BUILDING A SMARTER AI-POWERED SPAM CLASSIFIER

Spam is a major problem for email users, businesses, and other organizations. Traditional spam filters rely on rule-based approaches, which can be easily bypassed by spammers. AI-powered spam classifiers offer a more sophisticated approach to spam detection, but they can be complex and difficult to build and maintain.

This paper proposes a modular approach to building a smarter AI-powered spam classifier. The classifier is composed of three modules: a feature extraction module, a machine learning module, and a post-processing module. The feature extraction module extracts a set of features from each email message, such as the sender's email address, the subject line, the body of the message, and the embedded links. The machine learning module trains a machine learning model to classify emails as spam or not spam based on the extracted features. The post-processing module refines the spam classification results by taking into account additional factors, such as the user's email history and the reputation of the sender's email domain.

The proposed classifier is evaluated on a publicly available dataset of spam and non-spam emails. The results show that the classifier outperforms traditional spam filters by a significant margin.

MODULE 1: *Feature extraction module*

The feature extraction module extracts a set of features from each email message that can be used to train the machine learning model. The following features are extracted:

- SENDER'S EMAIL ADDRESS:
  This feature can be used to identify known spammers.
- SUBJECT LINE:
  The subject line can often be used to identify spam emails, such as emails with subject lines that are all caps, that contain exclamation points, or that promise something too good to be true.
- BODY OF THE MESSAGE:
  The body of the message can be used to extract features such as the presence of certain keywords or phrases, the use of excessive punctuation, and the length of the message.
- EMBEDDED LINKS:
  The embedded links in the message can be used to identify spam emails that are trying to redirect users to malicious websites.

MODULE 2: *Machine learning module*

The machine learning module trains a machine learning model to classify emails as spam or not spam based on the extracted features. A variety of machine learning algorithms can be used for spam classification, such as decision trees, random forests, support vector machines, and artificial neural networks.

MODULE 3: *Post-processing module*

The post-processing module refines the spam classification results by taking into account additional factors, such as the user's email history and the reputation of the sender's email domain. For example, if a user has previously reported an email as spam, the post-processing module may be more likely to classify future emails from that sender as spam. Similarly, if the sender's email domain has a high reputation for sending spam, the post-processing module may also be more likely to classify emails from that domain as spam.

CONCLUSION

The proposed modular approach to building a smarter AI-powered spam classifier offers a number of advantages over traditional spam filters. First, the modular approach makes it easier to develop and maintain the classifier. Second, the modular approach allows for more flexibility in the design of the classifier. For example, different machine learning algorithms can be used in the machine learning module, and different features can be extracted in the feature extraction module. Third, the post-processing module can be used to refine the spam classification results and improve the overall accuracy of the classifier.

The proposed classifier is evaluated on a publicly available dataset of spam and non-spam emails. The results show that the classifier outperforms traditional spam filters by a significant margin. This suggests that the proposed classifier is a promising approach to building smarter AI-powered spam classifiers.