مجلة جامعة دمشق للعلوم الهندسية
المجلد 39- العدد 4 (2023) 273-281

Damascus university Journal for engineering sciences
Vol 39 No.4 (2023) : 273-281

# نظام تصنيف الأوراق البحثية واقتراح المحكمين بالاعتماد على معالجة اللغات الطبيعية NLP والنموذج اللغوي AraBERT

**رزان حلاق[1]، د. محمد بسّام الكردي[2]**

[1] طالبة دراسات عليا في ماجستير تقنيات الويب MWT، الجامعة الافتراضية السورية.

[2] اختصاص ذكاء صنعي، المعهد العالي للعلوم التطبيقية، والجامعة الافتراضية السورية.

## الملخص:

تعتبر آليّة مراجعة الأوراق البحثية أداة التحكم بجودة الأبحاث العلمية المقدمة للنشر في المجلات والمؤتمرات العلمية، ومع ازدياد عدد الأبحاث العلمية المقدمة للنشر زاد الضغط على عملية المراجعة والتي تعتمد على الجهد البشري في فرز الأبحاث حسب موضوعها وإسنادها إلى المحكم الأنسب، ما زاد الحاجة لتطوير النظام التقليدي المتّبع لتفادي حدوث بطء أو تحيّز أو نقص في جودة العملية ولمواكبة التسارع العلمي المستمر. ومع تطور أدوات وتقنيّات الذكاء الصنعي وظهور العديد من الأدوات والمكتبات المساعدة في تحليل نصوص اللغة العربية ما سهّل استخدامها لتطوير النظم الحالية. تمّ في هذا البحث إنشاء مصنف متعدد التسميات للأبحاث حسب موضوعها (Multi-Label Classifier) بالاعتماد على نموذج التعلم العميق المدرب مسبقاً على اللغة العربية AraBERT وتم تخصيصه عبر تدريبه على بيانات مجلة محكّمة (المجلة السورية للبحوث الزراعية) بعد معالجتها أولياً باستخدام تقنيات معالجة اللغات الطبيعية، وأثبت هذا المصنف قدرته على تحديد موضوع البحث باستخدام العنوان والملخص فقط، وتم توظيف هذا المصنف في بناء نظام توصية بالمحكمين الأنسب لتحكيم ورقة بحثية، وهو عبارة عن نظام استرجاع المعلومات والذي يعمل على تصنيف وتحليل نصوص كل من البحث المقدم للتحكيم والأبحاث المنشورة سابقاً لمجموعة المحكمين في المجلة وتضمين النص دلالياً بالاعتماد على النموذج AraBERT للمقارنة بين النصوص وقياس نسبة التشابه لتقديم اقتراحات بأسماء عشر محكمين ذوي الاختصاص الأقرب للبحث، كما يوفر النظام اقتراحات إضافية كمحكمين جدد من ضمن الباحثين الذين نشروا مسبقاً في المجلة. تم اختبار النظام المنشأ على بيانات المجلة المحكمة، وأعطى نموذج التصنيف القيم 86.5%، 91.8%، 92.4%، 91.4% لكل من المقاييس Accuracy، F1-score، Precision، Recall على التوالي. كما وُجد أن 85% من المحكمين المقترحين من قبل النظام مناسبين تماماً للمراجعة حسب تقييم خبراء النشر في

تاريخ الإيداع:2023/8/29
تاريخ القبول:2023/10/24

حقوق النشر: جامعة دمشق –سورية، يحتفظ المؤلفون بحقوق النشر بموجب الترخيص **CC BY-NC-SA 04**

المجلة، وبتطبيق مقياس recall@ 10 أظهر أنّ 49% من الاقتراحات المقدمة من نظام التوصية طابقت المحكمين الفعليين للأبحاث المختبرة. تظهر أهمية هذا النظام كأداة فعّالة لمساعدة المحريين في المجلات والمؤتمرات العلمية في تصنيف الأبحاث واختيار المحكمين الأنسب بسرعة وفعاليّة ودون تحيّز.

**الكلمات المفتاحية:** أوراق بحثيّة، تصنيف، نموذج لغوي، معالجة اللغات الطبيعية، نظام توصية

# Research-Paper Classification And Reviewer Recommender System Based On NLP And Arabert Linguistic Model

**Razan Hallak[1], Dr. Muhammad Bassam Kurdy[2]**

[1]Syrian Virtual University, Master of Web Technologies MWT,

[2] Higher Institute for Applied Sciences and Technology, Syrian Virtual University, Artificial intelligence specialist,

## Abstract:

Peer-Review system is a tool for controlling the quality of scientific publications in scientific journals and conferences, with huge increase of submissions, the pressure on the review process has increased, which depends basically on human efforts in sorting papers according to its topic and assigning it to the most appropriate reviewer, which raised the need to develop the traditional systems to avoid slowness, bias, or lack of quality and to keep pace with the continuous scientific acceleration. Recently artificial intelligence techniques have been widely used in most fields, and many libraries were developed to analyzing Arabic texts, which facilitated their use for the development of current systems. In this paper, a multi-label classifier was created based on the deep learning pre-trained model for Arabic language, AraBERT, which was customized by training on data from peer-reviewed journal (Syrian Journal of Agricultural Research) after pre-processing using natural language processing techniques. This model has proven its ability to identify the topic using paper's title and abstract only, then it was used in developing a reviewers recommender system, which is information retrieval model works on classifying and processing Arabic text semantically based on the AraBERT Linguistic model, this leads to measure the similarity ratio between a paper and collection of publications by reviewers in the same subject to provide top ten suggestions of reviewers with the closest specialization to the paper. The system also provides additional suggestions as new reviewer among previously published articles. This system was tested on data of the journal, and the classification model reached the values 86.5%, 91.8%, 92.4%, and 91.4% for each of the Accuracy, F1-score, Precision, and Recall, respectively. It was also found that 85% of the reviewers suggested by the system are fully suitable for review process, according to the assessment of the journal's publishing experts, and by applying the recall@10 scale, it showed that 49% of the suggestions submitted by the system matched the actual reviewers. The importance of this system appears as an effective tool to assist editors in scientific journals and conferences in classifying research and selecting the most appropriate reviewers quickly, effectively and without bias.

**Keywords**: Paper, Classification, Linguistic Model, Natural Language Processing, Recommender System

Received: 29/8/2023
Accepted: 24/10/2023

**Copyright:** Damascus University- Syria, The authors retain the copyright under a CC BY- NC-SA

## 1. المقدمة:

يعدّ نظام مراجعة الأوراق البحثية العمود الفقري للنشر الأكاديمي في المجلات والمؤتمرات العلمية وهو الطريقة المثلى لتقييم الأبحاث العلمية، حيث يضمن هذا النظام مراجعة الأبحاث من قبل خبراء مختصين بالمجال العلمي التابع له هذا البحث، يعمل هؤلاء الخبراء على تقييم المحتوى العلمي والتأكد من بناء البحث على أسس علمية صحيحة بالإضافة إلى مراجعة النتائج قبل السماح بنشرها في الأوساط العلمية، ويكون للمحررين دور أساسي في العملية، بدءاً من إجراء المراجعة الأولية للأوراق البحثية الواردة للمجلة للتأكد من مطابقتها لشروط النشر العامة، ومن ثم اختيار المحكمين الأنسب لكل ورقة بحثية، وصولاً إلى إدارة المراسلات والتنسيق الزمني مع كلّ من المحكمين والباحثين، والتحديث المستمر لقوائم المحكمين وحتى إصدار شهادات النشر النهائية (Jana، 2019,152).

تعتمد هذه العملية على الجهد البشري بشكل أساسي لذا فهي عملية مستهلكة للوقت كما أنها معرضة للتحيز وعدم المساواة، ومع ازدياد عدد الأبحاث العلمية المقدمة للنشر بازدياد عدد الجامعات والمؤسسات العلمية والتفرع الكبير في التخصصات العلمية، زاد الضغط على عملية المراجعة بشكل عام. ومع تطور أدوات وتقنيّات الذكاء الصنعي ودمجها بمختلف جوانب الحياة برزت أهميّة استخدامها لتطوير الأنظمة المساعد لعملية مراجعة الأوراق البحثية، حيث استُخدمت في كشف الاستلال وجودة وأصالة الورقة البحثية، كما استُخدمت أنظمة التوصية القائمة على تعلم الآلة والمعتمدة على المحتوى (content-based filtering CBF) وكذلك الأنظمة المعتمدة على الفلترة التعاونية (Collaborative filtering CF) لفرز الأوراق البحثية في المؤتمرات الضخمة (Beel *et al*.، 2016، 305)، واعتُمدت عدة أساليب لربط الأوراق البحثية بالمحكمين المناسبين منها المعتمد على الأنطولوجيا والمظاهر الدلالية (Price *et al*.، 2017، 70)، كما استُخدمت تقنيات تعلم الآلة في تحليل نصوص البحث للتنبؤ بكونه قابل للنشر أم لا (Checco *et al*.، 2021,1)، بينما اعتمدت دراسات أخرى على استخدام الشبكات العصبونية لبناء منظومة التنبؤ بقبول أو رفض البحث (Ghosal *et al*.، 2019، 1120)،

تُستخدم تقنيات معالجة اللغات الطبيعية على نطاق واسع من التطبيقات أهمها تصنيف النصوص والبحث الدلالي وتحليل المشاعر (Ramachandran *et al.*, 2017, 534)، ويمر تحليل النصوص بعدّة مراحل تختلف باختلاف الهدف من التحليل وغنى اللغة، ففي اللغة العربية تتضمن عمليات معالجة اللغات الطبيعية تحليل اللغة الأساسية (BLA) بما فيها التحليل الصرفي والنحوي والإملائي والأنطولوجيا الدلالية، وبناء الموارد (BR) وتحديد اللغة (LI) وتحليل المستوى الدلالي (SemA) ويستخدم في الترجمة الآلية وتحليل المشاعر (Guellil *et al*.، 2021، 497) و(Rashidi *et al*.، 2020).

ظهرت العديد من الأدوات والنماذج التابعة لمفهوم معالجة اللغات الطبيعية أهمها word2vec والنماذج اللغوية مسبقة التدريب (Pre-trained models) والتي استخدمت بشكل موسع في تحليل النصوص دلالياً مثل نموذج BERT، وتمتاز هذه النماذج بأنها مدربة على كمية ضخمة من البيانات وجاهزة لتخصيصها لأداء مهام معينة بجهد قليل وباستخدام كمية قليلة من بيانات التدريب، كما تم تطوير العديد من النماذج المخصصة للغة محددة مثل نموذج AraBERT المدرب على كمية ضخمة من نصوص اللغة العربية (Antoun *et al*.، 2020، 9).

هدف هذا البحث إلى الاستفادة من أدوات وتقنيات الذكاء الصنعي ومعالجة اللغات الطبيعية في تطوير نظام مراجعة الأوراق البحثية العربي في مجلة علمية محكمة عبر بناء نظام تصنيف الأوراق البحثية حسب موضوعها ثم اقتراح أسماء المحكمين الأنسب لها حسب درجة تشابه منشوراتهم مع الورقة البحثية المقدمة للنشر.

## 2 . الدراسات المرجعية :

تتكون منظومة النشر العلمي من عدة مراحل، وتم استخدام أساليب الذكاء الصنعي وتقنيات معالجة اللغات الطبيعية في تطوير العديد من هذه المراحل في العديد من الدراسات، حيث تم إنشاء نظام التوصية الخاص بمجلات Elsevier، والذي يستخدم المقالات المنشورة سابقاً لديه لتحديد المجال العلمي


ISSN:2789-6854 (online)
http://journal.damascusuniversity.edu.sy

لكل مجلة ويقدم اقتراحات بالمجلة الأنسب لنشر ورقة بحثية معينة (Kang *et al*.، 2015، 231)، يستخدم نظام Elsevier معالجة اللغات الطبيعية لتوليد الميزات (features)، وخوارزمية (Okapi BM25 matching) كخوارزمية توصية. ومن الأنظمة المشابهة نظام التوصية Maglet لاقتراح المجلة الأنسب لنشر ورقة بحثية حسب موضوعها وهو خاص بالمجلات الفارسية، وأعطى دقة عالية حيث وُجد أن 78% من الحالات كان الناشر الفعلي هو المقترح الأول للنظام، وفي 89% من الحالات كان الناشر الفعلي من ضمن أول 10 مقترحات للنظام (Mohtaj *et al*.، 2018، 348).

وبما أن عملية ربط الأوراق البحثية بالمحكمين المناسبين من أهم مراحل منظومة المراجعة، والتي تتطلب معرفةً في موضوع كل بحث واختصاص كل محكم وإتاحته، ويقتصر الاختيار في كثير من المجلات على قائمة محددة من المحكمين يتم تحديثها حسب الاختصاصات اللازمة للمجلة، ظهرت عدة دراسات تركز على هذه المرحلة منها إنشاء نظام استرجاع معلومات للبحث عن أبحاث سابقة مشابه للورقة البحثية المدروسة لتحليل كل من الأصالة ونسبة الاستلال بالإضافة إلى اقتراح المحكمين الذين راجعوا هذه الأبحاث كمحكمين محتملين للورقة البحثية الجديدة، كما تم استخدام تحليل الرسم البياني للاقتباس، باستخدام خوارزمية (particle-swarm)، لاقتراح مراجعين محتملين لورقة بحثية على مبدأ أن موضوع البحث يُقيّم أفضل ما يمكن عبر المؤلفين الذين يستشهد بهم (Rodriguez *et al*.، 2008، 319). وفي دراسة أخرى تم بناء نظام اقتراح مراجعين Toronto Paper Matching System (TPMS) في مؤتمر NIPS 2010 وتم إضافته لاحقاً إلى أدوات إدارة المؤتمرات CMT من Microsoft، ويطلب هذا النظام من المحكمين إرفاق أبحاث علمية منشورة سابقاً لهم مع تقارير ومستندات نصية معبرة عنهم ليتم تحليلها وإنشاء الملف التعريفي الخاص بالمحكم، والذي يمكن استخدامه في مؤتمرات أو مجلات أخرى لنفس المحكم، كذلك تم استخدام بيانات الفهرسة ومخططات ربط المؤلفين المشاركين من الأوراق البحثية (Charlin، 2013). وفي دراسات مشابهة استخدمت الفلترة التعاونية للتنبؤ بارتباط محكم– ورقة بحثية وحساب علامته، كما بُني المخطط البياني لمؤلفي المقالات الواردة ضمن مراجع الأبحاث المقدمة للتقييم للمساعدة في اقتراح المراجعين (Price *et al*.، 2017، 70).

استُخدمت العديد من الأساليب في تصنيف النصوص وتحديد موضوعها أو تصنيف الآراء منها المعتمدة على المعرفة (knowledge-based) أو التعلم الآلي أو التعلم العميق أو معالجة اللغات الطبيعية. يتم عادة تمثيل الجمل أو كامل المستند عبر (n-gram features) وتصنيفها باستخدام النماذج المختلفة، وبما أن لأدوات التعلم العميق القدرة على التعلم التلقائي وتمثيل الجمل والنصوص دلالياً دون تدخل يدوي فقد حققت أفضل النتائج، حيث تم بناء نموذج التصنيف C-HAN باستخدام شبكة CNN لاستخراج ميزات النص ونموذج LSTM و attention model لتحليل هذه الميزات وتصنيف الآراء (Cheng *et al*.، 2019، 83)، وخلصت بعض الدارسات إلى أن استخدام البنية ثنائية الاتجاه BiLSTM دون الحاجة لاستخدام شبكات عصبونية معقدة أعطى دقة وقيمة F1-score تضاهي أو تفوق نتائج أدوات state-of-the-art بعد تطبيقها على عدة مجموعات بيانات (Adhikari *et al*.، 2019، 4046).

وبشكل عام تتوفر أدوات لدعم مهام تعيين محكمين للأوراق البحثية في مجلات عديدة منها Machine Learning journal وتدعى MLj Matcher، وفي المؤتمرات مثل SIGKDD تم استخدام أداة مشابهة تدعى SubSift tool، والتي تقوم بالتحري عن عناوين الأبحاث المنشورة من قبل الباحثين (فريق التحكيم) في موقع فهرسة أبحاث علوم الحاسوب DBLP وتجميع عناوينها لإنشاء ملف تعريف لكل محكم، وبعدها يتم إنشاء ملف تعريف لكل ورقة علمية مقدمة للنشر من نص الملخص، ثم يتم حساب زاوية جيب التمام بين الملفين لمعرفة نسبة التشابه وإنشاء قائمة درجات للأبحاث المرتبطة بكل محكم، وقائمة درجات للمحكمين لكل ورقة بحثية (Price *et al*.، 2017، 70).

أما بالنسبة للمجلات التي تنشر باللغة العربية فلا يوجد توظيف لأدوات الذكاء الصنعي إلا بشكل محدود يقتصر غالباً

على كشف الاستلال، ولم يتم العثور على بيانات تحكيم (أسماء محكمين ومراجعات وردود المؤلفين) بشكل متاح في أيّ من منصات النشر العلمي والأكاديمي وإنما اقتصرت هذه المنصات والمجلات على تحكيم ونشر الأبحاث العلمية بالطريقة التقليدية أو إنشاء قواعد بيانات لتجميع مقالات المجلات والمؤتمرات في بلد معين مثلاً، كمنصة Cybrarians البوابة العربية للمكتبات والمعلومات التي تنشر النتاج العلمي للجامعات المصرية، وبالتركيز على بناء الموارد تم تجميع العديد من الدواوين النصية القرآنية والتاريخية ومعالجتها لاستخدامها في تحاليل الانطولوجيا العربية (Belinkov *et al*.، 2016، 1612).

## 3. مواد البحث وطرائقه:

**مجموعة البيانات:**

وهي مجموعة من المقالات المنشورة في المجلة السورية للبحوث الزراعية وعددها 997 مقالة. تنشر المجلة في تسع مواضيع أساسية وهي (الاقتصاد الزراعي، المحاصيل الزراعية، البستنة، الموارد الطبيعية، البيئة، التقانات الحيوية، الإنتاج الحيواني، وقاية النبات، تكنولوجيا الأغذية) وكل مقالة تتبع موضوع أو أكثر أي أن مجموعة البيانات تعتبر متعددة المعلمات (multi-label dataset)، تم معالجة نصوص المقالات أولياً بإزالة تشكيل الحروف والأرقام وعلامات الترقيم وكلمات التوقف، تم تقسيم المقالات إلى بيانات تدريب وبيانات اختبار بنسبة (20/80)، كما تم تقسيم بيانات التدريب إلى قسمين قسم للتدريب وقسم للتقييم أثناء التدريب بنسبة (20/80)، وتم موازنة بيانات التدريب فقط بطريقة (Over sampling) بحسب الموضوع.

أما بالنسبة لمجموعة البيانات المستخدمة في إنجاز نظام اقتراح المحكمين فهي عبارة عن مجموعة من المقالات المنشورة سابقاً في مجلات مختلفة للمحكمين الذين يعملون حالياً في المجلة السورية للبحوث الزراعية، بالإضافة إلى المقالات المنشورة في المجلة، وتم معالجة النصوص أولياً بنفس الطريقة السابقة.

**الأدوات والمكتبات المستخدمة:**

تم إنجاز كافة مراحل البحث باستخدام لغة البرمجة Python 3-9-7 والأداة GoogleColab.

تم استخدام العديد من المكتبات والخوارزميات في هذا البحث أهما (farasa و camel_tools و Regex) لمعالجة النصوص أولياً، والنموذج اللغوي BERT (Bidirectional Encoder Representations from Transformers)، وهو نموذج تعلم عميق مكون من 12 طبقة مرمزات (encoders) يعمل على تمثيل النصوص بكفاءة عالية وتتميز النسخة العربية منه AraBERT بأنها مدربة على حوالي GB77 من النصوص العربية (Antoun *et al*.، 2020، 9)، وتم في هذا البحث تخصيصه (Fine-Tuning) لتصنيف الأوراق البحثية حسب موضوعها عبر تدريبه على بيانات التدريب المتوازنة، وبما أن دخل النموذج لا يتجاوز 512 كلمة لذا تم تقطيع البيانات إلى أجزاء أصغر كالتالي:

- بيانات التدريب 18418 مقطع متوازن من حيث المواضيع، بيانات التقييم 2649 مقطع، بيانات الاختبار 3407 مقطع.
- تمثلت خطوات بناء المصنف متعدد المعلمات (Multi-label classifier) بالخطوات التالية:
- تحديد المعلمات التسع الأساسية (Target)
- استيراد المكتبات اللازمة لبناء صفوف النموذج وأهمها PyTorch ونموذج AraBERT مسبق التدريب من المكتبة Transformers
- بناء صف معالجة البيانات والذي يقوم بتقطيع النص Tokenization وتحويله إلى مصفوفات رقمية PyTorch tensors وتحميل بيانات التدريب إليه.
- بناء مصنف BERT باستخدام النموذج مسبق التدريب مع إضافة طبقتي Dropout و Linear إلى شبكة الخرج.
- تحديد معامل loss function (BCEWithLogitsLoss)
- تحديد المحسن Adam optimizer
- إنشاء وظيفة حساب دقة التصنيف في كل دورة تدريب.
- تدريب النموذج على 20 دورة وحفظ أفضل مصنف.

كما تم بناء نموذج التوصية حسب الهيكلية التالية:

الدخل: هو عبارة عن عنوان وملخص الورقة البحثية المقدمة للتحكيم وأسماء المؤلفين وقائمة المحكمين الحاليين للمجلة.

المعالجة: يتم استخدام المصنف السابق في تحديد موضوع أو مواضيع الورقة البحثية، يلي ذلك تهيئة (courpus) بيانات المحكمين الذين نشروا بمواضيع مشابهة، بالإضافة إلى (courpus) المقالات المنشورة سابقاً في المجلة بمواضيع مشابهة، ثم تضمين النصوص في كلا المجموعتين باستخدام نموذج AraBERT كشعاع موزون عبر المكتبة (sentence transformers) تمهيداً لحساب التشابه بين نص الورقة البحثية وبقية النصوص عبر تشابه جيب التمام (cosine similarity) وترتيب النتائج حسب درجة التشابه.

الخرج: عشر أسماء محكمين مقترحين لتحكيم الورقة البحثية المدروسة، ويمكن للنموذج تقديم عشر أسماء إضافية لمحكمين جدد.

تم تقييم نتائج النماذج السابقة باستخدام بيانات الاختبار وبعدة طرق، فقد استُخدمت الطريقتين (Count Vectorizer و Tf-idf Vectorizer) لتمثيل النص بشعاع موزون، وتمت المقارنة مع نتائج خوارزميات التصنيف التالية: (RandomForest و NaïveBayes و KNeighbors و DecisionTree) باستخدام المقاييس: (Accuracy، F1-score، Precision، Recall)، كما تم حساب المقياس Recall@10 لتقييم أداء نظام التوصية بالمحكمين، بالإضافة لذلك تم تقييم النتائج من قبل خبراء النشر في المجلة.

## 4. النتائج والمناقشة:

أظهر نظام التصنيف والتوصية بالمحكمين نتائج جيدة ودقة عالية في تصنيف المقالات حسب موضوعها، وتم قياس دقة وكفاءة نموذج تصنيف الأوراق البحثية عبر (Confusion matrix)، حيث وصلت دقة النموذج على بيانات التقييم validation set أثناء التدريب إلى 92%.

وأظهر التقييم النهائي قدرة النموذج على التنبؤ بدقة بموضوع البحث سواءً كان مفرد أو متعدد المواضيع، ويعني ذلك وجود موضوع رئيسي وآخر فرعي للبحث فمثلاً في مجال العلوم الزراعية التي تم تدريب نموذج التصنيف في هذا البحث عليها، يمكن لورقة بحثية أن تحلل أثر آفة ما على محصول معين (قسم وقاية النبات) ويتم دراسة جينات هذه الآفة جزيئياً (قسم التقانات الحيوية)، أو قد يركز البحث على دراسة تربية نوع محدد من المواشي (قسم الإنتاج الحيواني) مع تحليل العناصر الغذائية في حليب تلك المواشي (قسم تكنولوجيا الأغذية)، ما يعني أن البحث يتناول موضوعين مختلفين ومتكاملين، ما يترتب عليه فيما بعد تعيين محكمين باختصاصات موافقة لجميع المواضيع المطروحة في البحث. ويبين الجدول (1) نتائج التقييم النهائي لنموذج التصنيف المقترح على بيانات الاختبار test set.

**الجدول (1) مقاييس دقة نموذج التصنيف لكل موضوع**

| | precision | recall | f1-score | support |
|---|---|---|---|---|
| **Agricultural Economics** | 0.99 | 0.97 | 0.98 | **501** |
| **Biotechnology** | 0.88 | 0.85 | 0.87 | **153** |
| **Crop production** | 0.94 | 0.78 | 0.85 | **596** |
| **Environmental Sciences** | 0.92 | 0.91 | 0.91 | **486** |
| **Food Technology** | 0.93 | 0.94 | 0.94 | **464** |
| **Horticulture** | 0.86 | 0.94 | 0.90 | **714** |
| **Livestock production** | 0.94 | 0.97 | 0.96 | **219** |
| **Natural Resources** | 0.88 | 0.95 | 0.91 | **786** |
| **Plant Protection** | 0.96 | 0.92 | 0.94 | **471** |
| | | | | |
| **micro avg** | 0.92 | 0.92 | 0.92 | **4390** |
| **macro avg** | 0.92 | 0.91 | 0.92 | **4390** |
| **weighted avg** | 0.92 | 0.92 | 0.92 | **4390** |
| **samples avg** | 0.93 | 0.93 | 0.92 | **4390** |
| | | | | |
| **Model Accuracy** | | | | **0.87** |
| **Total F1_score** | | | | **0.92** |
| **Total Precision** | | | | **0.92** |
| **Total Recall** | | | | **0.91** |

تم مقارنة أداء مصنف AraBERT مع أداء عدة خوارزميات تصنيف أخرى على نفس مجموعة البيانات وباعتماد طريقة multi-class classification أي التنبؤ بالموضوع الرئيسي للورقة البحثية فقط للتبسيط ومع ذلك تفوق مصنف AraBERT بكافة المقاييس على بقية المصنفات الجدول (2)، إضافة لأنه يتنبأ بالموضوع الرئيسي والثانوي للبحث.

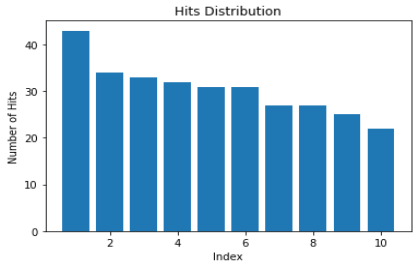
**الجدول (2) مقارنة بين مصنف AraBERT وخوارزميات تصنيف أخرى**

| | Ara BERT | Random Forest | Naive Bayes | KNN | Id3 |
|---|---|---|---|---|---|
| **Accuracy** | **0.87** | 0.87 | 0.8 | 0.67 | **0.66** |
| **Precision** | **0.92** | 0.88 | 0.83 | 0.68 | **0.65** |

| | | | | | |
|---|---|---|---|---|---|
| **Recall** | **0.91** | 0.87 | 0.8 | 0.67 | **0.66** |
| **F1-score** | **0.92** | 0.87 | 0.8 | 0.67 | **0.64** |

توافقت هذه النتائج مع دراسات سابقة توصلت إلى تفوق نماذج التعلم العميق على غيرها من خوارزميات التعلم الآلي في تحليل نصوص اللغة العربية واستخراج السمات وتصنيفها والبحث الدلالي ضمنها وتحليل المشاعر كما في (Alharbi *et al*.، 2018) و (Dahou *et al*.، 2016، 2418) وبعضها حقق أداءً عالياً وصل في نموذج تصنيف المشاعر إلى 91.1% لمقياس F1-score (Guellil *et al*., 2021, 497).

وفي مجال التوصية بالأوراق البحثية، تعتمد معظم أنظمة التوصية على الفلترة بناءً على المحتوى CBF-based، في حين يستخدم القليل منها الفلترة التعاونية CF-based وطرق أخرى، وتعد خوارزمية TF-IDF من أكثر الخوارزميات استخداماً في هذا المجال (Beel *et al*.، 2016، 305). وتم في هذا البحث إنشاء نظام توصية بالمحكمين يعتمد على الفلترة حسب المحتوى، وتم تمثيل النص دلالياً باستخدام النموذج اللغوي المدرب AraBERT بدلاً من استخدام خوارزمية TF-IDF الشائعة ما ساعد في إعطاء نتائج أقرب ما يمكن للواقع، وتم اختبار نموذج التوصية على مجموعة من المقالات المنشورة في المجلة المحكّمة، وتم مقارنة أفضل عشر توصيات ناتجة مع قائمة المحكمين الحقيقيين لكل مقالة وتطبيق المقياس Recall@10 والذي أعطى القيمة 0.49 أي أن في 49% من الحالات كان المحكمين الفعليين ضمن أول عشر مقترحات من قبل النظام، وبما أن مجموعة المحكمين في مجلة ما تتغير باستمرار حسب توافر المحكمين أو انشغالهم فيتم إضافة محكمين جدد وإزالة آخرين، هذا يعني أن المقالات المنشورة قبل ثلاث سنوات لم يكن لديها نفس قائمة المحكمين المتاحين للمقالات المنشورة هذا العام وهذا ما يفسر عدم توافق نتائج التوصية بشكل كامل مع النتائج الحقيقية، لذلك تم تقييم نتائج التوصية بالمحكمين من قبل خبراء النشر في مجلة البحوث الزراعية وهم المستخدم الفعلي والمستفيد الحقيقي من هذا النموذج، وبينت نتائج التقييم أن 85% من المقترحات الناتجة عن النظام هي مناسبة تماماً لتحكيم البحث المدروس علماً أن لكل مقالة محكمين أو ثلاثة محكمين على الأكثر ومع ذلك فإن النظام يقدم ثماني إلى تسع مقترحات مناسبة مع نسب التشابه المحسوبة، ويبين الشكل (1) توزع التوصيات الصحيحة ضمن أول عشر مقترحات لنظام التوصية لمجموعة من المقالات.




**الشكل (1) توزع التوصيات ضمن أفضل 10 مقترحات لنظام التوصية بالمراجعين**

يقدم النظام ميزة اقتراح عشر محكمين إضافيين من خارج قائمة المحكمين الحاليين للمجلة وذلك عن طريق البحث عن مشابهات نص البحث المدروس مع الأبحاث المنشورة سابقاً في المجلة مع الأخذ بعين الاعتبار تصنيف الأبحاث حسب مصنف AraBERT وكذلك الدرجة العلمية للباحثين لاختيار محكمين جدد باختصاصات دقيقة أقرب ما يمكن للبحث المدروس.

## 5. الخلاصة:

1) تم في هذا البحث إنشاء نظام توصية بالمحكمين يعتمد على الفلترة بناءً على المحتوى، يصنف الأوراق البحثية حسب موضوعها باستخدام عنوان وملخص البحث فقط، ويقترح المحكمين الأنسب لتحكيمه حسب اختصاصهم وتشابه منشوراتهم مع البحث المدروس.

2) يمثل النموذج المقترح نظام مساعد لهيئة التحرير في المجلات والمؤتمرات العلمية يساهم في تسهيل وتسريع مهامهم باختيار المحكم الأنسب لكل ورقة بحثية مهما كان موضوعها، وهو نموذج فريد على المستوى العربي فلا يوجد مجلات عربية تستخدم نظام مشابه بعد.

3) إن أهمية استخدام طرق التصنيف متعدد المعلمات Multi-label classification في تصنيف الأوراق البحثية تنبع من حقيقة أن الأبحاث العلمية غالباً ما تتناول أكثر من موضوع مختلف ومتكامل في نفس الوقت وبالتالي فهي تحتاج لمحكمين باختصاصات مختلفة تغطي المواضيع المستخدمة.

4) تتفوق النماذج اللغوية الحديثة مسبقة التدريب على الأساليب السابقة في تمثيل النصوص دلالياً، ويمكن

تخصيصها لأداء مهام محددة بكمية قليلة من بيانات التدريب وبجهد مقبول وإمكانيات متوفرة.

## References


1. Adhikari, A., Ram, A., Tang, R., & Lin, J. (2019). Rethinking complex neural network architectures for document classification. *In Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 1*, pp. 4046-4051.
2. Alharbi, R., Magdy, W., Darwish, K., Abdelali, A., & Mubarak, H. (2018). Part-of-speech tagging for Arabic Gulf dialect using Bi-LSTM. *In Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018).*
3. Antoun, W., Baly, F., & Hajj, H. (2020). Arabert: Transformer-based model for arabic language understanding. *arXiv preprint arXiv:2003.00104*, 9.
4. Beel, J., Gipp, B., Langer, S., & Breitinger, C. (2016). Research-paper recommender systems: a literature survey. *International Journal on Digital Libraries, 17*(4), 305-338.
5. Belinkov, Y., Magidow, A., Romanov, M., Shmidman, A., & Koppel, M. (2016). Shamela: a large-scale historical arabic corpus. *arXiv preprint arXiv*, (p. 1612.08989).
6. Charlin, L., & Zemel, R. (2013). he Toronto paper matching system: an automated paper-reviewer assignment system.
7. Checco, A., Bracciale, L., Loreti, P., Pinfield, S., & Bianchi, G. (2021). AI-assisted peer review. *Humanities and Social Sciences Communications, 8*(1), 1-11.
8. Cheng, Y., Ye, Z., Wang, M., & Zhang, Q. (2019). Document classification based on convolutional neural network and hierarchical attention network. *Neural Network World, 29*(2), 83-98.
9. Dahou, A., Xiong, S., Zhou, J., Haddoud, M. H., & Duan, P. (2016). Word embeddings and convolutional neural network for arabic sentiment classification. *In Proceedings of coling 2016, the 26th international conference on computational linguistics: Technical papers*, (pp. 2418-2427).
10. Ghosal, T., Verma, R., Ekbal, A., & Bhattacharyya, P. (2019). Deepsentipeer: Harnessing sentiment in review texts to recommend peer review decisions. *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, (pp. 1120-1130). Florence, Italy,.
11. Guellil, I., Saâdane, H., Azouaou, F., Gueni, B., & Nouvel, D. (2021). Arabic natural language processing: An overview. *Journal of King Saud University–Computer and Information Sciences, 33*(5), 497-507.
12. Jana, S. (2019). A history and development of peer-review process. *Annals of Library and Information Studies, 66*, 152-162.
13. Kang, N., Doornenbal, M. A., & Schijvenaars, R. J. (2015). Elsevier journal finder: recommending journals for your paper. *In Proceedings of the 9th ACM Conference on Recommender Systems*, (pp. 261-264).
14. Mohtaj, S., & Tavakkoli, F. (2018). Maglet: a Persian Journal Recommender System. *In 2018 9th International Symposium on Telecommunications (IST)* (pp. 348-352). IEEE.
15. Price, S., & Flach, P. A. (2017). Computational Support for Academic Peer Review: A Perspective from Artificial Intelligence. *Communications of The ACM, 60*, 70-79.
16. Ramachandran, L., Gehringer, E. F., & Yadav, R. K. (2017). Automated assessment of the quality of peer reviews using natural language processing techniques. *International Journal of Artificial Intelligence in Education, 27*(3), 534-581.
17. Rashidi, K., Sotudeh, H., Mirzabeigi, M., & Nikseresht, A. (2020). Determining the informativeness of comments: a natural language study of F1000Research open peer review reports. *Online Information Review*.
18. Rodriguez, M. A., & Bollen, J. (2008). n algorithm to determine peer-reviewers. *In Proceedings of the 17th ACM conference on Information and knowledge management*, (pp. 319-328).