Classifying Emails for Spam Detection

Background:

With the pervasive use of emails in both personal and professional communication, the issue of spam emails has become a significant concern. Spam emails not only clutter inboxes but also pose security risks and privacy threats to users. Traditional methods of spam filtering, such as rule-based systems and blacklisting, have limitations in adaptability and effectiveness. Therefore, the application of machine learning techniques for email classification has gained attention due to its potential to improve accuracy and efficiency in detecting spam.

Method:

In this project, we aim to develop a program for classifying emails into two categories: spam and legitimate (ham). The proposed method involves the following steps:

1. Data Collection and Preprocessing: We will utilize a dataset containing labeled email messages categorized as spam or ham. The dataset will undergo preprocessing steps, including text normalization, tokenization, stop word removal, and stemming or lemmatization.

2. Feature Extraction: Features will be extracted from the preprocessed text data to represent each email. Commonly used techniques for feature extraction in text classification tasks include bag-of-words representation, TF-IDF (Term Frequency-Inverse Document Frequency), and word embeddings.

3. Model Selection and Training: We will train machine learning models using the preprocessed and feature-engineered data. Initially, we will experiment with the Naive Bayes algorithm due to its simplicity and effectiveness in text classification tasks. Additionally, we may explore other algorithms such as Support Vector Machines (SVM) and ensemble methods for comparison.

4. Evaluation: The trained models will be evaluated using various metrics such as accuracy, precision, recall, and F1-score. We will also visualize the performance metrics using techniques like confusion matrices to gain insights into the model's behavior.

5. Interactive Prediction: The final model will be deployed in an interactive application where users can input email text, and the program will predict whether the email is spam or ham.

Experiment:

As of the midpoint of the project, we have completed data preprocessing and model training. The dataset has been cleaned, and features have been extracted from the email text. Initial experiments with the Naive Bayes algorithm have shown promising results in terms of accuracy and efficiency. However, further experimentation and evaluation are required to fine-tune the model and improve its performance.

Conclusion:

In conclusion, the development of a program for classifying emails into spam and ham categories using machine learning techniques has the potential to enhance email security and user experience. By systematically preprocessing the data, selecting appropriate features, and training models, we aim to create an efficient and accurate classification system. Ongoing experiments and refinements will contribute to the effectiveness of the proposed method in combating spam emails.

References:

[1] Kaggle Dataset: Email Spam Classification Dataset. Available at: <https://www.kaggle.com/datasets/purusinghvi/email-spam-classification-dataset>

تصنيف الرسائل الإلكترونية للكشف عن البريد المزعج

الخلفية:

مع الاستخدام الواسع النطاق للبريد الإلكتروني في كل من التواصل الشخصي والمهني، أصبحت مشكلة رسائل البريد المزعج مصدر قلق كبير. فالرسائل الغير مرغوب فيها لا تعبّر فقط عن البريد العشوائي بل تشكل أيضًا مخاطر أمنية وتهديدات للخصوصية للمستخدمين. تعاني الطرق التقليدية لفرز البريد المزعج، مثل الأنظمة المبنية على القواعد والتصنيف السوداء، من قيود في القابلية للتكيّف والفعالية. لذلك، اكتساب تقنيات التعلم الآلي لتصنيف البريد الإلكتروني قد اكتسب أهمية بسبب قدرتها الكبيرة على تحسين الدقة والكفاءة في اكتشاف البريد المزعج.

الطريقة:

في هذا المشروع، نهدف إلى تطوير برنامج لتصنيف الرسائل الإلكترونية إلى فئتين: البريد المزعج والشرعي (البريد الحقيقي). تتضمن الطريقة المقترحة الخطوات التالية:

1. جمع وتجهيز البيانات: سنستخدم مجموعة بيانات تحتوي على رسائل بريد إلكتروني مصنفة كبريد مزعج أو بريد حقيقي. ستخضع المجموعة البيانات لعمليات تجهيز البيانات، بما في ذلك تطبيع النص، وتجزئة الكلمات، وإزالة الكلمات الزائدة، والتجذيع أو التصغير.

2. استخراج الميزات: سيتم استخراج الميزات من البيانات النصية المجهزة لتمثيل كل رسالة بريد إلكتروني. تشمل تقنيات استخراج الميزات المستخدمة الإحصاء الكلي للكلمات، وتقنية تردد الكلمة المعكوسة (TF-IDF)، والتضمينات الكلامية.

3. اختيار النموذج والتدريب: سنقوم بتدريب نماذج تعلم الآلة باستخدام البيانات المجهزة والتي تم استخراج الميزات منها. في البداية، سنقوم بتجربة خوارزمية النموذج البسيط Bayes الساذجة بسبب بساطتها وفعاليتها في مهام تصنيف النصوص. بالإضافة إلى ذلك، قد نستكشف الخوارزميات الأخرى مثل الآلة الدعم الناقل (SVM) وطرق التجميع للمقارنة.

4. التقييم: سيتم تقييم النماذج المدربة باستخدام مقاييس مختلفة مثل الدقة والاستدعاء والتوازن الهرمي ومعدل F1. سنقوم أيضًا بتصور المقاييس الأدائية باستخدام تقنيات مثل مصفوفات الارتباك لكسب رؤى في سلوك النموذج.

5. التنبؤ التفاعلي: سيتم نشر النموذج النهائي في تطبيق تفاعلي حيث يمكن للمستخدمين إدخال نص البريد الإلكتروني، وسيقوم البرنامج بالتنبؤ ما إذا كان البريد الإلكتروني هو بريد مزعج أم لا.

التجربة:

في نصف المشروع، قمنا بإكمال تجهيز البيانات وتدريب النموذج بنجاح. تم تنظيف المجموعة البيانات، واستخراج الميزات من النصوص البريدية. أظهرت التجارب الأولية مع خوارزمية Bayes الساذجة نتائج واعدة من حيث الدقة والكفاءة. ومع ذلك، يتطلب الأمر إجراء تجارب وتقييم إض

افي لضبط النموذج وتحسين أدائه.

الاستنتاج:

في الختام، فإن تطوير برنامج لتصنيف الرسائل الإلكترونية إلى فئتي البريد المزعج والشرعي باستخدام تقنيات التعلم الآلي له القدرة على تعزيز أمان البريد الإلكتروني وتجربة المستخدم. من خلال تجهيز البيانات بشكل منهجي، واختيار الميزات المناسبة، وتدريب النماذج، نهدف إلى إنشاء نظام تصنيف فعال ودقيق. ستسهم التجارب المستمرة والتعديلات في تحسين فعالية الطريقة المقترحة في مكافحة رسائل البريد المزعج.

المراجع:

[1] مجموعة بيانات Kaggle: Email Spam Classification Dataset. متوفر على: https://www.kaggle.com/datasets/purusinghvi/email-spam-classification-dataset