Hook: Spam Email Detection

Imagine you've just been hired as a data scientist for UVA's IT Security team. Your first assignment? Develop a reliable system to protect students and faculty from the growing threat of malicious spam emails that can compromise personal information and university systems.

Your task is to evaluate two machine learning models—Logistic Regression and Random Forest Classification—to determine which provides the most effective spam detection. Using a dataset of pre-classified emails, you'll need to implement both models, compare their performance, and make a recommendation to the IT Security team about which model should be deployed to protect the university's email system.

With phishing attacks and email scams becoming increasingly sophisticated, your work will directly impact the digital safety of UVA community members. Again, these emails have the potential to expose personal information related to an individual or associated organization, which can be very harmful to UVA as a whole. As education becomes increasingly interwoven with technology, it is imperative that we ensure both the security and efficiency of online student-communication.

Your final deliverable will be a recommendation report that compares the models and justifies your final model selection.