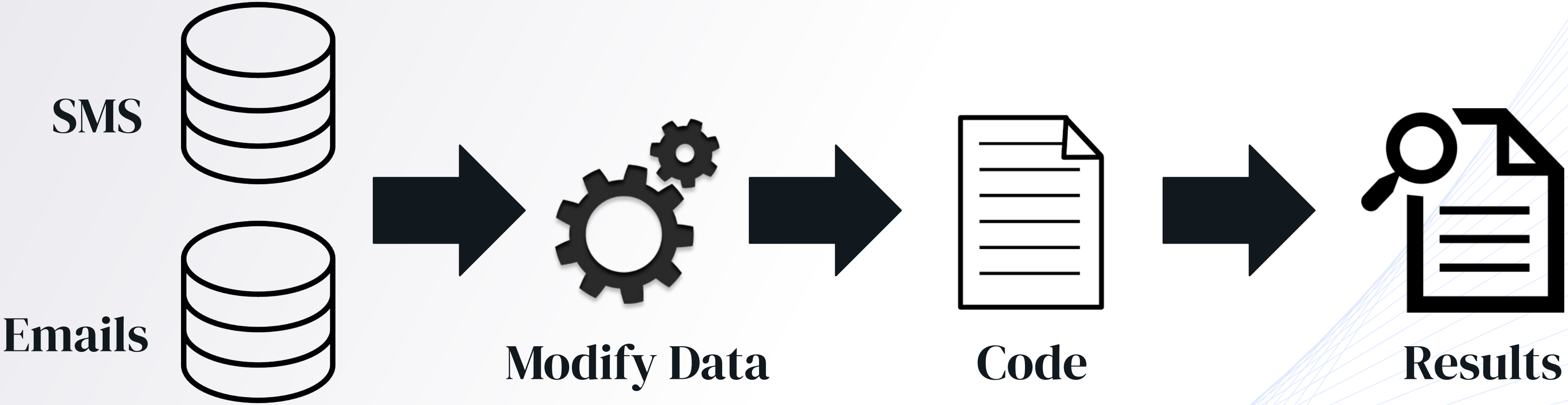
Term t not in the document: *PredictScore[class] = PredictScore x (1 -P[t][c])*

Multinomial Event Model

$$P(S \mid W) = P(S) * P(w1 \mid S) * P(w2 \mid S) * \ldots$$

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

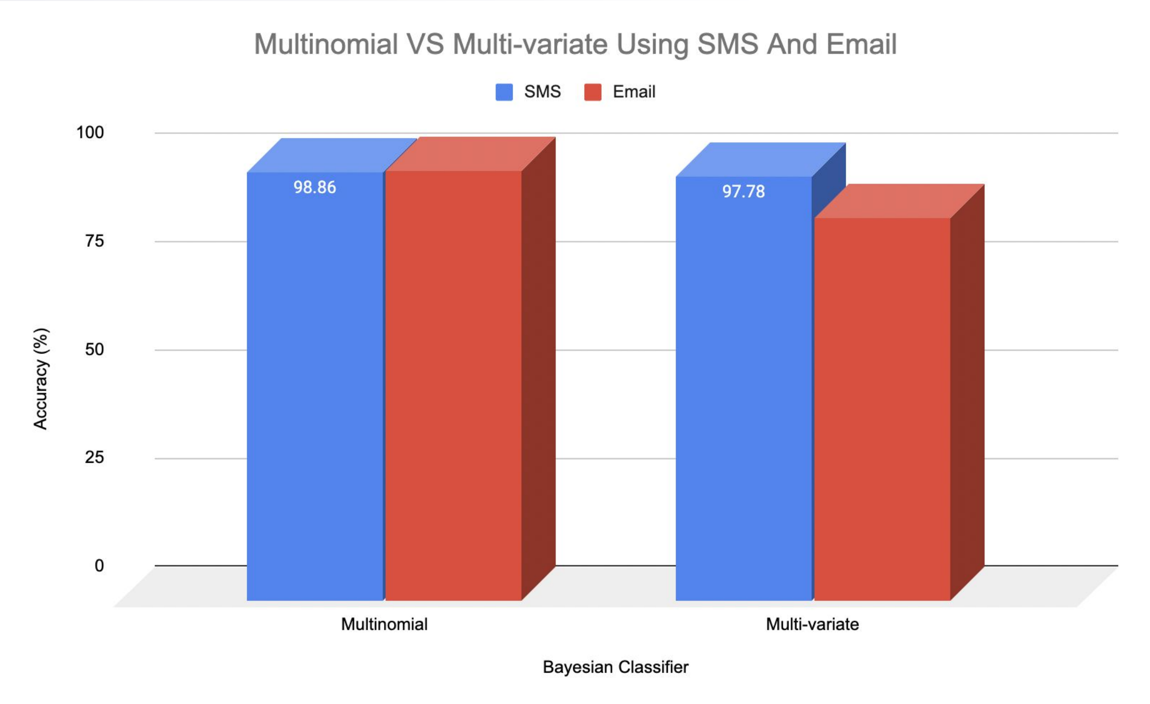
Our Approach - Thresholding



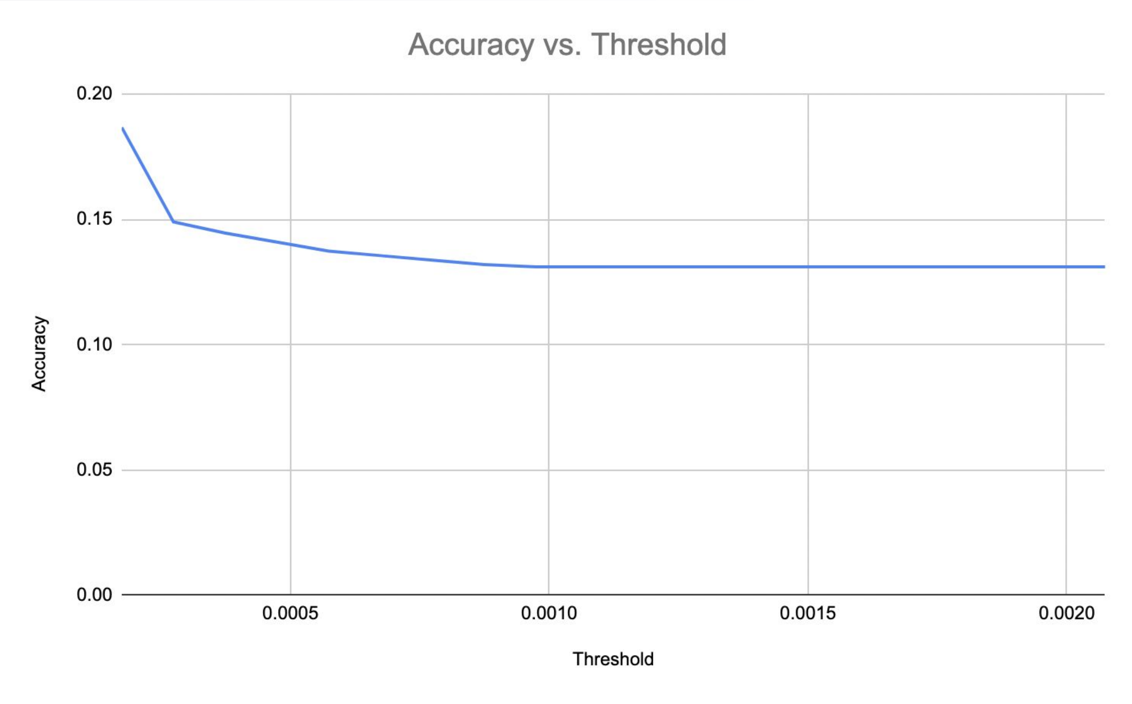
- Multinomial is a better approach
- Multinomial + Thresholding

$$P(S \mid W) = P(S) * P(w1 \mid S) * P(w2 \mid S) * \ldots$$

Performance and Testing

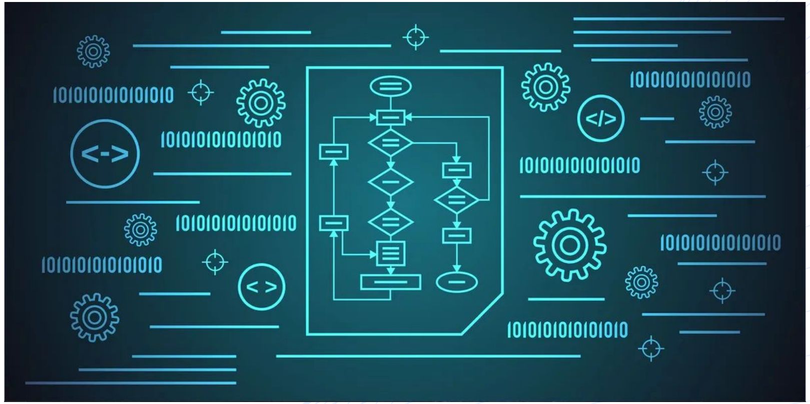


Performance and Testing



Conclusion

- (**scikit-learn**)
- We expected the multinomial implementation is to perform better for both datasets
- Thresholding did not improve the multinomial implementation
- Instead, thresholding did not improve the Naïve Bayes Multinomial implementation at all



Problem - Future Work

- Figuring out low accuracy score using Thresholding
- Other potential approaches:
    - Partitions
    - Information Gain
    - Remove stopped word or common words
    - Case sensitivity
    - Looking at other Naive-Bayesian model
        - Gaussian Model
        - Complement Naive Bayesian model
        - Categorical Naive Bayesian model
        - Out-of-core Naive Bayesian model fitting

References

Hanif Bhuiyan, Akm Ashiquzzaman, Tamanna Islam Juthi, Suzit Biswas & Jinat Ara. *A Survey of Existing E-Mail Spam Filtering Methods Considering Machine Learning Techniques.* 2018. *Global Journal of Computer Science and Technology; C Software & Data Engineering*. Global Journals, 2018

Sanghani, Gopi & Kotecha, Ketan. SUPPORT VECTOR MACHINE FOR PERSONALIZED E-MAIL SPAM FILTERING. 2018. *International Journal of Advanced Research in Engineering and Technology*, IAEME Publication, 2017, 108–120

Sanghani, Gopi & Kotecha, Ketan. SUPPORT VECTOR MACHINE FOR PERSONALIZED E-MAIL SPAM FILTERING. 2018. *International Journal of Advanced Research in Engineering and Technology*, IAEME Publication, 2017, 108–120

Nurul Fitriah Rusland, Norfaradilla Wahid1, Shahreen Kasim, and Hanayanti Hafit. 2017. Analysis of Naïve Bayes Algorithm for Email Spam Filtering across Multiple Datasets. *International Research and Innovation Summit*, Melaka, Malaysia. IOP Publishing Ltd.

S. Roy, A. Patra, S.Sau, K.Mandal, S. Kunar. 2013. An Efficient Spam Filtering Technique for Email Account. *American Journal of Engineering Research* Volume-02 Issue-10. 63-73

Hyunsoo Kim, Peg Howland, and Haesun Park. 2005. Dimension Reduction in Text Classification with Support Vector Machines. J. Mach. Learn. Res. 6 (12/1/2005), 37–53.

Androutsopoulos, Ion & Paliouras, Georgios & Karkaletsis, Vangelis & Sakkis, Georgios & Spyropoulos, Costantine & Stamatopoulos, Panagiotis. 2000. Learning to Filter Spam EMail: A Comparison of a Naive Bayesian and a Memory-Based Approach. ArXiv. Athen, Greece.