مجلة الآداب والعلوم الاجتماعية Journal of letters and Social Sciences (JLSS)
https://www.asjp.cerist.dz/en/PresentationRevue/4

الذكاء الاصطناعي في التحليل الآلي للمشاعر واكتشاف العواطف في النصوص العربية

Artificial intelligence-based automated sentiment analysis and emotion detection in arabic text

د. خديجة مدني [1]
Dr. Khedidja Medani
جامعة محمد لمين دباغين سطيف2، الجزائر
k.medani@univ-setif2.dz

د. إيمان زنبوط
Dr. Imene Zenbout
مخبر الشبكات والأنظمة الموزعة
جامعة فرحات عباس سطيف1، الجزائر
imene.zenbout@univ-setif.dz

معلومات حول المقال
تاريخ الاستلام 28-01-2025
تاريخ القبول 19-03-2025

الكلمات المفتاحية
المعالجة الآلية للغة الطبيعية
PLN
تحليل المشاعر
تنقيب الآراء
اكتشاف العواطف
اللغة العربية

ملخص

تستعرض هذه الدراسة دور الذكاء الاصطناعي في تحليل المشاعر واكتشاف العواطف في النصوص العربية، من خلال مراجعة منهجية شاملة للأدبيات العلمية ذات الصلة. يركز البحث على التحديات الخاصة باللغة العربية، بما في ذلك تعقيداتها الصرفية والنحوية وتعدد لهجاتها، ويستعرض التطبيقات المختلفة لهذه التقنيات في مجالات متنوعة، مثل التسويق، والرعاية الصحية، والتحليل السياسي والاجتماعي. كما يناقش الأدوات والخوارزميات المستخدمة في معالجة النصوص العربية، مع إبراز مدى فعاليتها في هذا المجال. أخيرًا، يقدم البحث توصيات للباحثين والمطورين لتحسين أداء التقنيات المستخدمة في هذا المجال، بما يتماشى مع أحدث التطورات الحديثة في معالجة اللغة الطبيعية (gnissecorP egaugnaL larutaN :PLN) والذكاء الاصطناعي (ecnegilletnI laicifitrA :IA).

مقدمة

يُعنَى تحليل المشاعر(SA: Sentiment Analysis)، والمعروف أيضا بالتنقيب عن الآراء (Opinion mining)، بدراسة آراء الناس -حاسوبيا- واتجاهاتهم المعبر عنها عادة في النصوص حول مواضيع مختلفة، وهو يشمل عدة أنواع؛ في حين أن تحليل المشاعر الأكثر شيوعا يهتم بتصنيف قطبية المشاعر إلى إيجابية، سلبية، أو محايدة لتحديد درجات/مستويات الرضا وقياس الرأي العام حول موضوع أو منتج معين على مواقع التواصل الاجتماعي مثلا، فإن تحليل المشاعر الدّقيق يهدف إلى تقييم النص حسب مقياس معين (من 1 إلى 5 نجوم) وهو ما يمكن استخدامه لتقييم التطبيقات والأفلام والروايات على سبيل المثال. من جهة أخرى، يسعى تحليل المشاعر بناء على الجوانب إلى الحصول على تحليلات حول جوانب أو نقاط معينة لموضوع ما، كالحصول على ملاحظات حول تقييم المستخدمين لتطبيق التنقل «يسير» من جانب الأمان والحفاظ على خصوصية المشتركين. أخيرا، يهدف تحليل المشاعر بناء على العواطف أو ما يعرف باكتشاف العواطف (Emotion Detection) إلى الكشف عن نبرات وانفعالات معينة في النص مثل الخوف والغضب والفرح والحزن والاشمئزاز وغيرها. فإذن، الكشف عن العواطف يعد مقاربة هامة توفر فهما لمشاعر المستخدمين وتوجهاتهم وآرائهم، الأمر الذي يكتسي أهمية متزايدة في العديد من التطبيقات مثل التسويق والبحث الاجتماعي والتحليل السياسي والأدبي.

1-أهمية البحث

في الآونة الأخيرة، أدى دمج تقنيات الذكاء الاصطناعي (AI: Artificial Intelligence) في مجال تحليل المشاعر واكتشاف العواطف إلى إحداث نقلة نوعية، مما أسهم في تحسين دقة التحليل واستخراج رؤى أعمق من كميات هائلة من البيانات النصية. وأصبح هذا المجال يكتسي أهمية متزايدة نظرًا للدور المحوري الذي يؤديه في تطبيقات المعالجة الآلية للغة الطبيعية (NLP: Natural Language Processing)، حيث

1 المؤلف المراسل

مكّن الباحثين والممارسين من تطوير نماذج قادرة على فهم النصوص البشرية بشكل أكثر دقة وفعالية.

من جهة أخرى، تكتسي اللغة العربية أهمية بالغة كونها تحتل المركز الرابع كأكثر لغة استعمالا وانتشارا في العالم، لاسيما على منصات التواصل الاجتماعي، حيث يتم إنتاج محتوى ضخم من قِبل المستخدمين على مدار الساعة. ونتيجة لذلك، توفر النصوص العربية بيئة غنية وأرضا خصبة لتحليل المشاعر واكتشاف العواطف، إلا أن خصائص اللغة العربية الغنية مورفولوجيا وقواعدها النحوية وتعدد لهجاتها تطرح تحديات جدية تفرض على الباحثين أخذها بعين الاعتبار. هذه التحديات تستدعي تطوير تقنيات متقدمة في المعالجة الآلية للغة الطبيعية وخوارزميات الذكاء الاصطناعي لتكييفها مع خصوصيات اللغة العربية. على غرار هذه التحديات، فقد نمت أهمية تحليل المشاعر واكتشاف العواطف في النصوص العربية المعتمدة على تقنيات الذكاء الاصطناعي، مما أدى إلى زيادة الأبحاث التي تركز على تكييف وتطوير هاته التقنيات لمواكبة ما وصلت إليه اللغات الطبيعية الأخرى من تقدم ونضوج مقارنة باللغة العربية في هذا المجال. وتشمل تطبيقات هذه التقنيات مجالات واسعة تمتد من الأعمال والتسويق إلى الرعاية الصحية والتحليل السياسي والأدبي (Abo et al., 2019; Alsayat & Elmitwally, 2020).

2-أهداف البحث

يهدف هذا البحث إلى القيام بمسح منهجي للأدبيات والوقوف على الحالة الحالية لتحليل المشاعر والكشف عن العواطف المعتمد على تقنيات الذكاء الاصطناعي في النصوص العربية. كما يسعى هذا البحث إلى استكشاف التقنيات والموارد والتحديات الخاصة التي تفرضها الخصائص الفريدة للغة العربية والسياق الثقافي الذي تُستخدم فيه. علاوة على ذلك، يساهم هذا البحث في مساعدة الباحثين والمطورين على حد سواء على فهم هذا المجال وتحديد الاتجاهات المستقبلية لهم.

3-أسئلة البحث

من خلال هذا البحث، نسعى إلى الإجابة على الأسئلة التالية:

-ما هي التطبيقات الرئيسية لتحليل المشاعر واكتشاف العواطف المعتمدة على الذكاء الاصطناعي باللغة العربية؟

-ما هي المنهجيات والتقنيات الحالية المستخدمة لتحليل المشاعر واكتشاف العواطف في اللغة العربية؟

-ما هي الموارد التي تم تطويرها لدعم أبحاث تحليل المشاعر والكشف عن العواطف باللغة العربية، وما مدى فعاليتها؟

-ما هي التحديات الفريدة التي تطرحها اللغة العربية لتحليل المشاعر والكشف عن العواطف وكيف حاول الباحثون معالجة هذه التحديات؟

-ما هي الاتجاهات البحثية المستقبلية والمجالات المحتملة للتحسين في تحليل المشاعر واكتشاف العواطف باللغة العربية ؟

للإجابة على هذه الأسئلة، يغطي هذا البحث المقالات والأعمال الأكاديمية التي تم نشرها بين عامي 2019 و2024 والتي تركز على تحليل المشاعر واكتشاف العواطف باستخدام الذكاء الاصطناعي في النص العربي.

تم تنظيم هيكل هذا البحث -كما هو موضح في الشكل التوضيحي 1 أدناه- على النحو التالي:

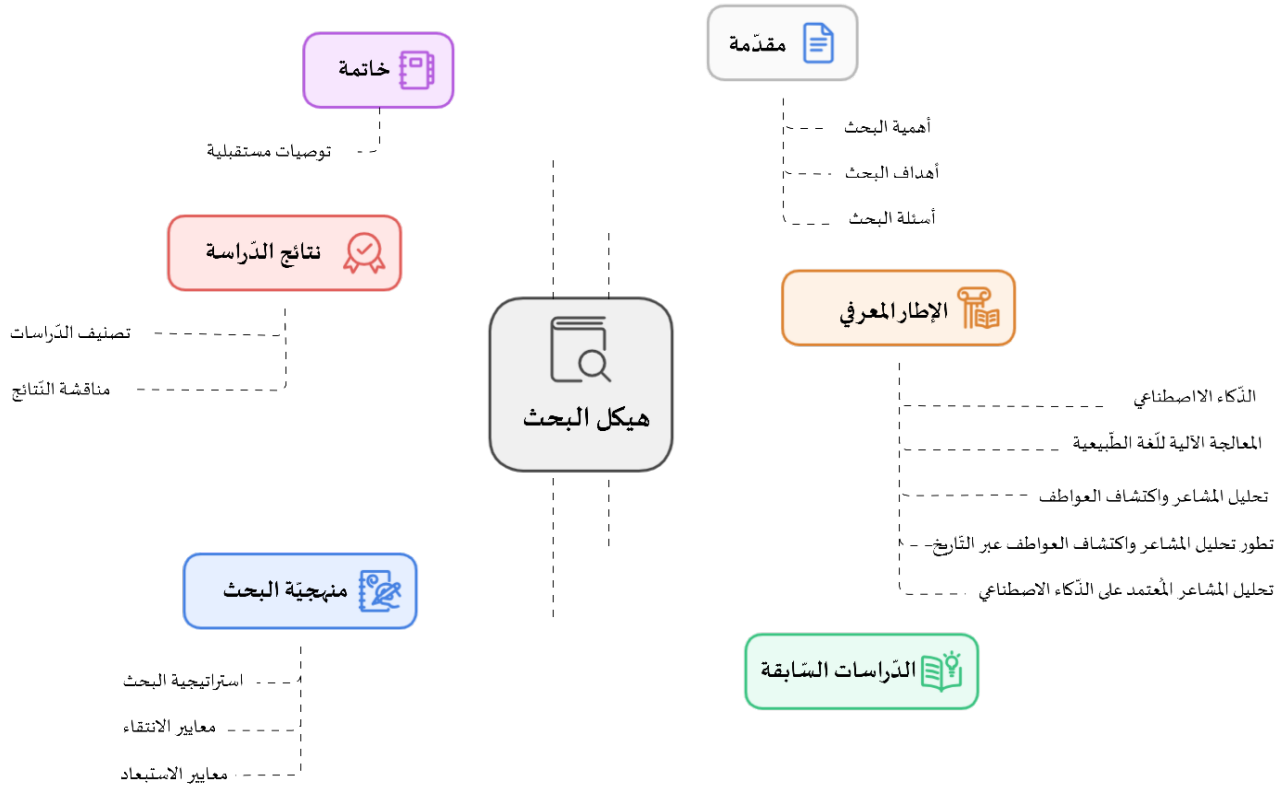
يصف القسم 2 الإطار المعرفي للدراسة مرورا بمفهوم تحليل المشاعر واكتشاف العواطف والمفاهيم المرتبطة به كالذكاء الاصطناعي والمعالجة الآلية للغات الطبيعية وصولا الى التطور التاريخي لتحليل المشاعر واكتشاف العواطف عبر التاريخ.

يعرض القسم 3 الأعمال السابقة المتعلقة بمراجعة أدبيات تحليل المشاعر واكتشاف العواطف في النص العربي.

يصف القسم 4 المنهجية المستخدمة في إجراء هذه الدراسة، بما في ذلك استراتيجية البحث ومعايير الإدراج والاستبعاد وعملية استخراج البيانات.

يناقش القسم 5 الحلول المقترحة الرئيسية لتحليل المشاعر العربية واكتشاف العواطف بغية الإجابة على أسئلة البحث.

أخيرا، يختتم القسم 6 المراجعة بتلخيص النتائج الرئيسية وتقديم توصيات للأبحاث المستقبلية.



الشكل (01): هيكل البحث

4-الإطار المعرفي للدراسة

4-1-الذكاء الاصطناعي

صاغ مصطلح الذكاء الاصطناعي لأول مرة جون مكارثي في عام 1955، الذي وصفه بأنه" علم وهندسة صناعة الآلات الذكية". ويُعرّف الذكاء الاصطناعي بأنه قدرة الآلة على محاكاة السلوك البشري الذكي. وهو يشمل مجموعة من التقنيات التي تمكّن الحواسيب والأنظمة من أداء المهام التي تتطلب عادةً ذكاءً بشريًا، مثل التفكير والتعلم وحل المشكلات والإدراك وفهم اللغة (Manning, 2020).

تمتد تطبيقات الذكاء الاصطناعي إلى مجالات وميادين عدة، مثل الصحة والنقل والتعليم والتجارة وغيرها. ومؤخرا، يشهد الذكاء الاصطناعي وتطبيقاته تقدما كبيرا، خاصةً مع تطور تقنيات تعلم الآلة (ML: Machine Learning) والشبكات العصبية (Neural Networks) والتعلم العميق (DL: Deep Learning). إن إنشاء خوارزميات تسمح لأجهزة الحاسوب بالتعلم وتوقع الأشياء بناءً على البيانات هو الهدف الرئيسي لتعلم الآلة، وهو من أهم فروع الذكاء الاصطناعي. وتوجد ثلاث فئات أساسية له: التعلم الخاضع للإشراف (Super-vised Learning)، والتعلم غير الخاضع للإشراف (Unsuper-vised Learning)، والتعلم المعزز (Reinforcement Leaning).

بينما يقوم التعلّم الخاضع للإشراف بتدريب النماذج على البيانات المُعنونة، يبحث التعلّم غير الخاضع للإشراف عن الأنماط في البيانات غير المُعنونة. في حين أن التعلم المعزز يتيح للوكيل اكتشاف أفضل مسار عمل يزيد من المكافآت من خلال التجربة والخطأ.

في حين تتكون الشبكات العصبية من عقد أو خلايا عصبية مترابطة تعالج المعلومات في طبقات، وهي مصممة على غرار بنية الدماغ البشري. هذه الشبكات ممتازة في التعرف على الأنماط المعقدة، مما يجعلها مثالية لمجموعة واسعة من تطبيقات الذكاء الاصطناعي، بما في ذلك التعرف على الكلام والصور. تُستخدم الشبكات العصبية العميقة ذات الطبقات المتعددة في التعلم العميق، وهو فرع متخصص من تعلم الآلة، لتفسير البيانات بمستويات تجريدية عالية. وقد أسفرت هذه الاستراتيجية عن تطورات مهمة في مجالات مثل معالجة اللغات الطبيعية والرؤية الحاسوبية (Computer Vi-sion)، مما سمح للآلات بتنفيذ مهام كان يُعتقد في السابق أنها تقتصر على الذكاء البشري.

4-2-المعالجة الآلية للغة الطبيعية

المعالجة الآلية للغة الطبيعية، هي مجال متعدد التخصصات يجمع بين الذكاء الاصطناعي واللسانيات وعلوم الحاسوب، يتيح لأجهزة الحاسوب فهم اللغة البشرية وتفسيرها وإنتاجها. وهو يغطي مجموعة متنوعة من المهام، مثل التعرف على الكلام وتصنيف النصوص وفهم اللغة الطبيعية (NLU: Natural Language Understanding) وتوليدها (NLG: Natural Language Generation). وقد شهدت المعالجة الآلية للغة الطبيعية ثورة كبيرة بفضل التطورات الأخيرة في خوارزميات تعلم الآلة، وخاصة في مجال التعلم العميق، بالإضافة إلى تزايد البيانات على شبكة الانترنت ومواقع التواصل الاجتماعي، والتي جعلت من الممكن إنشاء نماذج معقدة بشكل متزايد يمكنها فهم معاني النص المدخل وتقديم مخرجات متماسكة ومناسبة للسياق.

4-3-تحليل المشاعر واكتشاف العواطف في النص

4-3-1-تحليل المشاعر في النص

تحليل المشاعر هو عملية رقمية والمعروف أيضًا باسم" التنقيب عن الرأي"، لتقييم وتصنيف النبرة العاطفية الكامنة وراء نص ما. تهدف هذه التقنية إلى تحديد ما إذا كانت المشاعر المعبر عنها في النص إيجابية أو سلبية أو محايدة. ويُستخدم على نطاق واسع لتحليل الآراء والعواطف والمواقف المنقولة في أشكال مختلفة من التواصل، مثل مراجعات العملاء ومنشورات وسائل التواصل الاجتماعي والردود على الاستبيانات.

ويستفيد تحليل المشاعر من الذكاء الاصطناعي ومعالجة اللغة الطبيعية لتوفير رؤى متسقة وغير متحيزة حول آراء العملاء، مما يساعد المؤسسات على التخفيف من التحيزات الذاتية المتأصلة في التحليل البشري. وعلى الرغم من ارتباطه عادةً بالأعمال التجارية، إلا أن تطبيقاته تمتد لتشمل الرعاية الصحية والتحليل السياسي والبحوث الاجتماعية

4-3-2-اكتشاف العواطف في النص

الكشف عن العواطف هي عملية تحديد وتصنيف المشاعر الإنسانية التي يتم التعبير عنها من خلال أشكال مختلفة من التواصل، خاصةً في النصوص. حيث يتضمن استخدام تقنيات متقدمة مثل التعلم الآلي ومعالجة اللغة الطبيعية لتحليل البيانات النصية واستخلاص رؤى حول الحالات العاطفية للأفراد.

بعبارات أبسط، يهدف اكتشاف العواطف إلى التعرف على المشاعر المختلفة مثل الفرح والحزن والغضب والإحباط من المحتوى المكتوب. تُعد هذه العملية حاسمة في فهم التفاعلات البشرية وتحسين تجارب المستخدمين في مختلف التطبيقات، بما في ذلك تحليل ملاحظات العملاء، ومراقبة الصحة العقلية، وتعزيز التواصل في روبوتات الدردشة (Nandwani & Verma, 2021).

4-3-3-الفرق بين تحليل المشاعر واكتشاف العواطف في النص

تحليل المشاعر واكتشاف العواطف هما تقنيتان تستخدمان في معالجة اللغة الطبيعية لفهم المشاعر والمواقف البشرية المعبر عنها في النص، ولكن كل منهما يركز على جوانب مختلفة. حيث يهدف تحليل المشاعر إلى تحديد المشاعر أو الرأي العام المعبر عنه في النص. إذ عادةً ما يصنف النص على أنه إيجابي أو سلبي أو محايد بينما يسعى اكتشاف العواطف إلى تحديد العواطف المعبر عنها في النص، مثل السعادة، الغضب، الحزن، الخوف، الاشمئزاز، أو الدهشة.

تعمل تقنية تحليل المشاعر بشكل عام على مستوى أوسع، ملخصة الشعور العام للنص بدلاً من التعمق في العواطف والانفعالات المحددة. في حين تعمل تقنية اكتشاف العواطف على مستوى أكثر دقة، حيث تهدف الى تحديد مجموعة واسعة من العواطف التي قد تتواجد في النص الواحد. إذا أخذنا بعين الاعتبار النص التالي «يا للعجب! هذا العنكبوت عملاق! لا أصدق أنني أرى هذا! أبعدوه عني فورًا، لا أستطيع تحمله!» فالتقنية الاولى ستقوم بتصنيف النص على أنه سلبي بينما التقنية الثانية ستقوم باستكشاف العواطف بين سطور النص في هذه الحالة الخوف الهلع والاشمئزاز.

الشكل(02): تحليل المشاعر واكتشاف العواطف في النص

4-4-تطور تحليل المشاعر واكتشاف العواطف في النص عبر التاريخ

تحليل المشاعر هو علم يتعلق بفهم وجود الإنسان، آرائه، تفكيره، وكيفية تعامله مع مختلف مجالات الحياة ومواقفها. على الرغم من أن هذا العلم لم يكن معروفًا بمسمّاه الحالي عبر التاريخ، إلا أن له جذورًا عميقة في الفكر الفلسفي وتطور الحياة الاجتماعية والفكريّة للإنسان. يمكن تتبّع تاريخه عبر مراحل مختلفة تعكس تطوره عبر الزمن بتقسيم تطور تحليل المشاعر واستكشاف العواطف إلى أربع حقب زمنية رئيسية، تبدأ من العصور اليونانية (كما هو موضح في الشكل 3 أدناه). في تلك الفترة، ارتبط هذا العلم بمفهوم «الدُّوكْسَا»، وهو مصطلح يوناني يشير إلى المعتقدات أو الآراء الشائعة. أفلاطون اعتبر «الدُّوكْسَا» ذات دلالة سلبية، حيث ترتبط بإدراكات أو معتقدات لا تستند إلى العقل أو المعرفة. بينما رأى أرسطو أن هذه الآراء، على الرغم من أنها قد لا تؤدي دائمًا إلى الحقيقة، ضرورية في تكوين المعرفة الجماعية واتخاذ القرارات.

استمر التطور الفلسفي في دراسة الإدراك البشري ودور المشاعر والعواطف في اتخاذ القرارات والأحكام الأخلاقية حتى القرن العشرين، الذي شهد نقلة نوعية في المدارس النفسية، مما أدى إلى تحوّل تحليل المشاعر من مقاربة فلسفية نظرية إلى مقاربة نفسية تجريبية. كان لذلك أثر كبير على نظريات العاطفة التي قدمها فرويد ويونغ، واللذان ساعدا في فهم العواطف وتأثيرها على اللغة.

في أواخر الستينيات، تطورت اللسانيات الحاسوبية، حيث ظهرت أدوات مثل «المستفسر العام» لتحليل النصوص من حيث المشاعر. كانت هذه الأدوات حاسمة في تحديد الأنماط العاطفية في النصوص، وأسست للبحوث المستقبلية في هذا المجال. ومع انتشار الرقمنة والتطور الكبير في علوم الحاسوب في نهاية القرن العشرين، أصبحت الأبحاث المتعلقة بتحليل المشاعر أكثر تطورًا. كما تم ابتكار خوارزميات التعلم الآلي لاستخلاص وتحليل المشاعر.

مع استفحال الرقمنة وعلوم الحاسوب في نهاية القرن العشرين وسيادة العصر الرقمي على مجالات الحياة المختلفة، شهدت البحوث المتعلّقة بتحليل المشاعر تطوّرا سريعا زاده توسعا استحداث تكنولوجيا إنترنت متاحة للجميع بعد أن كانت حكرا على المؤسسات الدفاعية. كذلك شهدت هذه الفترة تقدما ملحوظا في مجال التعلم الآلي إذ عمل المطورون على استحداث خوارزميات التعلم الآلي لاستنباط المشاعر وتحليلها. في الفترة بين 2000 و2010، شهد تحليل المشاعر تطورًا ملحوظًا بفضل التقدم في تقنيات الذكاء الاصطناعي وتزايد استخدام الإنترنت ووسائل التواصل الاجتماعي. إذ وُسِمَت هذه الفترة بتقديم ناسوكا ويي مصطلح «تحليل المشاعر» سنة 2003، الذي يشير إلى استخراج المشاعر من النصوص. تزامن ذلك مع استخدام ديف لمصطلح «تنقيب الآراء»، مما عزز من قدرة الأنظمة على تحليل كميات ضخمة من البيانات النصية. في هذا السياق، تم تطوير قاعدة بيانات SentiWordNet التي تصنّف الكلمات بناءً على محتواها العاطفي (إيجابي، سلبي، أو محايد). ومع انتشار وسائل التواصل الاجتماعي، أصبحت هذه الأدوات ضرورية لفهم الرأي العام في الوقت الفعلي. منذ عام 2010، تطور تحليل المشاعر بشكل كبير بفضل التعلم العميق، حيث ظهرت تقنيات مثل BERT وGPT، التي تعتمد على نماذج اللغة لتحسين الفهم السِّياقي للنصوص. بفضل هذه التطورات، أصبحت الأنظمة قادرة على التعامل مع الفروق الدقيقة في المشاعر والمشاعر المختلطة التي كانت تشكل تحديًا للنماذج التقليدية. في عام 2017، أحدثت تقنيات المحولات (Transformers) ثورة في تحليل النصوص من خلال

تحسين الفهم السياقي للنصوص على نطاق واسع. نماذج مثل GEMINI وChatGPT، التي تعتمد على تلك المحولات، أصبحت أدوات قوية لفهم المشاعر وتحليل النصوص على منصات التواصل الاجتماعي.

شكل (03): تطور تحليل المشاعر واكتشاف العواطف في النص عبر التاريخ

4-5-تحليل المشاعر واكتشاف العواطف المعتمد على الذّكاء الاصطناعي

يعتمد تحليل المشاعر في اللغة العربية القائم على الذكاء الاصطناعي على استخدام تقنيات التعلّم الآلي والتعلّم العميق، لتحديد المشاعر المعبر عنها في نص عربي بشكل آلي. ويمكن تقسيم عملية تحليل المشاعر في النصوص العربية المعتمدة على الذكاء الاصطناعي إلى ستة مراحل رئيسية، نلخصها في الآتي (Alsayat & Elmitwally, 2020):

-جمع البيانات (Data collection): في هذه المرحلة، يتم جمع البيانات النصية من مصادر مختلفة، منها منصات التواصل الاجتماعي (تويتر وفيسبوك)، والمراجعات عبر الإنترنت، والمقالات الإخبارية، وملاحظات العملاء وغيرها. وتهدف هذه العملية إلى تكوين قاعدة بيانات يتم تدريب نماذج الذّكاء الاصطناعي عليها. توجد مجموعة من قواعد البيانات المتاحة على شبكة الإنترنت التي يمكن استخدامها، مثل قاعدة بيانات تغريدات المشاعر العربية (ASDT: Arabic Sentiment Tweets Dataset) ومعجم المشاعر العربية (ArSenL: Arabic Sentiment Lexicon).

-المعالجة الأولية للبيانات (Data preprocessing): تتضمن هذه المرحلة عدة عمليات. عملية التوكنة (Tokenization) أو الترميز، وفيها يتم تقسيم النص إلى كلمات أو رموز فردية مع الأخذ بعين الاعتبار الحروف المقطعية (على سبيل المثال، حروف العطف وحروف الجر المرتبطة بالكلمات). ثم تأتي عملية توحيد النص (Text normalization) عن طريق إزالة علامات التشكيل، وتوحيد الحروف الممدودة، والتعامل مع الأحرف الخاصة باللغة العربية (على سبيل المثال، تحويل "أ"و"إ" و"آ" إلى "ا"). بعدها يتم اختزال الكلمات إلى صيغتها الجذرية أو صيغتها الأساسية، وهو أمر ضروري في اللغة العربية نظرًا لغناها الصرفي. وتُعرَف هذه العملية

بعملية التجذير/التصريف (Stemming/Lemmatization). كما يتم خلال هذه المرحلة إزالة الكلمات غير المطلوبة (Stop word removal) كونها لا تساهم في تحليل الرأي أو الشعور (على سبيل المثال،" من"،" في").

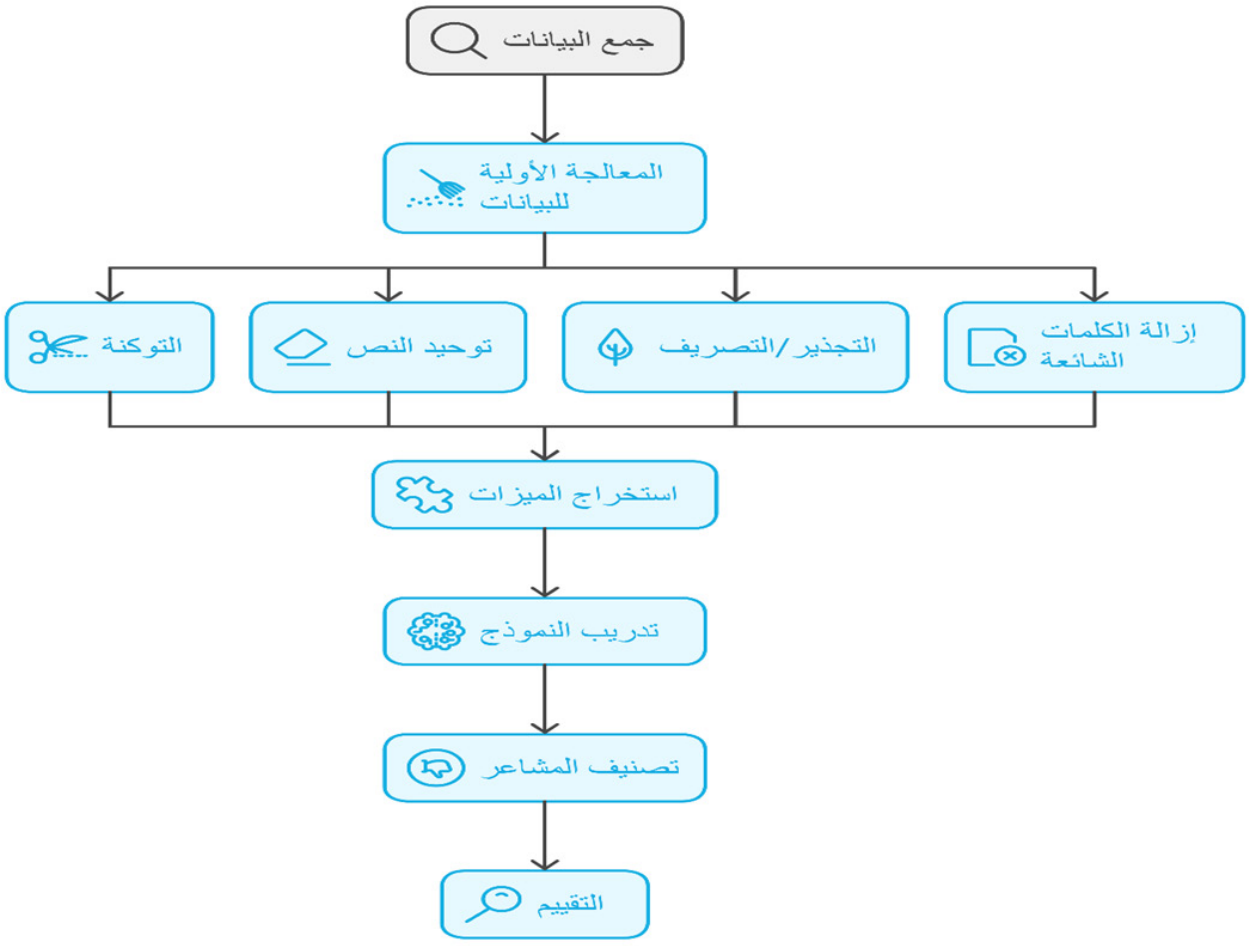
-استخراج الميزات (Features extraction): تتضمن هذه المرحلة تمثيل النص كمجموعة من ترددات الكلمات أو تكراراتها دون النظر إلى ترتيب الكلمات (BoW: Bag of Words)، وترجيح الكلمات بناءً على تكرارها في النص وندرتها في قاعدة البيانات (TF-IDF: Term Frequency-Inverse Document Frequency)، وكذا تحويل الكلمات إلى متجهات كثيفة (Word Embeddings) باستخدام نماذج محدّدة مثل AraVec.

-تدريب النماذج (Model training): يتم تدريب نماذج تعلّم الآلة، مثل آلات دعم المتجهات (SVM) و Naïve Bayes والانحدار اللوجستي على متجهات السمات المستخرجة من النص. أو باعتماد نماذج التعلّم العميق، يتم تدريب الشبكات العصبية، مثل الشبكات العصبية التلافيفية (CNNs) والشبكات العصبية المتكررة (RNNs)، لالتقاط الأنماط المعقدة في النص.

-تصنيف المشاعر (Sentiment classification): يتنبأ النموذج المُدَرّب بمشاعر النص الجديد المُدخَل. ويمكن أن تكون المخرجات تصنيفًا ثنائيًا (إيجابي/سلبي) أو تصنيفًا متعدد الفئات (إيجابي/سلبي/محايد) أو معتمدا على العواطف.

-التقييم (Evaluation): تُستخدم عدّة معايير لتقييم أداء نموذج تحليل المشاعر، مثل الدقة والإحكام. كما يتم استخدام مصفوفة الارتباك (Confusion matrix) في تصور الأداء من خلال إظهار التصنيفات الحقيقية مقابل التصنيفات المتوقعة.

يوضح المخطط الآتي الخطوات الأساسية المتبعة في تحليل المشاعر للنصوص العربية القائم على الذكاء الاصطناعي، ويقدم نظرة عامة واضحة عن كيفية تحويل النص من بيانات خام إلى مشاعر مصنفة.



شكل(04): خطوات تحليل المشاعر المعتمد على الذّكاء الاصطناعي

5–الدراسات السابقة

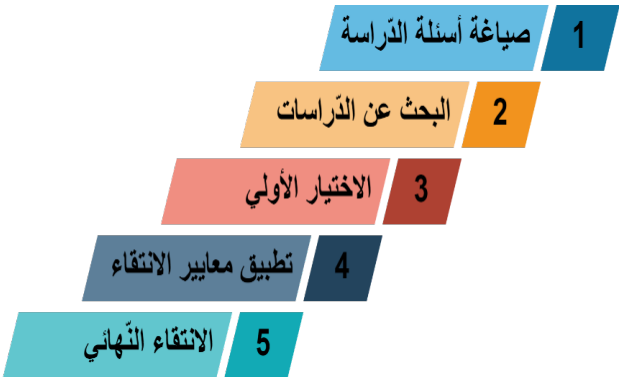
قَدَّم (Alsayat & Elmitwally, 2020) دراسة شاملة لتحليل المشاعر في النصوص العربية، حيث تناول الباحثون مراجعة أدبية لتطبيقات وتحديات تحليل المشاعر في النصوص من خلال المستويات الستة للغة العربية؛ الصوتية (Phonet-ic) والصرفية (Morphology) والنحوية (Syntax) والمعجمية (Lexicology) والدلالية (Semantics) والتصويرية (Figura-tive). وقد خلصت الدراسة إلى ضرورة تطوير منصة موحدة لاستخراج المشاعر وتحليلها تشمل انشاء مكنز لغوي ثري يدمج المستويات اللغوية الستة لتحسين دقة تصنيف المشاعر.

قام مطران وآخرون في (Matrane et al., 2023) بمراجعة وتصنيف أحدث التقنيات المستخدمة في تحليل المشاعر مركزين على مختلف اللهجات في النصوص المكتوبة باللغة العربية؛ حيث أثبتت الدّراسة أنّ نماذج التّعلم العميق خاصة تلك التي تدمج بين مستويات متعدّدة من تضمين الكلمات تحقق دقّة أعلى في تصنيف المشاعر. غير أن التّنوع اللّهجي يظل يشكّل تحديا رئيسياً خاصة في ظل نقص الموارد وقواعد البيانات الموجهة لتحليل اللغة العربية. وعليه، خلُصت هذه الدراسة إلى مجموعة من التوصيات أهمُّها ضرورة انشاء قاعدة بيانات شاملة للغة العربية ومختلف اللّهجات واللغة العربية المعرّبة (Arabizi)، وكذا إنشاء محلّلات لهجية مخصّصة لحلّ الغموض في معاني الكلمات (Ambiguity)، وأخذ عوامل تغيير التوجه العاطفي بعين الاعتبار مثل النّفي وصيغ المبالغة والحال وغيرها.

في الورقة الموسومة **معالجة اللغة الطبيعية لتحليل المشاعر باللغة العربية: مراجعة أدبية منهجية (2024)**، قام الباحثون بعرض التقنيات والأساليب المستخدمة في تحليل المشاعر في النصوص العربية خلال الخمس سنوات الماضية. تبين الدراسة أن تقنيات التعلم العميق وتعلم الآلة تهيمن على هذا المجال، حيث أظهرت نماذج التعلم العميق أداءً متفوقًا بشكل عام. توصي الدراسة بتعزيز الحلول المستخدمة في تحليل المشاعر باعتماد نماذج متقدمة لفهم أفضل للمشاعر الدقيقة المُعبر عنها في سياق معين وبواسطة لهجات متعددة مع الأخذ بعين الاعتبار التعامل مع تعابير السخرية والاستهزاء (Al Katat et al., 2024).

6–منهجية البحث

تبحث هذه الدراسة في كيفية استغلال تقنيات الذكاء الاصطناعي في تحليل المشاعر واكتشاف العواطف في النصوص العربية باستخدام المراجعة المنهجية للأدبيات (SLR: Systematic Literature Review). يوضح الشكل 5 أدناه الخطوات المتبعة في هذه الدراسة. تمر عملية المراجعة التي قمنا بها على خمس مراحل رئيسية: صياغة أسئلة الدراسة، وإجراء عملية البحث عن الأدبيات، والاختيار الأولي، وتصفية الأدبيات، والانتقاء النهائي.



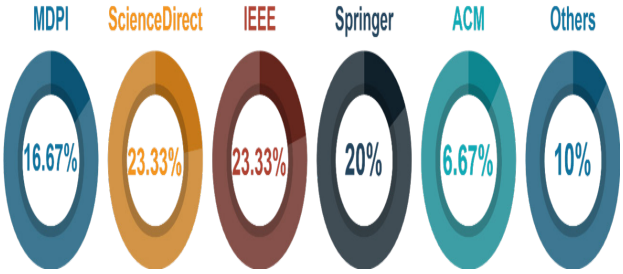
شكل(05): خطوات البحث وجمع الدّراسات

بعد صياغة أسئلة الدراسة، جرى البحث عن أحدث الأعمال في الأدبيات من خلال كلمات دالة على Google Scholar، وهو محرك بحث يستخدم في البحث العلمي على نطاق واسع وتضمن قاعدة بياناته الوصول إلى المقالات المتاحة على مستوى أهم دور النشر العالمية مثل IEEE، Elsevier و ACM وغيرها. ولهذا تم استخدام العبارات المنطقية AND لتحديد الكلمات الدالة الأساسية و OR لتضمين المرادفات والتهجئات البديلة لها كما هو موضح في الجدول التالي:

جدول (01): الكلمات المفتاحية للبحث

الكلمات الرئيسية	المرادفات والتهجئات البديلة			
الذكاء الاصطناعي	Artificial Intelligence			AI
تحليل المشاعر	Sentiment Analysis	SA	Opinion mining	Emotion detection
عربي				Arabic

بعد ذلك، اخترنا مبدئيا واحدا وثمانون (81) ورقة بحثية. يوضح الشكل 6 أدناه توزيع الأوراق البحثية بين مختلف دور النشر.



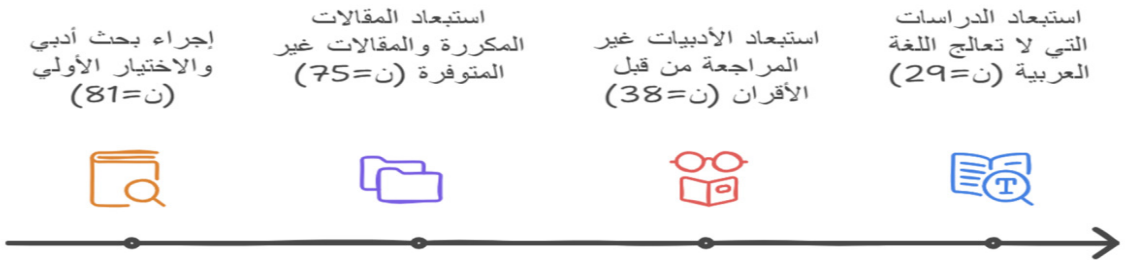
شكل(06): توزيع الأوراق البحثية بين مختلف دور النشر.

بعد الاختيار الأولي، من خلال تحليل عناوين وملخصات كل مقالة، قمنا بتطبيق معايير الانتقاء والاستبعاد الموضحة في الجدول 2.

جدول(02):معايير الانتقاء والاستبعاد

معايير الانتقاء	معايير الاستبعاد
المقالات المنشورة في المجلات العلمية المفهرسة ومداخلات الملتقيات	المراجعات والمقالات غير المتوفرة
المقالات المنشورة بين عامي 2019 و2024	المقالات المكرّرة وغير المُراجَعة من قِبَل الأقران
المقالات التي تركز على دراسة تحليل المشاعر واكتشاف العواطف في النصوص العربية المعتمدة على الذكاء الاصطناعي	المقالات التي لا تُعالج تحليل المشاعر واكتشاف العواطف في النصوص العربية المكتوبة باللغة العربية

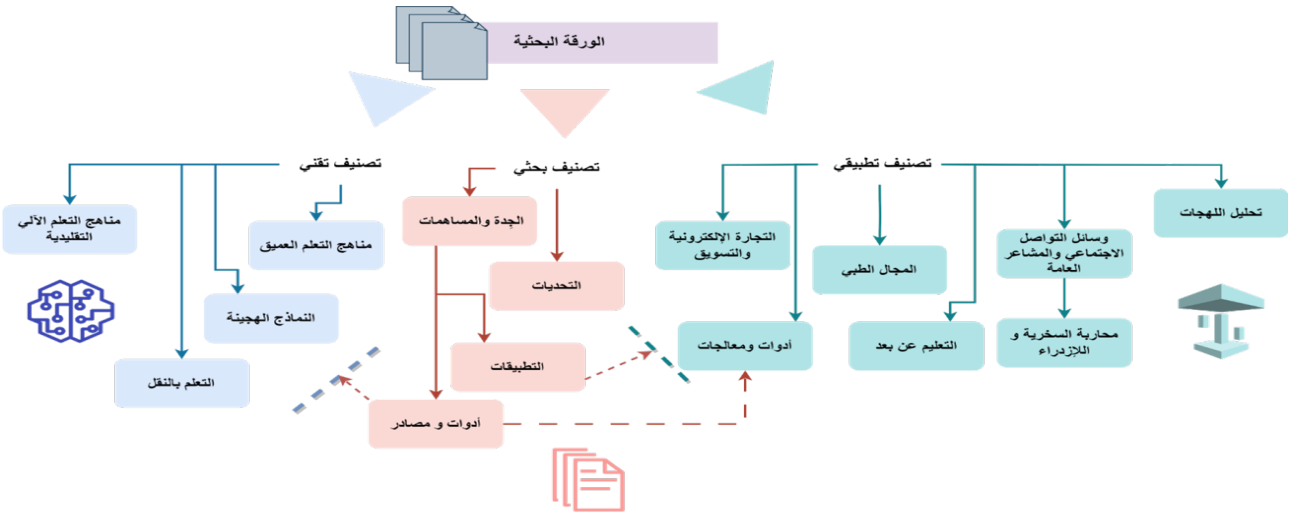
وأخيرًا، تم انتقاء ما مجموعه تسع وعشرون (29) من الأعمال للإجابة على أسئلة البحث المحددة ضمن المراجعة المنهجية. يوضح الشكل 7 أدناه عملية الانتقاء النهائية للدراسات الأدبية، بالاعتماد على جدول الانتقاء والاستبعاد.



شكل(07): عملية انتقاء الدّراسَات

7-نتائج الدراسة

7-1-تصنيف الدراسات



شكل(08): تصنيف الدّراسات

بعد تصفية مجموعة الأبحاث التي تم جمعها والمتعلقة باستخدام الذكاء الاصطناعي في تحليل المشاعر باللغة العربية، اعتمدنا منهجية تصنيف تهدف إلى تقسيم الأبحاث إلى مجالات فرعية. تعتمد هذه المنهجية على التوجه الأساسي للبحث، حيث تم الحصول على تصنيف هرمي كما هو موضح في الشكل أعلاه يتضمن ثلاثة مجالات فرعية رئيسية كما يلي:

التصنيف التقني: تجمع هذه الفئة الأوراق البحثية ذات التوجه التقني التي تركز على تقديم نماذج وخوارزميات ذكية جديدة في المجال. تشمل هذه الفئة الأبحاث التي تستكشف التعلم الآلي التقليدي أو النماذج المتقدمة مثل التعلم العميق، التعلم بالنقل، أو النماذج الهجينة.

التصنيف البحثي: تجمع هذه الفئة الأبحاث التي تتمحور حول البحث العلمي بشكل أكبر ولديها إمكانات تطبيقية محدودة. تسلط الضوء على المساهمات التي قدمتها الأبحاث الحديثة والتحديات التي واجهها الباحثون. كما تستكشف إمكانيات تطبيق النماذج الرائجة مثل الذكاء الاصطناعي القابل للتفسير وكيفية تحسين مجال تحليل المشاعر باللغة العربية. وتشمل أيضًا الأدوات والموارد الحديثة التي يمكن استخدامها في التطبيقات العملية.

التصنيف التطبيقي: تشمل هذه الفئة الأوراق البحثية التي تهدف إلى حل مشكلة تطبيقية محددة باستخدام نماذج الذكاء الاصطناعي، مثل التعامل مع قضايا التطرف أو السخرية أو التعليم في إطار مجموعة سكانية معينة أو خلال فترة أحداث محددة (كوفيد-19، الانتخابات ...).

فيما يلي نقدم مناقشة مفصلة لعشرة أوراق بحثية مختارة، بالإضافة إلى وصف توضيحي لكل ورقة في الجدول 1، بناءً على مساهماتها، التحديات التي تواجهها، والنماذج الذكية المستخدمة.

7–1–1– A Comparative Analysis for Arabic Sentiment Analysis Models In E-Marketing Using Deep Learning Techniques

«تحليل مقارن لنماذج تحليل المشاعر العربية في التسويق الإلكتروني باستخدام تقنيات التعلم العميق»

يقدم الباحثون (Almutairi & Alotaibi, 2023) عبر هذه الدراسة رؤى قيمة في مجال تحليل المشاعر باللغة العربية (ASA) في سياق التسويق الإلكتروني، مع التركيز على إمكانيات تقنيات التعلم العميق في التغلب على التحديات التي تفرضها تعقيدات اللغة العربية المتأصلة. يسد هذا البحث فجوة هامة من خلال معالجة الفروق المورفولوجية والاختلافات اللهجية في اللغة العربية، والتي كانت تمثل صعوبات تاريخية في مهام تحليل المشاعر. بين هذا البحث أن استخدام شبكات الذاكرة طويلة المدى (LSTM) بالتكامل مع الشبكات العصبية التلافيفية (CNN) يظهر تحسنًا كبيرًا في التعامل مع تعقيدات النصوص العربية. بالإضافة إلى ذلك، فإن إدخال مجموعة بيانات «سارة» (Sara-Dataset)، التي تم تصميمها خصيصًا لقطاع الاتصالات، يضيف قيمة كبيرة للمجال، حيث يوفر مجموعة بيانات أكثر تركيزًا لتدريب وتقييم نماذج تحليل المشاعر في سياق محدد للصناعة. هذا الإسهام مهم لتعزيز تطبيقات تحليل المشاعر في التسويق الإلكتروني، حيث يمكن أن يؤثر فهم آراء المستهلكين بشكل مباشر على استراتيجيات التسويق وأداء المبيعات. ومع ذلك، تعترف الدراسة أيضًا بوجود قيود، تتعلق بشكل أساسي بالتعامل مع مجموعات البيانات الضخمة وتعقيد الوقت اللازم لتدريب النماذج.

7–1–2–A Novel Hybrid Network for Arabic Sentiment Analysis using fine-tuned AraBERT model

«شبكة هجينة جديدة لتحليل المشاعر العربية باستخدام نموذج AraBERT المُحسّن»

تقدم الورقة (Habbat et al., 2021) اقتراح شبكة تعلم عميقة هجينة تدمج بين نموذج AraBERT المعدل ونماذج BiLSTM (ذاكرة قصيرة وطويلة المدى ثنائية الاتجاه) وGRU (وحدة متكررة بوابة). يتناول هذا النهج المبتكر تعقيدات اللغة العربية، خاصة في السياقات الديناميكية وغير الرسمية غالبًا مثل التفاعلات على وسائل التواصل الاجتماعي خاصة في جائحة كورونا، مثل التغريدات. التركيز على معالجة هذه التعقيدات ذو أهمية كبيرة، نظرًا للتحديات الفريدة المتعلقة بالصرف والنحو والدلالة التي تواجهها اللغة العربية. من خلال الاستفادة من نموذج هجين، ينجح البحث في التقاط الجوانب التسلسلية والغنية بالسياق في التعبير عن المشاعر باللغة العربية

يضيف اختبار النموذج على ثلاث مجموعات بيانات مدمجة، خاصة خلال جائحة COVID-19، مصداقية للعالم الواقعي للنتائج، مما يدل على متانة النموذج وقابليته للتطبيق في تحليل المشاعر من مجموعات بيانات كبيرة ومتنوعة. يعد سياق الجائحة ذا أهمية خاصة، حيث يبرز أهمية فهم مشاعر الجمهور بدقة خلال أوقات الأزمات، حيث يكون

حجم وتنوع المناقشات على الإنترنت مرتفعًا.

ومع ذلك، على الرغم من النتائج المثيرة للإعجاب، تعترف الدراسة ببعض التحديات الأوسع التي تواجه مهام تحليل المشاعر، مثل فهم السخرية، التي تظل عقبة كبيرة في تفسير المشاعر بدقة في التغريدات العربية. تتفاقم هذه المشكلة بسبب تعقيد النحو والدلالة العربية، مما قد يؤدي إلى غموض في تصنيف المشاعر. من بين القيود الملحوظة الأخرى الاعتماد على مجموعات بيانات كبيرة وموصوفة، والتي غالبًا ما يصعب الحصول عليها أو إنشاؤها لتحليل المشاعر باللغة العربية نظرًا لاختلافات اللهجات وطبيعة اللغة الأقل تنظيمًا.

7–1–3–ArabBert-LSTM: improving Arabic sentiment analysis based on transformer model and Long Short-Term Memory ArabBert-LSTM

"تحسين تحليل المشاعر العربية بالاعتماد على نموذج المحولات والذاكرة طويلة وقصيرة المدى"

تقدم هذه الورقة نموذج ArabBert-LSTM، وهو نموذج مبتكر لتحليل المشاعر باللغة العربية يدمج بين هياكل المحولات وشبكات الذاكرة قصيرة وطويلة المدى (LSTM) (Alosaimi et al., 2024). يهدف هذا الدمج إلى استغلال نقاط القوة في كلا النوعين من النماذج، مستفيدًا من قدرات Ar-aBERT للتعامل مع تعقيدات صرف ونحو اللغة العربية، بينما يستخدم كفاءة LSTM في نمذجة التسلسلات. تقارن الدراسة بين طرق مختلفة للتتّوجيه المتّجهي (Vectoriza-tion)، بما في ذلك تكرار المصطلحات - تكرار المستند العكسي (TF-IDF)، وصندوق الكلمات المستمر (CBOW)، و-Skip Gram. هذه المقارنة مهمة لتحديد تقنيات استخراج الميزات الأكثر فعالية المصممة خصيصًا للنصوص العربية. من خلال تحسين استخراج الميزات، يمكن للنموذج تحسين جودة المدخلات، وبالتالي تعزيز دقة تصنيف المشاعر.

أدى تطبيق نموذج ArabBert-LSTM عبر عدة مجموعات بيانات مرجعية عربية إلى تحقيق نتائج مثيرة للإعجاب، حيث تجاوزت معدلات الدقة 97%. تُظهر هذه الأداءات تقدمًا كبيرًا مقارنة بأساليب التعلم الآلي والتعلم العميق التقليدية، مما يثبت أن النموذج يعد نهجًا متطورًا لتحليل المشاعر باللغة العربية. تشير النتائج إلى أن النموذج الهجين لا يُحسن فقط دقة التصنيف، بل يوفر أيضًا إطارًا قويًا لفهم تفاصيل المشاعر العربية. ولكن بالرغم من النتائج الواعدة للنموذج، تعترف الورقة بالتعقيدات الكامنة في اللغة العربية التي لا تزال تمثل تحديات في تحليل المشاعر.

7–1–4–Arabic sentiment analysis about online learning to mitigate covid-19

«تحليل المشاعر العربية حول التعلم عبر الإنترنت للتخفيف من آثار COVID-19»

قدمت منال مصطفى علي في هذه الورقة تحليلًا مفصلًا لمشاعر اللغة العربية تجاه التعلم عبر الإنترنت خلال جائحة كوفيد-19، مستخدمة منهجية منظمة تتضمن جمع البيانات، ومعالجة البيانات المسبقة، وتحليل المشاعر، واستخراج العواطف (Ali, 2021). تقوم بجمع التغريدات باستخدام واجهات برمجة تطبيقات تويتر، مما يوفر لمحة فورية عن المشاعر العامة خلال فترة حاسمة. إذ تعد مرحلة معالجة البيانات المسبقة مهمة بشكل خاص بسبب تعقيدات اللغة العربية، التي تتميز بالغنى في الصرف وتنوع اللهجات. قامت منال باختيار خوارزميات تعلم الآلة، مثل نايف بايز و KNN والانحدار اللوجستي وآلة الدعم الناقل (SVM)، لإجراء تحليل مقارن لفعاليتها في تصنيف المشاعر. أظهرت النتائج أن خوارزمية SVM حققت أعلى دقة بنسبة 89.6%، مما يؤكد فعاليتها مع البيانات ذات الأبعاد العالية {تعدد الميزات}. كما ساعد استخدام «معلومات الكسب» (IG) في تحسين الأداء التصنيفي من خلال تصفية الميزات غير ذات الصلة، وزيادة موثوقية النتائج. أضافت الدراسة بعدًا إضافيًا باستخراج العواطف باستخدام قاموس NRC، حيث تم تحديد الغضب والخوف كأهم العواطف، مما يساهم في فهم التأثير النفسي للتعلم عبر الإنترنت وتقديم رؤى لتطوير تدخلات تستهدف تخفيف المخاوف. لا تتردد الدراسة في الاعتراف بالتحديات التي واجهتها، ولاسيما فيما يتعلق بتعقيدات اللغة العربية وندرة توفر مجموعات بيانات مشروحة. يمكن أن تؤثر هذه التحديات بشكل كبير على موثوقية نتائج تحليل المشاعر وتبرز الحاجة إلى موارد أكثر شمولاً في معالجة اللغة العربية.

7–1–5–Arabic Sentiment Analysis for ChatGPT Using Machine Learning Classification Algorithms: A Hyperparameter Optimization Technique

«تحليل المشاعر العربية لـ ChatGPT باستخدام خوارزميات تصنيف التعلم الآلي: تقنية تحسين المعلمات الفائقة»

تقدم الدراسة (Nasayreh et al., 2024)مساهمة هامة في تحليل المشاعر المعبر عنها باللغة العربية تجاه ChatGPT، كونها من أوائل الأبحاث التي تعنى باللغة العربية وهذا النموذج. من

خلال استخدام مجموعة بيانات تتكون من 2,247 تغريدة، تهدف الدراسة إلى تصنيف الآراء على أنها إيجابية أو سلبية، وبالتالي تقديم رؤى قيمة حول مشاعر المستخدمين الناطقين بالعربية تجاه هذا النموذج من التعلم الآلي. تؤكد الدراسة أيضًا على الحاجة إلى مزيد من الدراسات في هذا المجال، نظرًا لندرة الأدبيات الموجودة التي تحلل مشاعر اللغة العربية حول ChatGPT. يشير المؤلفون إلى أن بعض الدراسات السابقة لم تستخدم مصفوفات الالتباس (confusion matrices) للتحقق من النتائج، مما يعقد تقييم دقة تصنيف المشاعر. وهذا يبرز فجوة منهجية حرجة يجب أن تتناولها الأبحاث المستقبلية.

7–1–6–Arabic Sentiment Analysis with Noisy Deep Explainable Model

«تحليل المشاعر العربية باستخدام نموذج عميق قابل للتفسير مع الضوضاء»

تناول محمد عتاب الزمان وآخرون (Atabuzzaman et al., 2023) مجالًا جد حيوي في تحليل المشاعر باللغة العربية (SA) من خلال اقتراح إطار عمل جديد يهدف إلى تعزيز كل من الأداء وقابلية التفسير لنماذج التعلم العميق. يحدد الباحثون تحديات كبيرة تواجه النماذج الحالية، خصوصًا مشكلتي الإفراط في التوافق وغياب القابلية للتفسير، وهما من القضايا البارزة في اللغات ذات الموارد المحدودة مثل العربية. من خلال دمج طبقة ضوضاء في هياكل التعلم العميق، يهدف الإطار إلى التخفيف من هذه التحديات، مما يعزز تقدم هذا المجال. تقدم الدراسة طريقتين مختلفتين مبنيتين على التعلم العميق، وهما BiLSTM وCNN-BiLSTM، مما يعكس استخدامًا مدروسًا لمزايا هياكل الشبكات المختلفة. إذ يتم تدريب هذه النماذج على مراجعات عربية مُعالجة لفحص كيفية تحسين طبقة الضوضاء لأداء النموذج وصلابته. بعد تحسين أداء النموذج يُستخدم LIME لتفسير توقعاته، مما يضيف طبقة من الشفافية ويتيح للباحثين فهم أسباب تصنيفات المشاعر. تضع الفوائد المزدوجة المتمثلة في تحسين الأداء وقابلية التفسير النماذج المقترحة في موقع يجعلها أكثر شفافية وقابلية للمساءلة، وهو أمر ضروري لاعتمادها في التطبيقات العملية. ومع ذلك، تعترف الورقة بوجود قيود في تركيزها على تحليل المشاعر باللغة العربية.

7–1–7–Artificial Intelligence-Based Approach for Misogyny and Sarcasm Detection from Arabic Text

«نهج قائم على الذكاء الاصطناعي للكشف عن كراهية النساء والسخرية في النصوص العربية»

عمل عبد الله ي. معاذ وباحثون من أنحاء العالم على تطوير نموذج قائم على الذكاء الاصطناعي للكشف تلقائيًا عن الكراهية ضد النساء والسخرية في النصوص العربية، خاصة على منصات التواصل الاجتماعي مثل تويتر(Muaad et al., 2022). توظف هذه الدراسة تقنيات متنوعة من التعلم الآلي (ML) والتعلم العميق (DL) من خلال تدريب سبع مصنفات متقدمة في معالجة اللغة الطبيعية. استخدم الباحثون مجموعتين من البيانات مجموعة بيانات الكراهية ضد النساء ومجموعة بيانات أبو فرح لتدريب النماذج واختبارها في سيناريوهات ثنائية ومتعددة الفئات. التطبيق الرئيسي هو في مراقبة وسائل التواصل الاجتماعي، حيث يهدف النموذج إلى تحديد وتصنيف المحتوى الضار، مما يسهم في خلق بيئات أكثر أمانًا على الإنترنت.

تشير النتائج إلى أن النهج المقترح للذكاء الاصطناعي فعال، حيث حقق دقة عالية بنسبة 91% في مهام التصنيف الثنائي و88% في كشف السخرية. حيث يظهر مصنف AraBERT كأكثر النماذج فعالية، مما يبرز أهمية استخدام تقنيات معالجة اللغة الطبيعية المتقدمة المصممة للغة العربية. ومع ذلك، تعترف الدراسة ببعض القيود إذ أن ندرة الأدوات ومجموعات البيانات المرجعية لكشف النصوص العربية تشكل تحديات كبيرة في تطوير نماذج قوية. علاوة على ذلك، يمكن أن تؤدي التعقيدات الكامنة في السخرية والكراهية ضد النساء في اللغة إلى تصنيفات خاطئة، مما يشير إلى أن هناك حاجة إلى طرق كشف أكثر دقة ومزيد من الأبحاث.

7–1–8–CAMeL Tools: An Open-Source Python Toolkit for Arabic Natural Language Processing

«أدوات CAMeL: مجموعة أدوات بايثون مفتوحة المصدر لمعالجة اللغة الطبيعية باللغة العربية»

تلعب أدوات CAMeL دورا مهمًا في معالجة اللغة الطبيعية (NLP) باللغة العربية، حيث تقدم مجموعة من الأدوات المتخصصة لمعالجة التحديات الفريدة للغة العربية. يستخدم هذا المحلل الصرفي نهجًا قائمًا على المعجم لتفكيك الكلمات العربية إلى أجزائها الأساسية، وهو أمر ضروري لفهم المعاني والأدوار النحوية. كما يدعم هذا التحليل الأساسي العديد من المهام، بما في ذلك تصنيف النص والترجمة الآلية (Obeid et al., 2020).

7–1–9–Sentiment analysis of Arabic social media texts: A machine learning approach to deciphering customer perceptions

«تحليل المشاعر لنصوص وسائل التواصل الاجتماعي العربية: نهج تعلم الآلة لفهم تصورات العملاء»

تتناول الورقة تحليل المشاعر (SA) للنصوص العربية على وسائل التواصل الاجتماعي، مع التركيز على تصورات العملاء بشأن منتجات القهوة (Alsemaree et al., 2024). استخدم الباحثون نماذج التعلم الآلي التقليدية مع مجموعة بيانات تضم 10,646 مراجعة من تويتر، حيث تم استخدام تقنيات مثل تردد الكلمات-عكس تردد الوثائق (TF-IDF) لاستخراج الميزات. درب الباحثون عدة مصنفات مثل KNN وSVM و DT وRF، وتم دمج النتائج باستخدام طرق التصويت القاسية والناعمة، محققةً دقة تنبؤية بلغت 95.95% و94.51% على التوالي. إذ تشير النتائج إلى أهمية فهم مشاعر العملاء لتوجيه استراتيجيات تطوير المنتجات والتسويق في صناعة القهوة. مثل جل الأبحاث في هذا المجال فعلى الرغم من تحقيق دقة عالية، يقترح الباحثون استكشاف طرق التعلم العميق في المستقبل، ويؤكدون على التحديات المستمرة بسبب تعقيد اللغة العربية.

7–1–10–Transfer learning and sentiment analysis of Bahraini dialects sequential text data using multilingual deep learning approach

« التعلم بالنقل وتحليل المشاعر للنصوص المتسلسلة باللهجة البحرينية باستخدام نهج التعلم العميق متعدد اللغات»

تتناول الدراسة تحليل المشاعر (SA) في معالجة اللغة الطبيعية (NLP) العربية، مع التركيز على اللهجات البحرينية (Omran et al., 2023). حيث قام أحد السمري وآخرون بإنشاء مجموعة بيانات متوازنة من خلال ترجمة مراجعات المنتجات من الإنجليزية إلى العربية الفصحى ثم إلى اللهجات البحرينية. تستخدم الدراسة نموذج تعلم عميق (LSTM) لتحليل البيانات الموازية، مع اعتماد تقنيات تعلم آلي متنوعة. إذ حقق النموذج دقة تتراوح بين 96.72% و97.04%، مثبتا فعاليته في تحليل المشاعر في اللهجات البحرينية. تساهم هذه الدراسة في مجال معالجة اللغة العربية بتوفير مجموعة بيانات ونموذج يمكن استخدامه في تطبيقات مختلفة. ومع ذلك، تعترف بالتحديات مثل نقص البيانات المعلّمة والغنى الصرفي للغة العربية، مما يؤثر على التحليل وبدوره على التعلم الدقيق للنماذج الذكية.

جدول(03): جدول تحليلي حول التحديات والمساهمات والتطبيقات وتقنيات التعلم الآلي في مجال تحليل المشاعر للنصوص العربية.

البحث	تقنيات التعلم الآلي	التطبيقات	الجِدة/المساهمات	التحديات
1	التعلم العميق (LSTM, CNN)	التسويق الإلكتروني	نموذج التعلم العميق الهجين، مجموعة بيانات سارة	تعقيد اللغة
2	نموذج هجين (AraBERT, BiL-STM, GRU)	وسائل التواصل الاجتماعي (COVID-19)	نموذج هجين (Ara-BERT-BiLSTM-GRU)	تعقيد اللغة
3	التعلم العميق (ArabBERT, LSTM)	تحليل المشاعر العام	دمج المحولات LSTM	استخراج الميزات، مشكلات اللغة
4	التعلم الآلي التقليدي (Naive Bayes, SVM, KNN)	التعليم، COVID-19	استخدام SVM لتصنيف المشاعر	معالجة البيانات المسبقة العربية
5	التعلم الآلي (تحسين المعلمات الفائقة)	تحليل المشاعر حول ChatGPT	تحسين المعلمات الفائقة لتحليل مشاعر ChatGPT	ندرة البيانات، تقنيات التقييم
6	التعلم العميق (BiLSTM, CNN-BiLSTM)	تحليل المشاعر العام	إضافة الضوضاء، LIME للتفسير	الإفراط في التوافق، قابلية التفسير
7	التعلم الآلي والتعلم العميق	وسائل التواصل الاجتماعي (كراهية النساء والسخرية)	الكشف عن كراهية النساء والسخرية	موارد NLP العربية، حجم البيانات
8	أدوات معالجة اللغة الطبيعية (NLP)	معالجة اللغة الطبيعية العربية	مجموعة أدوات مفتوحة المصدر لـ NLP العربية	تحديات معالجة اللغة
9	التعلم الآلي (SVM, KNN, DT, RF)	وسائل التواصل الاجتماعي، مراجعات المنتجات	مقارنة أداء النماذج في تحليل المشاعر	تعقيد اللغة، استخراج الميزات
10	التعلم العميق (LSTM)	تحليل المشاعر في اللهجات	التعلم بالنقل لتحليل اللهجات	نقص البيانات، تحديات اللهجة

7-2-تحليل النتائج

تُظهر الأبحاث المجموعة تقدمًا كبيرًا في مجال معالجة اللغة الطبيعية وتحليل المشاعر في اللغة العربية، إلا أن التحديات لا تزال قائمة، وخاصة فيما يتعلق بالتعامل مع الاختلافات اللهجية، والكشف عن السخرية، ونقص مجموعات البيانات الشاملة. إذ يمكن تلخيص التحديات الرئيسية في تحليل المشاعر في النصوص العربية كالآتي:

المورفولوجيا الغنية: اللغة العربية لغة ذات تصريفات عالية مع أشكال كلمات معقدة.

الاختلافات اللهجية: تحتوي اللغة العربية على العديد من اللهجات، والتي تختلف اختلافًا كبيرًا عن اللغة العربية الفصحى الحديثة.

الغموض وتعدد المعاني: غالبًا ما يكون للكلمات في اللغة العربية معانٍ متعددة، مما يجعل تحليل المشاعر أمرًا صعبًا.

غياب التوحيد القياسي: يمكن أن تؤدي الاختلافات الإملائية واستخدام علامات التشكيل إلى تعقيد معالجة النصوص.

وبِعدِّ تحليل المشاعر جزءا من المعالجة الآلية للغة الطبيعية، أظهرت نماذج التعلم العميق، وخاصة النماذج المعتمدة على المحولات مثل AraBERT، إمكانيات كبيرة في تطوير تحليل المشاعر باللغة العربية، ومواجهة التحديات المطروحة نسبيا حيث تقدم نتائج أكثر دقة وتفصيلاً مقارنة بتقنيات التعلم الآلي التقليدية.

في الختام يمكن للأبحاث المستقبلية أن تركز على تحسين دعم اللغات المتعددة واللهجات، وإنشاء مجموعات بيانات أكثر ثراءً، ودمج تقنيات الذكاء الاصطناعي القابل للتفسير لتقديم نماذج أكثر وضوحًا وفهمًا. علاوة على ذلك، أثبتت الأساليب القائمة على التكيف مع المجال والتعلم بالنقل أنها فعالة في تحسين أداء تحليل المشاعر، ويجب توسيع نطاقها في الدراسات المستقبلية. بشكل عام، تسلط هذه المجموعة من الأعمال الضوء على أهمية مواصلة تحسين وتطوير تقنيات معالجة اللغة الطبيعية التي تتناسب مع هيكل اللغة العربية الفريد وتنوعها، وكذلك دراسة التأثيرات الأوسع لاستخدام هذه النماذج في مختلف الصناعات والتطبيقات.

خاتمة

يوفر الذكاء الاصطناعي أدوات قوية لتعزيز أتمتة عملية تفسير وتصنيف المشاعر الانسانية والعواطف المعبر عنها في النصوص من خلال تنفيذ المهام بسرعة ودقة واتساق والتعامل مع كم هائل من البيانات يفوق القدرات البشرية مما يساهم في فهم اللغة وتحسين اتخاذ القرارات المعقدة في الوقت الفعلي. ومع ذلك، يجب مراعاة الفهم الدقيق للتفاصيل اللغوية والعاطفية مثل خصائص وتعقيدات اللغة والسياقات الثقافية المختلفة، العربية منها أخصُّ ذكرا. من خلال هذه الدّراسة، تَأكّدت أهمية تطوير موارد لغوية شاملة للغة العربية تشمل كافة اللهجات المختلفة، مع التركيز على تحسين الخوارزميات المستخدمة لتحليل المشاعر واكتشاف العواطف. كما تأكّدت الحاجة إلى تطوير منصات موحدة تعزز دقة التصنيف والاستفادة من مستويات اللغة المختلفة مثل الصوتية والنحوية والدلالية. كما تمت الإشارة إلى ضرورة البحث في كيفية تحسين أداء الخوارزميات عند التعامل مع المحتوى الغني من حيث اللهجات والتعابير الثقافية المعقدة. مستقبلًا، يمكن أن تتوجه الأبحاث نحو دمج الذكاء الاصطناعي التوضيحي (Explainable AI) لتحسين فهم وتحليل النصوص العربية بشكل أكثر شفافية وموثوقية، إضافة إلى تعزيز استخدام البيانات الكبيرة في مجالات جديدة مثل التعليم وتحليل البيانات الاجتماعية والسياسية.

المراجع

1. Abo, M. E. M., Raj, R. G., & Qazi, A. (2019). A Review on Arabic sentiment analysis: State-of-The-Art, taxonomy and open research challenges. In *IEEE Access* (Vol. 7, pp. 162008–162024). Institute of Electrical and Electronics Engineers Inc. https://doi.org/10.1109/ACCESS.2019.2951530
2. Al Katat, S., Zaki, C., Hazimeh, H., Bitar, I., Angarita, R., & Trojman, L. (2024). Natural Language Processing for Arabic Sentiment Analysis: A Systematic Literature Review. *IEEE Transactions on Big Data*, *10*(5), 576–594.
3. Ali, M. M. (2021). Arabic sentiment analysis about online learning to mitigate covid-19. *Journal of Intelligent Systems*, *30*(1), 524–540. https://doi.org/10.1515/jisys-2020-0115
4. Almutairi, S., & Alotaibi, F. (2023). A Comparative Analysis for Arabic Sentiment Analysis Models In E-Marketing Using Deep Learning Techniques. *Journal of Engineering and Applied Sciences*, *10*(1).
5. Alosaimi, W., Saleh, H., Hamzah, A. A., El-Rashidy, N., Alharb, A., Elaraby, A., & Mostafa, S. (2024). AraBert-LSTM: improving Arabic sentiment analysis based on transformer model and Long Short-Term Memory. *Frontiers in Artificial Intelligence*, *7*. https://doi.org/10.3389/frai.2024.1408845
6. Alsayat, A., & Elmitwally, N. (2020). A comprehensive study for Arabic Sentiment Analysis (Challenges and Applications). *Egyptian Informatics Journal*, *21*(1), 7–12. https://doi.org/10.1016/j.eij.2019.06.001
7. Alsemaree, O., Alam, A. S., Gill, S. S., & Uhlig, S. (2024). Sentiment analysis of Arabic social media texts: A machine learning approach to deciphering customer perceptions. *Heliyon*, *10*(9). https://doi.org/10.1016/j.heliyon.2024.e27863
8. Atabuzzaman, M., Shajalal, M., Baby, M. B., & Boden, A. (2023). Arabic Sentiment Analysis with Noisy Deep Explainable Model. *ACM International Conference Proceeding Series*, 185–189. https://doi.org/10.1145/3639233.3639241
9. Habbat, N., Anoun, H., & Hassouni, L. (2021). A Novel Hybrid Network for Arabic Sentiment Analysis using fine-tuned AraBERT model. *International Journal on Electrical Engineering and Informatics*, *13*(4), 801–812. https://doi.org/10.15676/ijeei.2021.13.4.3
10. Manning, C. (2020). Artificial intelligence definitions. *HAI Stanford Univ*, *September*, 1.
11. Matrane, Y., Benabbou, F., & Sael, N. (2023). A systematic literature review of Arabic dialect sentiment analysis. In *Journal of King Saud University - Computer and Information Sciences* (Vol. 35, Issue 6). King Saud bin Abdulaziz University. https://doi.org/10.1016/j.jksuci.2023.101570
12. Muaad, A. Y., Jayappa Davanagere, H., Benifa, J. V. B., Alabrah, A., Naji Saif, M. A., Pushpa, D., Al-Antari, M. A., & Alfakih, T. M. (2022). Artificial Intelligence-Based Approach for Misogyny and Sarcasm Detection from Arabic Texts. *Computational Intelligence and Neuroscience*, *2022*. https://doi.org/10.11557937667/2022/
13. Nandwani, P., & Verma, R. (2021). A review on sentiment analysis and emotion detection from text. *Social Network Analysis and Mining*, *11*(1), 81.
14. Nasayreh, A., Al Mamlook, R. E., Samara, G., Gharaibeh, H., Aljaidi, M., Alzu'Bi, D., Al-Daoud, E., & Abualigah, L. (2024). Arabic Sentiment Analysis for ChatGPT Using Machine Learning Classification Algorithms: A Hyperparameter Optimization Technique. *ACM Transactions on Asian and Low-Resource Language Information Processing*, *23*(3). https://doi.org/10.1145/3638285
15. Obeid, O., Zalmout, N., Khalifa, S., Taji, D., Oudah, M., Alhafni, B., Inoue, G., Eryani, F., Erdmann, A., & Habash, N. (2020). CAMeL Tools: An Open-Source Python Toolkit for Arabic Natural Language Processing. In N. Calzolari, F. Béchet, P. Blache, K. Choukri, C. Cieri, T. Declerck, S. Goggi, H. Isahara, B. Maegaard, J. Mariani, H. Mazo, A. Moreno, J. Odijk, & S. Piperidis (Eds.), *Proceedings of the Twelfth Language Resources and Evaluation Conference* (pp. 7022–7032). European Language Resources Association. https://aclanthology.org/2020.lrec-1.868
16. Omran, T. M., Sharef, B. T., Grosan, C., & Li, Y. (2023). Transfer learning and sentiment analysis of Bahraini dialects sequential text data using multilingual deep learning approach. Data and Knowledge Engineering, 143. https://doi.org/10.1016/j.datak.2022.102106

Artificial intelligence-based automated sentiment analysis and emotion detection in arabic text

Abstract

This paper explores the application of Artificial Intelligence (AI) in Arabic sentiment analysis and emotion detection through a comprehensive systematic literature review (SLR). It examines the unique linguistic challenges associated with Arabic, including its morphological and syntactic complexity as well as the diversity of its dialects. Additionally, it highlights the various applications of sentiment analysis and emotion detection in key domains such as marketing, healthcare, and political and social analysis. This study also reviews the tools and algorithms employed for Arabic text processing, evaluating their effectiveness and limitations. Finally, it provides research-driven recommendations to enhance existing technologies, to ensure compatibility with the latest advancements in natural language processing (NLP) and AI.

Keywords
Natural language processing
NLP
sentiment analysis
emotion detection
opinion mining
Arabic

L'Intelligence Artificielle à l'analyse automatisée des sentiments et à la détection des émotions dans les textes arabes

Résumé

Cet article présente une revue systématique approfondie de la littérature scientifique sur l'application de l'intelligence artificielle à l'analyse des sentiments et à la détection des émotions dans les textes arabes, mettant en lumière les défis spécifiques inhérent à la langue arabe, notamment sa complexité morphologique et syntaxique ainsi que la diversité de ses dialectes. En outre, il examine les différentes applications de l'analyse des sentiments et de la détection des émotions dans des domaines clés tels que le marketing, la santé et l'analyse politique et sociale. Le travail passe également en revue les outils et algorithmes utilisés pour le traitement des textes arabes, en évaluant leur efficacité et leurs limites. Enfin, il propose des recommandations scientifiques visant à améliorer ces technologies afin d'accompagner les progrès récents en traitement automatique du langage naturel (TALN) et en intelligence artificielle.

Mots clés
Traitement automatique du langage naturel
TALN
analyse des sentiments
détection des émotions
langue arabe

Competing interests

The author(s) declare no competing interests

تضارب المصالح

يعلن المؤلف (المؤلفون) لا تضارب في المصالح

Author copyright and License agreement

Articles published in the Journal of letters and Social Sciences are published under the Creative Commons of the journal›s copyright. All articles are issued under the CC BY NC 4.0 Creative Commons Open Access License).
To see a copy of this license, visit:
http://creativecommons.org/licenses/by/4.0/
This license allows the maximum reuse of open access research materials. Thus, users are free to copy, transmit, distribute and adapt (remix) the contributions published in this journal, even for commercial purposes; Provided that the contributions used are credited to their authors, in accordance with a recognized method of writing references.

© The Author(s) 2023

حقوق المؤلف واذن الترخيص

إن المقالات التي تنشر في المجلة تنشر بموجب المشاع الإبداعي بحقوق النشر التي تملكها مجلة الآداب والعلوم الاجتماعية. ويتم إصدار كل المقالات بموجب ترخيص الوصول المفتوح المشاع الإبداعي CC BY NC 4.0.
للاطلاع على نسخة من هذا الترخيص، يمكنكم زيارة الموقع الموالي :
http://creativecommons.org/licenses/by/4.0/
إن هذا الترخيص يسمح بإعادة استخدام المواد البحثية المفتوحة الوصول إلى الحد الأقصى. وبالتالي، فإن المعنيين بالاستفادة أحرار في نسخ ونقل وتوزيع وتكييف (إعادة خلط) المساهمات المنشورة في هذه المجلة، وهذا حتى لأغراض تجارية؛ بشرط أن يتم نسب المساهمات المستخدمة من طرفهم إلى مؤلفي هذه المساهمات، وهذا وفقًا لطريقة من الطرق المعترف بها في كتابة المراجع.

© المؤلف (المؤلفون) 2023