**تصنيف الرسائل الإلكترونية للكشف عن البريد المزعج**

* **مقدمة:**

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، 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.

في هذا التقرير، سنقوم بدراسة استخدام النماذج الإحصائية Naive Bayes وLogistic Regression لتصنيف البريد الإلكتروني إلى فئات "spam" و "ham" (غير مؤذي). سنقوم بتحليل أداء النماذج وفقًا لمعايير مثل الدقة ومصفوفة الارتباك.

* **الطريقة:**

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

1. استيراد المكتبات اللازمة.
2. يتم تعريف نموذجين للتصنيف:

* الأول يستخدم Naive Bayes.
* الثاني يستخدم Logistic Regression.

1. تعريف الكلاس الرئيسي EmailClassifier: يقوم بقراءة البيانات، تنظيفها، تدريب النماذج، وطباعة النتائج.
2. قراءة وتنظيف البيانات: يتم قراءة البيانات من ملف CSV ومعالجتها لإزالة القيم المفقودة.
3. تدريب النماذج: يتم تدريب النماذج باستخدام تقنية K-Fold Cross Validation.
4. تقييم الأداء: يتم قياس الدقة ومصفوفة الارتباك لكل نموذج.

* **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 Naive Bayesنتائج واعدة من حيث الدقة والكفاءة. ومع ذلك، يتطلب الأمر إجراء تجارب وتقييم إضافي 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 Email Spam Classification Dataset. متوفر على Kaggle:

<https://www.kaggle.com/datasets/purusinghvi/email-spam-classification-dataset>

**البرنامج:**

