AI proposal

Spam Email Detection

Introduction:

Email is one of the most common methods of communication, but it’s also a frequent target for spam, and harmful emails. Spam emails can clutter inboxes and sometimes contain threats, leading to privacy issues or even financial loss. A spam detection system that uses Natural Language Processing (NLP) and machine learning can enhance email security by filtering out unwanted emails and improving user experience.

Problem Statement

Email spam detection is the task of classifying emails as either "spam" or "ham" based on their content. Traditional rule-based spam filters often struggle to keep up with new tactics used by spammers, which makes them ineffective in the long run. By integrating machine learning and NLP techniques, we can build a detection system that is not only accurate but also adaptive, capable of learning and evolving as new spam methods arise.

Goals:

1. **Develop a Robust Classifier:** Create a model that can accurately identify spam emails using NLP techniques like tokenization, stop-word removal, and text vectorization.
2. **Minimize False Positives:** Ensure legitimate emails are not mistakenly flagged as spam by incorporating advanced NLP methods, which help understand email context.

3. **Real-time Filtering:** Enable fast classification to ensure a smooth user experience, using efficient NLP preprocessing steps that don’t impact performance.
4. **Adaptability:** Implement techniques that allow the model to recognize and adapt to new spam tactics over time, leveraging NLP models that can capture evolving language patterns.

Related Work:

Numerous studies have explored machine learning and NLP techniques in email spam detection, with varying degrees of success:

- **Naive Bayes with NLP Preprocessing:** Naive Bayes is a widely used method for spam detection due to its simplicity and effectiveness with text. When combined with NLP preprocessing techniques like tokenization, lemma, and removing stop words, Naive Bayes can improve the accuracy of spam classification.

- **Logistic Regression and Support Vector Machines (SVM) with NLP Feature Engineering:** Both logistic regression and SVM have shown strong results, especially when NLP feature engineering techniques are applied. Methods like TF-IDF and word embeddings help capture the significance and context of words, boosting model performance.

- **Ensemble Methods with NLP Feature Selection:** Ensemble techniques such as Random Forest and AdaBoost improve classification performance by combining multiple models. When paired with NLP-based features like (NER) or (POS) tagging, ensemble models can achieve better results by considering different aspects of email text.

- Transformer-based Models (BERT, GPT) for Contextual Understanding: Transformer models like BERT and GPT are advanced NLP techniques that can understand context at a deeper level. These models are particularly useful for detecting spam messages that may have more nuanced or disguised language, improving accuracy and adaptability.

By using a combination of these machine learning and NLP techniques, we can build a comprehensive and robust spam detection system that maintains high accuracy and adapts to changing spam tactics.