مجلة جامعة طرطوس للبحوث والدراسات العلمية _ سلسلة العلوم الهندسية المجلد (٩) العدد (٧) ٢٠٢٥
Tartous University Journal for Research and Scientific Studies - engineering Sciences Series Vol. (9) No. (7) 2025

استخدام خوارزميات التعلم الآلي لتحسين دقة كشف الرسائل النصية المزعجة

د. راغب طعمه *

(تاريخ الإيداع ٢٠٢٥/١/٢٧ . قُبِل للنشر في ٢٠٢٥/٦/١٢)

□ ملخّص □

التقدم التكنولوجي السريع وزيادة الاعتمادية على الوسائط الالكترونية في كافة مجالات الحياة دفعت الى حماية المستخدمين من التأثيرات السلبية لرسائل ال SMS المزعجة، في هذا البحث تم استخدام خوارزميات التعلم الآلي لتمييز الرسائل غير المزعجة من الرسائل المزعجة، نستعرض النتائج التي تم الحصول عليها باستخدام خوارزميات الانحدار اللوجستي Logistic regression و آلة متجهة الدعم (SVM) Support vector machineوالغابة العشوائية Random Forest، عرضت النتائج قيم اختبار الدقة والاستدعاء ومقياس F1-score حيث حققت الغابة العشوائية أفضل أداء من حيث الدقة والاستدعاء في تصنيف الرسائل المزعجة (spam) والرسائل غير المزعجة (ham). يؤكد البحث أن استخدام الغابة العشوائية يعد أكثر فعالية في تصنيف رسائل الـ SMS، حيث أظهرت نتائج دقيقة وفعالة في التحديد بين الرسائل المرغوب بها وغير المرغوب بها.

الكلمات المفتاحية: التعلم الآلي – الانحدار اللوجستي(LR) – الغابة العشوائية(RF) – آلة دعم المتجهات(SVM)

*مدرس في قسم هندسة تكنولوجيا المعلومات – كليّة هندسة تكنولوجيا المعلومات والاتصالات – جامعة طرطوس– سوريا

مجلة جامعة طرطوس للبحوث والدراسات العلمية _ سلسلة العلوم الهندسية المجلد (٩) العدد (٧) ٢٠٢٥
Tartous University Journal for Research and Scientific Studies - engineering Sciences Series Vol. (9) No. (7) 2025

Using of Machine Learning Algorithms to Enhance the Accuracy of Spam Messages Detection

Dr. Ragheb Toemeh *

(Received 27/1/2025 . Accepted 12/6/2025)

□ ABSTRACT □

The rapid technological advancement and the increasing reliance on electronic media in all aspects of life have necessitated the protection of users from the negative impacts of unwanted SMS messages. In this research, machine learning algorithms were employed to distinguish between non-spam and spam messages. The results obtained using Logistic Regression, Support Vector Machine (SVM), and Random Forest algorithms are presented. The evaluation included accuracy, recall, and F1-score metrics, where the Random Forest algorithm demonstrated the best performance in terms of both accuracy and recall when classifying spam and ham messages. The study confirms that the use of the Random Forest algorithm is more effective for SMS classification, as it yielded precise and efficient results in differentiating between desired and undesired messages.

Keywords: machine learning - logistic regression (LR) - random forest (RF) - support vector machine (SVM)

* Lecturer, Information Technology Engineering Department, Information and communication Technology Engineering, Tartous University, Syria.

١. المقدمة:

في ظل التوسع الكبير في الاعتماد على التطبيقات الرقمية وتوليد كميات هائلة من البيانات، أصبحت تقنيات تحليل البيانات والتعلم الآلي أدوات أساسية لاستخلاص المعرفة من هذا الكم المتزايد من المعلومات. ومن بين أكثر وسائل الاتصال استخداماً، تبرز خدمة الرسائل القصيرة (SMS) التي، رغم شعبيتها، لا تزال تعاني من مشكلة الرسائل المزعجة(Spam) ، والتي تمثل تحدياً أمنياً يتطلب حلولاً ذكية وفعالة.

ورغم وجود تقنيات أولية مثل القوائم السوداء **[١]** وبعض الأساليب الإحصائية، فإن هذه الحلول لم تصل بعد إلى درجة النضج الكافية لضمان الحماية الكاملة. لذلك، اتجهت الأبحاث إلى استخدام تقنيات أكثر تطوراً تعتمد على التعلم الآلي والتنقيب في البيانات. وقد أظهرت عدة دراسات نتائج واعدة في هذا المجال، حيث حقق نموذج (LSTM) المستخدم في **[٢]** دقة تصنيف بلغت ٩٨,٥% باستخدام مجموعة بياناتUCI ، بينما اعتمدت دراسة أخرى **[٣]** على تدريب نموذج انحدار لوجستي باستخدام خوارزمية مستعمرة النحل الاصطناعية (ABC)لتحسين أداء الكشف عن الرسائل المزعجة.

كما قدمت دراسة **[٤]** إطاراً متعدد المراحل للكشف عن الرسائل المزعجة على منصةYouTube ، باستخدام خوارزميات Naïve Bayes والانحدار اللوجستي، وحققت دقة تجاوزت ٨٧% في بعض الأدوات مثلWeka. في المقابل، استخدمت دراسة **[٥]** خوارزميات Support Vector Classifier ومصنف الغابة العشوائية على مجموعة بيانات UCL Spambase ، وأظهرت أداءً أعلى بدقة بلغت ٩١,٣٦%. من جهة أخرى، ركزت دراسة **[٦]** على استخدام آلة متجه الدعم (SVM) واقترحت نوى متقدمة مثل نواة السلسلة لتحسين أداء الفلترة في الزمن الحقيقي.

بناءً على ذلك، يهدف هذا البحث إلى تقييم أداء بعض خوارزميات التعلم الآلي في تصنيف الرسائل النصية القصيرة وتمييزالرسائل الغير مرغوب بها عن الرسائل المرغوبة، بالاعتماد على معايير الدقة والاستدعاء والدقة الإيجابية(Precision) و F1-score.

٢. أهمية البحث وأهدافه:

في ظل تصاعد استخدام وسائل الاتصال الرقمية، أصبحت الرسائل النصية القصيرة (SMS) وسيلة شائعة للتواصل، سواء على مستوى الأفراد أو المؤسسات. ومع هذا الانتشار، تصاعدت أيضاً ظاهرة الرسائل المزعجة والغير مرغوب بها، التي قد تتضمن محاولات احتيال أو انتهاكات للخصوصية أو تسويق عدواني غير مرغوب به. والتحدي الأكبر في كشف هذه الرسائل يكمن في تطور أساليب التمويه فيها، مما يجعل الأساليب التقليدية غير كافية للتمييز الفعّال بينها وبين الرسائل المرغوب بها.

تبرز أهمية هذا البحث في تطوير آلية فعالة لاكتشاف وتصنيف الرسائل المزعجة ضمن خدمة الرسائل القصيرة، وذلك من خلال توظيف تقنيات تعلم آلي قادرة على تحسين دقة التصفية وتقليل نسبة الخطأ في التصنيف. ويسعى البحث إلى تزويد المهتمين بتقنيات الاتصالات وأنظمة الحماية الرقمية بنتائج قابلة للتطبيق في بيئات حقيقية.

يهدف هذا البحث إلى:

- تطبيق ومقارنة أداء ثلاث خوارزميات تعلم آلي في تصنيف الرسائل النصية القصيرة، وهي: الانحدار اللوجستي وآلة متجه الدعم، ومصنف الغابة العشوائية.
- تحليل نتائج الخوارزميات وفقاً لمقاييس الأداء القياسية مثل: الدقة(Accuracy) ، الاستدعاء(Recall) ، ومعامل F1 (F1-score) و الدقة الإيجابية(Precision) ، باستخدام مجموعة بيانات معيارية.
- تحديد الخوارزمية الأكثر كفاءة بناءً على النتائج التجريبية، وتقديم توصيات لإمكانية استخدامها في أنظمة تصفية الرسائل القصيرة المزعجة.

٣. خوارزميات التعلم الآلي والتصنيف

خوارزميات التعلم الآلي (Machine Learning Algorithms – ML) تمثل مجموعة من الأساليب الرياضية القوية التي تتيح للحواسيب القدرة على التعلم من البيانات واكتشاف الأنماط، ومن ثم تعميمها للتنبؤ بالنتائج أو اتخاذ قرارات دون برمجة صريحة. تُستخدم هذه الخوارزميات في مجالات متعددة مثل التعرف على الصور، تحليل النصوص، التوصية بالمنتجات، والتنبؤ بسلوك المستخدم. ومن أبرز تقنياتها الشبكات العصبية الاصطناعية، الانحدار اللوجستي، وخوارزميات التصنيف المختلفة.

خوارزميات التصنيف تعتبر الأساس في التمييز بين الرسائل النصية المزعجة (Spam) وغير المزعجة (Ham) .وقد تم اعتماد ثلاث خوارزميات تصنيف شائعة وفعالة لاختبار أدائها على مجموعة البيانات وهي:

٣-١ الانحدار اللوجستي(Logistic Regression – LR)

يُعد الانحدار اللوجستي من النماذج الإحصائية المستخدمة في التصنيف الثنائي، حيث يقوم بتقدير احتمال انتماء العينة إلى إحدى الفئتين باستخدام الدالة اللوجستية، بناءً على المتغيرات المدخلة **[3]**. ويمتاز هذا النموذج ببساطته وكفاءته خاصة في الحالات التي تكون فيها البيانات عالية الأبعاد.

٣-٢ الغابة العشوائية(Random Forest – RF)

الغابة العشوائية هي خوارزمية تجميع تعتمد على إنشاء عدة أشجار قرار مدربة على عينات مختلفة من البيانات، ودمج نتائجها لاتخاذ القرار النهائي. تُعرف هذه التقنية بفعاليتها في التعامل مع البيانات المعقدة، كما أنها تقلل من مخاطر فرط التخصيص**[7]** .

٣-٣ آلة متجه الدعم(Support Vector Machine – SVM)

تُعد آلة متجه الدعم من الخوارزميات المتقدمة التي تسعى إلى إيجاد المستوى الفائق (Hyperplane)الأمثل الذي يفصل بين الفئات بأكبر هامش ممكن. وتدعم هذه الخوارزمية كلا من التصنيف الخطي وغير الخطي باستخدام نوى مخصصة، مما يجعلها فعالة في حالات البيانات غير المنفصلة بوضوح **[8]**.

٤. استخراج الميزات

تُعدّ عملية استخراج الميزات من المراحل الجوهرية في معالجة اللغة الطبيعية(NLP) ، حيث تهدف إلى تحويل البيانات النصية إلى تمثيلات عددية قابلة للفهم والمعالجة من قبل الخوارزميات الذكية، سواء

الإحصائية أو نماذج التعلم الآلي العميق. ومن التقنيات الشائعة لتحقيق هذا الهدف تقنية تردد المصطلح – تردد المستند العكسي (TF–IDF –Term Frequency–Inverse Document Frequency) ، التي تسهم في تحديد أهمية الكلمات داخل المستندات النصية.

تمر عملية استخراج الميزات بمراحل أساسية هي:

١) تجزئة النص إلى عناصر (Tokenization)

وهي العملية التي يتم فيها تقسيم السلسلة النصية إلى كلمات منفصلة (رموز مميزة أو(Lemmas ، مما يسمح بتكوين قائمة من العناصر التي يمكن تحليلها لاحقاً.

٢) تحويل الرموز إلى متجهات (Vectorization)

بعد تجزئة النص، يتم تمثيل كل كلمة في شكل عددي باستخدام تقنيات مثل نموذج حقيبة الكلمات (Bag–of–Words).ويُستخدم في هذا السياق CountVectorizerلحساب عدد مرات تكرار كل كلمة في المستندات النصية.

٣) حساب أوزان الكلمات باستخدام TF–IDF

تُستخدم هذه التقنية لمنح وزن لكل كلمة بناءً على أهميتها داخل المستند، وليس فقط عدد مرات تكرارها. فالكلمات الشائعة والتي تظهر في العديد من المستندات تحصل على وزن أقل، بينما تُمنح الكلمات النادرة وزناً أعلى.

يتم حساب قيمة TF–IDF لكل كلمة عن طريق ضرب معدل تكرار الكلمة داخل المستند (TF)بمعدل تكرارها العكسي في كل المستندات (IDF). هذا يسمح بتحديد أهمية كلمة معينة في مستند ما بناءً على تكرارها في ذلك المستند وندرتها في المجموعة الكاملة من المستندات**[9]**

حساب TF–IDF للمصطلح i ضمن المستند jهو:

$$TF - IDF = TF(i,j) \times IDF(i) \qquad (١)$$

- IDF : تردد المستند العكسي
- TF: تردد المصطلح (term) ، وهو يعبر عن عدد مرات ظهور المصطلح i في المستندj .

تردد المصطلح (TF)، ويعرف أيضا باسم. عدد مرات ظهور الكلمة في المستند، مقسوماً على إجمالي عدد الكلمات في هذا المستند، أي أنه يقيس مدى تكرار ظهور المصطلح في المستند**[9]** .

$$TF(i,j) = \frac{Term\ i\ frequency\ in\ document\ j}{Total\ words\ in\ document\ j} \qquad (2)$$

تكرار المستند العكسي (IDF)، يتم حسابه على أنه لوغاريتم عدد المستندات الموجودة في المجموعة مقسوماً على عدد المستندات التي يظهر فيها المصطلح المحدد، أي أنه يقيس مدى أهمية المصطلح **[9]**.

$$IDF(i) = \log_2(\frac{Total\ documents}{documents\ with\ term\ i}) \qquad (3)$$

حيث:

- TF (i,j) :تردد المصطلح i في المستند j، وهو يتم حسابه كنسبة لعدد مرات ظهور المصطلح i في المستند j إلى إجمالي عدد المصطلحات في المستند j .
- IDF (i) :تقييم تردد المستند العكسي للمصطلح i، وهو يحسب كـ لوغاريتم للنسبة بين إجمالي عدد المستندات في المجموعة وعدد المستندات التي تحتوي على المصطلح i.

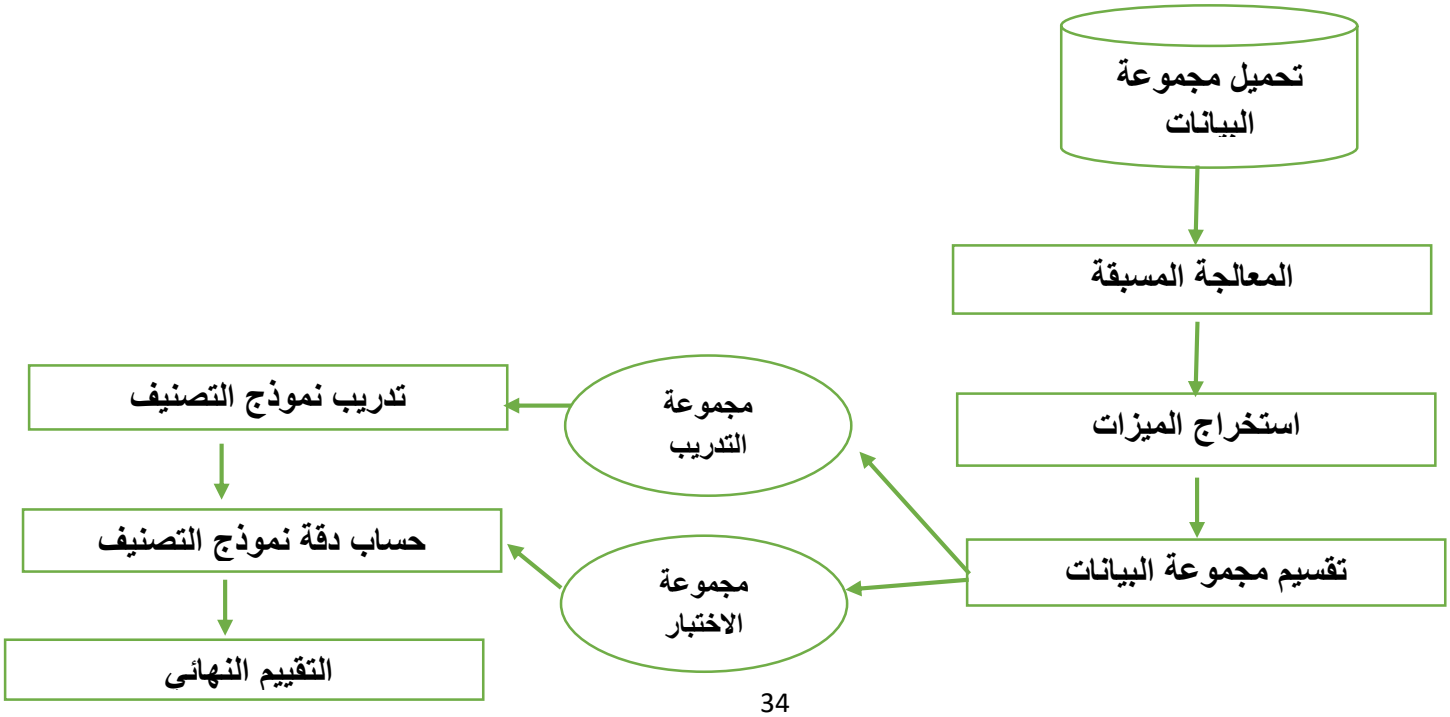
أثناء حساب TF، تعتبر جميع المصطلحات على نفس القدر من الأهمية. ومع ذلك، فمن المعروف أن بعض المصطلحات، مثل "هو"، و "من"، و"ذلك"، قد تظهر في كثير من الأحيان ولكن ليس لها أهمية كبيرة.

٥. مجموعة البيانات

إن مجموعة البيانات المستخدمة هي مجموعة الرسائل النصية القصيرة المزعجة وهي عبارة عن مجموعة من الرسائل النصية القصيرة التي تحمل علامات والتي تم جمعها لأبحاث الرسائل القصيرة المزعجة (spam sms). تحتوي على مجموعة واحدة من الرسائل النصية القصيرة باللغة الإنجليزية مكونة من ٥,٥٧٤ رسالة، تم تصنيفها على أنها رسائل غير مزعجة (ham) أو رسائل مزعجة (spam). تم استخراج مجموعة من ٤٢٥ رسالة نصية مزعجة يدوياً من موقع الويب Grumbletext وهو منتدى في المملكة المتحدة يقدم فيه مستخدمو الهواتف المحمولة شكاوى عامة حول الرسائل المزعجة، ومعظمها دون الإبلاغ عن نفس الرسالة المزعجة المستلمة.

٦. منهجية البحث وأدواته:

يركز هذا البحث على تصنيف رسائل SMS إلى رسائل مزعجة ورسائل غير مزعجة باستخدام خوارزميات التعلم الآلي ولقد تم توضيح مراحل العمل التي قمنا بها لتصميم النظام ضمن المخطط الصندوقي في الشكل (١).



الشكل (١): المخطط الصندوقي لمنهجية البحث

في المرحلة الأولى يتم تحميل مجموعة البيانات المستخدمة في البحث يلي هذه المرحلة إجراء مجموعة من عمليات المعالجة المسبقة عليها، في المرحلة التالية يتم استخراج الميزات من النص وبعد ذلك تم تقسيم البيانات الى مجموعتين مجموعة بيانات للتدريب ومجموعة بيانات للاختبار وفي مرحلة لاحقة تؤخذ بيانات التدريب وتطبق على نموذج التصنيف المصمم من اجل تدريبه على أنماط النصوص المختلفة مع التسميات الخاصة بها، وبينما تستخدم مجموعة بيانات الاختبار من اجل حساب دقة النموذج ، وفي المرحلة الأخيرة وهي مرحلة تقييم النماذج المستخدمة.حيث تم تنفيذ هذه المراحل من خلال استخدام لغة البرمجة بايثون ضمن بيئة العمل google colab .

6-١ تحميل مجموعة البيانات:

تم تحميل مجموعة بيانات المكونة من ٥٥٧٢ سطر وعمودين (label, message) ضمن كود البايثون يظهر جزء من مجموعة البيانات المستخدمة في الشكل (٢):

	label	message
0	ham	Go until jurong point, crazy.. Available only ...
1	ham	Ok lar... Joking wif u oni...
2	spam	Free entry in 2 a wkly comp to win FA Cup fina...
3	ham	U dun say so early hor... U c already then say...
4	ham	Nah I don't think he goes to usf, he lives aro...

الشكل (٢): مجموعة البيانات

يعرض الشكل (٢) جزءاً من مجموعة بيانات الرسائل النصية، حيث يظهر عمودان: الأول هو التصنيف (رسالة مزعجة spam أو غير مزعجة ham)، والثاني نص الرسالة نفسه.

6-2 المعالجة المسبقة للبيانات

تنطيف البيانات

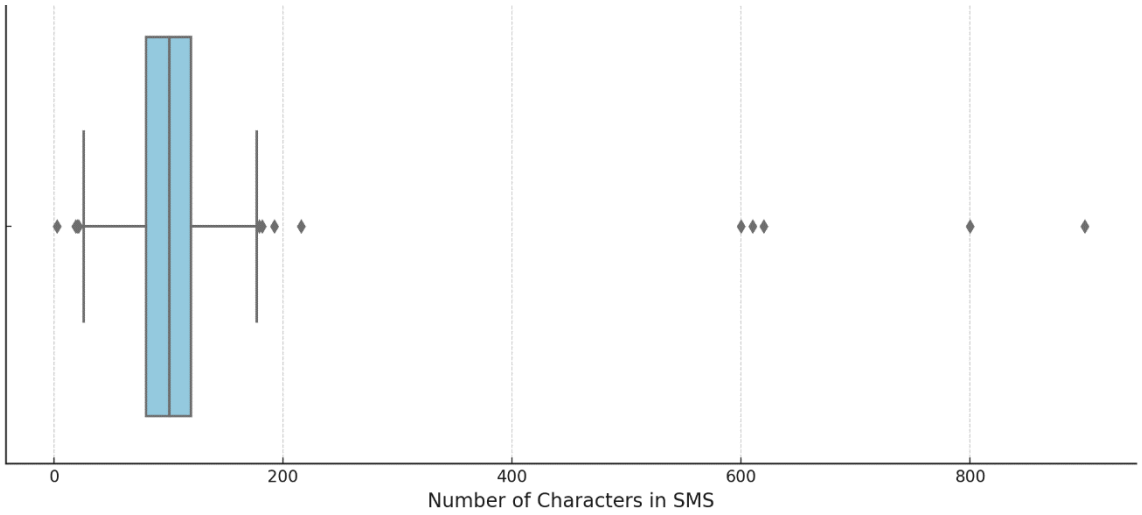
تفيد عملية تنظيف البيانات في إزالة البيانات المكررة والبيانات غير الكاملة وتحويل البيانات الى شكل موحد، تعد هذه المرحلة ضرورة أساسية قبل القيام بعملية تدريب الشبكة العصبونية الهدف منها تقليل زمن المعالجة وزمن التصنيف بالإضافة الى زيادة دقة التصنيف كون البيانات المكررة والسمات غير الهامة غالبا ماتشوش على خوارزمية التصنيف.

- **القيم المفقودة:** يمكن أن تسبب البيانات المفقودة انخفاضاً في حجم العينة، وبالتالي هناك حاجة إلى التأكد من أن البيانات المفقودة ليست متحيزة وتخفي حقيقة مزعجة.
- **القيم المكررة:** تعتبر السجلات المكررة من قيود تصنيف سجلات قاعدة البيانات ،إذ أن القيام بهذه الخطوة يضمن عدم تحيز خوارزمية التعليم للسجلات المكررة والذي يؤدي الى نتائج غير صحيحة وبالتالي تحسين دقة الكشف بالإضافة الى التخفيض من متطلبات مساحة التخزين وبالتالي تقليل زمن المعالجة. تم إزالة ٤٠٣ قيمة مكررة من قاعدة البيانات.

- **القيم المتطرفة:** يمكن أن تؤثر القيم المتطرفة بشكل ملحوظ على نماذجنا، ومن اجل أغراض استكشاف البيانات، سنقوم بإنشاء ميزات جديدة من خلال كود برمجي بلغة البايثون.

تشير القيم المتطرفة من حيث عدد الحروف وعدد الكلمات وعدد الجمل إلى نقاط بيانات تختلف بشكل كبير عن غالبية البيانات في هذه الفئات المعنية.

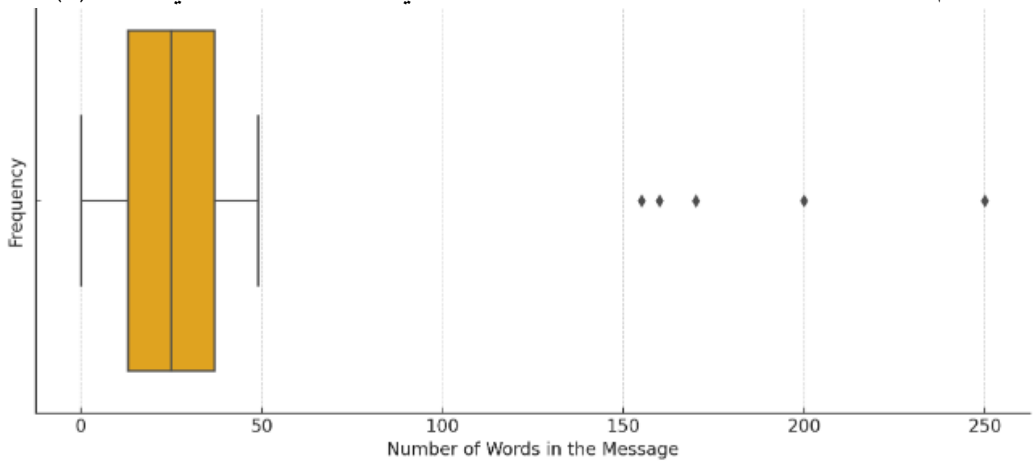
قد تكون القيمة المتطرفة في عدد الأحرف عبارة عن رسالة نصية تحتوي على عدد كبير بشكل غير عادي من الأحرف مقارنة بالرسائل الأخرى، تم تنفيذ كود برمجي لمعرفة القيم المتطرفة من حيث عدد الأحرف في مجموعة البيانات كما يظهر في الشكل (٣):



الشكل (٣)، القيم المتطرفة من حيث عدد الأحرف

يبين الشكل (٣) أنه يوجد ثلاث رسائل تحتوي على حوالي ٦٠٠ حرف، وواحدة تحتوي على ٨٠٠ حرف وواحدة تحتوي على ٩٠٠ حرف وهذه الرسائل تعتبر ذات قيم متطرفة من حيث عدد الاحرف مقارنة بغالبية الرسائل الأخرى.

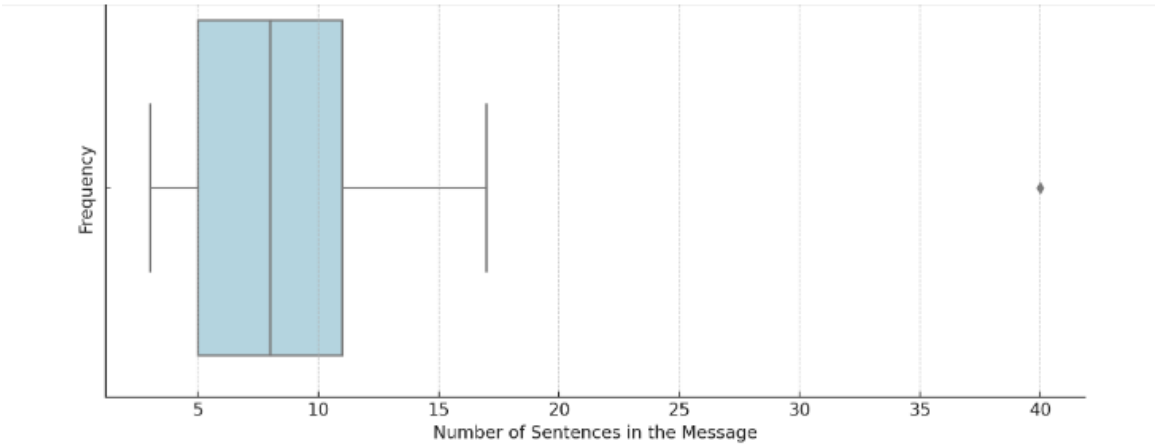
بينما القيم المتطرفة لمجموعة البيانات من حيث عدد الكلمات في نص الرسالة تظهر في الشكل (٤):



الشكل (٤) القيم المتطرفة من حيث عدد الكلمات

يوضح الشكل (٤) وجود بعض الرسائل تحتوي على عدد كلمات مرتفع حيث يوجد ثلاث رسائل تحوي ١٥٠–١٧٥ كلمة، وواحدة ٢٠٠ كلمة، وأخرى ٢٥٠ كلمة، وهي تعتبر قيماً متطرفة مقارنة بباقي الرسائل.

بينما تظهر القيم المتطرفة لمجموعة البيانات من حيث عدد الجمل في نص الرسالة في الشكل (٥):



الشكل (٥) القيم المتطرفة من حيث عدد الجمل

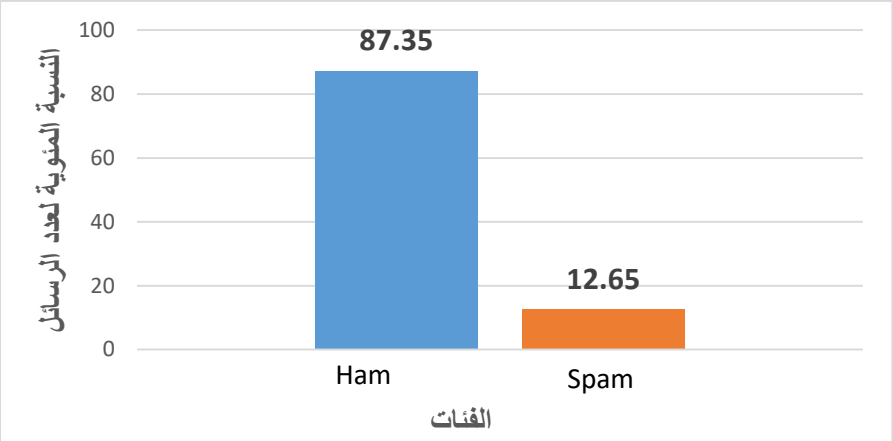
يعرض الشكل (٥) توزيع عدد الجمل في الرسائل النصية، حيث أن معظم الرسائل تحتوي أقل من ١٥ جملة، مع وجود رسالة استثنائية تحتوي على ٤٠ جملة.

تتم إزالة جميع هذه القيم المتطرفة حيث تتم إزالة ال sms التي يكون عدد محارفها أكبر من ٥٠٠ وعدد الكلمات ضمنها أكثر من ١٤٠ كلمة وعدد الجمل أكثر من ٢٠.

تحليل البيانات الاستكشافية (Exploratory Data Analysis– EDA):

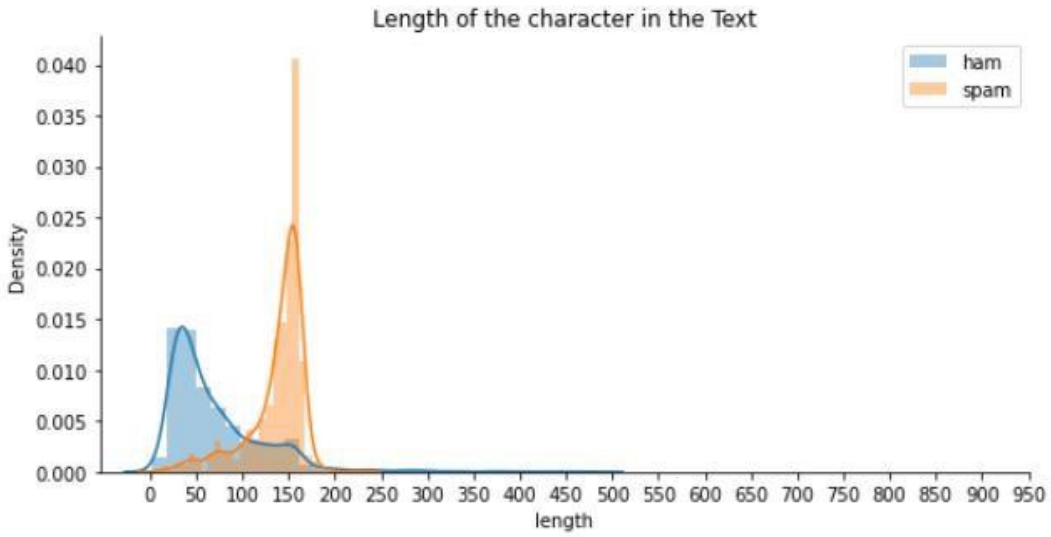
يعرف تحليل البيانات الاستكشافية (EDA) بأنه نهج إحصائي حيوي يكمل تحليل البيانات المؤكدة (Confirmed Data Analysis – CDA)من خلال المساعدة في اكتشاف الأنماط وتحسين الفرضيات [٩]. يساعد EDAأيضاً في اكتشاف التحيز المحتمل في مجموعة البيانات من خلال فحص التوزيع النسبي للفئات المختلفة. في حال كان هناك انحراف واضح في عدد أو نسبة الرسائل بين الفئات (مثل وجود عدد أكبر بكثير من الرسائل غير المزعجة مقارنة بالرسائل المزعجة أو العكس)، فهذا قد يشير إلى وجود تحيز في البيانات.

تحوي مجموعة البيانات التي تم تحميلها على ٤٥١١ رسالة غير مزعجة (ham) و٦٥٣ رسالة مزعجة (spam)، كما يظهر في الشكل (٦) توزيع الرسائل المرغوبة والرسائل غير المرغوبة في مجموعة البيانات الخاصة بنا بالنسبة المئوية. يشير هذا التوزيع إلى وجود تحيز نحو الرسائل المرغوبة، حيث أن النسبة المئوية للرسائل الغير مرغوبة أقل بكثير من الرسائل المرغوبة.



الشكل (٦) نسبة توزع الرسائل المرغوبة والرسائل غير المرغوب فيها ضمن مجموعة البيانات

يُظهر الشكل (٦) أن نسبة الرسائل المزعجة (spam) ضمن مجموعة البيانات تبلغ ١٢,٦٥%، في حين تشكل الرسائل غير المزعجة (ham) نسبة ٨٧,٣٥%، مما يدل على وجود خلل واضح في توازن الفئات. ومن ناحية أخرى، يوضح الشكل (٧) أن متوسط طول الرسائل يختلف بشكل ملحوظ بين الفئتين، إذ يبلغ متوسط طول الرسائل غير المزعجة حوالي ٤٠ حرفاً، مقارنة بمتوسط ١٦٠ حرفاً للرسائل المزعجة. يشير هذا التفاوت إلى أن طول النص قد يُعد عاملاً مهماً يمكن استغلاله في تحسين أداء نماذج تصنيف الرسائل.



الشكل (٧): طول الحرف في النص

إزالة كلمات التوقف وجعل الهدف رقمي

الكمبيوتر لا يفهم النصوص كما يفهمها البشر، فهي بالنسبة له مجرد مجموعة من الرموز. لذلك، من الضروري إجراء مزيد من المعالجة لتحويل النصوص إلى صيغة قابلة للفهم من قبل نماذج تعلم الآلة، مما يجعل البيانات أكثر نظافةً ووضوحاً.

أحد أهم خطوات هذه المعالجة هو جعل الهدف رقمياً، وذلك لأن معظم خوارزميات تعلم الآلة لا تتعامل مع البيانات النصية مباشرة عند تدريب نموذج التصنيف. بل تتطلب أن تكون الفئات (Labels) ممثلة بقيم عددية لتتمكن من حساب الدوال الرياضية وتحديد الحدود بين الفئات. لذلك، تم تحويل الفئة hamإلى

القيمة العددية 0والفئة spamإلى القيمة العددية 1، وذلك ضمن عمود جديد باسم (label) مضاف إلى قاعدة البيانات.

في الخطوة التالية، وهي استخراج الأحرف الأبجدية فقط، حيث تتم إزالة علامات الترقيم وكلمات التوقف، وهي كلمات شائعة جداً مثل (is, an, the) تستخدم في بناء الجمل ولكنها لا تحمل معلومات مفيدة للنموذج، وبالتالي فإن إزالتها تساعد على تقليل الضجيج في البيانات وتحسين الأداء.

لغرض إزالة كلمات التوقف، تم استخدام مكتبة NLTK، وهي إحدى أدوات بايثون المتخصصة في مجال معالجة اللغة الطبيعية، وتوفر مجموعة افتراضية من هذه الكلمات. بعد إزالة علامات الترقيم وكلمات التوقف، نحصل على نصوص أكثر تركيزاً على الكلمات ذات المعنى، كما يظهر في الشكل(8) :

	label	message	clean_message
0	0	Go until jurong point, crazy.. Available only ...	Go jurong point crazy Available bugis n great ...
1	0	Ok lar... Joking wif u oni...	Ok lar Joking wif oni
2	1	Free entry in 2 a wkly comp to win FA Cup fina...	Free entry wkly comp win FA Cup final tkts 21s...
3	0	U dun say so early hor... U c already then say...	dun say early hor c already say
4	0	Nah I don't think he goes to usf, he lives aro...	Nah think goes usf lives around though

الشكل (٨) مجموعة البيانات بعد إزالة علامات الترقيم وكلمات التوقف وإضافة عمود الهدف label

يعرض الشكل (٨) نسخة مصفّاة من البيانات بعد إزالة علامات الترقيم وكلمات التوقف وتحويل التصنيفات إلى أرقام.

6-3 استخراج الميزات

بعد إعداد البيانات وتجهيزها للمعالجة، تم تنفيذ مرحلة استخراج الميزات النصية باستخدام أدوات معالجة اللغة الطبيعية في مكتبة scikit-learn في بيئةPython ، وذلك بغرض تحويل الرسائل النصية إلى تمثيل عددي قابل للاستخدام من قبل نماذج التعلم الآلي. وتم استخراج الميزات وفق المراحل الثلاث :

٦-٣-١ تجزئة النصوصTokenization ، حيث تم تقسيم كل رسالة نصية إلى مجموعة من الكلمات (tokens) باستخدام أدوات المعالجة المتاحة في مكتبة nltk و.sklearn

٦-٣-٢ تحويل النصوص إلى شعاع (Vectorization) باستخدام CountVectorizerلتحويل الرسائل النصية إلى مصفوفة عددية تمثل عدد مرات تكرار كل كلمة في كل رسالة. يُنتج هذا النموذج مصفوفة ثنائية الأبعاد حيث تمثل الصفوف الرسائل، وتمثل الأعمدة الTokens .

٦-٣-٣ حساب أوزان الكلمات باستخدامTF-IDF بعد الحصول على تمثيل أولي بالنموذج العددي، تم تطبيق TfidfVectorizer لحساب وزن كل كلمة بناءً على أهميتها النسبية في كل رسالة مقارنة بباقي الرسائل.

تمثل مصفوفة X الناتجة التمثيل العددي النهائي للنصوص، وهي تُستخدم كمدخل للمرحلة التالية لتدريب نموذج التصنيف.

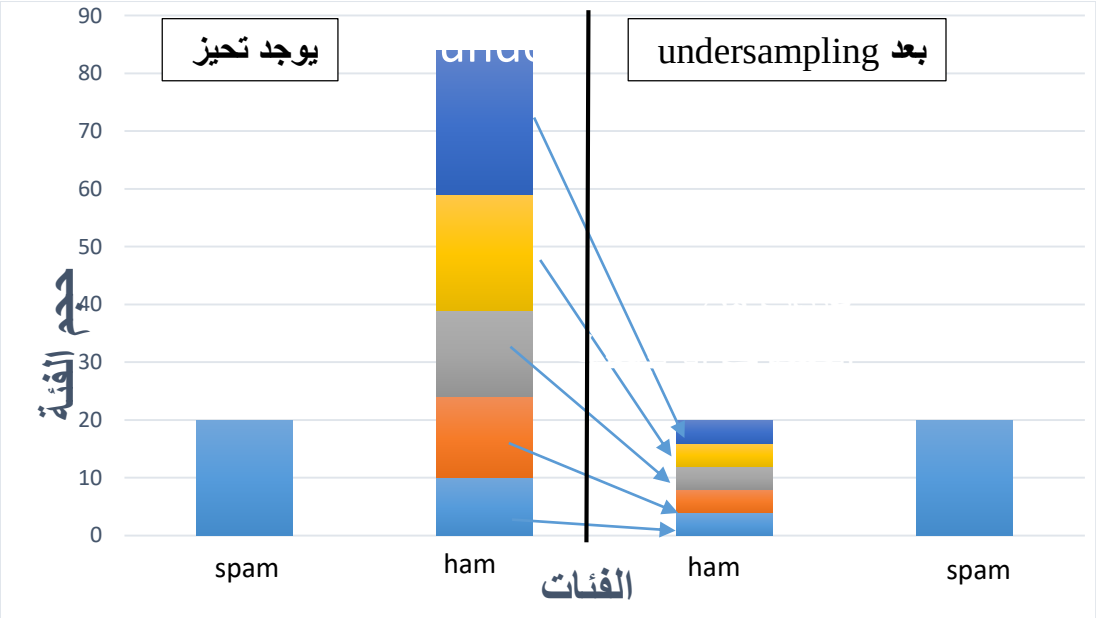
4-6 تقسيم مجموعة البيانات

خوارزميات التعلم الآلي تتعلم من مجموعة التدريب، ولتقييم مدى جودة أداء خوارزميات التعلم الآلي المدربة، يتم إجراء التنبؤات على مجموعة الاختبار. ولذلك تم تقسيم البيانات لدينا إلى مجموعات التدريب والاختبار اذ استخدم 80% من العينات للتدريب و٢٠% من العينات للاختبار.

تقنية Smote:

تقنية SMOTE هي تقنية تستخدم لجعل مجموعة البيانات متوازنة من خلال تولد عينات اصطناعية من فئة الأقلية لتصبح مساوية لفئة الأكثرية (oversampling) أو العكس أخذ عينات من فئة الأكثرية مساوية لفئة الأقلية.(undersampling) يتم استخدامه للحصول على مجموعة تدريب متوازنة بشكل صناعي أو شبه متوازنة**[10]** ، والتي يتم استخدامها بعد ذلك لتدريب المصنف.

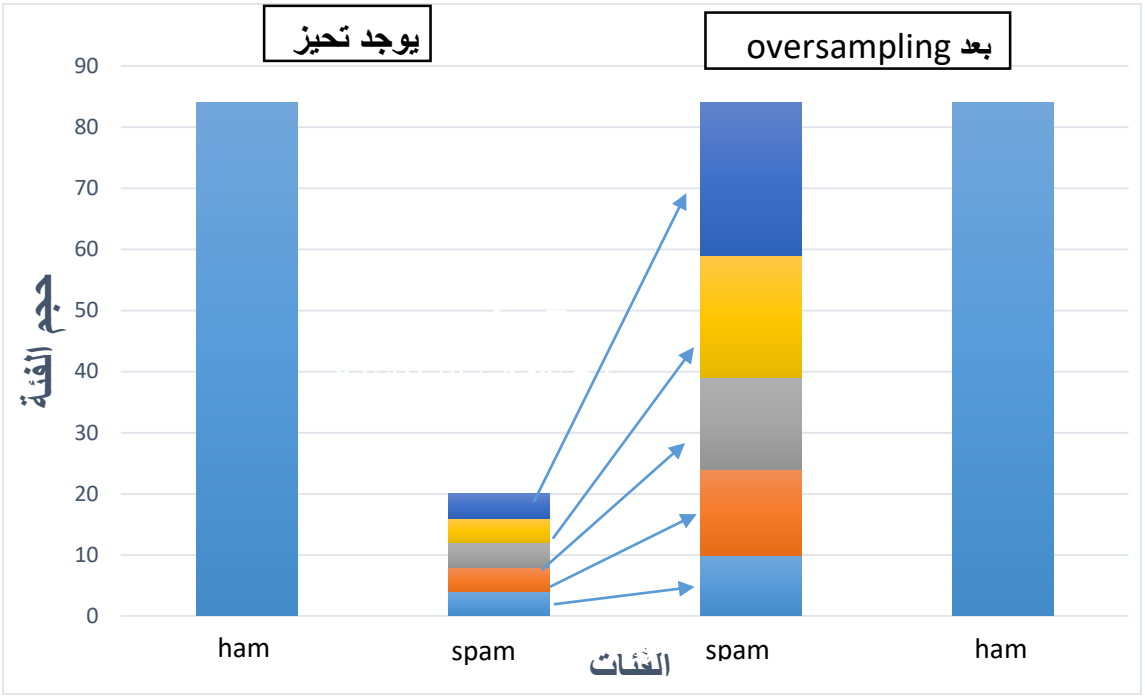
في طريقة smote– undersampling نأخذ حجم بيانات من الصنف الأكثر يساوي حجم البيانات من الصنف الأقل كما يظهر في الشكل (٩):



الشكل (٩): طريقة موازنة مجموعة البيانات باستخدام تقنية smote– undersampling

يوضّح الشكل (٩) كيفية تقليل حجم بيانات الفئة الأكبر (Ham) لتتناسب مع الفئة الأصغر (Spam) لتحقيق توازن في البيانات.

في طريقة smote–oversampling نقوم بتكرار صنف البيانات الأقل ليصبح مساوياً لحجم البيانات في الصنف الأكثركما يظهر في الشكل (١٠).



الشكل (١٠): طريقة موازنة مجموعة البيانات باستخدام تقنية smote–oversampling

يبين الشكل (١٠) طريقة زيادة عدد العينات من الفئة الأقل (Spam) عبر توليد عينات اصطناعية حتى تساوي حجم الفئة الأكبر.

6-5 بناء نموذج التصنيف واختباره

تم تطبيق الخوارزميات الثلاثة على مجموعة بيانات تتضمن رسائل نصية قصيرة (SMS) مصنفة مسبقاً إلى رسائل مزعجة وغير مزعجة. تم استخدام مكتبة scikit–learn في بيئة Python لتدريب النماذج، وتحليل أدائها بناءً على دقة التصنيف ومصفوفة الارتباك لكل نموذج.

من أجل الحكم على فعالية نماذج التصنيف التي تم دراستها، سنأخذ في الاعتبار ثمانية مؤشرات يمكن تحقيقها: المعدل الإيجابي الحقيقيTP ، المعدل الإيجابي الكاذبFP ، المعدل السلبي الحقيقيTN، المعدل السلبي الكاذبFN، درجةF1 ، الدقة(Accuracy) ، الدقة الإيجابية(Precision) ، والاستدعاء(Recall) . تمثل هذه المؤشرات مقاييس الجودة لقياس أداء أي نظام للكشف عن الرسائل المزعجة (Spam) ويمكن تلخيصها ضمن مصفوفة الارتباك التالية:

الجدول ١: مصفوفة الارتباك

		Actual
Predict	TP	FP
	FN	TN

TP (True Positive)تم تصنيف رسالة مزعجة بشكل صحيح كـSpam.

TN (True Negative) تم تصنيف رسالة غير مزعجة بشكل صحيح كـHam.

FP (False Positive) رسالة غير مزعجة تصنف خطأً كرسالة مزعجة.

FN (False Negative) رسالة مزعجة تصنف خطأً كرسالة غير مزعجة.

- الدقة(Precision): تشير إلى النسبة المئوية للرسائل التي كانت وتم تصنيفها فعلياً على أنها رسائل مزعجة بواسطة خوارزمية التصنيف. ويظهر الصواب الدقيق. وتعطى على النحو التالي **[11]**:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \qquad (4)$$

- الاستدعاء(Recall): يشير إلى النسبة المئوية للرسائل التي كانت عبارة عن رسائل مزعجة وتم تصنيفها على أنها رسائل مزعجة. ويظهر الاكتمال. وتعطى على النحو التالي **[11]**:

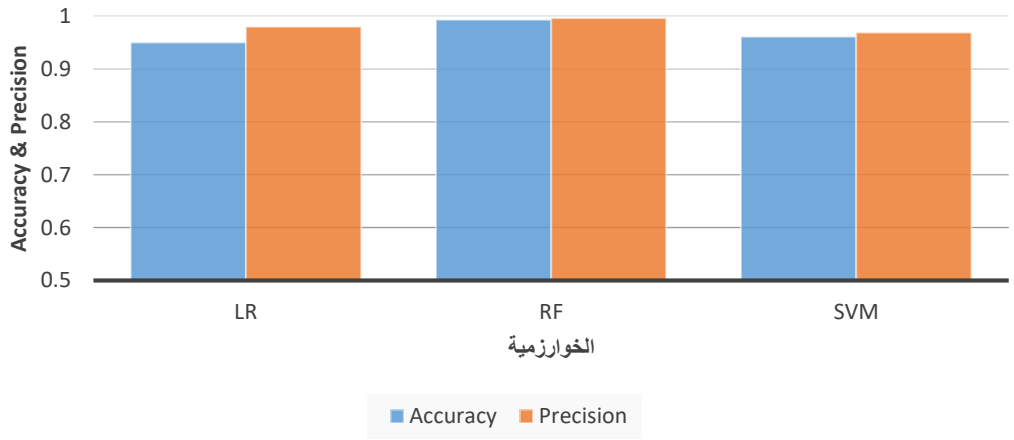
$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \qquad (5)$$

- F1−score:يتم تعريفه على أنه الوسط التوافقي للدقة والاستدعاء. وتعطى على النحو التالي**[11]** :

$$\text{F1}_{\text{score}} = \frac{2(\text{Precision} \times \text{Recall})}{\text{Precision} + \text{Recall}} \qquad (6)$$

٧. النتائج والمناقشة:

تم تنفيذ كود بلغة Python لحساب قيم كل من الدقة (Accuracy) والدقة الإيجابية (Precision) للمصنفات الثلاثة، كما هو موضح في الشكل(11)



الشكل (11): مقارنة بين المصنفات الثلاثة من حيث الدقة العامة والدقة الإيجابية

يوضح الشكل (١١) مقارنة بين المصنفات الثلاثة من حيث الدقة العامة (Accuracy) والدقة الإيجابية(Precision) ، حيث يُظهر مصنف الغابة العشوائية (RF) تفوقاً ملحوظاً بتحقيقه أعلى القيم في كلا

المؤشرين (٠,٩٩٣ و٠,٩٩٦ على التوالي). وبذلك فأن مصنف RF يتمتع بأعلى قدرة على التمييز بين الرسائل المزعجة وغير المزعجة، مما يجعله الخيار الأكثر فعالية.

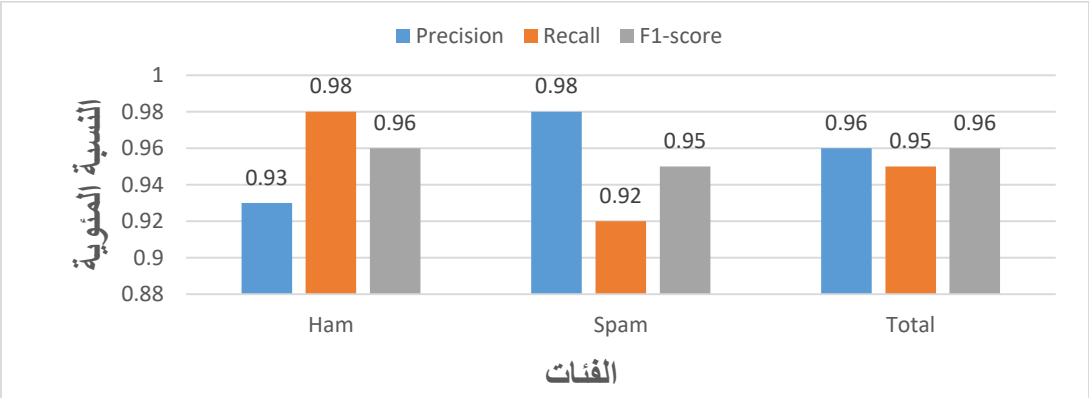
٧-١ الانحدار اللوجستي LR:

تم تدريب نموذج الانحدار اللوجستي على مصفوفة الميزات المستخرجة باستخدام تقنيةTF-IDF ، وتم تقييم النموذج على بيانات الاختبار. توضح مصفوفة الارتباك في الشكل (12) أداء النموذج من حيث عدد التنبؤات الصحيحة والخاطئة لكل فئة.

	ham	spam
ham	909	14
spam	60	822

الشكل (12): مصفوفة الارتباك ل LR

يبين الشكل (12) أن نموذج LRبسيط وسريع، لكنه قد يفشل في التعامل مع تداخل البيانات ويوجد أخطاء محتملة في تصنيف بعض الرسائل المزعجة كرسائل غير مزعجة كما هو واضح حيث صنف ٦٠ رسالة مزعجة على أنها غير مزعجة.



الشكل (13): نتائج الPrecision وRecallوF1-score لمصنف الانحدار اللوجستي

يشير الشكل (13) إلى أداء جيد في اكتشاف كل من الرسائل المزعجة وغير المزعجة، مع توازن واضح بين الدقة والاستدعاء.

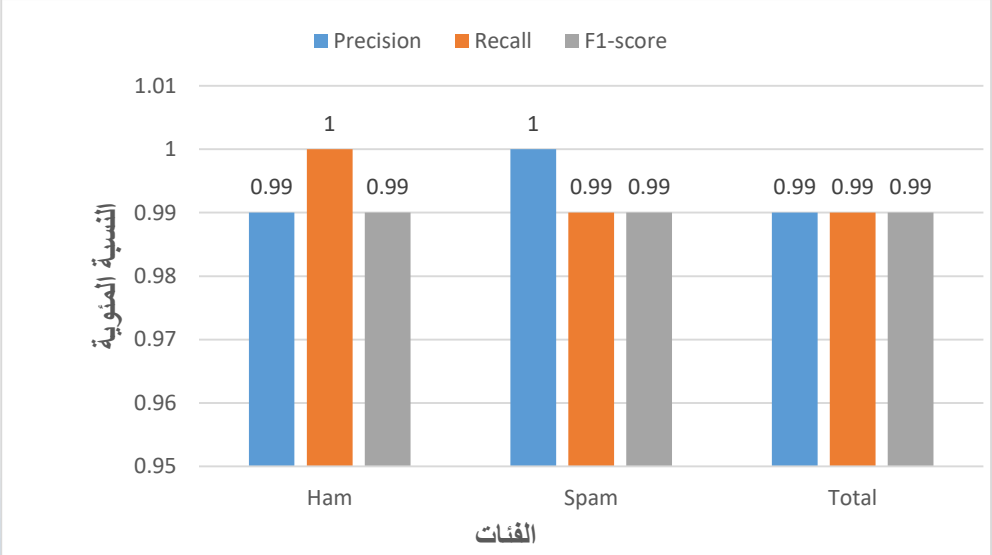
٧-٢ مصنف الغابة العشوائي RF:

تم استخدام RandomForestClassifier لتدريب النموذج على البيانات نفسها. أظهرت النتائج قدرة هذه الخوارزمية على تحقيق توازن جيد بين الدقة وتقليل الأخطاء، كما هو موضح في الشكل.(14)

	ham	spam
ham	920	3
spam	8	874

الشكل (14): مصفوفة الارتباك للغابة العشوائية

يظهر الشكل (14) قيم TP و TN مرتفعة جداً، بينما FP و FN منخفضة للغاية، هذا يدل أن مصنف Random Forest قوي جداً في التمييز بين Spam و Ham.



الشكل (١٥): نتائج الPrecision وRecallوF1-score لمصنف الغابة العشوائية

يظهر الشكل (١٥) أداء ممتاز لمصنف الغابة العشوائية، مع معدلات قريبة جداً من الكمال في تصنيف الرسائل حيث بلغت الدقة الايجابية للصنف Ham (٩٩%) و لصنف Spam(١٠٠%).

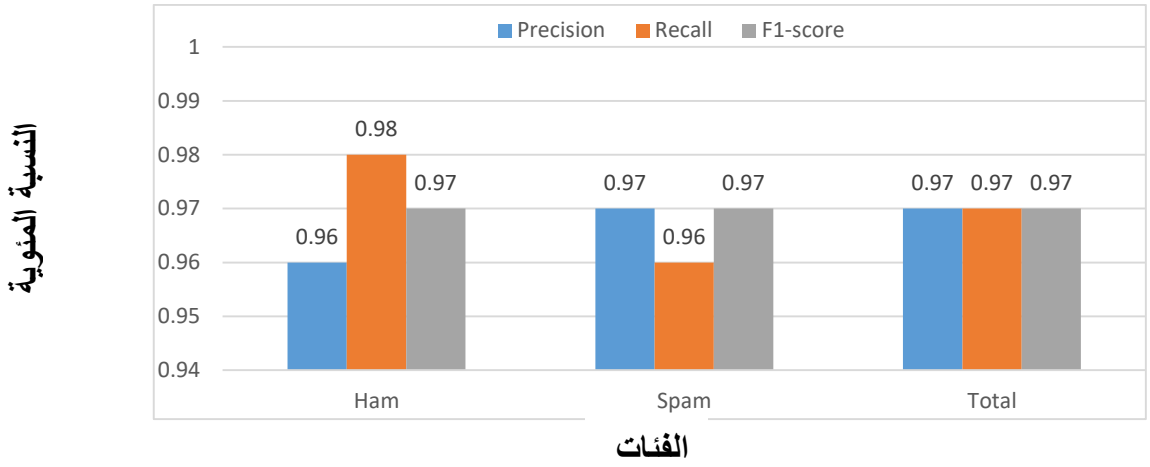
٧-٣ آلة متجه الدعم SVM:

تم تدريب نموذج SVM باستخدام نواة خطية، وأُجري التنبؤ على مجموعة الاختبار. تعكس مصفوفة الارتباك في الشكل (16) أداء النموذج في التفريق بين الرسائل.

	ham	spam
ham	897	26
spam	43	839

الشكل (16): مصفوفة الارتباك لـ SVM

يظهر الشكل (16) أن SVM يتمتع بأداء عالي الدقة أيضاً، لكن أقل قليلاً من الغابة العشوائية حيث أن قيم FPو FNمرتفعة مقارنة ب RF وبالتالي زيادة احتمالية تصنيف الرسائل بشكل خاطئ،



الشكل (١٧): نتائج الPrecision وRecallوF1-score لمصنف SVM

تظهر النتائج في الشكل (١٧) قدرة عالية للمصنف على التمييز بين الفئتين بدقة واستدعاء متقاربين لكل من الفئتين.

من خلال مقارنة أداء النماذج الثلاثة، يتضح أن مصنف الغابة العشوائية يقدم أفضل أداء من حيث الدقة والاستدعاء، يليه مصنف SVM، ثم الانحدار اللوجستي .وتعد هذه النتائج مؤشراً قوياً على أهمية اختيار الخوارزمية المناسبة بناءً على خصائص البيانات وطبيعة المشكلة.

٨. الاستنتاجات والتوصيات:

يهدف هذا البحث إلى مقارنة أداء نماذج التعلم الآلي لاكتشاف الرسائل القصيرة المزعجة وتأمين الاتصالات البريدية والإلكترونية. يُعتبر الكشف الدقيق عن الرسائل المزعجة تحدياً كبيراً، وقد اقترح الباحثون العديد من الطرق للكشف، لكنها غالباً ما تكون غير فعالة. في هذا البحث، تم اقتراح ثلاث نماذج للكشف الدقيق عن الرسائل المزعجة، وأظهرت النتائج التجريبية فعالية هذه النماذج بدقة تصل إلى ٩٩%. بالإضافة إلى ذلك، يتيح البحث تحسين أداء تصنيف الرسائل المزعجة وغير المزعجة باستخدام توصيات متعددة، مثل توسيع نطاق التجارب وتحسين عملية تقسيم البيانات، مما قد يسهم في تطوير وتحسين الأداء والفعالية في مجال تصنيف الرسائل.

المراجع:

1) Ramachandran, A., Feamster, N., & Vempala, S. (2007, October). Filtering spam with behavioral blacklisting. In Proceedings of the 14th ACM conference on computer and communications security (pp. 342-351).

2) Gadde, S., Lakshmanarao, A., & Satyanarayana, S. (2021, March). SMS spam detection using machine learning and deep learning techniques. In 2021 7th International Conference on Advanced Computing and Communication Systems (ICACCS) (Vol. 1, pp. 358-362). IEEE.

3) Dedeturk, B. K., & Akay, B. (2020). Spam filtering using a logistic regression model trained by an artificial bee colony algorithm. *Applied Soft Computing*, *91*, 106229.

4) Othman, N. F., & Din, W. I. S. W. (2019). Youtube spam detection framework using naïve bayes and logistic regression. *Indonesian Journal of Electrical Engineering and Computer Science*, *14*(3), 1508-1517.

5) Taylor, O. E., & Ezekiel, P. S. (2020). A model to detect spam email using support vector classifier and random forest classifier. *Int. J. Comput. Sci. Math. Theory*, *6*(1), 1-11.

6) Amayri, O., & Bouguila, N. (2010). A study of spam filtering using support vector machines. *Artificial Intelligence Review*, *34*, 73-108.

7) Reddy, K. N., & Kakulapati, V. (2021). Classification of Spam Messages using Random Forest Algorithm. *Journal of Xidian University*, *15*(8), 495-505.

8) Sculley, D., & Wachman, G. M. (2007, July). Relaxed online SVMs for spam filtering. In *Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval* (pp. 415-422).

9) Ramos, J. (2003, December). Using tf-idf to determine word relevance in document queries. In *Proceedings of the first instructional conference on machine learning* (Vol. 242, No. 1, pp. 29-48).

10) Behrens, J. T. (1997). Principles and procedures of exploratory data analysis. *Psychological methods*, *2*(2), 131. 7

11) Chawla, N. V. (2010). Data mining for imbalanced datasets: An overview. Data mining and knowledge discovery handbook, 875-886. 9