# استخدام التعلم العميق لتحليل المشاعر في اللغة العربية

**طالبة الدكتوراه: م. ريم ناصر**

**كلية الهندسة المعلوماتية– جامعة البعث**

**اشراف الدكتور: كمال السلوم**

**المشرف المشارك الدكتور: ناصر أبو صالح**

الملخص:

أصبح موضوع تحليل المشاعر يشغل حيزاً مهماً في المجالين الأكاديمي والصناعي وذلك بسبب التزايد الكبير في أعداد مستخدمي مواقع التواصل الاجتماعي وبالتالي كمية التعليقات والمراجعات حول المنتجات، الأحداث، الأخبار، الخ.

تم اقتراح العديد من التقنيات لحل مشكلة تحليل هذا الكم الهائل من البيانات وتصنيفها، حيث أظهر مجال التعلم العميق نتائجاً واعدةً في مجال تحليل المشاعر وخاصة بعد ازدياد أعداد العرب المستخدمين للانترنت وبالتالي ازدياد كمية التعليقات التي شجعت الباحثين للخوض في تطبيق التعلم العميق لتحليلهم. تقدم هذه المقالة دراسة حول خوارزميات التعلم العميق الأكثر استخداماً في مجال تحليل المشاعر، ومن ثم تطبيق هذه الخوارزميات (CNN,LSTM,CNN+LSTM)على مجموعتين من البيانات النصية باللغة العربية والمقارنة بينها من حيث الأداء باستخدام عدة مقاييس للوصول للخوارزمية الأفضل وبالتالي يتحقق الهدف المرجو من بحثنا.

**الكلمات المفتاحية:** تحليل المشاعر، التعلم العميق، تضمين الكلمات، استخراج الميزات.

# Using deep learning for sentiment analysis in Arabic

**Abstract**

Sentiment analysis has become an important topic in the academic and industrial fields due to the large increase in the number of users of social networking sites and thus the amount of comments and reviews about products, events, news, etc. Several techniques have been proposed to solve the problem of analyzing and classifying this huge amount of data, as the field of deep learning showed promising results in the field of sentiment analysis, especially after the increase in the number of Arabs using the Internet and thus the increase in the amount of comments that encouraged researchers to delve into the application of deep learning for their analysis. This article presents a study on the most commonly used deep learning algorithms in the field of sentiment analysis(CNN,LSTM,CNN+LSTM), and then applying these algorithms to two sets of text data in arabic and comparing them in terms of performance using several measures to reach the best algorithm that achieves the desired goal of our research.

**Keywords:** Sentiment analysis, deep learning, word embedding, feature extraction.

## 1. مقدمة:

لسنواتٍ طويلةٍ جربت الشركات مجموعة واسعة من الأفكار لتحليل سلوك العملاء والتنبؤ به، وذلك بهدف بناء جسر تواصل معهم لمعرفة ما لديهم من مشاكل أوعقبات أو حتى مديح لما يتم تقديمه. وهذا ما أدى لظهور مصطلح تحليل المشاعر الذي يهتم بتحليل الآراء، التقييمات، المواقف تجاه خدمة معينة، أفراد، قضايا مهمة، الخ. وذلك بهدف تصنيف هذه المشاعر إلى عدة فئات.

يعرف مصطلح تحليل المشاعر sentiment analysis أو التنقيب في الآراء opinion mining بأنه تعريف أتوماتيكي للآراء المُعبر عنها في النص تجاه موضوع معين [1] .

يمكن القول إن السبب الرئيسي وراء ازدياد الاهتمام بتحليل المشاعر هو انتشار استخدام شبكة الانترنت التي شكلت بيئة مميزة لتفاعل البشر على اختلاف أماكنهم وأفكارهم، حيث زاد حجم البيانات المتبادلة بين مستخدمي الانترنت وهذا ما ساعد على زيادة التواصل وتبادل المعارف والمشاعر حول ظاهرة أو منتج معين، وبذلك أيضاً توفر للشركات وللمحللين بيانات خام لإجراء الإحصائيات اللازمة لمعرفة المنتجات الأكثر رغبة من قبل المستهلكين أو معرفة أراء الناس حول إحدى الظواهر بين إيجابي أو سلبي.

تجري عملية تحليل المشاعر على مستويات عدة من التقسيمات: الكلمة، العبارة، الجملة، المستند، أو المفهوم. كما توجد طريقتان أساسيتان لتطبيق تحليل المشاعر: الطرق التي تعتمد على القواميس وطرق التعلم الآلي.

تعتمد الطرق الأكثر شيوعاً لتحليل المشاعر على خوارزميات التعلم الآلي التقليدية والتي بدورها تستخرج السمات من النصوص، وتحدد السمات الأكثر أهمية ثم تتم عملية التصنيف باستخدام إحدى تلك الخوارزميات. إذاً تعتبر المشكلة الفعلية للطرق التقليدية أنها تعتمد على استخراج السمات يدوياً وهذا مايعد استهلاكاً للوقت. إن استخراج هذه السمات في اللغة العربية مكلف ويحتاج الكثير من عمليات المعالجة الأولية إضافةً إلى

تعدد اللهجات فيها هذا ما أدى إلى عدم وجود أي طريقة فعالة وموحدة للوصول إلى السمات أو معالجة النصوص. تعد عملية مراقبة وتحليل مواقع التواصل الاجتماعي والمدونات ومواقع الآراء أمراً صعباً وذلك بسبب الانتشار الواسع وكمية البيانات الهائلة، لذلك تتم عملية تحليل المشاعر باستخدام العديد من التقنيات بهدف الوصول للدقة الأعلى والأداء الأفضل.

## 2. الهدف من البحث:

تستخدم أساليب التعلم الالي بشكل متزايد في مجال تحليل النصوص ولكن هذه الأساليب قد لا تكون الأمثل في كل الحالات. سنوضح في هذا البحث أهمية استخدام التعلم العميق مع البيانات النصية وهو أسلوب قادر على استخراج الميزات آلياً وذلك بهدف إثراء فهمنا وتحسين قدرتنا على التنبؤ بالمشكلة المطروحة.

سنستعرض في هذا البحث بعض الجهود السابقة في مجال تحليل المشاعر، ونحاول تفادي بعض الثغرات واقتراح عدد من الحلول بهدف تحسين أداء الخوارزميات المستخدمة.

يهدف هذا البحث فعلياً إلى تقديم دراسة عامة حول التعلم العميق وعملية تضمين الكلمات

إضافةً إلى مقارنة تحليلية لعدة خوارزميات تعلم عميق على مجموعتين من البيانات باللغة العربية بهدف تحديد الخوارزمية الأفضل.

### 3. أدوات البحث:

يتم التعامل في هذه الورقة مع نظام تشغيل windows 10/64 bit منصب على جهاز له المواصفات التالية:

- المعالج: intel core i7-10510U 2.30 GHz
- الذاكرة: 16 غيغا بايت.

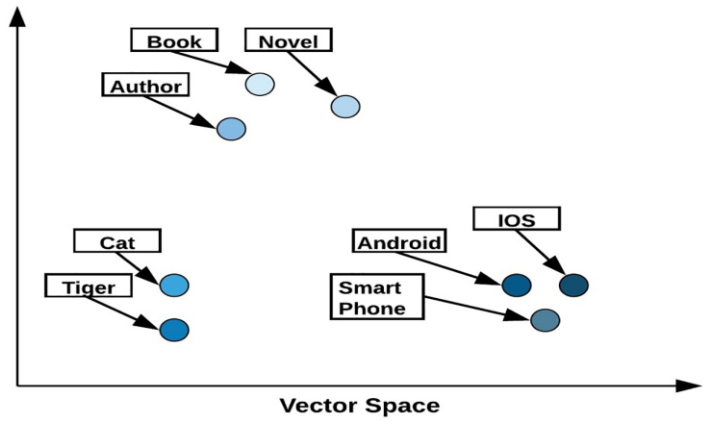
كما تم استخدام google colab لتطبيق الخوارزميات.[2]

### 4. تضمين الكلمات Word Embedding:

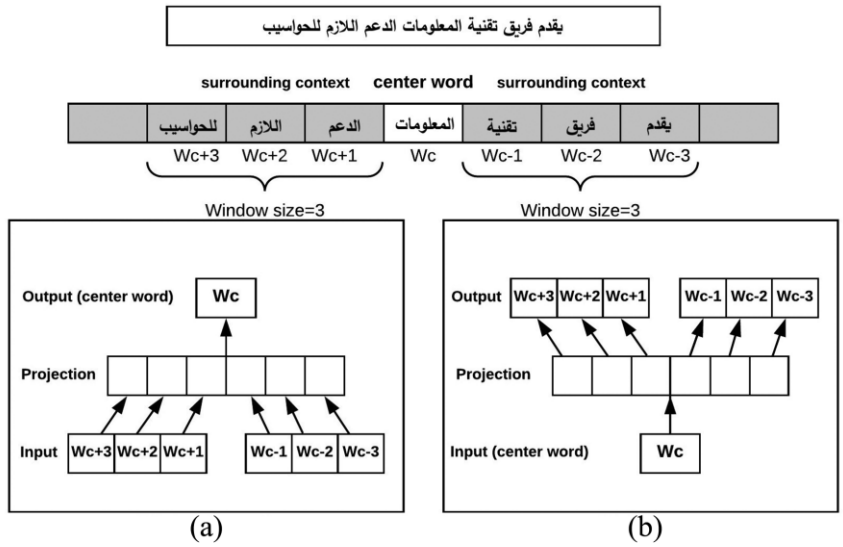
إن الهدف الأساسي لتضمين الكلمات هو تمثيل هذه الكلمات كأشعة مؤلفة من قيم رقمية مستمرة، لذلك الكلمات المتشابهة المعنى تكون أشعتها متقاربة كما يظهر في الشكل(1). يحقق مبدأ تضمين الكلمات تمثيلاً مكثفاً (dense) للسمات بدلاً من الأشعة المبعثرة (sparse vectors) التي تعتمد على التردد frequency التي يقدمها TF-IDF وغيرها، حيث إن التمثيل السابق يعطي أداءً أفضل مع الشبكات العصبونية. يعتبر word2vec و Global Vectors (GloVe) أشهر نماذج تضمين الكلمات.[3]

نلاحظ في الشكل(2) طريقتين لتحقيق word2vec: الأولى Skip-gram والثانية CBOW (Continuous Bag-Of-Word). [4]

تعمل CBOW على توقع الكلمة الهدف (الحالية) اعتماداً على السياق أي الكلمات المحيطة بها (الكلمات السابقة والتالية)، أي أن هذه الكلمة هي الخرج بينما باقي الكلمات هي الدخل. أما فيما يخص Skip-gram فهي تعمل بالعكس أي تأخذ الكلمة كدخل وتتوقع من خلالها الكلمات المحيطة.




شكل(1): Word Embedding




شكل(2): بنية word2vec

(a) CBOW and (b) SG

استخدمنا في هذا البحث نموذجاً مدرباً مسبقاً على مجموعة ضخمة من النصوص العربية لتضمين الكلمات هو Aravec [5] وقد تم تدريب هذا النموذج على بيانات من

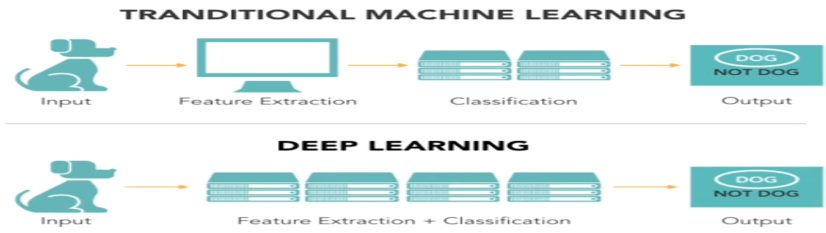
مصادر متعددة كتويتر، ويكيبيديا، وصفحات الويب، حيث ستنتج مصفوفة تضمين الكلمات اعتماداً على نموذج CBOW ذات الابعاد 300 (بعد التجريب والاطلاع على الأبحاث السابقة كلما زادت الابعاد تحسن الاداء). سنمثل كل جملة بشعاع من الشكل n x d حيث أن n هي عدد الكلمات في كل جملة و d هي طول شعاع التضمين (هنا d=300)

### 5. التعلم العميق: [6]

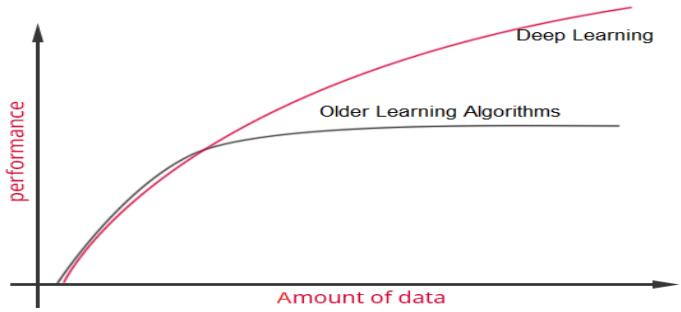
أحد أشكال التعلم الآلي، يتضمن خوارزميات تتيح للآلة التعلم بنفسها عن طريق محاكاة الخلايا العصبية.

يمكن أن يقدم التعلم العميق دقة عالية قريبة من دقة الأداء البشري وذلك لأنه يحاكي الخلايا العصبية، وتستخدم نماذج التعلم العميق لمجموعات البيانات الكبيرة المحددة الصفوف حيث تتعلم المميزات مباشرةً من البيانات دون الحاجة لعملية الاستخراج اليدوي لها. يوضح الشكل(3) الفرق بين التعلم الالي التقليدي الذي يقوم باستخراج الميزات يدوياً ثم تتم عملية التصنيف أما في التعلم العميق فإنّ الميزات يتم توليدها تلقائياً.

نلاحظ من الشكل (4) أن أداء شبكات التعلم العميق يزداد بازدياد حجم البيانات بينما أداء خوارزميات التعلم الآلي التقليدية يزداد إلى حد معين ثم يصبح ثابتاً مهما ازداد حجم البيانات.

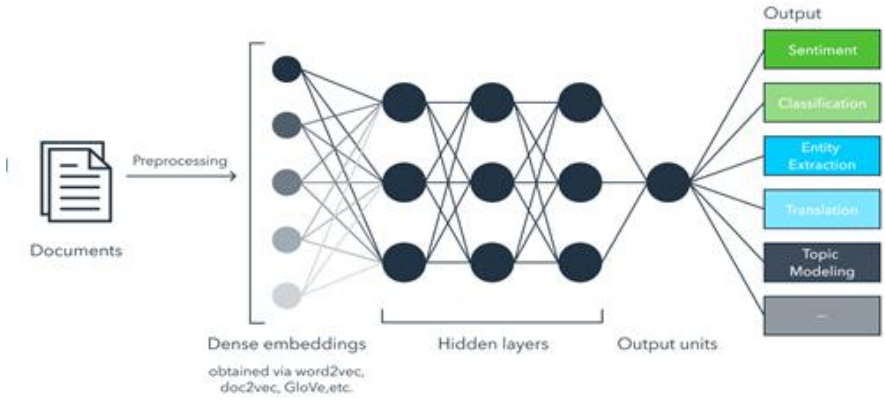



الشكل –3– : الفرق بين التعلم الالي التقليدي والتعلم العميق.




الشكل (4) : أداء التعلم العميق والخوارزميات التقليدية بازدياد حجم البيانات

يوضح الشكل (5) أن بنى التعلم العميق ماهي إلا شبكات عصبونية ذات طبقات مخفية كثيرة العدد (كلما كان عددها أكبر كلما كانت الشبكة أعمق)، يتم معالجة المستندات وتمرر إلى طبقة الدخل للشبكة العصبونية حيث يعبر عن كل كلمة بشعاع مستخدماً موقع كل كلمة في الجملة لإيجاد التشابه العاطفي بين الكلمات وهذا ما يسمى Word2vec أو يتم حفظ ترتيب الكلمات في الجملة إضافةً إلى ما سبق (Doc2vec) وغيرها الكثير من الأساليب لأنظمة تضمين الكلمات word embedding. ثم ترسل للطبقات المخفية ومنها لطبقة الخرج التي تعطي الخرج النهائي (الشعور، التصنيف، الخ).



الشكل (5) : معالجة اللغات الطبيعية باستخدام شبكة تعلم عميق.

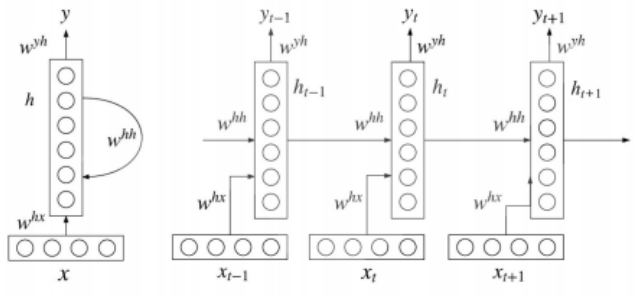
إنّ أشهر خوارزميات التعلم العميق المستخدمة في مجال تحليل النصوص هي: [7]

- **الشبكات العصبونية التكرارية Recurrent Neural Network (RNN):**

أحد أنواع شبكات التعلم العميق تحتوي على حلقات داخل الشبكات (تستخدم المعلومات السابقة–الذاكرة– لتقدير القيمة التي تليها)، تفيد في التنبؤ بالقيمة التالية من الجملة.

هناك حدود لهذه الشبكة حيث إنه من المفترض أن تحمل RNN المعلومات حتى وقت طويل، لكن من الصعب نشر كل هذه المعلومات عندما تكون الخطوة الزمنية طويلة جداً. وعندما تكون الشبكة تحوي عدداً كبيراً من الطبقات المخفية تصبح غير قابلة للتدريب وتسمى هذه المشكلة: مشكلة اختفاء التدرج vanishing gradient problem ، حيث إن الشبكة تقوم بتحديث الوزن باستخدام خوارزمية النسب المتدرجة [7]، تنمو التدرجات بشكل أصغر عندما تتقدم الشبكة إلى الطبقات السفلية. إذا بقي التدرج ثابتاً يعني أنه لا يوجد مساحة للتحسين فإذا كان هذا الفرق صغيراً(تتغير الأوزان قليلاً) فلن تتعلم الشبكة أي شيء.

يوضح الشكل (6) شبكة عصبونية متكررة حيث أن x يمثل الدخل و y الخرج بينما w فهو يمثل الوزن.



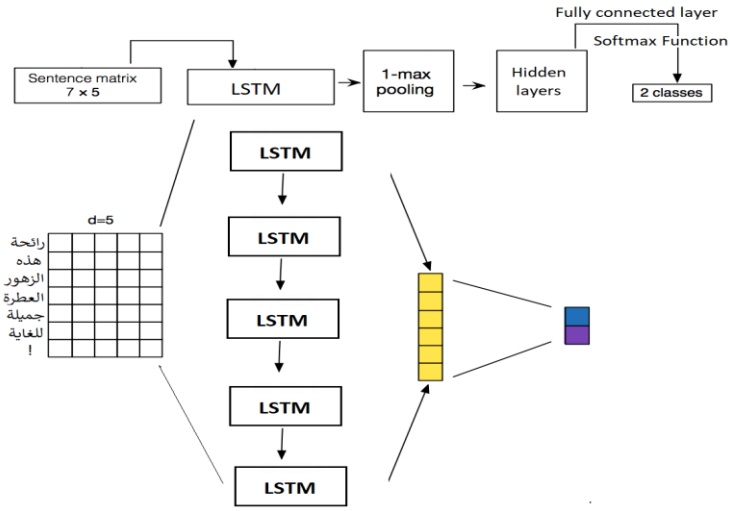
الشكل (6): شبكة عصبونية متكررة.

وهناك نوع خاص من RNN هو LSTM.

- **الذاكرة طويلة قصيرة الأمد (LSTM)**: تعد نوعاً خاصاً من شبكات RNN من فوائدها بالإضافة إلى تعلم التسلسلات الطويلة أنها يمكن أن تتعلم كيفية عمل تنبؤ متعدد الخطوات من خطوة واحدة والتي قد تكون مفيدة للتنبؤ بسلسلة زمنية ومناسبة لتصنيف النصوص. تكمن الصعوبة فيها أنه قد يكون من الصعب تكوينها وقد تتطلب الإعداد للحصول على البيانات بالتنسيق الصحيح للتعلم.

إنّ شبكة LSTM تقرأ الدخل من اليسار إلى اليمين وذلك حسب اللغة طبعاً، بينما BI–LSTM تتألف من طبقتين الأولى تقرأ من اليسار إلى اليمين والثانية تقرأ بالعكس، يعد النوع الثاني جيداً عندما يكون السياق( context )للدخل النصي مهماً على سبيل المثال عندما يأتي مصطلح النفي بعد المصطلح الإيجابي.

يظهر الشكل(7) بنية مصغرة لشبكة LSTM. إنّ خرج طبقة التضمين هو قائمة من الأشعة لكل الكلمات في الجملة المراد تحليلها. الطبقة التالية هي طبقة LSTM التي تعمل بالتغذية العودية حيث إن خرج الشبكة السابق ماهو إلا دخل للشبكة الحالية. فيما بعد يتم تحويل الخرج من 3D إلى 2D. ثم يتم استخدام طبقة Max Pooling والتي تقلل أبعاد البيانات لدينا، حيث يتم اختيار القيمة العظمى لكل جزء (patch) من البيانات. تعتبر المجموعة المكونة من القيم العظمى التي تم الحصول عليها مجموعة البيانات الجديدة (الأقل حجماً) والتي ستستخدم في الطبقة التالية(Dropout Layer).

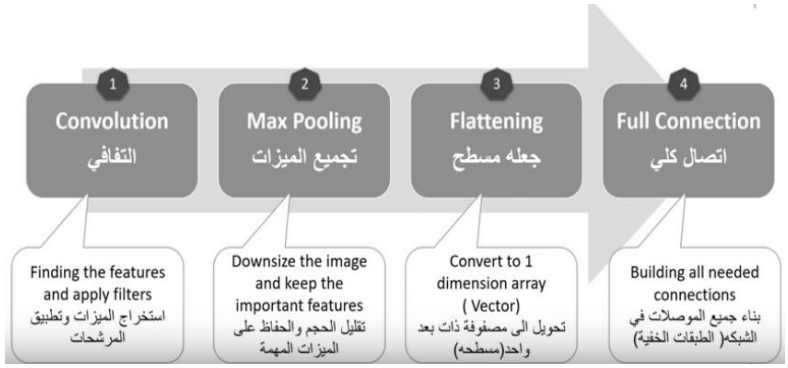



شكل(7): شبكة LSTM تستخدم للتصنيف الثنائي [8] .

- **الشبكات العصبونية الالتفافية Convolutional Neural Networks(CNN):** نوع خاص من الشبكات العصبونية ذات التغذية الأمامية المستخدمة في مجال الرؤية الحاسوبية. تتكون هذه الشبكات من العديد من الطبقات الالتفافية والتي توجد المميزات المهمة. هناك أنواع مختلفة من الطبقات في CNN فمثلاً طبقة الدخل تحمل البيانات الخام بينما الطبقة الالتفافية فهي مسؤولة عن حساب الخرج من خلال عملية الجداء النقطي ( dot product) بين الفلاتر وأجزاء الدخل [9]. تعد مسؤولية طبقة تابع التنشيط تطبيق أحد توابع التنشيط على طبقات الالتفاف، ويمكن أن يكون تابع التنشيط من الشكل (Sigmoid، ReLU، tanh، Softmax،..). يعتبر ال Sigmoid تابع تنشيط غير خطي يحول القيم بين 0 و1 ، إضافةً إلى ال ReLU (Rectified Linear Unit) الذي يعتبر تابع تنشيط غير خطي أيضاً. يستخدم تابع ال Softmax بشكل كبير في مسائل التصنيف الثنائي والتصنيف المتعدد أيضاً. تستخدم طبقة التجميع (Pool) لتقليل حجم البيانات وزيادة الفعالية، إلا أن الهدف الأساسي من إضافتها في شبكة ال CNN هو منع التلاؤم المفرط

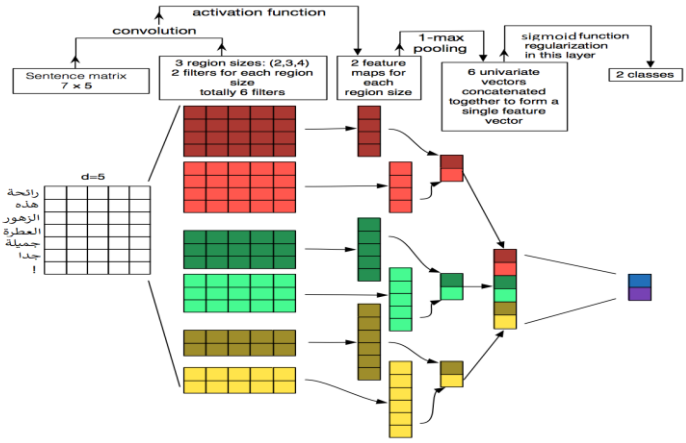
Overfitting. هناك عدة أشكال لهذه الطبقة يمكن أن يكون max pooling أو average pooling . تعد طبقة Fully Connected Layer أخر طبقة في الشبكة وتأخذ دخلها من الطبقة السابقة وعملها الأساسي هو حساب نتيجة التصنيف ويكون خرجها عبارة عن مصفوفة أحادية البعد حجمها يساوي عدد الصفوف.

يظهر الشكل(8) مكونات الشبكة الالتفافية CNN ووظيفة كل منها في كل مرحلة من المراحل.




الشكل(8): مكونات الشبكة الالتفافية ومهامهم.

يظهر الشكل(9) أن البيانات تدخل إلى الطبقة الالتفافية التي تعمل على إزاحة(تحريك) الفلتر أو window على بيانات الدخل وبحركته يتم تطبيق بعض العمليات على تلك البيانات.




الشكل(9): بنية CNN لتصنيف الجمل تعمل على مستوى الكلمة [8].

## 6. الدراسات المرجعية:

نركز في بحثنا هذا على الدراسات التي تناولت اللغة العربية باستخدام التعلم العميق .

### ❖ دراسات باللغة العربية:

استخدم الباحثون [10] تقنيات التعلم العميق لتحليل المشاعر باللغة العربية على مجموعة بيانات كبيرة مؤلفة من 40K تغريدة عربية تتناول عدة مجالات منها الرياضة، الصحة، السياسة، الخ. حيث تم تطبيق ثلاث تقنيات وهي CNN، LSTM ، RCNN (وهي عبارة عن تطبيق CNN ثم طبقة LSTM ) . وذلك بمساعدة تقنيات تضمين الكلمات كطبقة دخل للنماذج الثلاثة. لُوحظ تفوق خوارزمية LSTM بدقة وقدرها 81.3% على CNN التي أعطت دقة 75.72% و RCNN بدقة 78.46%.

استخدم Alwehaibi وآخرون[11] LSTM لتصنيف النصوص العربية اعتماداً على تقنيات مختلفة لتضمين الكلمات مثل Aravec ، ArabicNews ، FastText لدراسة تأثير هذه التقنيات على دقة النموذج. بداية تم استخدام مجموعة بيانات عبارة عن تغريدات تويتر تم معالجتها باستخدام التقنيات المذكورة سابقاً لتوليد شعاع لكل كلمة. ثم يتم إدخال هذه الأشعة إلى طبقة LSTM مؤلفة من 128 عقدة في الطبقة المخفية

لتصنيف كل تغريدة ضمن ثلاثة صفوف (إيجابي، سلبي، محايد). تفوق FastText بالدقة على باقي التقنيات المستخدمة وذلك في حال كانت مجموعة البيانات متوازنة.

قدم El-Kilany [12] نموذجاً للتعرف على المشاعر وطبّق على مجموعة تغريدات باستخدام طبقتين: طبقة تضمين للمحرف و word2vec و Bi-LSTM مع طبقة تصنيف CRF. تستخدم Bi-LSTM الأشعة المتولدة عن الطبقة الأولى لتوليد الميزات من أجل المصنف الذي يستخدمها لتوقع الصف الذي تنتمي له الكلمة اعتماداً على الكلمة السابقة والتالية في الجملة. أثبتت التجارب أن النموذج المقترح تفوق على نفس النموذج والذي يستخدم فقط تضمين المحرف بدون word2vec .

❖ **دراسات باللغة الإنكليزية:**

قدم الباحثون في هذه الدراسة [13] شبكة تتألف من طبقة CNN وطبقة LSTM لتوقع المشاعر في مراجعات المستخدمين. حيث إن أهم ميزات هذه الشبكة أنها غير موجهة لمجال معين وليس هناك أية تفاصيل متعلقة باستخراج الميزات وذلك بفضل استخدام نموذج تضمين الكلمات لتمثيل أشعة ميزات الدخل، ثم يتم استخدام شبكة CNN لتحديد الميزات المهمة وهذا ما يحسن زمن التدريب.

## 7. القسم العملي:

### 7-1- إعدادات التجربة:

- **تقسيم البيانات:** تم تدريب واختبار النماذج باستخدام بايثون، وتقسيم كل مجموعة بيانات إلى مجموعات تدريب train بنسبة 70% وتحقق validation بنسبة 10% واختبار test بنسبة 20%.

- **المكتبات المستخدمة:** تم استخدام مكتبة Keras لتنفيذ التجربة[14]، واستخدمنا Keras checkpoint function لحفظ الأوزان الأفضل للنموذج. تم اختيار البارمترات التي تعطي النتائج الأفضل بعد عدة تجارب لكل نموذج وتم تحديدها باللون الغامق في كل جدول.

**7-2- مجموعة البيانات:** سنستخدم في هذا البحث مجموعتي بيانات باللغة العربية الأكثر استخداماً في مجال تحليل المشاعر:

- المجموعة الأولى(LABR):[15] وهي مجموعة بيانات كبيرة لمراجعات باللغة العربية الفصحى (LABR)، تتضمن حوالي 63000 مراجعة لكتب جُمعت من موقع goodreads.com. تتألف كل مراجعة review من: تقييم المستخدم rating، رقم المراجعة review id، رقم المستخدم user id، رقم الكتاب book id ، ونص المراجعة review. تم تقييم كل كتاب على هذا الموقع ضمن المجال من 1 إلى 5 ومن ثم تم تحويلها إلى قطبية مشاعر محددة (سلبي أو إيجابي).

-المجموعة الثانية: [16] وهي مجموعة بيانات باللغة العربية العامية مأخوذة من مواقع التواصل الاجتماعي تتضمن حوالي 1200 سجل، ومقسمة إلى ثلاث فئات (إيجابي، سلبي، أو محايد).

| | إيجابي | سلبي | محايد |
|---|---|---|---|
| LABR(balanced) | 6580 | 6578 | - |
| Dataset2(balanced) | 434 | 438 | 405 |

الجدول(1): توزيع البيانات ضمن الفئات في مجموعتي البيانات.

**7-3- عمليات المعالجة الأولية:** سيتم تطبيق عمليات المعالجة الأولية على مجموعات البيانات المستخدمة وذلك على عدة مراحل:

- التقطيع tokenization : فيما يخص عملية التقطيع سنستخدم Natural Language Toolkit (NLTK ).

- التقييس Normalization :بدايةً ستتم عملية التقييس في بحثنا هذا تبعاً لعدة قواعد وهي: حذف إشارة "_" ،علامات الترقيم، الأعداد، المحارف الخاصة، الأحرف غير العربية، والأحرف المكررة. مثلاً سنستبدل كل المحارف ( أ، إ، آ) بالحرف ( ا ) والأحرف ( ئ، ؤ) بالحرف (ء).

- إزالة كلمات التوقف: أيضاً سنستخدم قائمة كلمات التوقف العربية الخاصة بNLTK والتي تحوي 248 كلمة.

- إعادة الكلمات إلى أصلها stemming: سنستخدم المجذع الخفيف( تاشفن tashaphyne) الذي يقوم بتجذيع الكلمة، استخراج الجذر، تقطيع الكلمة إلى جميع الحالات الممكنة [17].

- lemmatization: وتستخدم لتحديد الكلمات الثانوية حيث تقوم بتحويل الصيغة البديلة إلى الصيغة الأساسية وبذلك يقل عدد الكلمات.

**7-4- طبقة تضمين الكلمات لشبكة التعلم العميق:** في هذه المرحلة نستبدل كل مراجعة (review) بشعاع تضمين من النموذج AraVec المدرب مسبقاً. في حال كانت الكلمات الموجودة في النص ليس لها تضمين في النموذج المستخدم، نبحث عن الكلمات الأكثر تشابهاً ونستخدمهم لتمثيل شعاع التضمين الخاص بالكلمة. قمنا بتثبيت الطول الأعظمي للكلمات في الجملة على الطول 100 وأي مراجعة تحوي أقل من العدد الأعظمي للكلمات سيضاف لها padding لنحصل على نفس الطول لكل المراجعات.

**7-5- تطبيق CNN:** يوضح الجدول(2) طبقات شبكة CNN المستخدمة، أما فيما يخص إعدادات التجربة التي قمنا بها فهي موضحة في الجدول(3).

| 1- طبقة الالتفاف الأولى: |
|---|

| | |
|---|---|
| عدد الفلاتر | 512 |
| أبعاد الفلتر Kernel size | 3 |
| تابع التنشيط activation | ReLU(Rectified Linear Unit) |
| Strides | 1 |
| 2- طبقة التجميعMax pooling | |
| 3- طبقة الالتفاف الثانية: | |
| عدد الفلاتر | 256 |
| أبعاد الفلتر Kernel size | 3 |
| تابع التنشيط activation | ReLU |
| Strides | 1 |
| 4- طبقة التجميعMax pooling | |
| 5- Fully connected Layer (Dense Layer) | |
| 6- تابع التصنيف softmax | |

الجدول

(2): طبقات CNN المستخدمة.

| | |
|---|---|
| Batch size | 20,**30**,50 |
| Pooling | 2 |
| Epoch | 10,40,**50**,100 |
| Optimizer | Adam |
| Learning rate | 0.001 |

الجدول (3): إعدادات التجربة.

**7-6- تطبيق LSTM :**

بعد عدة تجارب للوصول للمعاملات الأفضل استخدمنا طبقتي LSTM الأولى حجمها 128 والثانية 64 متبوعةً بطبقة متصلة بالكامل fully connected layer تستخدم تابع التنشيط Relu وطبقة التسريب dropout بقيمة (0.5) ونهاية سنستخدم التابع softmaxللتوقع. يوضح الجدول (4) المعاملات المستخدمة لشبكة LSTM .

| | |
|---|---|
| Batch size | 20,**30**,50 |
| Epoch | 10,**40**,50,100 |
| Dropout | 0.5 |
| Optimizer | Adam |
| تابع التنشيط Activation Function | Softmax |

الجدول (4): المعاملات المستخدمة لشبكة LSTM.

### 7-7- تطبيق خوارزمية CNN مع LSTM:

سنقدم في هذه الطريقة نموذجاً يعتمد على دمج CNN مع LSTM ،تم استخدام شبكة CNN لتستخرج الميزات أما LSTM تهتم بتسلسل ترتيب الكلمات وهذا الأمر مهم في حالة الجمل المنفية أو السخرية Sarcasm ، قمنا بزيادة عدد الفلاتر المستخدمة في طبقة الالتفاف للحصول على ميزات مختلفة وزيادة عدد خلايا LSTM لتتناسب مع خرج طبقة الالتفاف. أيضاً قمنا بحذف طبقة Max-pooling (المسؤولة عن تقليل أبعاد الدخل) لتجنب فقدان أي ميزة قبل إرسال هذه الميزات إلى طبقة LSTM. (عدد ال Epoch المستخدم هنا هو 50).

- طبقة الالتفاف: إنَّ أبعاد الفلتر المستخدم في هذه الطبقة هو (3)، أما عدد الفلاتر فهو(n-1) حيث إن n هي عدد الكلمات في الجملة في حين أن ال strides قيمتها (1). يتحرك كل فلتر على طبقة الدخل من البداية إلى النهاية لاستخراج الميزات الأفضل لكل مراجعة review باستخدام تابع التنشيط ReLU . ثم تدمج هذه الميزات على شكل أشعة لتشكل خرائط الميزات عددها يساوي عدد الفلاتر.

- طبقة التسريب dropout: بقيمة (0.5) يعد هدف هذه الطبقة بشكل أساسي منع ال overfitting .
- طبقة LSTM : سنستخدم ثلاث طبقات LSTM كل طبقة تحوي خلايا LSTM بنفس عدد الفلاتر المستخدمة في طبقة الالتفاف، ثم سنجرب استخدام طبقتين فقط بهدف المقارنة.
- طبقة Fully Connected Layer : يتم هنا دمج خرج الطبقة السابقة في شعاع واحد وتحويله لقيمة واحدة ومن ثم تمرير هذه القيمة إلى طبقة التنشيط لتحديد الصف الذي تنتمي له هذه المراجعة وذلك باستخدام التابع softmax .

## 7-8- التجربة والنتائج:

سنستخدم أربعة معايير شائعة لقياس دقة التصنيف وهي:[18 ]

(1) Accuracy: يقيس هذا المعيار كم عينة تم توقعها بشكل صحيح نسبةً إلى عدد التوقعات الكلي ويُعطى وفق الصيغة التالية:

$$Accuracy = \frac{TP + TN}{TP + FN + FP + TN}$$

(2) Precision: يقيس هذا المعيار نسبة العينات التي تم تصنيفها بشكل صحيح إلى العدد الكلي للعينات التي تم تصنيفها بشكل صحيح والعينات التي تم تصنيفها بشكل خاطئ misclassified، ويُعطى وفق الصيغة التالية:

$$Precision = \frac{TP}{TP + FP}$$

(3) Recall: يقيس هذا المعيار نسبة العينات التي تم تصنيفها بشكل صحيح إلى العينات التي تم تصنيفها بشكل صحيح والعينات التي لم يتم تصنيفها, ويُعطى وفق الصيغة التالية:

$$Recall = \frac{TP}{TP + FN}$$

(4) F-Measure: دمج للمعيارين السابقين Recall و Precision, حيث إنه كلما كانت قيمته أكبر كان الأداء أفضل، ويُعطى وفق الصيغة التالية:

$$F - Measure = 2 * \left(\frac{Precision * Recall}{Precision + Recall}\right)$$

يمكن القول إن الدقة هي مقياس جيد عندما تكون مجموعة البيانات متوازنة، بينما AvgRec او f1 هو المقياس الأفضل في حال مجموعات البيانات غير المتوازنة. تم تنفيذ النموذج أكثر من مرة وهذه النتائج هي متوسط لنتائج التنفيذ أكثر من مرة[19].

- العدد الأمثلي لطبقات LSTM في المنهجية الثالثة:

بداية سنطبق المنهجية الثالثة باستخدام طبقتي LSTM ثم ثلاث طبقات لدراسة تأثيرها على أداء عملية التصنيف، نلاحظ من الجدول(5) أن استخدام ثلاث طبقات متتالية من LSTM يحسن الأداء بنسبة أكبر من استخدام طبقتين فقط، لذلك سيتم اعتماد منهجية الطبقات الثلاث لإنتاج النموذج الأفضل. هذه النتيجة تتوافق مع [20] الذي استنتج أن استخدام طبقات LSTM متتالية يحسن من دقة التصنيف.

| عدد الطبقات | مجموعة البيانات | Accuracy | Precision | Recall | F1-score |
|---|---|---|---|---|---|
| 2 | LABR | 82.2 | 81.9 | 84.1 | 82.9 |
| | Dataset2 | 86.8 | 85.3 | 88 | 86.6 |

| | | | | | |
|---|---|---|---|---|---|
| 3 | LABR | 83.4 | 82.1 | 85.2 | 83.5 |
| | Dataset2 | 87.1 | 85.9 | 89 | 87.4 |

الجدول(5): تأثير عدد طبقات LSTM المستخدمة.

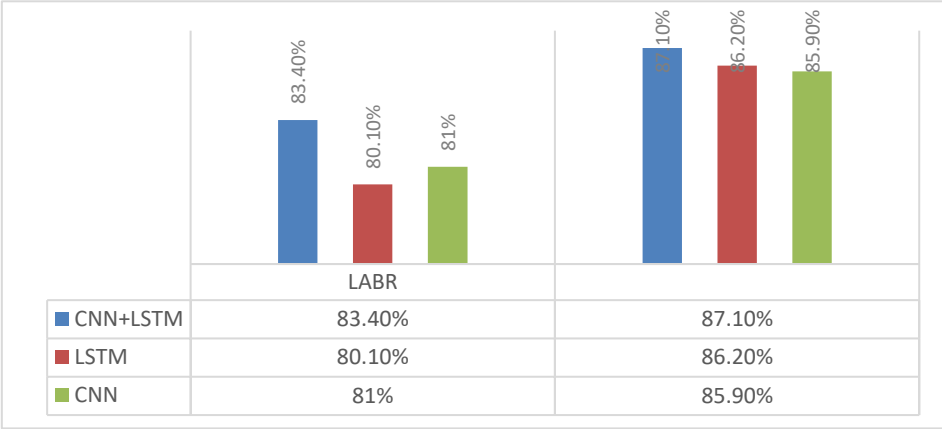
| DataSet | CNN | LSTM | CNN+LSTM |
|---|---|---|---|
| LABR | 81% | 80.1% | **83.4%** |
| Dataset2 | 85.9% | 86.2% | **87.1%** |

الجدول(6): مقارنة الدقة للخوارزميات الثلاثة.

| مجموعة البيانات | الخوارزمية المستخدمة | Precision | Recall | F1-score |
|---|---|---|---|---|
| LABR | CNN | 80.4% | 83.11% | 81.7% |
| | LSTM | 80% | 82.6% | 81.27% |
| | CNN+LSTM | 82.1% | 85.2% | 83.5% |
| Dataset2 | CNN | 84.6% | 86% | 85.2% |
| | LSTM | 85% | 87.1% | 86% |
| | CNN+LSTM | 85.9% | 89% | 87.4% |

الجدول(7): نتائج مقارنة الطرق الثلاثة على مجموعتي البيانات.

نلاحظ من الشكل(10) أن المنهجية CNN+LSTM الأفضل من حيث الدقة بين الخوارزميات.




الشكل–10– : مقارنة الدقة للطرق الثلاثة.

أما بالنسبة لزمن التدريب المستهلك من قبل المنهجيات الثلاثة فهذا ما يوضحه الجدول(8) حيث نلاحظ أن CNN هي الخوارزمية الأسرع من ناحية زمن التدريب وبذلك تكون الأفضل في حالة مجموعة البيانات الكبيرة.

| DataSet | CNN | LSTM | CNN+LSTM |
|---|---|---|---|
| LABR | **2h 27m** | 3h 10m | 3h 2m |
| Dataset2 | **43m 3s** | 50m 6s | 48m 4s |

الجدول(8): مقارنة الزمن للمنهجيات الثلاثة.

## 8. الاستنتاجات والتوصيات:

قدمنا في هذا البحث دراسة عامة عن خوارزميات التعلم العميق وعملية تضمين الكلمات، وقمنا بتحليل بعض الأعمال السابقة باللغة العربية والإنكليزية.

فيما بعد طبقنا ثلاث منهجيات وخوارزميات للتعلم العميق وهي CNN و LSTM والمنهجية الأخيرة تعتمد على دمج الخوارزميتين السابقتين CNN+LSTM على التتالي. لنجد أن CNN+LSTM هي الأفضل بالنسبة لمجموعة البيانات الأولى والثانية.

أيضاً يمكن ملاحظة أن CNN أفضل من LSTM من حيث الدقة في حالة مجموعة البيانات الأولى LABR والتي صنفت البيانات فيها إلى صنفين (إيجابي، سلبي) إضافةً إلى أنها استغرقت زمن تدريب أقل من باقي الخوارزميات، ويجدر بنا أن نذكر أن مجموعة البيانات الأولى أكبر حجماً من المجموعة الثانية بفارق كبير وهذا ما ذكر سابقاً.

إلا أن LSTM كانت أفضل من CNN من حيث الدقة في حالة مجموعة البيانات الثانية والتي صنفت البيانات فيها إلى ثلاثة صفوف(إيجابي، سلبي، محايد).

يمكن لنا في بحثنا القادم الدمج بين خوارزميات التعلم الآلي التقليدية والتعلم العميق بهدف الوصول لمنهجية أفضل أو دمج خوارزميات التعلم العميق ولكن بأسلوب مختلف عن الطرق المستخدمة في هذا البحث.

## 9. المراجع:

[1] Bing Liu. Sentiment Analysis and Opinion Mining, Morgan &Claypool Publishers, May 2012.

[2] https://colab.research.google.com/

[3] Alharbi, A., Taileb, M. and Kalkatawi, M., 2021. Deep learning in Arabic sentiment analysis: An overview. Journal of Information Science, 47(1), pp.129-140.

[4] Mikolov T, Sutskever I, Chen K et al. Distributed representations of words and phrases and their compositionality. In:Proceedings of the advances in neural information processing systems, 2013

[5] A. B. Soliman, Eissa, K., and El-Beltagy, S. R. Aravec(2017): A Set of Arabic Word Embedding Models for Use in Arabic Nlp. Procedia Computer Science.

[6] Kaur, H. and Mangat, V., 2017, February. A survey of sentiment analysis techniques. In 2017 International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud)(I-SMAC) (pp. 921-925). IEEE.

[7] Jha S, Kumar R, Abdel-Basset M, Priyadarshini I, Sharma R, Long HV (2019) Deep learning approach for software maintainability metrics prediction. IEEE Access 7:61840–61855

[8] Ye Zhang and Byron Wallace. 2015. A sensitivity analysis of (and practitioners' guide to) convolutional neural networks for sentence classification.

[9] Priyadarshini I, Puri V (2021) A convolutional neural network (CNN) based ensemble model for exoplanet detection. Earth Sci Inf 1–13

[10] Mohammed, A. and Kora, R., 2019. Deep learning approaches for Arabic sentiment analysis. Social Network Analysis and Mining, 9(1), pp.1-12.

[11] Alwehaibi A, Roy K (2018) Comparison of pre-trained word vectors for Arabic text classification using deep learning approach. In: 2018 17th IEEE international conference on machine learning and appli-cations (ICMLA), pp 1471–1474

[12] El-Kilany A, Azzam A, El-Beltagy SR (2018) Using deep neural networks for extracting sentiment tar-
gets in Arabic Tweets. In: Shaalan K, Hassanien AE, Tolba F (eds) Intelligent natural language
processing: trends and applications. Springer, Cham, pp 3–15

[13] R. K. Behera, M. Jena, S. K. Rath, and S. Misra, "Co-lstm: convolutional lstm model for sentiment analysis in social big data," Information Processing & Management, vol. 58, no. 1, Article ID 102435, 2021.

[14] Gulli, A., Pal, S.: Deep Learning with Keras. Packt Publishing Ltd., Birmingham
(2017)

[15] http://www.mohamedaly.info/datasets/labr

[16]http://saifmohammad.com/WebDocs/Arabic-Sentimnt Corpora/bbn_shared-2.xls

[17]T.Zerrouki,Tashaphyne,Arabic lights stemmer,

https://pypi.python.org/pypi/ Tashaphyne/0.2

[18] Fawcett T (2006) An introduction to roc analysis. Pattern Recognit Lett 27(8):861–874

[19]Brownlee J. Imbalanced classification with python: Better metrics, balance skewed classes, cost-sensitive learning. Machine Learning Mastery; 2020

[20] Pal S, Ghosh S, Nag A (2018) Sentiment analysis in the light of LSTM recurrent neural networks. Int J Synth Emot 9(1):33–39. https :// doi.org/10.4018/ijse.20180 10103