Does Emotional Sentiment Assist Email Spam Classification?

Every day, approximately 347.3 billion emails are sent and received across the world. Major email providers like Gmail and Outlook work around the clock to block hundreds of malicious emails a day. Despite powerful spam filters, however, scams still find their way into users' inboxes—ranging from fake job offers to malware-laden attachments. These threats are more than just annoying; they're the most common entry point for cyberattacks today.

You've just been hired by one of these major email providers. Your team has been tasked with an ambitious challenge: uncover new ways to improve spam detection models. Traditional spam filters primarily focus on key words (content), known malicious URLs, or metadata (headers). Your company wants to an explore a new idea: **does the emotional tone of an email—the way it feels—help signal whether it's safe or dangerous for our clients?**

Think about it. Scammers rely heavily on manipulating intense emotions, like fear, urgency, or even excitement. Phishing emails are known to create a sense of crisis ("Your account has been compromised!") or irresistible opportunity ("I want to send you a million dollars!"). Legitimate communications, on the other hand, might sound more neutral, professional, calm, or politely positive. By analyzing the *emotional sentiment* of emails, we hope to uncover hidden patterns that traditional keyword patterns miss.

Your mission is to investigate whether emotional sentiment can improve spam classification. You'll work with a real-world dataset of thousands of emails, some safe and some phishing attempts. After engineering features that capture what the email says and how it feels, you will train and assess a machine learning model to classify emails based off these patterns. You will also develop an understanding of how to represent text data for modeling by combining emotional tone with important word-based features extracted from the emails.

A major part of your work will involve statistical testing. You should evaluate whether sentiment scores—such as positive or negative emotional tones—are significant predictors of phishing behavior. Are emails that sound too good to be true more likely to be scams? Are highly negative or alarming messages a warning sign? You'll test these ideas using real data and provide clear, evidence-based conclusions.

Finally, you'll visualize your results to convey your findings to a non-technical audience. Your work will help answer critical questions and help shape the direction of email spam filters as they evolve to combat new cyberattacks.

Ready to defend inboxes with data? Start here: https://github.com/w-mayer/DS4002-CS3