



القياس الأسلوبى في الحديث النبوي الشريف

Stylometry in the Prophet's statements

د. مولاي عبد الله قاسمي*¹، جمال والزين²

Dr. Moulay Abdellah KASSIMI¹, Jamal OUAZZINE²

المدرسة الوطنية للعلوم التطبيقية بأكادير (المغرب)، m.kassimi@uiz.ac.ma

الكلية المتعددة التخصصات تازة (المغرب)، jamal.ouazine@usmba.ac.ma

تاريخ القبول: 2023/06/18

تاريخ الإرسال: 2023/06/08

ملخص

يعد الحديث النبوي الشريف المصدر الثاني من مصادر التشريع الإسلامي والتطبيق العملي لما جاء في القرآن الكريم. في ضوء هذه الأهمية، وفي عصر الرقمنة، يتم تداول العديد من الأحاديث النبوية الشريفة على الشبكة العنكبوتية تغيب فيها المرجعية والتوثيق في ظل عدم وجود تطبيقات متخصصة في النقد الآلي للحديث النبوي، لذلك يظل نقد الحديث وتصنيفه آليا من أهم المشكلات التي تواجه الباحثين والمهتمين. يهدف هذا البحث إلى بناء نموذج ذكي لنقد وتصنيف الحديث النبوي الشريف باستخدام التطبيقات المرتبطة بتقنيات معالجة اللغات الطبيعية باعتباره أحد أهم مجالات الذكاء الاصطناعي. اعتمد البحث على المنهج التجريبي، وذلك بالقيام باختبارات مختلفة على قاعدة بيانات الأحاديث النبوية، وأيضاً تم استخدام المنهج التحليلي الذي يركز على قراءة تحليلية لمخرجات النماذج والخروج منها بنتائج تؤكد أهمية توظيف الذكاء الاصطناعي في علم الحديث. أكد تدريب النماذج فاعلية التعلم العميق وإسناد التأليف في مجال النقد الآلي للحديث النبوي. فقد حقق النموذج المدرب مسبقاً على البيانات العربية الأداء الأفضل بنسبة 89%.

الكلمات المفتاحية: القياس الأسلوبي، إسناد التأليف، لفظ الحديث، الذكاء الاصطناعي، معالجة اللغات الطبيعية.

Abstract

The Prophet's hadith is the second source of Islamic legislation and the practical application of what was stated in the Holy Qur'an. In view of this importance, and in the age of digitization, many of the noble Prophet's hadiths are circulated on the Internet without verifying its authenticity in the absence of specialized applications in the Prophet's hadith that rely on automated analysis and elicitation. Therefore, the criticism and classification of hadith remain one of the most important problems facing researchers. This research aims to build an intelligent model for criticizing and classifying the hadith of the Prophet automatically using applications related to the techniques of natural language processing as one of the most important fields of artificial intelligence. The paper relied on the experimental method, by conducting various experiments on the database of the Prophet's hadiths, and also the descriptive analytical approach was used, which focuses on an analytical reading of the outputs of the models and producing results confirming the importance of employing artificial intelligence in the science of hadith. The training of the models confirmed the effectiveness of the hybrid approach (deep learning and authorship attribution) in the field of automated criticism of the Prophet's hadith. The pre-trained model on Arabic data achieved the best performance by 89%.

Keywords: Stylometry, Author Identification, Writing Style, Artificial intelligence, Natural Language Processing

1. مقدمة

يعنى علم مصطلح الحديث، باعتباره واحداً من أهم فروع علوم الحديث، بدراسة الحديث النبوي من حيث السند والمتن. فقد اعتنى علماء الحديث بنقد الحديث بشقيه، فوضعوا له قواعد شرعية قائمة على منهج علمي متين لرد الروايات المخالفة للقرآن والسنة الصحيحة حتى أصبح من أوثق ما يعتمد عليه. ومع الطفرة التي شهدتها استخدام التقنيات الحديثة، استعان العلماء بوسائل التواصل الاجتماعي والمتديات لتعزيز التواصل مع أكبر عدد ممكن من فئات المجتمع. غير أن هذه الوسائل أتاحت لغير المتخصصين وأصحاب الأهواء نشر الأحاديث النبوية الشريفة بدون إسناد، بل إن مما يروج عبر هذه الوسائل أحاديث موضوعة ومكذوبة عن النبي صلى الله عليه وسلم. ففي هذا العصر الإلكتروني، ومع انتشار خدمات التواصل الاجتماعي التي اقتحمت حياة الإنسان، أصبح تبادل الرأي والتعبير عنه أمراً سهلاً وذلك من خلال المدونات أو التغريدات؛ كما أصبحت المؤسسات والأفراد تعتمد على شبكات التواصل الاجتماعي من أجل ترويج الأفكار والإشاعات، مما أدى إلى إنتاج بيانات نصية ضخمة تحتوي على الخطأ والصواب. ومن الملفت أن المنشورات اليومية المرتبطة بالدين الإسلامي الحنيف خاصة الأحاديث النبوية والتفاسير والأدعية، تغيب فيها المرجعية والتوثيق.

1.1. أهمية الموضوع

تنبثق أهمية هذا البحث بدايةً من أهمية دراسة أصالة الحديث النبوي الشريف في عصر الرقمنة وفي غياب تطبيقات مبنية على التحليل والاستنباط الآلي والمتخصصة في الحديث النبوي، فقد أدى افتقارنا للأدوات الحديثة، التي تعتمد علوم البيانات والذكاء الاصطناعي، إلى ازدياد شراسة الهجمة المعلوماتية على السنة النبوية، لذا أصبحت الحاجة ماسة إلى دعم جهود علماء الحديث المنافحين عن السنة بتطبيقات سهلة تمكنهم من الرد المتزامن آلياً على صفحات الإنترنت لصده هذه الهجمة الشرسة. هذه التطبيقات تمكن علماء الحديث من سهولة التحقق من درجة الحديث، والرد على الهجمة المعلوماتية وحملات التشويه التي تريد النيل من الحديث النبوي الشريف.

1.2. مشكلة البحث

تبقى كيفية تحليل الأحاديث النبوية وتصنيفها آليا أهم إشكال يواجه الباحثين والمهتمين؛ كما يقارب البحث الإشكالية التالية:

هل يمكن توظيف الذكاء الاصطناعي في نقد الأحاديث والحكم عليها بالصحة أو الضعف؟

وتتفرع عن هذه الإشكالية مجموعة من التساؤلات:

- هل يمكن للذكاء الاصطناعي أن يساعد في تقليد عمل المحدثين وتصنيف الحديث آليا حسب حكمه: (المقبول والمردود)؟
- هل يمكن بناء قاعدة بيانات ألفاظ الحديث كما جاءت في أشهر الكتب الصحاح؟
- هل يمكن تطوير البصمة الأسلوبية لسيدنا محمد صلى الله عليه وسلم، من خلال دراسة وتحليل لغة الخطاب الموجود في المتون الخاصة بالحديث الشريف؟

1.3. سبب البحث:

جاءت فكرة هذا البحث إيمانا منا أن بلاغة النص النبوي ليست كبلادة النصوص الأخرى لأنه بيان رباني لا يعبر عن ذات فردية، ومن ثم الاستفادة من تقنية القياس الأسلوبي في تصنيف المتون باستخدام التعلم العميق من خلال التعلم على كمية كبيرة من البيانات النصية ودون الرجوع لقواعد اللغة.

1.4. أهداف البحث:

نهدف بالخصوص في هذا البحث إلى تصنيف "لفظ الحديث" آليا باستخدام خوارزميات التعلم العميق والتمهيد لمشروع كبير يعمل على جمع الحديث وتصنيفه. لقد كان بلوغ هذا الهدف فيما سبق، يحتاج منا، باعتماد البرمجة التقليدية، إلى سنوات عديدة، غير أن الطفرة المعلوماتية مكتنتنا في هذا العصر من استخدام أسلوب جديد للتأكد من الحديث بتوظيف الذكاء الاصطناعي وتقنيات الحاسوب وذلك لدقة نتائجها وسرعتها الفائقة. وترتكز فكرة البحث على بناء نموذج ذكي لتصنيف ونقد

المتون آلياً؛ وهذه التقنية قد تم استعمالها في عدة مجالات حيوية أخرى، حيث يتم تصميم منظومة تقوم بفرز وتصنيف وفحص ونقد البيانات باستخدام الذكاء الاصطناعي.

1.5. منهج البحث:

اتبعنا في هذه الدراسة المنهج التجريبي التحليلي، فجمعنا ألفاظ النبي صلى الله عليه وسلم آلياً ويدوياً في قاعدة بيانات، مصنفة تبعاً لمراتبها إلى مقبول (صحيح وحسن) أو مردود (ضعيف وموضوع). هدفنا من هذا التصنيف هو محاولة الكشف عن الأسلوب النبوي في الأحاديث الصحيحة، وهذا الأسلوب قد لا يظهر إذا درست الأحاديث مجمعة. ثم أدخلنا هذه البيانات لعدة تطبيقات تعلم الآلة لتتعلم منها وتعطي حكماً بعد ذلك على أحاديث غير موجودة في قاعدة البيانات. كما قمنا باستخدام نموذج التعلم العميق كالمحولات Transformers تم تدريبها على بيانات ليتعلم النموذج السياق داخل الجملة نفسها أو بين الجمل. وفي الأخير قمنا برصد ومقارنة أهم النتائج التجريبية التي توصلنا إليها.

1.6. أهم المباحث

يشتمل البحث على النقاط التالية:

المقدمة: وتشمل أهمية الموضوع، وأسباب اختياره، ومشكلته، وأهدافه، ومنهجه، وأهم المباحث

الدراسات السابقة: يقتصر على بعض الدراسات التي تحاول الإجابة عن بعض الإشكالات المطروحة في البحث مع بيان الملاحظات حولها

مصطلحات البحث: يشمل بعض التعاريف لمصطلحات الذكاء الاصطناعي وفروعه

الذكاء الاصطناعي ومعالجة اللغة الطبيعية: بيان بعض المفاهيم وتقنيات الذكاء الاصطناعي مع أهمية توظيفه في خدمة علم مصطلح الحديث

القياس الأسلوبيّ ولفظ الحديث: بيان دور تقنية إسناد المؤلف في تحديد الأسلوب النبوي مع الأدوات والميزات المستعملة لتحديد البصمة المعرفية للنبي صلى الله عليه وسلم في الحديث النبوي

تطبيقات ونتائج: بيان صعوبة تجميع البيانات وتقديم إحصائيات ونتائج تعلم الآلة والتعلم العميق

خاتمة: تشمل خلاصة النتائج التي توصلنا إليها، وكذا بعض الصعوبات ثم التوصيات ذات الصلة بموضوع البحث.

2. الدراسات السابقة:

ما يزال استخدام تقنية الذكاء الاصطناعي متواضعا في مجال علم مصطلح الحديث، فبعد الاطلاع على العديد من الدراسات التي تناولت الذكاء الاصطناعي في نقد الحديث، تم تجميع الأهم منها والتي تحاول الإجابة عن بعض تساؤلاتنا وهي:

- توظيف الذكاء الاصطناعي في خدمة السنة النبوية (السيفاو، 2014): حاول المؤلف في هذا البحث تقديم بديل للطريقة التقليدية في نقد الأحاديث، فقد اعتبر أن الحكم مبني على العلم البشري المحدود وبالتالي وجود أخطاء من حين لآخر، بالإضافة إلى استنزاف الوقت والجهد وأحيانا تغيير الحكم على الحديث. ويرى الباحث أن الحل يكمن في توظيف الذكاء الاصطناعي في معالجة البيانات والاستفادة منها في خدمة السنة النبوية، ويمكن أن يقوم النظام الخبير بنقد الحديث والحكم عليه بالصحة أو الضعف بناء على ما استوعبه من مصادر السنة وخبرة العلماء.

- توظيف الذكاء الاصطناعي في خدمة الحديث الشريف وعلومه (خضر، 2018): في هذه المقالة للدكتور محمد زكي خضر أكد أنه يمكن للذكاء الاصطناعي كذلك أن يساعد في تقليد عمل المحدثين الأوائل وتطبيق قواعد الجرح والتعديل، كما أكد أن تطبيقات منطق الغموض Fuzzy Logic يمكن استعمالها في تطبيقات علم الحديث، خاصة إذا توفرت بيانات ضخمة مبوبة ومخدومة. ويبقى ما جاء في البحثين السابقين نظري يفتقر إلى تطبيق وبناء نموذج النظام الخبير لنقد الحديث.

- نحو نظام آلي لعلم الجرح والتعديل (عبد القادر، 2016): كشف الباحث في مقالته أنه قام بإعداد قاعدة بيانات لرواة الحديث بالاعتماد على كتب التراجم وكتب الجرح والتعديل، فأنجز بعد ذلك خوارزميات تقوم بتحليل السند وإصدار الحكم عليه. وفي مقالة لنواف محمد¹ أبرز أنه يمكن أن يقوم النظام الخبير باستخلاص المعلومات واستخدامها كمادة جاهزة لإصدار الحكم على السند أو المتن، وذلك بالرجوع إلى الخبرة الملقنة له مسبقاً على هيئة أمثلة يرجع إليها عند نقد الأسانيد أو المتن. الباحث في المقالة (عبد القادر، 2016)، كغيره من مجموعة من المؤلفين، ركّز على إنجاز خوارزميات تقوم بتحليل السند فقط وإصدار الحكم عليه.

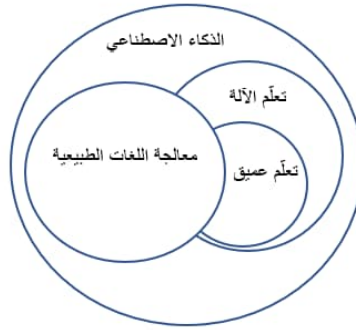
- استعمل المؤلف أعمور في المقالة (أعمور، 2016) في تجربة قياس الأسلوب لإسناد التأليف مجموعة مصنفات تعلم الآلة على مقاطع القرآن ومقاطع الحديث. ورغم أن القرآن والحديث قد نسبا دون خطأ، فإن المؤلف قد استخدم الحديث بشقيه: السند والمتن، مما أثار ملاحظات نقدية كبيرة في هذه الدراسة.

في هذه المقالة قمنا بإحداث قاعدة بيانات الألفاظ النبوية باستخراج المتن من الحديث آلياً ثم اللفظ من المتن يدوياً. وقمنا بعد ذلك، باستعمال مصنفات تعلم الآلة وتطبيقات التعلم العميق ومقارنة بعضها البعض.

3. مصطلحات البحث

1.7. الذكاء الاصطناعي: هو الآلة القادرة على أداء مهام تتطلب عادةً ذكاءً بشرياً، ومن أهم خصائصه القدرة على التعلم والاستنتاج ورد الفعل على أوضاع لم ترمج في الآلة. سنتناول في هذا الجزء بعض فروع الذكاء الاصطناعي ويبين الرسم التوضيحي¹ العلاقة بين الذكاء الاصطناعي ومختلف هذه المصطلحات.

1 <https://www.nmthgiat.com> /توظيف-الذكاء-الاصطناعي-في-خدمة-السنة-1/



رسم توضيحي 1: العلاقة بين الذكاء الاصطناعي والتعلم الآلي والتعلم العميق ومعالجة اللغات الطبيعية

التعلم الآلي: هو فرع من فروع الذكاء الاصطناعي تتعلم خوارزمياته من البيانات دون الحاجة لبرمجتها، بحيث تتمكن من استخراج المعلومات واستخدامها لإجراء التوقعات أو اتخاذ القرارات، بينما يعتمد التعلم العميق، وهو أيضا فرع من فروع الذكاء الاصطناعي، على الشبكات العصبونية الاصطناعية لاستخلاص السمات القيمة في البيانات باستخدام عدة طبقات مخفية في الشبكات العصبية لحل المشكلات المعقدة عن طريق تحديد أهم الخصائص الأساسية لبيانات الإدخال.

شبكة المحولات العصبية: ظهرت شبكة المحولات العصبية أو باختصار المحولات Transformers لحل مشكلة تمثيل سياق الجملة عوضا عن تمثيل كلماتها بشكل مستقل. وهو بنية شبكة عصبية لمعالجة البيانات المتسلسلة بالاعتماد على آلية الانتباه الذاتي لإعطاء الكلمات داخل النص تمثيلا يعبر عن سياق النص. فنجد معنى الكلمة داخل جملة ما مختلفا عن معناها داخل جملة أخرى. وظهرت بعد ذلك أساليب عدة للتعلم العميق الذي يعتمد على بنية المحولات لمعالجة اللغات الطبيعية مثل "تمثيلات الترميز الثنائية الاتجاه من المحولات (بيرت)"¹: Bidirectional

1 تم تطوير نموذج تمثيلات أداة الترميز ثنائية الاتجاه من المحولات BERT من قبل جوجل. تم تدريب هذا النموذج مسبقاً على مهام نمذجة اللغة باستخدام بيانات غير موسومة تتضمن نصوصاً من موقع ويكيبيديا.

Encoder Representations From Transformers (BERT). وأثبتت النماذج القائمة على BERT فعاليتها في فهم اللغة العربية وذلك عند تدريبها المسبق على كمية كبيرة من النصوص.

الضبط الدقيق: يقوم الضبط الدقيق Fine-tuning لشبكة ما بتعديل معاملات شبكة مدربة مسبقاً بحيث تتلاءم مع المهمة الجديدة المراد تنفيذها، ويجنبنا الضبط الدقيق البيانات الضخمة من أجل التدريب، كما يوفر الوقت للباحث.

معالجة اللغات الطبيعية (Natural Language Processing (NLP: هو علم تطبيقي يعنى باستخدام خوارزميات تعلم الآلة والتعلم العميق واتباع أساليبها كالأسلوب المبني على الدلالات التوزيعية لمعالجة النصوص والعثور على الأنماط.

4. الذكاء الاصطناعي ومعالجة اللغة الطبيعية

1.1. معالجة اللغات الطبيعية NLP

علم معالجة اللغات الطبيعية فرع من علوم اللغويات والذكاء الاصطناعي، مختص بالتفاعل والتواصل بين أجهزة الحاسوب واللغة الطبيعية. يُقصد باللغة الطبيعية اللغة البشرية السائدة المكتوبة منها أو المنطوقة والناجمة عن التطور الطبيعي بدون أن يكون هناك تخطيط للقواعد ولا غيرها (ديانا، 2019). ويهدف علم معالجة اللغات الطبيعية الى تهيئ الحاسوب لكي يكون قادراً على تمييز وفهم اللغات الطبيعية سواء كانت مكتوبة أو منطوقة، وذلك لاتخاذ إجراء بناء على ما تم فهمه من البيانات. وتحظى اللغة العربية كإحدى اللغات الطبيعية بعناية الباحثين في الذكاء الاصطناعي؛ فقد أثبتت خوارزميات التعلم العميق قدرتها على معالجة اللغات الطبيعية بشكل هائل، إذ تسهل معالجة اللغات وتطبيقاتها وذلك باعتماد بناء الشبكات العصبية الاصطناعية (رَشوان، 2019) لاستنباط القواعد والأنماط بشكل آلي وبدقة عالية، ويتم ذلك من خلال التعلم على كمية كبيرة من البيانات النصية ودون الرجوع لقواعد اللغة. أهم ما يميز اللغات الطبيعية أنها غالباً ما يكون فيها الاشتراك أو الغموض اللغوي نتيجة لتطورها الطبيعي الذي يحتم وجود الالتباس فيها. والتطبيقات التي تم

إعدادها أخيرا في مجال معالجة اللغات الطبيعية، مثل روبوتات المحادثة chatbots والأدوات الذكية الأخرى، مُبرجة على التواصل بين البشر وأجهزة الحاسوب وتجنب الغموض عن طريق استخراج معلومات مفيدة من البيانات. وتتميز هذه التطبيقات بالقدرة على الفهم ومساعدة الإنسان، وهي عبارة عن برامج مصممة لمحاكاة المحادثة مع الإنسان بشكل منطوق أو مكتوب.

في عام 1950، نشر آلان تورنغ¹ Alan Turing مقالته الشهيرة "آلات الحوسبة والذكاء" التي قدمت ما يسمى الآن اختبار تورنغ ليكون معيارا للذكاء الاصطناعي، ويعتمد هذا المعيار على قدرة برنامج الحاسوب على انتحال صفة إنسان في محادثة كتابية مع حكم بشري في زمن حقيقي، إذ يكون الحكم غير قادر على التمييز بشكل موثوق بين البرنامج وإنسان حقيقي. ولا تزال كفاءة أغلب خوارزميات إنتاج النصوص محدودة، ولم تتمكن بعد من الوصول لكفاءة متميزة أو النجاح في اختبار آلان تورنغ، وبذلك تكون أجهزة الحاسوب حتى الآن غير قادرة على فهم اللغة البشرية بالطريقة التي يفهمها الناس، ومع ذلك يمكنها فعل الكثير.

يهدف علم معالجة اللغات الطبيعية أساسا إلى جعل الأجهزة ذكية وتفهم اللغات كما يفهمها الإنسان، وهي التكنولوجيا الأساسية المستخدمة في آبل سيري Apple (Siri)، وكوكل اسيستانت Google Assistant، وأمازون الكسا Amazon Alexa، وتطبيقات الترجمة Google Translator، Microsoft Translator والعديد من التطبيقات الذكية الأخرى. ولكي نتمكن من التعامل مع هذا العلم، لابد من فهم اللغات بشكل صحيح ليتم كتابة قواعد اللغة بشكل برمجي؛ وهو أسلوب يُعتمد في البرمجة التقليدية، بينما يعتمد التعلم الآلي على البيانات الضخمة وبالأخص عملية التعلم العميق Deep Learning الذي لا نحتاج فيه إلى معرفة عميقة بالمجال.

1 آلان ماتيسون تورنغ هو عالم رياضيات وحاسوب شكلت أوراقه العلمية الأساس للبحث في موضوع الذكاء الاصطناعي.

يندرج تحت مجال معالجة اللغات الطبيعية عدة علوم مثل:

- الترجمة الآلية Machine Translation
- علم استرجاع المعلومات Information Retrieval
- علم استخراج المعلومات Information Extraction
- الإجابة على الأسئلة Question Answering
- علم تحليل الرأي أو العاطفة Sentiment Analysis
- تمييز الأصوات Speech Recognition

وتبقى المعالجة الآلية للغة من حيث الدلالة تحدياً ضخماً، وتعد الآن الحقل الأكثر خصوصيةً في مجال المعالجة الطبيعية للغات؛ وفي هذا الصدد نحن بحاجة ماسة إلى تعاون كل من اللسانيات والمهارة التقنية من أجل استعمال التقنيات الحديثة في المعالجة الآلية للغة.

1.2. صعوبة فهم اللغة

قال المتنبي في إحدى أبياته الشعرية: "أَلَمْ أَلَمْ أَلَمْ بِدَائِهِ إِنَّ أَنْ أَنْ أَنْ أَنْ وَأَوَانِهِ". والمعنى العام لهذا البيت الشعري "لقد أحاط بي ألم لم أعرفه من قبل إن جاء وقت شفائه من الله فقد آن ذلك له"¹. يوضح هذا البيت الشعري، الذي ربما لا يفهم دون تفسير، صعوبة فهم اللغات بشكل عام، لذلك يعد علم معالجة اللغات الطبيعية من أهم وأعقد علوم الذكاء الاصطناعي حيث يتعامل مع النصوص وهي التي يُساء فهمها في كثير من الأحيان بين البشر أنفسهم. كما أن أجهزة الحاسوب يسهل لها التعامل فقط مع الأرقام والقيام بالعمليات الحسابية عليها، لكن لا يسهل عليها التعامل مع النصوص، خاصة وأن أغلب البيانات غير مهيكلة unstructured data، وتأتي في أشكال ونماذج ولغات مختلفة، لهذا نحتاج لأشخاص متمكنين من لغة البرمجة واللغة التي ينشئ لها التطبيقات.

1 <https://el-ma3loma.com/2017/03/motnabi.html>

ومن التحديات الأخرى التي تواجه علم المعالجة الطبيعية للغات:

- الاختلاف الكبير في اللهجات واللغة غير الرسمية المستخدمة في كثير من المحادثات؛
- التقطيعات غير الصحيحة في الجمل؛
- التعبيرات المجازية والاستخدامات الساخرة للكلمات؛
- المصطلحات الجديدة؛
- كلمات متعددة الاستخدامات والمعاني.

بالإضافة إلى ذلك، فالكمية الكبيرة من البيانات غير المهيكلة، وغياب استخلاص المعنى الصحيح من الكلمات وفقا للسياق جعل مهمة فهم أجهزة الحاسوب للغة البشرية مهمة أكثر تعقيداً. وتقوم خوارزميات تعلم الآلة بدور كبير في تطوير الأجهزة لكي تفهم اللغات، وذلك من كثرة التعلم من البيانات للتعرف على أنماط المعلومات حتى يصبح أكثر دقة في التعرف على كلام الإنسان المكتوب أو المنطوق؛ لذلك فإن البيانات الضخمة مهمة لفهم لغة الإنسان الطبيعية باستخدام الذكاء الاصطناعي.

1.3. معالجة اللغات الطبيعية والمعاني الدلالية في الأحاديث النبوية الشريفة

كان صلى الله عليه وسلم يمتلك معجماً لغوياً مهماً نجده في أحاديثه بألفاظ مناسبة، ورغم اشتراك بعضها في معنى واحد لكنها مختلفة في الدقة في الدلالة على هذا المعنى، كما أن لكل لفظ خاصية يتميز بها عن اللفظ الآخر في بيان المعنى والقدرة على التعبير عنه، وهو ما يبرز انفرادية الأسلوب النبوي خاصة في استخدام الألفاظ القليلة والواضحة للتعبير عن المعاني الكثيرة، مع الدقة في التعبير؛ فقد أوتي صلى الله عليه وسلم جوامع الكلم واختصر له الكلام اختصاراً، أي أنه أوتي القدرة على إيجاز اللفظ مع سعة المعنى وتلك غاية البلاغة. إن الحديث عن الأسلوب النبوي في الحديث الشريف؛ معناه البحث في بلاغة الحديث وخصائصه وسماته البلاغية، باستخدام الذكاء الاصطناعي. وهي كلها سمات أسلوبية تُخدم القياس الأسلوبية Stylometry

باستخلاص الأسلوب من قاعدة بيانات الألفاظ النبوية باستخدام تقنيات تعلم الآلة كوحدة TF-IDF، وهي تقنية تستخدم على نطاق واسع في استخلاص المعلومات وذلك بالقياس الإحصائي لمدى أهمية الكلمة أو اللفظ في مجموعة من الوثائق. وإذ تركز هذه التقنية على الألفاظ في الدلالة على المعنى، إلا أن التعلم العميق Deep Learning أثبت قدرته على معالجة اللغات الطبيعية بشكل يفوق تعلم الآلة؛ إذ يعتبر أن تحليل الخطاب النبوي ينطلق من تحليل نصّ في سياق، وأن للكلمة المفردة تعبير عن معنى إذا استخدمت في سياق ما، ولها علاقة بالكلمات السابقة لها أو اللاحقة بها داخل العبارة أو الجملة.

تهدف هذه المقالة إلى بناء نموذج لدراسة وتصنيف الأحاديث لغويا بغية تمييز الصحيح منها من الضعيف، إذ غالبا ما تكون الأحاديث الضعيفة مختلفة من حيث تركيبها وسياقها عن الأحاديث الصحيحة. فعلى الرغم من أن الكاذب في الحديث الضعيف يستطيع السيطرة على المحتوى الملفق، إلا أنه دائما يترك فجوات في سياق الحديث. لذلك نحتاج الى قاعدة بيانات الأحاديث تحتوي على عبارات أو جمل، ثم نستخدم النموذج ونقوم بتدريبه على هذه البيانات. غير أن قاعدة بياناتنا تفتقر إلى هذا السياق في بعض الأحاديث وذلك بعد حذف ما هو ليس بلفظ النبي صلى الله عليه وسلم، ومن أمثلة ذلك قوله صلى الله عليه وسلم: «هذا ركس¹»، في حديث البخاري «أتى النبي صلى الله عليه وسلم الغائط فأمرني أن آتيه بثلاثة أحجار، فوجدت حجرين، والتمست الثالث فلم أجده، فأخذت روثه فأتيت بها، فأخذ الحجرين وألقى الروث وقال: هذا ركس²». وكذلك قوله صلى الله عليه وسلم: «أطولكن يدا³»، في حديث مسلم «أن بعض أزواج النبي صلى الله عليه وسلم، قلن للنبي صلى الله عليه وسلم: أئنا أسرع بك لحوقا؟ قال: أطولكن يدا، فأخذوا فصبة يذرعوها، فكانت سودة أطولهن يدا، فعلمنا بعد أنما كانت طول يدها الصدقة، وكانت أسرعنا لحوقا به

1 حديث رقم 154 - صحيح البخاري - كتاب الوضوء

2 حديث رقم 1365 - صحيح البخاري - باب فضل صدقة الشحيح الصحيح

وَكَاثَتْ تُحِبُّ الصَّدَقَةَ.» ولحل هذه الإشكالية، نجد أن النبي صلى الله عليه وسلم استخدم في أحاديثه الشريفة أسلوب التكرار للمعنى بألفاظ مختلفة، مما أحدث تنوعاً في النصوص وزيادة في قاعدة البيانات، ومثال ذلك قوله صلى الله عليه وسلم في صحيح مسلم¹: «أَسْرَعُكُمْ لِحَاقًا بِي أَطْوَلُكُمْ يَدًا.» فنجد في هذا الحديث زيادة اللفظ بعبارات إضافية، ومثال التكرار في المعنى دون اللفظ: «لَا يَحْكُمُ أَحَدٌ بَيْنَ اثْنَيْنِ وَهُوَ غَضْبَانٌ²» وجاء في حديث البخاري: «لَا يَقْضِيَنَّ حَكْمٌ بَيْنَ اثْنَيْنِ وَهُوَ غَضْبَانٌ³». هذه الأساليب النبوية تزيد في المعاني الدلالية في الأحاديث الشريفة وتحل مشكلة تمثيل سياق الجملة عوضاً عن تمثيل كلماتها.

5. القياس الأسلوبى ولفظ الحديث

1.1. القياس الأسلوبى

يعتبر القياس الأسلوبى من التطبيقات المهمة في معالجة اللغات الطبيعية التي تسمح بإمكانية التأكد من نسبة مقال أو كتاب لكاتب معين؛ فهو يقوم بدراسة الأنماط اللغوية الفريدة وسلوكيات الكتابة للأفراد من أجل تحديد المؤلف (الفلاحي، 2015). ويمكن تقسيم هذا المجال إلى ثلاثة أقسام:

- تحديد المؤلف (إسناد المؤلف): البحث عن المؤلف الحقيقي لمقال أو وثيقة؛
- تحديد سمات المؤلف (التوصيف): الحصول على خصائص المؤلف مثل الجنس والعمر؛
- الكشف عن التشابه: إيجاد التشابه بين النصوص لتحديد إمكانية إنتاجها من قبل كاتب واحد، دون إيجاد المؤلف الحقيقي بالضرورة. ويشيع استخدام هذا المجال في الكشف عن السرقة الأدبية.

1 حديث رقم 4617 - صحيح مسلم - باب: بَابُ مَنْ فَصَّلَ زَيْنَبَ أُمِّ الْمُؤْمِنِينَ

2 حديث رقم 3343 - صحيح مسلم - باب: بَابُ كَرَاهَةِ قَضَاءِ الْقَاضِي وَهُوَ غَضْبَانٌ

3 حديث رقم 7158 - صحيح البخاري - باب: هل يقضي القاضي أو يفتي وهو غضبان

يعرض المؤلف في الأساس عادات الكتابة التي تميزه، والتي يتم عرضها في سمات مثل استخدام مفردات معينة أو عرضها بتعبير معقد. ويُستخدم تحليل الخطاب كذلك لبناء كثيرٍ من التطبيقات من خلال تحليل النصوص التي كُتبت سابقاً لمعرفة أسلوب الكتابة، ومن هذه التطبيقات تحديد المؤلف التي سنحاول من خلال هذا البحث تطوير البصمة التقنية المعرفية لسيدنا محمد صلى الله عليه وسلم من خلال تحليل أقواله في المتون الخاصة بالحديث الشريف.

ومن أجل استخلاص استنتاجات خاصة بالحكم على الحديث والتنبؤ بمرتبته، قمنا بدراسة أداة TF-IDF وأداة N-Grams والميزات النحوية في نص الحديث:

أداة TF-IDF: وهي اختصار Term Frequency-Inverse Document Frequency "تردد المصطلح - تردد المستند العكسي". تعتمد هذه الأداة على قيمة تردد المصطلح، أي معدل تكرار المصطلح المطلوب، وعلى تردد المستند العكسي، والذي يشير إلى مدى ندرة أو شيوع هذا المصطلح. وتستخدم أداة TF-IDF لهيئة البيانات النصية قبل عملية التصنيف وذلك بتحديد مدى تواجد كل مصطلح في النص، ثم حساب قيمة TF-IDF التي تعتبر خاصية مهمة أثناء تدريب المصنفات.

أداة N-Grams: تستخدم أداة N-Grams في أنظمة التعرف على الأنماط لتقييم احتمالية ظهور تسلسل عدد من المصطلحات (N) قبل وبعد المصطلح المطلوب في النص. في هذه الدراسة، قمنا باستعمال نظام bigram الذي يتعامل مع كلمتين ونظام trigram الذي يتعامل مع ثلاث كلمات. ويشكل N-Gram خاصية النص ومن مدخلات خوارزميات التعلم الآلي.

الميزات النحوية: طول النص، وعدد الكلمات، ومتوسط طول الكلمات، وعدد الكلمات القصيرة ونسبة الحروف.

أردنا من خلال هذا العمل أن نضع بين أيدي المهتمين بالبحث العلمي البيانات التي جمعناها والاستفادة من التقنية الحديثة في أبحاث العلوم الشرعية، لتكون عوناً لهم لاستثمارها في أبحاثهم لتحقيق أهدافها والوصول إلى نتائجها في وقت معقول.

1.2. الأسلوب النبوي في لفظ الحديث

لا شك أن البيان النبوي الشريف خال من ركافة اللفظ فهو في قمة البلاغة والوضوح، لأنه يجري وفق قوانين اللغة العربية وأساليبها في التعبير. فقد تربي صلى الله عليه وسلم تربية ربانية ونشأ في بيئة لغوية نقية حيث ولد في قريش أفصح القبائل العربية، وعاش صلى الله عليه وسلم في السنوات الأولى من عمره في مأمن من الاختلاط بالشعوب في قبيلة بني سعد. أمر صلى الله عليه وسلم بتبليغ رسالته بأرقى مستويات الأداء اللغوي لفظاً وصوتاً ودلالة، كما أن للنبي صلى الله عليه وسلم أسلوباً جميلاً بالفاظ قليلة بعيدة عن الإطناب والتطويل. هذا البيان الذي يسير على نسق واحد يمنحه الديمومة والتفرد في أسلوب متون الحديث يمكن من خلاله تطوير بصمته المعرفية صلى الله عليه وسلم. ويُمكِّننا استثمار الذكاء الاصطناعي من تطوير هذه البصمة النبوية والتي ستساعد في التحقيق الآلي للمتون الموجودة على الإنترنت.

6. تطبيقات ونتائج

1.1. تجميع البيانات

تعتبر عملية تجميع البيانات من أصعب وأطول مراحل بناء أنظمة الذكاء الاصطناعي، إذ أن مجال معالجة الأحاديث لا يزال يفتقر إلى البيانات الجيدة والمفتوحة المصدر، والتي تمكننا من تطوير تطبيقات معالجة اللغات الطبيعية في هذا المجال. وعلى الرغم من وجود بيانات مهمة في الحديث، فإنه - حسب علمنا - لا وجود لبيانات لفظ الحديث مستقلاً عن المتن؛ حيث إن المتن هو نص الحديث الذي يأتي بعد السند من قول أو فعل أو تقرير. وإذ نعتمد في هذا البحث على الأسلوب النبوي في لفظ الحديث، فإننا نعتد أيضاً على الأشخاص لتصنيف القول عن الفعل أو التقرير في المتن لتقديم البيانات الصحيحة للذكاء الاصطناعي والعمل معها بالطريقة

المطلوبة. وهذه العملية مكلفة وتحدث بشكل يدوي، كما أنه لا بد من بذل الوقت والجهد قبل البدء في عملية التدريب في تهيئة البيانات لتكون صالحة للاستخدام.

في هذه الدراسة، استخدمنا تقنيات زاحف الويب Web Crawler لتصفح وفهرسة محتوى مواقع الويب المتاحة على الإنترنت بغرض جمع بيانات الحديث من مواقع موثوقة. وقد قامت مؤسسة الدرر السنية في موقعها ببناء أضخم قاعدة بيانات إلكترونية شاملة لميراث الرسول صلى الله عليه وسلم، ويسر الوصول إليها من خلال التقنيات الحديثة. فالموسوعة الحديثة في هذا الموقع تضم مئات الألوف من الأحاديث مع أحكام المحدثين مستخرجة من كتبهم، ويمكن البحث فيها عن حكم أي حديث من خلال خدمة API¹، مع إمكانية اختيار البحث ضمن الصحيح أو الضعيف. وقد تم الحصول على إجمالي 34324 حديثاً مصنفاً حسب الراوي (انظر الجدول 1).

قمنا بتجميع البيانات اللازمة وتكونت لدينا مجموعة مناسبة من الأحاديث النبوية فعملنا بعد ذلك على إزالة ما ليس بلفظ الحديث يدويا ونذكر هنا نماذج لما زيد على اللفظ للشرح: يَعْني...، يُريد...، أو نَحْوًا مَّا قَالَ، ثم تلا هذه الآية....، وَذَكَرَ....، كَانَ يَقُولُ....، وأشار إلى....، وأومأ إلى وجهه وكفيه....، وفي رواية....، حَسِبْتُ أَنَّهُ قَالَ...

جدول 1: إجمالي الأحاديث قبل المعالجة المسبقة

المقبول		المردود	
العدد	الراوي	العدد	الراوي
329	السيوطي	100	السخاوي
1944	الموطأ	102	ابن الملقن
2605	مسلم	105	شعيب الأرنؤوط

1 واجهة برمجة التطبيقات

الزرقاني	132	الألباني	3302
النووي	194	ابن ماجة	5069
محمد جار الله الصعدي	261	البخاري	8920
السيوطي	1056	أخرى	660
الألباني	8465		
أخرى	1080		
إجمالي المردود	11495	إجمالي المقبول	22829
		34324	الإجمالي

تمت تنقية البيانات كما يلي:

- عزل السند عن المتن ثم اللفظ عن المتن؛
 - ترك مسافة فراغ قبل وبعد علامات التنقيط؛
 - تم تقسيم النص إلى جمل، ووضع كل جملة في سطر واحد (مما زاد في عدد العينات)؛
 - فصل البيانات إلى مجموعات تدريب واختبار: جدول 2.
- جدول 2: تقسيم البيانات إلى مجموعة خاصة بالتدريب ومجموعة خاصة بالاختبار

قرآن - حديث مقبول	حديث مردود - حديث مقبول	
16038	19019	عدد العينات
12830	15215	التدريب
3208	3804	الاختبار

لتنظيف النصوص العربية ومجموعة من مهام المعالجة المسبقة preprocessing، قمنا باستخدام مكتبة تنقيح (زيد، 2020). وقمنا بعد ذلك بنفس المهام على آيات القرآن الكريم لتسهيل المقارنة والتأكد من الاختيار الأنسب للتقنيات.

تحتوي قاعدة البيانات النهائية على حوالي 10000 حديث مقبول (الحديث الصحيح والحديث الحسن) مأخوذة من صحيح البخاري ومسلم، وعلى حوالي 10000 حديث مردود (الحديث الضعيف والموضوع والمنكر والشاذ...) مأخوذة من كتب أبي عبد الرحمن محمد ناصر الدين الألباني وغيره. ولأن قاعدة البيانات المتوفرة لدينا غير متوازنة، فقد قمنا باستعمال مقياس $F1^1$ إلى جانب الكفاءة لتقييم النماذج.

هناك صعوبة كبيرة في هذه المرحلة، حيث أن أداء مجموعة من الأدوات المستعملة في اللغة العربية ضعيف، بالمقارنة مع اللغات الأخرى كالإنجليزية مثلاً، علاوة على ذلك، فإن الخطوة الأولى أنجزت كلها يدوياً.

1.2. التجارب والمقارنات

بعد تهيئة البيانات لتكون جاهزة كمدخلات لخوارزميات تعلم الآلة، قمنا باستعمال أداة Python المتخصصة في علوم البيانات والذكاء الاصطناعي لتحليل النتائج وتدريب النماذج المقترحة ومقارنتها من حيث الكفاءة ومقياس $F1$. وكما يتضح من الجدول 3، تمت مقارنة المصنفات التي تشمل الخوارزميات الشائعة التي يمكن استخدامها للتصنيف الثنائي: (غابة عشوائية Random Forest، الانحدار اللوجستي Logistic Regression، أقرب الجيران Nearest Neighbors، أشجار القرار Decision Trees، آلة المتجهات الداعمة Support Vector Machine، بايز بسيط Naive Bayes). ولتقييم النتائج، تم تقسيم البيانات إلى مجموعة خاصة بالتدريب ومجموعة خاصة بالاختبار. وطُبقت خوارزميات التعلم

$F1$ 1: تعني النتيجة الكاملة لكفاءة ودقة النموذج.

الآلي كما استخدمت TF-IDF لتهيئة البيانات النصية قبل عملية التصنيف لاختيار المصنف الأنسب لاستعماله في التجارب المقبلة فكانت النتائج كما في الجدول 3.

جدول 3: نتائج التصنيف باستعمال TF-IDF

النماذج	الكفاءة		الفارق
	قرآن - حديث مقبول	قرآن - حديث مردود	
أقرب الجيران	0.91253	0.90608	-0.00645
غابة عشوائية	0.94124	0.93745	-0.00379
أشجار القرار	0.8942	0.89961	0.005408
الانحدار اللوجستي	0.91366	0.92059	0.006928
آلة المتجهات الداعمة	0.83053	0.83137	0.000842

يتضح من النتائج الموضحة أعلاه أن أداء خوارزمية الغابة العشوائية والانحدار اللوجستي كان إلى حد ما جيدا، وكانت نسب التقييم 94% و92% على التوالي، وذلك بعد تطبيقها على قاعدة بيانات القرآن ونص الحديث لإصدار الحكم في مهمة إسناد التأليف. كما يبين الجدول صعوبة التنبؤ بمرتبة الحديث، حيث كان الفارق ضعيفا جدا عند مقارنة القرآن بألفاظ الحديث المقبول والمردود. كما تم فحص النتائج المختلفة بناء على الخوارزميات السابقة والتي تم تطبيقها على بيانات الحديث باستعمال TF-IDF وسمات النص (جدول 4).

جدول 4: نتائج التصنيف باستعمال TF-IDF وسمات النص

النماذج

TF-IDF		خصائص النص		
الكفاءة	F1	الكفاءة	F1	
0.625131	0.557625	0.578601	0.578601	أقرب الجيران

0.531283	0.531283	0.84547	0.845689	غابة عشوائية
0.750263	0.750263	0.806383	0.807308	أشجار القرار
0.563617	0.563617	0.753754	0.754206	الانحدار اللوجستي
0.630389	0.630389	0.720043	0.720557	آلة المتجهات الداعمة

عند مقارنة نتائج الجدول أعلاه، يتضح أن مصنف الغابة العشوائية هو أفضل مصنف عند مقارنته بمصنفات تعلم الآلة الأخرى. وقبل إدماج هذا المصنف في التجربة المقبلة، قمنا بمقارنة " تمثيلات الترميز الثنائية الاتجاه من المحولات (بيرت) BERT العربية كما في الجدول الموالي (5):

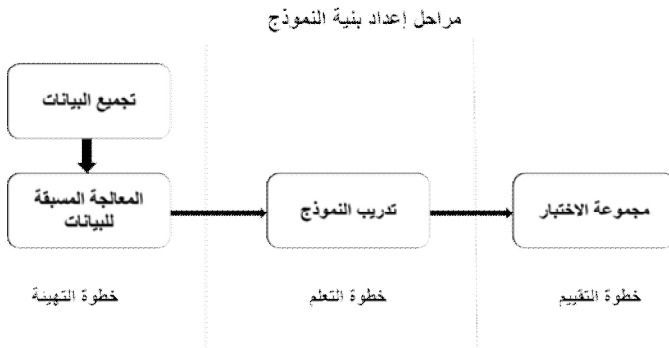
جدول 5: مقارنة بيرت BERT العربية

F1	الكفاءة	بيرت العربية
0.882394	0.883018	camelbert-ca
0.86591	0.866193	camelbert-msa
0.846912	0.847266	qarib
0.834791	0.835174	arabicbert
0.821739	0.822818	ARBERT
0.814865	0.814932	arabertv02
0.581478	0.58193	multilingual

استخدمنا نفس مجموعة بيانات الحديث لتنفيذ نموذج التعلم العميق ورسمنا كفاءة التدريب ودرجة F1 كما هو موضح في الجدول 5. ويُظهر الجدول البيانات التي تسمح لنا بمعاينة مقياس الكفاءة لكل نموذج مما يسمح لنا باختيار الأنسب منها. كما يمكن ملاحظة أن camelbert-ca يكتسب ما يقارب 89٪ من الكفاءة ودرجة F1

ويعتبر الأفضل، بينما يحقق نموذج المتعدد اللغات multilingual الأضعف بنسبة 58%، وبالتالي تسمح النماذج العربية (أحادية اللغة) بتحسين مهم مقارنة بالنموذج متعدد اللغات. ويمكن تفسير ذلك بكون نموذج camelbert-ca مدرباً مسبقاً على اللغة العربية الفصحى فقط. كما أن camelbert-msa يكتسب ما يقارب 87% من الكفاءة ودرجة F1 لكونه أيضاً مدرباً مسبقاً على اللغة العربية الحديثة، بينما النماذج الأخرى فهي مدربة مسبقاً على اللغة العربية إلى جانب اللهجات العامية والويكيبيديا وغيرها.

النموذج المقترح في هذه الدراسة يجمع بين خوارزميات تعلم الآلة المتمثل في المصنف الغابة العشوائية مع مدخلات بميزات الميزات المستخدمة في حل مشكلة القياس الأسلوب، وخوارزميات التعلم العميق المتمثل في بيرت المدرب مسبقاً على اللغة العربية الفصحى camelbert-ca. وتم دمج كل العناصر في منظومة واحدة لبناء نموذج لتعليم الآلة على معرفة أسلوب النبي صلى الله عليه وسلم واستخلاصه عبر دراسة مجموعة من الأحاديث واستنباط بصمته الأسلوبية ليتم من خلال ذلك التنبؤ بمرتبة الحديث. وتقوم فكرة تدريب الخوارزميات على الخطوات التالية (رسم توضيحي 2):



رسم توضيحي 2: مراحل إعداد النموذج

- قراءة كمية كبيرة من النصوص المحددة؛
- تنظيف البيانات وإزالة الرموز غير المعروفة؛

• تحويل الكلمات إلى أرقام حتى يتم التعامل معها في الشبكة العصبية RNN.

يتم تكرار هذا الأمر عددا من المرات epochs حتى يتمكن النموذج من التوقع.

قمنا بتعديل طبقات الطّي المدربة مسبقاً camelbert-ca، باستخدام نهج الضبط الدقيق Fine-Tuning؛ فعلى الرغم من أن BERT كبير جداً ومعقد ويحتوي على ملايين من الإعدادات، إلا أننا نحتاج في ضبطه إلى عدد قليل فقط من الفترات أو دورات تدريب Epochs.

يبين الشكل التالي نتائج تدريب النموذج بعد خمس دورات تدريب.

جدول 6: نتائج النموذج المقترح

البيانات	الكفاءة	F1
قرآن - حديث	0.99096	0.990504
مقبول - مردود	0.890641	0.890344

يمكن أن نستنتج أنه عندما نستخدم مجموعة بيانات الحديث التي أنشأناها ثم ضبطها مع بيرت، نحصل على أفضل أداء 89.81%، وهي نتيجة جيدة بالنظر إلى عدد ألفاظ الحديث. وعلى الرغم من ذلك فإننا نرى أن الأخذ بخصائص اللغة العربية كالتشكيل والعمل على مجموعة بيانات أكبر، سيزيد من نسبة هذه الكفاءة. هذا لأن النموذج تم تدريبه في الأصل على المفردات العربية باستخدام مصادر متنوعة باللغة العربية الفصحى (classical Arabic ca). كما أن بيرت يُعرف بأداء جيد في استخراج المعلومات الدلالية والنحوية، لذلك أظهر camelbert-ca أداءً ممتازاً عند استخدامه كطريقة لاستخراج الميزات في نص الحديث. كما يؤدي إضافة ميزات القياس الأسلوبي على بيرت إلى تحسين أداء النموذج المقترح.

7. خاتمة

عرضنا في هذه الدراسة نماذج تعلم الآلة والتعلم العميق وكيفية اختيار النموذج الأفضل منها لنقد الأحاديث. ومن أهم النتائج التجريبية التي توصلنا إليها:

1. بناء نموذج لتعليم الآلة على معرفة الأسلوب النبوي عبر دراسة مجموعة من الأحاديث واستنباط بصمته الأسلوبية ليتم من خلال ذلك إصدار الحكم على أحاديث آليا.

2. تم إنشاء قاعدة بيانات جديدة لألفاظ الحديث النبوي قد تساعد الباحثين في هذا المجال.

3. تم تحقيق أداء ممتاز لمهمة إصدار الحكم على الأحاديث في تجربتين

- ففي التجربة الأولى، تم بناء نموذج للتنبؤ بمرتبة الأحاديث بالاعتماد على خوارزميات تعلم الآلة، وكان أفضل أداء في التصنيف من نصيب الغابة العشوائية حيث وصل إلى كفاءة 84% وهو أقل من الهدف المنشود.

- وفي التجربة الثانية، أكد تدريب النماذج فاعلية التعلم العميق في هذا المجال، خاصة النماذج التي تم تدريبها على البيانات العربية التي أعطت نتائج أفضل من تلك التي تم تدريبها على البيانات المتعددة اللغات. فقد حقق camelbert-ca الأداء الأفضل بنسبة 89% وهي نتيجة محفزة ومشجعة للغاية، لذلك فمن المرجح أن الأداء سيكون أفضل عندما يتم تطبيقه في مجال محدد هو علم مصطلح الحديث.

1.3. صعوبات في البحث

ورغم هذه النتائج المشجعة فإن توظيف الذكاء الاصطناعي في نقد الحديث النبوي يواجه جملة من الصعوبات أهمها:

- التشابه الكبير بين الأحاديث الصحيحة والضعيفة (الجدول 3)؛
- الحكم على الأحاديث عند إزالة أو زيادة لفظ فقط؛
- الحكم على الأحاديث من خلال السند؛

- رواية بعض الأحاديث بالمعنى وليس باللفظ؛
- رواية الراوي لنفس الحديث بألفاظ مختلفة؛
- نوع التجديع stemming المستعمل في استخلاص المفردات؛
- الشح في الموارد، فكلما كان حجم البيانات أكبر، كلما كان معنى المصطلح أدق....

علاوة على ذلك، فإن مجموعة من الأدوات والمكتبات المستعملة في لغات البرمجة لا تدعم اللغة العربية، وبالتالي فأدائها ضعيف بالمقارنة مع اللغات الأخرى. كما أن اللغة الحالية المنتشرة على الإنترنت تحتوي على مشكلات كثيرة يصعب التعامل معها كالأخطاء الإملائية الشائعة (الهمزة، الهاء، التاء، الياء والألف المقصورة) واستخدام اللهجات العامية.

1.4. توصيات

إن أهمية النتائج التي تم التوصل إليها باستعمال نموذج camelbert-ca المدرب مسبقاً على اللغة العربية الفصحى تجعلنا نقترح في الأعمال المستقبلية ما يلي:

1. أن يتم الاعتماد في حل المشاكل أعلاه على استخدام هيكلية نموذج بيرت BERT بدون أوزان وتدريبه من البداية على بياناتنا الخاصة:

- كتب المتون
 - كتب التاريخ
 - التراجم المكتوبة باللغة العربية الفصحى ...
2. أن يتم الاعتماد كذلك على قاعدة بيانات كبيرة بإضافة أحاديث من غير الصحيحين،

3. ضبطها مع بيرت باستخدام نهج الضبط الدقيق Fine-Tuning،
4. كما يستحسن تجريب الاستراتيجيات الأخرى في مهمات أخرى وبيانات مختلفة.

5. ضرورة الاستعانة باللسانيين والمهندسين في المعلوماتية باعتبارهم المصدر الأساس لمرحلة إعداد البيانات؛

وبذلك سنساهم في إثراء المحتوى الرقمي على شبكة الانترنت في مجال فتي مثل الذكاء الاصطناعي مع علم مصطلح الحديث، وعليه فالدعوة هنا عامة من أجل تضافر الجهود والالتفات بجد لمثل هذه المشاريع، لتكون ركيزة أساس لمشاريع مستقبلية من شأنها صد الهجمات المعلوماتية ونشر وتحقيق صحيح السنة النبوية والدفاع عنها.

8. المراجع

- أحمد الفلاح، إسنادية التأليف في الشعر العربي، المؤتمر الدولي العاشر للأنظمة الذكية: النظريات والتطبيقات IEEE. (SITA)، 2015. ص. 1-6.
- ديانا ماينارد، معالجة اللغات الطبيعية للويب الدلالي، 23/1، 2019.
- زيد اليافعي وماجد سعيد، tkseem: مكتبة معالجة للغة العربية، GitHub repository، 2020.
- سهام أعمور، تأثير حجم النص على أسلوب القياس - التطبيق على النصوص الدينية العربية، وقائع المؤتمر الدولي الرابع لعلوم الكمبيوتر والرياضيات التطبيقية والتطبيقات، فيينا، النمسا، 2016، ص 215-228.
- عبد الكريم عبد القادر وعبد الرحمن عثمان، نحو نظام آلي لعلم الجرح والتعديل، Communications of the Arab Computer Society، المجلد 9، العدد 1، 2016.
- كاشور أفلاح السيفاو، توظيف تقنيات الذكاء الاصطناعي في خدمة السنة النبوية، ماجستير، جامعة المدينة العالمية، ماليزيا، 2014.
- محسن رشوان، المعالجة الآلية للنصوص العربية، 65/2، 2019.

- محمد زكي خضر، توظيف الذكاء الاصطناعي في خدمة الحديث الشريف وعلومه، المجلة الدولية للتطبيقات الإسلامية في علم الحاسب والتقنية، المجلد 6، العدد 4، كانون الأول/ديسمبر 2018
- نواف محمد، توظيف الذكاء الاصطناعي في خدمة السنة النبوية باستخدام النظم الخبيرة، <https://www.nmthgiat.com/> /توظيف-الذكاء-الاصطناعي-في-خدمة-السنة-ا، 07 يونيو 2023.