**Defending Against Spam: A Comparison Between Random Forest and Naive Bayes Algorithm**

Colin M. Hopkins

Department of Computer Science, Binghamton University

CS 301: Ethical, Social, and Global Issues in Computing

Dr. George Weinschenk

December 5, 2022

**Abstract**

The term “spam” refers to unsolicited bulk email; this mail usually has malicious intent to phish personal information or receive money from the recipient. Computing professionals acknowledged this matter and noticed a need for an automated email filtration system. The effects of spam and phishing have prompted the creation and use of classification algorithms to reduce the number of victims phished. Commonly used spam detection algorithms include Naïve Bayes and Random Forest. The Random Forest classification algorithm outperforms the Naïve Bayes with a 94 percent detection accuracy. This accuracy concludes the need for computing professionals to use Random Forest when faced with a spam detection matter. Though Random Forest is robust, attackers continue to profit off the six percent of emails not detected. These remaining attacks create a need to implement alternate actions, such as government regulation and modification of designs. The United States has few restrictions on email spam, and frequently companies and spammers work around the loose guidelines. Professionals argue that computing research and programmers must increase focus on security to reduce attacks on the public. Pre-planning for potential abuse hinders the damage caused by attackers. Without updated policies and practices, the sole reliance on the high accuracy of classification algorithms will cause stagnation in increasing spam defense.

**Defending Against Spam: A Comparison Between Random Forest and Naive Bayes Algorithm**

Most online communications take place through email. Research shows that approximately 300 billion emails are sent and received daily; over half of those emails are considered spam. Email spam refers to unwanted junk mail sent out in bulk, usually maliciously. These messages attempt to phish personal data, steal money, or trick recipients into visiting websites hosting malware. This ongoing problem has prompted the development of automated spam detection. Using the Random Forest (RF) classification algorithm instead of the Naïve Bayes (NB) classification algorithm for machine learning validates emails more accurately, which decreases the number of victims phished by malicious attachments. Researchers have tested various detection algorithms to determine which performed the highest. Though commonly favored for its time complexity, Naïve Bayes remains inferior to other machine learning algorithms.

**Alternate Technology**

The Naïve Bayes (NB) algorithm assigns data into categories through supervised machine learning classification. Supervised learning models use labeled datasets to train or “supervise” algorithms in classifying data (Delua, 2021, para.4). This classifier uses Bayes’ Theorem with a “naïve” assumption that the classified features are independent of each other. Independent means the algorithm processes a combination of features the same as each feature appears by itself. Bayes’ Theorem determines conditional probability, the likelihood of an outcome occurring based on previous occurrences (Chauhan, 2022, para.4). This Theorem uses prior probabilities to generate posterior (updated) probabilities for the occurrence of an event. The formula is which tells us the probability of A given B (, knowing the probability of B given A (, the probability of A occurring (, and the probability of B occurring (. Bayes’ equation uses known probabilities to determine a new probability.

Implementing Bayes’ Theorem in machine learning without the Naïve assumption would prohibit accurate classification of inputs. Without the independent assumption, Bayes’ rule would need measurements with identical feature sets, which would only allow for the classification of exact duplicates (Dinant, 2018, para.3). Assuming independent features enables the algorithm to take each feature separately, eliminating the need for exact duplicates. Applying Bayes’ equation with independent features allows for the accurate classification of an input. In a recent test, Naïve Bayes detected 88.69 percent of email spam (Amin, 2019, p.171). Despite its accuracy, NB has drawbacks. If NB encounters features not present in the training data, the algorithm will not accurately classify the input; additionally, situations occur where features form dependencies the algorithm cannot read, resulting in decreased accuracy**(**Chauhan, 2022, para.10). Though NB has above-average performance, alternate methods achieve more accurate results.

**Support**

**Random Forest**

The Random Forest classifier operates under supervised machine learning, much like Naïve Bayes. Decision trees lay the foundation for the Random Forest model. Decision trees function similarly to regular trees, except each node on a decision tree represents a feature, and each leaf displays a class (decision). In email practice, the nodes of the trees represent features determining text validity; the corresponding leaves display “spam” or “not spam”. Random Forest employs a variety of decision tree algorithms to classify categorical variables appropriately (Yiu, 2019, para.9). Each tree works on an independent sample from the data set to produce a class prediction; the most common class results as the final prediction.

The idea behind Random Forest closely resembles the wisdom of the crowd – the most common response from a large group of people provides more accuracy than a single response. Although a single decision tree might appropriately categorize an email, a Random Forest algorithm performs better. According to Tony Yiu (2019), Random Forest eliminates problems such as overfitting and low accuracy created by decision trees (para.30). Since Random Forest implements numerous decision trees, the run time is slower, but the prediction accuracy increases compared any individual tree. The accuracy of Random Forest relies on diversity between the individual trees. Each tree applies the same random sample from the test data but with minor changes for each set. The size of the data set remains the same for each tree, but the set alters elements to create diversity with duplicate values. Yiu defines this minor alteration of the data as bagging (para.23). Random Forest responds drastically to minor changes in data, creating diverse decision trees. Applying these diverse trees, a Random Forest grows to generate highly accurate results.

Ruhul Amin, Md Moshiur Rahman, and Nahid Hossain (2019), professors in the computer science department at United International University (Dhaka, Bangladesh), studied supervised machine learning algorithms’ ability to discover Bangla spam mail. The detection of western languages (English) has received copious attention, but Amin proposed a robust spam email detection for Bangla. Amin explored six detection models, notably Naïve Bayes and Random Forest. The dataset contained two types of emails split into independent sets.

**Table 1**

*Size and Breakdown of the Train and Test Datasets*

Table

Description automatically generated

Admin illustrates in table 1 the even split in the datasets (Admin et al., 2019, p.196). This experiment measured the performance of the algorithms in three categories; accuracy, sensitivity, and specificity. Amin’s et al. results show that the Radom Forest classifier performed the highest, with an accuracy of 93.60 percent. The Naïve Bayes had the weakest performance with 88.69 percent accuracy.

**Table 2**

*Summary of Experimental Results*

*Table

Description automatically generated*

Random Forest also dominates the sensitivity and specificity categories as shown in table 2 (Admin et al., 2019, p.171). Amin, Rahman, and Hossain et al. used the Random Forest classifier for their Bangla spam email detection as it performed the highest.

Hassan Najadat and Islam Qaraz from Jordan University of Science and Technology, performed with assistance from Mohammad Alazubaidi (2022) from Yarmouk University. Najadat et al. used supervised machine learning classification algorithms to detect Arabic spam on social networks. These algorithms include Decision Tree and Naïve Bayes. The data set contained 3,000 Facebook comments divided into two categories. The first category comprised 2,500 spam comments, while the second held 500 non-spam (para.35).

**Figure 1**

*Accuracy of Classifier Evaluated by Different Filter Methods*

*Chart, bar chart

Description automatically generated*

Najadat et al. ran multiple experiments with variations; the Decision Tree produced the highest accuracy of 92.53 percent, while Naïve Bayes placed last with an accuracy of 87.30 percent as shown in figure 1 (Najadat et al., 2022, para.47). Two separate studies validate the superior performance of the Random Forest algorithm over the Naïve Bayes. Email spam has severe impacts on society. Implementing a more robust spam detection algorithm reflects on the public’s online safety.

**Social Impact**

Email spam comes in many forms, such as unwanted marketing, malicious phishing, or unsolicited bulk mail. Though algorithms such as Random Forest can filter out most spam, people still interact enough for spammers to continue their activities. According to Amin (2019), 55 percent of the 300 billion emails transmitted daily contain spam (p.169). In 2003, the United States Federal Trade Commission (FTC) passed the CAN-SPAM Act: A Compliance Guide for Business. The CAN-SPAM act restricts businesses from sending unsolicited mail and gives recipients the right to stop receiving such mail (opt out); violation of this law results in harsh monetary penalties (CAN-SPAM, 2003). Corporations work around this act’s fine print to continue sending marketing mail. The law does not ban companies from flooding an inbox; it states that companies must provide an option for recipients to opt out. This option often hides as small text at the bottom of an email to increase the difficulty of unsubscribing. James Veitch (2018), esteemed film editor and author of *Dot Con: The Art of Scamming a Scammer*, describes an interaction where unsubscribing from an email list did not work. Veitch states, “I scrolled to the bottom of the email and clicked unsubscribe. I thought that would be the end, but two weeks later, I received another email” (Veitch, 2018,1:42). Veitch’s interaction reinforces the possible difficulties of trying to opt out from an email list.

The CAN-SPAM act only applies to corporations, not individuals; this law does not protect against the most significant spam threat (phishing). In 2021, the FTC reported more than 2.3 billion dollars in consumer losses solely through email phishing scams; nearly doubling the amount from the previous year (FTC, 2022, para.4). Sampsa Rauti (2019), a Doctoral Student and University professor at the University of Turku studied the localization of email spam. Rauti explains that spammers have started targeting smaller countries, such as Finland, with malicious spam in their native language. Ratui analyzed over 300 unsolicited messages from three long existing email accounts (p.42).

**Table 3**

*Spam Categories and Their Proportions*

*Table

Description automatically generated*

Table 3 shows that 69.3 percent (adding percentages together) of the spam mail intended to obtain money from the recipient (Rauti, 2021, p.44-46). Through his studies, Ratui concluded that spammers are becoming more dangerous and challenging to trace; requiring novel technical solutions to address the concern with spam (p.47).

**The Future of Spam**

Though implementations of algorithms and laws try to aid the pressing spam matter, data shows the problem is still prevalent. Dr. Emilio Ferrara (2019), researcher, and associated professor at the University of Southern California, has concluded that spam will continue to plague society for the foreseeable future. Ferrara believes the current defenses for spam will continue to fail society; new policies and research programs will fight against spam (p.91). Ferrara proposed designing technology with potential abuse in mind to guide efforts in the spam endeavor. Often, technology functions in ways unintended by the designers. Countless examples show new technology abused through a system failure or bug. Designers do not pre-plan for abuse of their software, especially in defending against attacks or spam. Ferrara suggests a greater focus on potential attacks and ways to defend against threats. Moreover, researchers need to pre-plan for the abuse of their technologies by spammers (p.91).

**Conclusion**

Failure to change methodology in spam defense research will allow attackers to continue their efforts. Using an appropriate algorithm to detect email spam can mitigate the number of victims affected by malicious email attacks, up to 94 percent in some cases. Notable professionals have demonstrated the effects different algorithms have on detecting spam. Through research and testing, developed algorithms can detect spam accurately in the 90th percentile. Billions of emails are transmitted daily; this highlights professionals' importance in choosing the most robust algorithm for spam detection systems. The Random Forest algorithm proves to outclass the commonly favored Naïve Bayes by validating more emails, with a higher accuracy ranging from five to ten percent. This difference in percentages may affect millions of people and their online safety. Professionals needing a spam detection algorithm should implement Random Forest due to its precision and reliability. Dr. Najadat and Dr. Amin validate the performance of Random Forest by producing results in the 90th percentile for accuracy; these results were obtained in separate experiments. Though Random Forest produces excellent results, certain aspects hinder the total elimination of spam; the government fails to have strong intervention to prevent companies and scammer from spamming users. Moreover, the small percentage of emails missed by Random Forest allows attackers to profit and continue their actions. Dr. Emilio Ferrara stressed the importance of computing professionals in accounting for abuse within programs and algorithms. Random Forest will continue to aid the spam matter, but spam detection will stagnate without updated government policies and new design implementations such as ones suggested by Ferrara.

**References**

# Amin, R., Rahman, M. M., & Hossain, N. (2019, December 26-28). *A Bangla spam email detection and datasets creation approach based on machine learning algorithms.* Paper presented at **2019 3rd International Conference on Electrical, Computer & Telecommunication Engineering (ICECTE),** Rajshahi, Bangladesh. 10.1109/ICECTE48615.2019.9303525

# Chauhan, N. (2022, April 8). *Naïve Bayes Algorithm: Everything you need to know.* KDN. <https://www.kdnuggets.com/2020/06/naive-bayes-algorithm-everything.html>

# Delua, J. (2021, March 12). *Supervised vs. unsupervised learning: What’s the difference?* IBM. <https://www.ibm.com/cloud/blog/supervised-vs-unsupervised-learning>

# Dinant, C. (2018, May 8). *What’s so naïve about naïve Bayes’?* TDS. <https://towardsdatascience.com/whats-so-naive-about-naive-bayes-58166a6a9eba>

# Ferrara, E. (2019, July 24). The history of digital spam. *Communications of the ACM, 62*(8), 82-91. 10.1145/3299768

# FTC. (2022, February 22). *New data shows FTC received 2.8 million fraud reports from consumers in 2021.* FTC <https://www.ftc.gov/news-events/news/press-releases/2022/02/new-data-shows-ftc-received-28-million-fraud-reports-consumers-2021>

# Hassan, N., Mohammad, A., & Islam, A. (2022). Detecting Arabic spam reviews in social networks based on classification algorithms. *ACM Transactions on Asian and Low-Resource Language Information Processing, 21*, (pp.1-13). 10.1145/3476115

# Sampsa, R. (2019, June 21-22). *An inquiry into localized email spam*. In V. Tzvetomir, & A. Smrikarov. International Conference Proceeding Series*.* Paper presented at CompSysTech '19: 20th International Conference on Computer Systems and Technologies, Ruse, Bulgaria (pp.42-48). Association for Computing Machinery. New York, NY, United States. 10.1145/3345252.3345298

# TED. (2018). *James Veitch – The agony of trying to unsubscribe.* [Video]. YouTube. <https://www.youtube.com/watch?v=H9EwW8Tl2CE&ab_channel=ElijahMiniuk>

# Yiu, T. (2019, June 8). *Understanding random forest.* TDS. <https://towardsdatascience.com/understanding-random-forest-58381e0602d2>