محاكاة تطبيق إسنادية التأليف على نصوص الأحاديث النبوية الشريفة

د. محمد سعيد دسوقي[1]، م. حسن المعريني2

[1]الجامعة العربية الدولية، اتستراد درعا، غباغب، سورية

[2]المعهد العالي للعلوم التطبيقية والتكنولوجيا، ص. ب. 31983 دمشق سورية

[1]s-desouki@aiu.edu.sy [2]hasan.maarini@hiast.edu.sy

الخلاصة: إسناديّة التّأليف هي مسألة تحديد الكاتب الأصلي لنصّ مُختَلَف عليه، وتعتمد على استخراج خصائص وسمات من النّصوص، واستخدامها لتمييز أسلوب الكتابة بين كاتب وآخر. يمكن أن تكون الخصائص أو السمات المستخرجة حرفيّة أو معجميّة أو بنيويّة أو نحويّة أو دلاليّة. نهدف في هذا البحث إلى محاكاة تطبيق إسنادية التأليف على نصوص الأحاديث النبوية الشريفة، حيث يمكن بالاعتماد على الأحاديث النبوية الشريفة الصحيحة تعلُّم أسلوب الرسول محمد ﷺ ومن ثم تمييز الأحاديث النبوية الشريفة الصحيحة من الأحاديث الموضوعة أو الضعيفة أو المشكوك في صحتها.

الكلمات الجوهرية. إسنادية التأليف، الحديث النبوي الشريف، الحديث الصحيح، المعالجة الآلية للغة العربية، النصوص القصيرة.

1. المقدمة

إسنادية التأليف هي مسألة تحديد الكاتب لنص مجهول أو مشكوك فيه عن طريق خصائص وسمات تُؤخذ من داخل النّص نفسه، وبدون العودة إلى أي مراجع خارج النّص مثل تاريخ النشر أو مكان النشر أو مسودّات النّص، ويكون ذلك عن طريق تعلم أسلوب كتابة كاتب معين بهدف التحقق من نسبة نصوص معينة إليه، أو تعلم أسلوب عدد من الكُتّاب المرشّحين بهدف التمكن من معرفة الكاتب الفعلي لنصّ جديد لأحد هؤلاء الكُتّاب [1].

تتنوّع الخصائص والسمات الّتي يمكن استخراجها من النّص، فيمكن الاعتماد على الجمل المكوّنة للنص وطولها، أو على الكلمات ونمط تكرارها، أو على الحروف وإحصائيّات تتعلق بها، أو على علامات التّرقيم [2].

تتناسب سهولة المسألة طرداً مع حجم النّصوص الّتي يتم التدريب عليها، فكلّما كانت النّصوص كبيرة وكثيرة كان احتمال كشف أسلوب الكاتب أكبر. من جهة أخرى في المسائل التي تدرس أسلوب عدد من الكتاب، يؤثّر عدد الكُتّاب المرشّحين على دقّة الإسناد وتتناسب دقّة النّتائج عكساً مع عدد المرشّحين.

نهدف في هذا البحث إلى محاكاة تطبيق إسنادية التأليف على نصوص الأحاديث النبوية الشريفة، حيث يمكن بالاعتماد على الأحاديث النبوية الشريفة الصحيحة معرفة أسلوب الرسول محمد ﷺ ومن ثم تمييز الأحاديث النبوية الشريفة الصحيحة من الأحاديث الموضوعة أو الضعيفة أو المشكوك في صحتها.

2. الحديث النبوي الشريف

الحديث النبوي الشريف هو كل ما ورد عن النبي محمد ﷺ من قول أو فعل أو إقرار، لكننا نقتصر في بحثنا هذا على نصوص أقوال النبي محمد ﷺ.كثرت الدّراسات التي تهتم بتفريق الأحاديث الصّحيحة عن غيرها عبر التّاريخ الإسلامي، لكن معظم هذه الدّراسات تعتمد على موثوقيّة سند الحديث وهي سّلسلة الأشخاص الّذين نقلوا المتن حتى الوصول إلى الرّسول محمد ﷺ. توجد كذلك دراسات تنظر إلى متن الحديث (نصه)، وتناقش صحّته بناءً على بلاغة اللّفظ، أو وجود مخالفة للقرآن الكريم، أو مخالفة لأحكام الدين، أو مخالفة للعقل، أو غيرها–.

نطرح ضمن إطار هذا العمل فكرة تقييم صحّة الحديث اعتماداً على خصائص وقياسات تؤخذ من متن الحديث، وفقط من كلمات المتن المنسوبة إلى الرسول محمد ﷺ، بغضّ النّظر عن السند، أو معنى المتن، أو الأجزاء الأخرى ضمن المتن التي تحتوي كلمات لا تُنسب إلى الرّسول محمد ﷺ.

اطّلعنا على عدّة دراسات إسلاميّة تناقش صحّة الأحاديث النّبويّة الشريفة، وتبين عدد الأحاديث الصّحيحة منها. من أهم الدّراسات التي توضح ذلك دراسة للشّيخ صالح الشّامي في كتابه (معالم السّنة النّبويّة) [3]. تجمع دراسة الشّيخ الشّامي الأحاديث الموجودة في 14 كتاب من كتب الحديث: صحيح البخاري، صحيح

مسلم، سنن أبي داود، وغيرها. حيث وصل عدد الأحاديث إلى 114194 حديثاً. عالج الشّيخ الشّامي هذه الأحاديث وحذف منها المكرّر والضّعيف والآثار (أقوال وأفعال الصّحابة والتّابعين)، فأصبح عدد الأحاديث بحسب الشّيخ الشّامي 3878 حديثاً فقط. وكان عدد أحاديث الصّحيحين (صحيح البخاري، وصحيح مسلم) ضمن هذه الأحاديث النّاتجة هو 2131 حديثاً، أي أكثر من نصف الأحاديث الّتي جمعها الشّيخ الشّامي. لم نعثر على دراسات تُقدّر أطوال الأحاديث بعدد الكلمات أو بعدد الحروف، فلجأنا إلى دراسة أطوال مجموعة من الأحاديث الصحيحة ووجدنا أن معظم الأحاديث يتراوح طولها بين 40 إلى 400 حرف.

3. إسنادية التأليف

مسألة إسنادية التأليف هي مسألة قديمة قدم الكتابة والكلمات، جرى التعامل معها في البداية من منظور إحصائي، حيث نشرت أول دراسة إحصائيّة في إسنادية التأليف في اللّغة الإنكليزيّة في عام 1887، وطرحت طول الكلمة كَسِمة مميّزة للكاتب. تتالت بعد ذلك الدّراسات في هذا المجال، وانتقل الباحثون إلى طرح أفكار لسمات من أنماط متعددة، مثل أسلوب استعمال جمل طويلة أو كلمات طويلة في النّص. فاعتمدوا على مقاييس، مثل متوسّط طول الجملة، ومتوسّط طول الكلمة، وعدد الحروف الصوتيّة في الكلمة، وتوزّع أقسام الكلام بين حروف وأفعال وأسماء، وصولاً إلى الغِنى اللّغوي وعدد وتنوع المفردات الّتي يستخدمها الكاتب، وغيرها [2]. على التّوازي مع المسار الإحصائي السّابق، انتهج بعض الباحثين أسالیباً مختلفةً نوعاً ما، فحاولوا استخراج بصمة الكاتب عن طريق المفردات التي يستعملها، فبعض الكلمات مثل "Big" و"Large" في اللغة الإنكليزية، تتشابه بالمعنى ولا توجد قيود قواعديّة في اللّغة تُحتّم استخدام إحداها بدلاً عن الأخرى. تطوّرت الفكرة السّابقة لتكوّن مساراً عاماً يهتم بدراسة كلمات محدّدة من النّص، وهي الكلمات التي لا تضيف معنىً حقيقيّاً، وسميت الكلمات الوظيفيّة، مثل حروف الجر والعطف والأسماء الموصولة والأفعال المساعدة، وأدوات التعريف، وغيرها [2]. وعلى الرّغم من أنّ عدد هذه الكلمات الوظيفيّة قليل نسبةً إلى عدد كلمات اللّغة إلّا أنّها تشكّل نسبةً كبيرةً من الكلمات المستعملة فعليّاً، وبما أنّها مستقلّة عن محتوى النّص المطروح فإن وجودها بتواتر معيّن يمكن أن يدل على الكاتب [1] .

4. إسنادية التأليف في اللغة العربية

موضوع إسناديّة التّأليف في اللّغة العربيّة هو موضوع جديد نسبيّاً، بدأ في عام 2012 وتضمن سلسلة من الأبحاث والدّراسات:

أ. إسنادية التأليف لعدد من الكتـّاب

تعد رسالة الدكتوراه للباحث كريم شاكر [1] أوّل دراسة في مجال إسناديّة التّأليف باللّغة العربيّة. عمل الدّكتور كريم على إنشاء مجموعة بيانات خاصّة به، حيث قام بتحميل 14 كِتَاباً مختلفاً تعود لـ 6 كُتّاب مختلفين. يحتوي أصغر الكتب على 14 ألف كلمة، ويحتوي أكبرها على 38 ألف كلمة. تعتمد أطروحة الدكتور شاكر على مفهوم الكلمات الوظيفيّة لتحديد أسلوب الكاتب. حيث تتمّ تجزئة كل كتاب إلى أجزاء، وبعدها يجري حساب تكرار كل كلمة وظيفيّة في كل جزء. قام الدّكتور كريم بتغيير أحجام الأجزاء وإعادة التّجربة من أجل أجزاء بحجم يساوي 1000 كلمة، و2000 كلمة، و3000 كلمة. يجري بعد ذلك اختيار عدد من الكلمات الوظيفيّة ليتم اعتمادها والتدرب عليها باستعمال نموذج خطي معمم Generalized Linear Model. تبدأ دقّة التّصنيف على مستوى الأجزاء من 80%، أمّا على مستوى الكتب كاملةً فكانت نسبة النّجاح 100%.

ب. إسناد نص شعري مجهول مؤلفه إلى شاعره الحقيقي

تندرج تحت هذا العنوان عدد من الأوراق البحثية، تَشَارك في كتابتها ثلاثة باحثين: أحمد الفلاحي، ومحمد الرّمضاني، ومصطفى بلفقيه، بدأت تجربتهم عام 2014 وما زالت مستمرّة حتّى الآن. يهدف العمل إلى الإسناد التّلقائي لنص شعري عربي مجهول مؤلفه إلى شاعره الحقيقي، وهي عمليّة مهمّة جدّاً بالنّسبة لأولئك الّذين يدرسون الأسلوبيّة في الشّعر العربي، أو لمساعدة الشّعراء في إثبات حقّهم الإبداعي ومعرفة النّصوص المنتحَلة من غيرها. وقد تم نشر آخر ورقة بحث في هذا الموضوع في تموز 2019 [4]، ويعد أنضج عمل يتناول إسناديّة التّأليف في اللّغة العربيّة. جرى تدريب البرنامج على مجموعة من الدواوين الشّعريّة تتبع لـ 114 شاعر من مختلف العصور، ثم جرى اختبار البرنامج على 6784 قصيدة. يعمل البرنامج وفق خوارزميّات Naïve Bayes NB, Support Vector Machine SVM, Markov Chain M.C، ويعتمد على معظم أنواع الخصائص: خصائص حرفيّة، وخصائص معجميّة، وخصائص بنيويّة، وخصائص نحويّة، وخصائص دلاليّة، وخصائص شعريّة، وخصائص على مستوى الكلمات المميّزة. كانت أفضل دقّة تم الحصول عليها 99.12%.

ج. إسنادية التأليف لنصوص عربيّة تاريخيّة قصيرة

في محاولة لتطبيق إسنادية التأليف على النصوص القصيرة، قام الباحثان وعمور وسيود [5] باستعمال أجزاء من عشرة كتب تاريخيّة قديمة لعشرة كُتّاب، متوسّط طول كل جزء 550 كلمة. واستعملوا مجموعة من خوارزميّات التصنيف وتقنيات الدمج. واعتمدوا على خصائص على مستوى الكلمات مثل Word n-grams، مع تغيير قيم n خلال التّجارب، وعلى خاصة الكلمات النّادرة، وخصائص أخرى على مستوى الحروف مثل

Char n-grams. حصلت التّجربة على نسبة نجاح 100%، وقد تكون الأقرب لدراستنا من حيث عدد الكتب المستخدمة.

د. إسناديّة التّأليف لتّغريدات عربيّة

تتشابه هذه الدّراسة [6] مع دراستنا من حيث حجم النّصوص المختارة، إذ إنّ حجم التّغريدات كان محدوداً ضمن التّجارب بـ 140 حرفاً (بدون احتساب الروابط وأسماء المستخدمين). اعتمدت الدراسة على مجموعة بيانات مكوّنة من أكثر من 37 ألف تغريدة تعود إلى 12 مغرّداً مختلفاً تم اختيارهم من قائمة أكثر المغرّدين العرب، لكلّ مغرّد حوالي 3100 تغريدة. وقد تمّ استخدام خوارزميّات NB, Decision Tree, SVM، والاعتماد على خاصة سلسلة الكلمات Bag-of-Words BOW التي يمثَّل فيها النص بالكلمات المختلفة الموجودة فيه مع تواتر تكرارها بغض النظر عن ترتيبها. وقد كانت أفضل نتيجة حققتها هذه الدراسة 68.67% وذلك باستخدام خوارزمية SVM.

5. محاكاة إسنادية التأليف على نصوص الأحاديث النبوية الشريفة

تعد مسألة تمييز الأحاديث النبوية الشريفة الصحيحة من أهم وأشهر الأمثلة الّتي يمكن تطبيق إسناديّة التّأليف في اللّغة العربيّة عليها، لكن التحدي الأكبر في هذه المسألة هو التعامل مع نصوص الأحاديث النبوية الشريفة التي يمكن أن تكون قصيرة جداً مما يزيد من صعوبة تحديد أسلوب الرسول محمد ﷺ من خلالها، أو صعوبة التأكد من نسبتها إليه ﷺ. يمكن تجاوز الصعوبة الأولى في تحديد أسلوب الرسول محمد ﷺ بالاعتماد على عدد كبير من الأحاديث النبوية الشريفة، في حين يمكن الوصول إلى مستوى دقة معقول يتعلق بطول الحديث النبوي الشريف عند محاولة التصدي لمسألة نسب حديث نبوي شريف ما له ﷺ.

في عمل مشابه أجرى حليم سيود دراسة تُقارن بين القرآن الكريم والأحاديث النّبويّة من حيث الأسلوب [7]، تتناول هذه الدّراسة أسلوب القرآن الكريم وأحاديث صحيح البخاري، في محاولة لإثبات أن القرآن الكريم وصحيح البخاري ليسا للكاتب نفسه. قام الباحث في هذه التّجربة بتقسيم القرآن الكريم إلى 29 قسماً، وتقسيم صحيح البخاري إلى 8 أقسام. بحيث يحتوي كل قسم على 2900 كلمة تقريباً. استعملت التّجربة خوارزميّة SVM مع خصائص Char 4-grams، وحققت نسبة نجاح 100%.

أ. الحل المقترح

يعتمد الحل الذي نقترحه على مجموعة من الخصائص التي يمكن تحديدها في النّصوص القصيرة، وعلى استعمال خوارزميات تتناسب مع هذه الخصائص. استخدمنا خوارزميّة Support Vector Machine (SVM) مع خاصّيّة على مستوى سلاسل الكلمات Bag-of-Words (BOW)، اعتماداً على نجاح عدّة تجارب سابقة لهما على اللّغتين العربيّة والإنكليزيّة، وكررنا التّجربة عدّة مرّات لدراسة أثر تغيير حجم مجموعة النصوص على دقّة النّتائج. تتأثّر خاصّيّة BOW بالمحتوى؛ لكن ما يُخفّف من وقع هذه السّلبية احتفاظنا بالكلمات الوظيفيّة دون حذفها، وبذلك نعطي دوراً أكبر للأسلوب، حيث ترتبط الكلمات الأكثر تكراراً، والّتي تشكل الكلمات الوظيفيّة جزءاً منها، ارتباطاً وثيقاً بالأسلوب.

ب. مجموعة النصوص المعتمدة

اخترنا مجموعة نصوص مؤلفة من عشرة كتب مختلفة لعشرة مؤلفين مختلفين بحيث تكون الكتب المختارة من طبيعة واحدة ومتقاربة في المحتوى، بحيث لا تحتوي على عدد كبير من الآيات القرآنيّة أو الأحاديث النّبويّة أو الشّعر المقتبس، لأن وجود مثل هذه النّصوص سيؤثّر على أسلوب الكاتب الّذي سيستنتجه النّظام. قمنا بتحميل هذه الكتب بصيغة نصّيّة من موقع المكتبة الشاملة[1]، ونستعرض ضمن الجدول الآتي بعض المعلومات عنها:

الجدول 1: الكتب المستخدمة لإنشاء مجموعات النصوص

اسم الكتاب	اسم الكاتب	التّاريخ التقريبي	عدد الجمل التقريبي
الضوء اللامع	شمس الدين السخاوي	900هـ	32000
بغية الطلب في تاريخ حلب	كمال الدين ابن العديم	650هـ	15000
تهذيب اللّغة	محمد بن أحمد بن الأزهري الهروي	350هـ	81000
أساس البلاغة	محمود بن عمر الزمخشري	500هـ	29000
العبر في خبر من غبر	محمد بن أحمد بن عثمان الذهبي	700هـ	14000
حاشية البجيرمي على المنهاج	سليمان البجيرمي	1150هـ	29000
الكامل في التّاريخ	ابن الأثير الجزري	600هـ	34000
الكتاب	سيبويه	170هـ	14000
المنطق	ابن سينا	400هـ	18000
موسوعة اليهود واليهودية	المسيري	1420هـ	69000

[1] https://shamela.ws/index.php/page/download-shamela

لن نقوم بالتّجارب العمليّة على الكتب نفسها لمحاولة تحديد الكاتب الأصلي لكتاب كامل، لأن استخراج أنماط لنصوص بهذا الحجم الكبير أمر سهل، بل ستكون دراستنا على مستوى الجمل حيث نعدّ كل جملة هي نصًا مستقلًا ونستخرج العدد نفسه من الجمل من كل كتاب بهدف الحصول على مجموعة متوازنة. حيث نهدف في هذه الدّراسة إلى معرفة مدى تأثير عدد النّصوص أي الجمل المأخوذة من كل كتاب على دقّة النّتائج، وإيجاد الحدود التقريبية الّتي يحدث عندها هبوط في الدّقّة أو استقرار فيها.

بهدف محاكاة إسنادية التأليف للأحاديث النبوية الشريفة بدقة، حددنا في إحدى مجموعات النصوص (المجموعة 12) عدد الجمل من كل كتاب بـ 3878 جملة (وهو عدد الأحاديث المرشحة لتكون صحيحة والنّاتجة عن دراسة الشّيخ صالح الشّامي)، وبحيث يتراوح طول الجملة بين 40 و400 حرف (وهو تقديرنا لطول الحديث النّبوي الشريف). ثم قمنا في تجربة أخرى (مجموعة النصوص 11) باختيار عدد من الجمل قدره 2131 (وهو عدد الأحاديث الموجودة في الصحيحين ضمن الأحاديث الناتجة دراسة الشّيخ الشّامي).

أ. خطوات العمل

- تقسيم مجموعة النصوص عشوائياً إلى مجموعة تدريب تحتوي 80% منها ومجموعة اختبار تحتوي 20% منها.
- إنشاء حقيبة الكلمات BOW من مجموعة التدريب عن طريق استخراج الكلمات المختلفة، وتحديد عدد مرّات تكرارها، بالاعتماد على مفهوم Term Frequency – Inverse Document Frequency (TF–IDF).
- تحويل مجموعة التدريب إلى مصفوفة أسطرها جمل مجموعة التدريب، وأعمدتها الكلمات المختلفة من مجموعة التدريب، تمثّل كل قيمة فيها عدد مرّات تكرار الكلمة في الجملة.
- تحويل مجموعة الاختبار إلى مصفوفة أسطرها جمل مجموعة الاختبار وأعمدتها الكلمات المختلفة من مجموعة التدريب.
- تدريب النّموذج على مجموعة التدريب.
- اختبار النّموذج على مجموعة الاختبار.
- حساب مقياس الدّقّة الذي يعبّر عن مدى نجاح التّجربة، وهو نسبة حالات الاختبار المصنّفة بشكل صحيح إلى العدد الكلّي لحالات الاختبار.

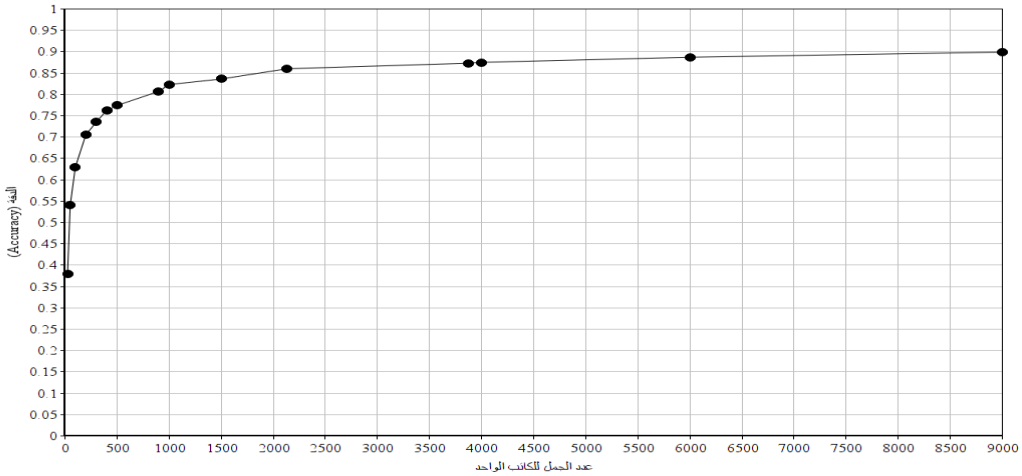
ب. النتائج

نستعرض في الجدول التالي النتائج التي حصلنا عليها على المجموعات المختلفة من النصوص مع بيان عدد جمل الكاتب الواحد في كل مجموعة والدقة التي حصلنا عليها بنتيجة التجربة:

الجدول 2: نتائج تطبيق خوارزمية SVM على 15 مجموعة بيانات مختلفة

مجموعة النصوص	عدد الجمل للكاتب الواحد	الدقة
1	25	0.38
2	50	0.54
3	100	0.63
4	200	0.705
5	300	0.735
6	400	0.7625
7	500	0.774
8	900	0.807
9	1000	0.823
10	1500	0.836
11	2131	0.860
12	3878	0.873
13	4000	0.875
14	6000	0.887
15	9000	0.899

نرسم في الشكل التّالي العلاقة بين عدد الجمل للكاتب الواحد ودقّة النّتائج:



الشكل 1: العلاقة بين عدد الجمل للكاتب الواحد ودقة النتائج

ج. تفسير النّتائج

كان الهدف من المحاكاة الإجابة عن التساؤلات التالية:

- هل يمكننا معرفة أسلوب كاتب بالاعتماد على نصوص تشابه الأحاديث النّبويّة الصّحيحة من حيث العدد والطول؟
- هل بمقدورنا مستقبلاً الاعتماد على الأحاديث الصّحيحة فقط مع التّقنيّات المطروحة في البحث للحكم على صحّة مئات أو آلاف الأحاديث المشكوك بها؟

تثبت النتائج التي حصلنا عليها أن التّقنيّات المطروحة قادرة على استخراج بصمة الكاتب من نصوص مشابهة للأحاديث النّبويّة من حيث الطول والعدد، ويمكن الاعتماد عليها في الحكم على نصوص لم يتدرب النّظام عليها تحاكي الأحاديث المشكوك بصّحتها.

قامت جميع التّجارب الخمسة عشر على جمل بطول بين 40 و400 حرف، وهو الطول الّذي نحاكي فيه طول الأحاديث النّبويّة. وبالتّالي يمكن إسقاط أي تجربة على نصوص الأحاديث النبوية واعتبار أنّها تحاكي الأحاديث النّبويّة بعدد أحاديث يُطابق عدد جمل الكاتب الواحد ضمن مجموعة النصوص المقابلة.

مثلاً إذا طبّقنا قيودًا وشروطًا إضافية على الأحاديث النّبويّة للوصول إلى أحاديث صحيحة بضمانات أكبر، حيث تعتمد بعض الدراسات على 1000 حديث صحيح فقط، فهل يمكننا من خلال 1000 حديث صحيح

الحكم على باقي الأحاديث المشكوك بها؟ نحصل على الإجابة من الدقة الناتجة عن تجربتنا على المجموعة 9. هل يمكن الاعتماد على 200 حديث صحيح فقط لمعرفة أسلوب كتابة الرّسول صلى الله عليه وسلم؟ نتائج التجربة على المجموعة 4 تعطينا فكرة واضحة عن ذلك، وهكذا يمكن القياس على جميع مجموعات النصوص.

6. الخاتمة

قمنا في هذا البحث بمحاكاة تطبيق إسنادية التأليف على نصوص الأحاديث النبوية الشريفة عن طريق تطبيق إسنادية التأليف لنصوص قصيرة مشابهة للأحاديث النبوية من حيث الطول والعدد. وحصلنا على دقة تبلغ 0.860 باستعمال خوارزمية SVM مع خاصة BOW على عدد من النصوص يوافق عدد تقديري أصغري للأحاديث الصحيحة، وتزداد بازدياد عدد هذه الأحاديث. نأمل في مراحل لاحقة البناء على عملنا هذا للوصول إلى تطبيق مباشر لإسنادية التأليف على الأحاديث النّبويّة الشريفة يمكّننا من التعرف على أسلوب الرسول محمد ﷺ وتعلُّمه، ومن ثم الحكم على صحة الأحاديث النبوية الشريفة وتمييز الأحاديث النبوية الشريفة الصحيحة من الأحاديث الموضوعة أو الضعيفة أو المشكوك في صحتها، بحسب موافقتها لأسلوبه صلى الله عليه وسلم.

7. المراجع

[1]

Kareem Shaker, Investigating Features and Techniques for Arabic Authorship Attribution, Edinburgh, Heriot-Watt University, 2012.

[2]

Patrick Juola, Authorship Attribution, 2008.

[3]

صالح أحمد الشامي، معالم السنة النبوية، دار الروائع، دمشق 2013.

[4]

Ahmed Al-Falahi, Mohamed Ramdani and Mostafa Bellafkih, Arabic Poetry Authorship Attribution using Machine Learning Techniques, 2019.

[5]

Siham Ouamour and H. Sayoud. Score Fusion Based Authorship Attribution of Ancient Arabic Texts, 2017.

[6]

Abdullateef Rabab'ah, Mahmoud Al-Ayyoub, Yaser Jararweh and Monther Aldwairi, Authorship attribution of Arabic tweets, IEEE, 2016.

[7]

Halim Sayoud, Authorship discrimination on Quran and Hadith using discriminative leave-one-out classification, 2017.

8. جدول الألفاظ

English	عربي
Authorship Attribution	إسنادية التأليف
Tweets	تغريدات
Bag-of-Words	سلاسل الكلمات
Support Vector Machine	آلة الشعاع الداعم
Term Frequency	تواتر الكلمات
Inverse Document Frequency	مقلوب تواتر الوثيقة

9. السيرة الذاتية للباحثين

د. محمد سعيد دسوقي

أستاذ وباحث يعمل في مجال نظم المعلومات وقواعد المعطيات والمعالجة الآلية للغة العربية، حاصل على دكتوراه في المعلوماتية من جامعة جوزيف فوريه في غرونوبل، فرنسا، 1997.

م. حسن المعريني

مهندس معلوماتية يعمل في إدارة عدد من المشاريع البرمجية. حاصل على ماجستير في نظم المعطيات الكبيرة من المعهد العالي للعلوم التطبيقية والتكنولوجيا في دمشق، سوريا، 2019.

10. الخلاصة باللغة الإنجليزية

Simulating the Application of Authorship Attribution for Hadith Sharif Texts

Mohammad Said Desouki [1], Hasan Al-Maarini[2]

[1]Arab International University, Daraa Highway Ghabagheb, Syria
[2] Higher Institute for Applied Sciences and Technology, P.B. 31983 Damascus, Syria
[1] s-desouki@aiu.edu.sy , [2] hasan.maarini@hiast.edu.sy

Abstract
Authorship attribution is the problem of identifying the true author of a disputed text, and it depends on extracting properties and features from the studied texts and using them to distinguish the writing style between authors. These properties are divided into several main types: literal, lexical, structural, syntactic, and semantic. The aim of this research is to simulate the application of author attribution on the texts of Hadiths Sharif, where we can, relying on the correct Hadiths, learn the Style of the Prophet Muhammad peace be upon him, and thus distinguish the correct Hadiths Sharif, and judge about the correctness of the doubtful Hadiths.

Keywords. Author Attribution, Hadith Sharif, Hadith Sahih, Arabic Language Processing, Short Texts.