تحليل المشاعر بتقنيات الذكاء الاصطناعي

تحت إشراف:

أ.د. صلاح عبية- مدينة زويل للعلوم والتكنولوجيا والخبير بالمجمع

بالاشتراك مع:

لجنة الذكاء الاصطناعي – بالمجمع

تحليل المشاعر بتقنيات الذكاء الاصطناعي

المقدمة(*)

ترتبط اللغة بالمشاعر ارتباطًا وثيقًا يبدأ من أولى لحظات تعبير الإنسان عن مكنون صدره، فأول حروف تُلامس شفاه الرضيع في مهده في أولى كلماته البشرية هي حروف عربية تعكس أنهارًا من المشاعر المجردة، ثمّ يكبر قليلًا ليعبر عن مشاعره وآرائه المركبة بتلك الحروف نفسها. فثمة ثمانية وعشرون حرفًا احتضنته فكانت له المصدر في كل آنٍ وحين، فكيف للإنسان أن يُعبّر عن عاطفته التي تُطْرُقُ صمتَ العالم دون أن يغوص بين لآلئ لغته التي حوته وآنست وحشته. علاوةً على رابطها الوثيق مع مشاعر المرء فإن هذه الحروف تشكِّل لغة تمنح قارئها وكاتبها ثراءً عَزَّ نظيرُهُ في أي لغةٍ أخرى، لغة تتميز بقدرتها على التكيّف والإبداع؛ لذلك يستخدمها متحَدِّثوها للتعبير عن مشاعرهم في مختلف المجالات والأمدية. تلك المشاعر إن استُخرجت بدقة من مكنون الكلام العربي فإن لها القدرة على توظيف الكلام في مكانه الأمثل، فيما يعرف بتحليل المشاعر.

أ) التعريف: ما هو تحليل المشاعر؟

يهتم مجال تحليل المشاعر أو تعدين الرأي بمعالجة النصوص المكتوبة من أجل استخراج الشعور والانطباع الغالب عليه مثل السخرية، أو الحزن، أو الغضب، أو السعادة وغيرها. يمس تحليل المعنويات العديد من التطبيقات العملية، ونظرًا للكمِّ المتزايد من النصوص التي تعبر عن الآراء عبر الإنترنت، والتي تتركز في مواقع الشراء والمدونات التقييمية ومجموعات التواصل الاجتماعي، فإنه يمكن بمساعدة نظم تحليل المعلومات تحويل هذه النصوص غير المهيكلة آليًّا إلى بياناتٍ منظمة. تعد هذه البيانات ذات قيمة عالية لتطبيقات الأعمال على وجه التحديد، مثل تحليل السوق، وتعليقات المنتج، وخدمة العملاء. تعمل أنظمة تحليل المشاعر في إطار ثلاث مراحل:

- استخراج المعلومات العاطفية.
- تصنيف المعلومات العاطفية.
- استرجاع المعلومات العاطفية وتوظيفها.

(*) شارك في عمل هذا البحث:
– وفاء أشرف صلاح الدين، الطالبة بالفرقة النهائية بجامعة العلوم والتكنولوجيا بمدينة زويل.

ب) أنواع تحليل المشاعر

تستخرج هذه الأنظمة المشاعر في عدة صور ويتراوح عمق التحليل ونوعه بناءً على نوع التطبيق:

- **القطبية والقطبية الدقيقة:** تعد أبسط صور تحليل المشاعر وتصنف النصوص إلى آراء إيجابية/سلبية بوجه عام، في حين أن القطبية الدقيقة تصنف النصوص إلى خمس فئاتٍ: إيجابية جدًّا، وإيجابية، ومحايدة، وسلبية، وسلبية جدًّا.
- **كشف العاطفة:** يسمح بالحصول على تبصر أعمق لأبعاد الآراء القطبية ويسمح بتصنيف المشاعر لفئاتٍ أكثر، مثل الحزن، و الغضب، و الرضا... إلخ. إضافة إلى إتاحة استخراج الأسباب وراء عاطفة معينة.
- **تحليل الجانب:** يهتم هذا النوع بتحديد الجانب/الجوانب المعينة التي يناقشها كما يهتم بالنص، ويتم تحليل الانطباع عنها، ويكون هذا التحليل مفيدًا بوجه خاصٍّ في تحليل تقييمات مستخدمي المنتجات.
- **تحليل النية:** يعتبر من أعمق صور تحليل المشاعر ويُعنى بالتنبؤ بسلوك القارئ المستقبلي / احتياجاته بناءً على النص المعبّر عن رأيه، والذي يتطلب معرفة أكبر بالسياق الكلي.

ج) نطاق تحليل المشاعر

بجانب درجات/أنواع أنظمة تحليل المشاعر فإنه يمكن استخدام النوع الواحد في عدة نطاقات تعتمد أيضًا على التطبيق:

- **مستوى الجملة:** يهدف هذا النطاق إلى استنتاج العاطفة المسيطرة على الجملة الواحدة وهو الأكثر استخدامًا في المجال التجاري؛ حيث إن معظم الآراء والتقييمات تكون مقتضبة ومختصرة.
- **مستوى النص:** يهدف هذا النوع إلى استنتاج العاطفة أو فئة العواطف المسيطرة على النص الكلي، لذلك يتطلب معالجة مزدوجةَ الاتجاه حتى يتسق مع السياق الكلي والترابط بين الجمل.
- **مستوى المستند/مجموعة النصوص:** يتشابه هذا النوع مع النوع الثاني إلا أنه يعمل على مقياسٍ أكبر، ويأخذ في اعتباره وجود اختلافات بين العاطفة

المسيطرة على كل نص عند حساب العاطفة المسيطرة على مجموعة النصوص. يعتبر هذا النوع مجال بحثٍ ناشئ، ولم يشهد بعد تطورًا مماثلًا للنوعين الأولين.

د) تطبيقات تحليل المشاعر

إن معدل إنتاج المحتوى النصي في العالم يزداد باطراد كبير. وردود الأفعال عليه تكاد تكون لحظية؛ إذ إن قضية ما أو حادثًا ما قد ينتشر ويتصدر قوائم المشاهدة في بضع سويعات قليلة، وقد يؤدي إلى قرارات مصيرية للأشخاص المعنيين، مثل العقوبات الصارمة أو عملات المقاطعة. وبناءً علي ذلك وُلدت الحاجة لإيجاد طريقة لمعالجة وتصنيف البيانات الصادرة بشكل دائم ومستمر لإدراك ما يحدث واستغلاله، أو تجنب آثاره السلبية في أسرع وقت. وفيما يلي بعض الاستخدامات الأساسية لأنظمة تحليل المشاعر:

- **متابعة ومراقبة منصات التواصل الاجتماعي:** يوجد العديد من الجهات التي قد ترغب في مراقبة مواقع التواصل الاجتماعي من أجل معرفة ما يشغل الرأي العام الآن أو للبحث عن شكوى، أو كره موقف حادث، أو شركة بعينها. تفتح أنظمة تحليل المشاعر طريقة سريعة لإيجاد هذه الأحداث وتصنيفها في وقت زمني قصير جدًّا.
- **متابعة العلامات التجارية:** تهتم الشركات والمنشآت الضخمة بمعرفة رأي المستخدمين في منتجاتها بصفة خاصة، والشركة بصفة عامة. قد يؤدي رأي العامة إلى أخذ قرارات محورية في مستقبل العلامة التجارية، مثل الحملات الدعائية أو إطلاق منتج ما يُمَكّن تحليلُ المشاعر العلاماتِ التجاريةَ من مراجعة آراء العملاء في المنتجات وتصنيفها بكفاءةٍ عالية وتكلفة أقل من التحليل والتصنيف البشري.
- **الأبحاث السوقية:** تمكنك أنظمة تحليل المشاعر من معرفة رأي السوق في سلعة ما والسمات المستحبة والمكروهة بها، ويكون ذلك في غاية الأهمية عند بدء تصنيع منتج ما أو دخول سوق جديدة مع شركات منافسة ذات خبرة. كما يتيح تحليل المشاعر القيام بتحليل إحصائي للاستطلاعات وهو بدوره يتيح للباحثين في مجالات الأعمال والتسويق استنتاج المعلوماتِ والقرارات استنتاجًا

صحيحًا لأنه يكون ممثلًا لشرائح واسعة من المستهلكين.

- **خدمة العملاء:** تُعَدُّ أنظمة تحليل المشاعر أداة لمعرفة حالة العميل والتي قد تكون مؤشرًا للحاجة إلى تنفيذ إستراتيجياتٍ معينة أو تفعيلها للتعامل مع كل حالة للعميل، كما قد تكون مؤشرًا للحاجة إلى تدخلٍ بشريٍ عاجل في حالة خدمة العملاء الآلية أو خدمات الانتظار.
- **العلاقات العامة:** يمكن استخدام أنظمة تحليل المشاعر في متابعة تقبل/ رفض العامة لشخصية أو علامة تجارية.

هـ) التحديات التي تواجه أنظمة تحليل المشاعر

هناك الكثير من التحديات التي تواجه أنظمة تحليل المشاعر، والتي تعكس القدرة غير المسبوقة للغة العربية في التعبير عن نطاق واسع من المشاعر، فيما يلي بعضٌ من تلك التحديات:

- **التعامل مع الذاتية والموضوعية:** تعدُّ التفرقة بين التعبيرات الموضوعية من أولى الخطوات الأساسية في تحليل المشاعر، وهي المعتمدة على الحقائق الكائنة بالموضوع المعني والجمل الشخصية المعبرة عن رأي الكاتب فقط. يسمى هذا التصنيف "تحليل الذاتية".
- **السياق:** من المهم معرفة السياق والوقت والمكان الخاصين بالنص المكتوب. توجد معايير أخرى مثل كون النص قد كُتب ردًّا على نص آخر أم لا. من المهم أيضًا معرفة المجال الذي كُتب فيه النص، فمثلًا يمكن للنص الذي يعد إيجابيًّا في المجال السياسي أن يكون سلبيًّا في الفنون أو الرياضة.
- **محور التصنيف الشعوري:** هل يجب تصنيف كل المقال تحت شعور واحد، ومن الأفضل التعامل مع كل فقرة على حدة... بل يوجد من الطرق ما تُفَتِّتُ في التصنيف حتى يصل إلى تصنيف كل جملة، بل وكل كلمة على حدة. ما هي المشاعر التي يجب التصنيف بناءً عليها؟ أبسط الأنواع ما ينبغي فيها التفرقة بين المشاعر الإيجابية والمشاعر السلبية ليعبر عن تحليل المشاعر الثنائي. وقد وُضِعَتْ تصنيفات أكثر، مثل تصنيف الرسائل إلى مهمة وغير

مهمة أو مضللة. يمكن أيضًا وضع تدرج للمشاعر الإيجابية والسلبية مثل تقييم المنتجات من (١ إلى ١٠).

- **المفارقة:** يمثل الفرق بين المعنى الحرفي والمعنى المقصود تحديًا كبيرًا لأنظمة تحليل المشاعر. أكبر مثال على ذلك هو السخرية، حيث إن المصطلحات المُستخدمة في حالة السخرية قد يتم خلطها بسهولة بقطبية معاكسة، ويُتَطلَّبُ لتجنب ذلك تحليلٌ عميق لخلفية النص.
- **اللهجات:** يزيد تعددُ اللهجات الصعوبةَ على تصنيف المشاعر، إذ ينبغي تحديد اللهجة أولاً من أجل الحصول على أفضل نتائج ممكنة، حيث يساعد على تحديد سياق الحديث، ويمثل هذا الجانب صعوبة إضافية في المنطقة العربية لتعدد اللهجات وتداخلها في النص الواحد.
- **المصدر:** يحتاج التعرف على المشاعر في بعض المجالات إلى معرفة مصدر الكتابة ووجهته إذ إنها نقطة محورية لا يمكن الاستغناء عنها في بعض التطبيقات. فمثلًا لا يمكن الاستغناء عن تحديد المصادر في منصات التواصل الاجتماعي لمعرفة ما إذا كان القائل جهة دعائية أم فردًا مستهلكًا لمنتج معين أم شخصًا يقوم بكتابة رسائل مضللة.

أمثلة على التحديات التي تواجه أنظمة تحليل المشاعر

- "حدث زلزال في مصر"

هل يجب أن يكون هذا شعورًا سلبيًّا أم إيجابيًّا (تعبير مجازي عن خبر مهم)، ويجب إدراجه تحت الجمل المحايدة؟

- "فشل الفريق في إحراز الهدف "

يعتمد تصنيف الشعور على قائل العبارة ما إذا كان مع/ أو ضد الفريق الخاسر.

- "لقد وعدونا بالانتصار"

كيف يتم تصنيف تلك الجملة ومثيلاتها إذا كان المتحدث ينقلها عن شخص غيره ؟

- "لقد وعدونا بالانتصار. وقد أوفوا بالوعد وانتصروا"
- "لقد وعدونا بالانتصار. لكنهم أخلفونا الوعد".

توضح الجملتان السابقتان أهمية السياق في تصنيف شعور الجمل.

- "لقد قال لي إنه في غاية السعادة مما حدث، وهذا ما أثار غضبي".

كيف يتم التعامل مع الجمل المحتوية على تعبير عن الكلام المنقول، والذي لم يوافق شعور القائل؟

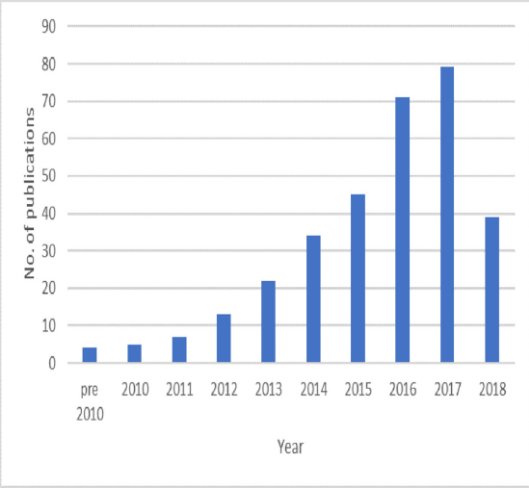
و) أهداف الدراسة

إن الاستخدامات العملية لهذا المجال تكاد لا تنتهي، وبناءً عليه ظهرت أبحاث عديدة عكفت على تطويره ودراسته والتوسع في تغطيته اللغوية. إن التقدم في الأبحاث العلمية المختصة بتحليل المشاعر في الإنجليزية أكبر بكثير منها في العربية، كما هو الحال في معظم الدراسات المختصة بعلوم الحاسب. ولكنَّ هذا العمل يهدف إلى تسليط الضوء على المنطقة العربية آخذًا في الاعتبار الطبيعة المعقدة للغة العربية وتعدد لهجاتها وسعتها اللغوية. كما يهدف إلى تعريف القارئ بمجال تحليل المشاعر عن طريق عمل بحث بيني للمقارنة بين النماذج المختلفة المستخدمة في عملية تحليل المشاعر، بالإضافة إلى تعريف القارئ بالطرائق المختلفة لقياس صحة المشاعر المتعرَّف عليها وتطبيقها على النماذج المتداولة للوصول إلى تقييماتٍ ودوافع لاختيار كلٍّ منها. أيضًا سنتطرق ببعض التفصيل إلى الخوارزميات المستخدمة في تحليل المشاعر، فهي تنقسم إلى خوارزميات قائمة على القواعد، وخوارزميات تحليلية وإحصائية، وخوارزميات هجينة. كذلك تناقش هذه الدراسة بعض التقنيات المتوافرة، وهي Smart Reader، تطبيق مشاعر، Amazon Comprehend، مع توضيح بعض التعديلات والتحسينات الممكنة للحصول على معدلات أداءٍ أفضل.

وسائل البحث

أ) أنواع الخوارزميات المستخدمة في أنظمة تحليل المشاعر

❖ خوارزميات قائمة على القواعد:



صورة ١: عدد المنشورات المتزايد في مجال تحليل المشاعر

تعتمد على مجموعة من القوانين المصممة عن طريق محللي البيانات.

❖ خوارزميات تحليلية وإحصائية: تعتمد على تحليل البيانات من خلال البيانات واستنباط نموذج قادر على التنبؤ بشعور الآلة. يعتمد هذا النظام على تعلم الآلة لإيجاد النموذج المعني.

❖ خوارزميات هجينة: تعتمد هذه الأنظمة على كل من البيانات والقواعد المصممة على حد سواء.

١- الخوارزميات القائمة على القواعد

تعتمد هذه القواعد على علم اللغات الطبيعية وأدواته، مثل إيجاد الجذر اللغوي وتحليل الجمل. يعتمد هذا النظام أيضًا على مكتبة كبيرة من الكلمات الخاصة والمصنفة لكل شعور. تتبع هذه الخوارزميات الخطوات التالية أو شبيهاتها أو مزيجًا بينهما:

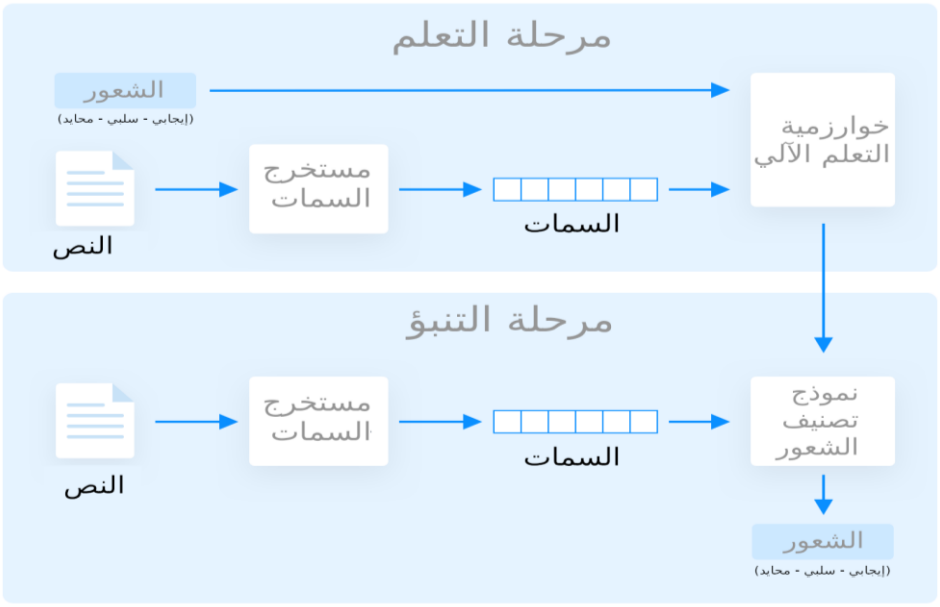
- حذف جميع الكلمات التي لا تتعلق بالموضوع (مثل أدوات التعريف وحروف الجر وغيرها).
- حذف علامات الترقيم مثل الفواصل والنقط وغيرها.
- إيجاد الجذر اللغوي لكل كلمة فتتحول كلمات مثل (يلعب – لعبت – يلعبون – ملعب) إلى (لعب)؛ وذلك لتقليل التنوع في التفاصيل حتى يتسنى لنا الحفاظ على مكتبة كلمات صغيرة نسبيًّا.
- نقوم بعَدِّ كل الكلمات الموجودة الخاصة بشعور معين حسب وجودها في قوائم لغوية مُصنِّفة لها. فإذا كان عدد الكلمات الإيجابية أكثر من أي عدد آخر لشعور آخر عُدَّ هذا النص إيجابيًّا في مجمله.

يمكن إضافة طرق تقييم أخرى غير عَدِّ الكلمات البسيط، ويمكن إعطاء كل طريقة تقييم أولوية ووزنًا مختلفين، ولكن هذا سيزيد درجة التعقيد وسيتطلب بحثًا وتحسينًا دقيقًا لكل المعايير وترتيب الأولويات بشكل يدوي. يصعب على هذا النوع التقييم بناءً على السياق أو ترتيب الكلمات عامة. ولذلك فإن هذا النوع ضعيف نسبيًّا.

٢- الخوارزميات التحليلية والإحصائية

تعتمد الخوارزميات التحليلية والإحصائية على تنبؤ المشاعر من خلال حزمة بيانات تحتوي على نصوص مكتوبة سابقًا مع كتابة الشعور المصاحب للنص يدويًّا، يُعالَج النص ويُحوَّلُ إلى مجموعة من السمات لإعطائها لخوارزمية تعلم الآلة فيما بعد. ينتج من عملية

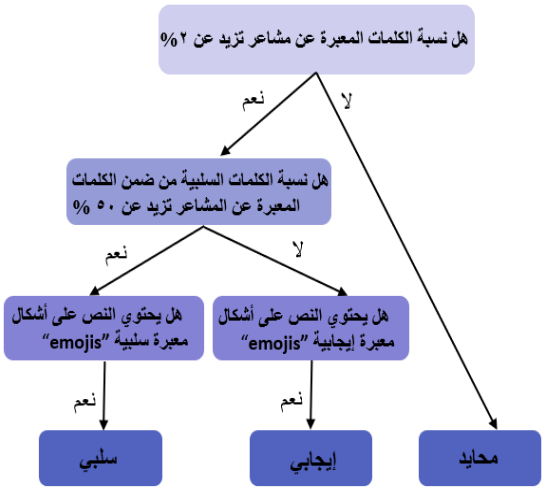
التعلم نموذج قادر على التنبؤ بالشعور من خلال نص معطى. لا تحتاج هذه الطريقة إلى قواعد مصممة من قبلُ للتنبؤ بالشعور على عكس الخوارزمية السابقة، لكنها تحتاج إلى كمية كبيرة من الأمثلة التي تم وسمها بالشعور المناسب يدويًّا وهو ما يشكل **تحدّيًا في ذاته.**



من أكبر التحديات الموجودة في جمع حزم البيانات اختلاف الحكم عليها من شخص إلى آخر، فهي عملية نسبية؛ فقد يَسِم شخص ما جملةً بأنها تعبر عن فرح، وقد يرى آخر أنها تعبر عن سخرية. وهو الأمر الذي يزيد صعوبة التعلم على النموذج. من أكثر الخوارزميات المستخدمة في تطبيق تحليل المشاعر: Naïve Bayes, SVM, Decision Trees والشبكات العصبية العميقة.

- شجرة القرارات (Decision Tree)

تستخدم شجرة القرارات تمثيلًا بيانيًّا يشبه الشجرة وتمثل كل عقدة في هذا التمثيل سمة من السمات المستخرجة من النص، وعليه فإن كل تفرعٍ يمثل قيمة من قيم هذه السمة.



على سبيل المثال وكما يُرى في شجرة القرارات المجاورة التي تعتبر مثالًا مبالغًا في بساطته لسهولة التوضيح- في التصنيف الثلاثي للمشاعر (إيجابي – محايد – سلبي) فإن السمات المستخدمة تتضمن ما يلي: هل نسبة الكلمات المعبرة عن مشاعر تزيد عن ٢٪؟ هل نسبة الكلمات السلبية من ضمن الكلمات المعبرة عن المشاعر تزيد عن ٥٠٪؟ هل يحتوي النص على أشكال معبرة "emojis" سلبية ؟ هل يحتوي النص على أشكال معبرة "emojis" إيجابية؟ .(كما يمكننا أن نرى أن كل تفرعٍ يمثل قيمة لهذه السمات التي تنحصر في هذه الحالة في (نعم أو لا). اختيار النسب المعينة لكل سمة وترتيبها وتعيينها ليس عشوائيًّا ألبتة، ولكن يتم عن طريق عملية حسابية نظامية توظف إنتروبيا شانون (Shannon Entropy) والمتوسطات الحسابية في اختيار ما يثري عملية التعلم ثم التنبؤ.

- بايز البسيطة (Naïve Bayes)

صورة ٣: توضيح لآلية عمل شجرة القرارات

تتبنى هذه الخوارزميات نظرية بايز المشهورة (Bayes' theorem) التي تنص على أنه يمكن حساب الاحتماليات البعدية (posterior probabilities) المشروطة باستخدام الاحتماليات القبلية (prior probabilities) والقابلية (likelihood) والدلائل تبعًا لهذه الصيغة الرياضية:

$$ح(أ \mid ب) = \frac{ح(أ) \times ح(ب \mid أ)}{ح(ب)}$$

حيث تقرأ هذه الاحتمالية "احتمالية وقوع الحدث أ بشرط وقوع الحدث **ب**".

يمكن تعميم هذه الاحتمالية في سياق تحليل المشاعر لتعني "احتمالية تصنيف نصٍّ في فئة مشاعر معينة **ف** من ضمن عدة فئات **{ف١، ف٢، ...، فن}** بشرط تحقيقه لقيم معينة في سمات النص **{س١، س٢، ...، سن}**". وبالتالي فإننا نهدف إلى وسم النص بالفئة التي تحقق أعلى احتمالية تبعًا لهذه الصيغة:

$$ح(ف) \times \prod^{ل} ح(س \mid ف)$$

حيث ل عدد الكلمات في النص.

تفترض هذه الخوارزميات فرضيتين غير حقيقيتين – وهما السبب في تسميتها بالبسيطة- وهما:

١- أن احتماليات حصول قيمة كل سمة على فئة مشاعر معينة تحقق

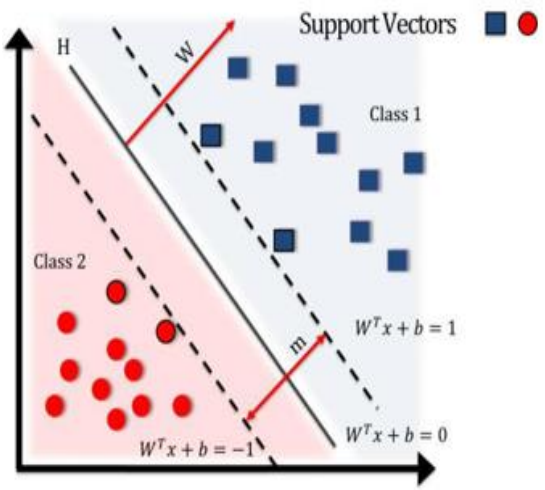
الاستقلال الشرطي (conditional independence).

٢- أن تصنيف النص لا يعتمد على موضوع الكلمات، ولكن يعتمد على مجرد وجود الكلمات أو عدم وجودها.

بالرغم من خطأ كلتا الفرضيتين إلا أن خوارزميات بايز البسيطة تحقق معدلات أداء جيدة للغاية، فهي تتعرض لتحسينات مستمرة بفضل تقنيات التقوية الحاسوبية (Computational Optimization Techniques).

- آلة المتجهات الداعمة (Support Vector Machine – SVM)

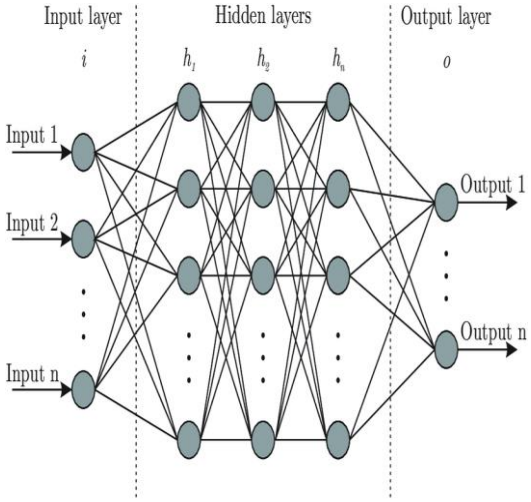
آلة المتجهات الداعمة هي خوارزمية تندرج تحت خوارزميات تعلم الآلة، وتعتبر من أكثر الخوارزميات شعبية في مهام التصنيف باختلاف أنواعها. تُبنى فكرة آلة المتجهات الداعمة على إيجاد مستوًى فائق (hyperplane) قادر على تقسيم مجموعة البيانات التي لم تكن قابلة للتقسيم من قبلُ. على سبيل المثال، تقوم الخوارزمية بتحويل البيانات وتوزيعها من بعد واحد غير قابل للتصنيف إلى بعدين يحتويان على مستوى فائق قادر على فصل نقاط البيانات بعضها من بعض، كما نرى في الشكل المواجه. يعتبر الخط المائل هو المستوى الفائق القادر على تمييز الفئة الأولي (المربعات الزرقاء) من الفئة الثانية (الدوائر الحمراء)، وعليه فعند انتهاء تدريب الخوارزمية فإنها تكون قادرة على الوصول لأفضل مستوى قادر على فصل نقاط بيانات من نفس الفئات حتى وإن لم تُر من قبل. يتم اختيار هذا المستوى الفائق عن طريق مجموعة معادلات تهدف إلى تعظيم المسافة بين المستوى وأقرب نقطة بياناتٍ من كل فئة.



صورة ٤: توضيح لآلية عمل آلة المتجهات الداعمة

- الشبكات العصبية العميقة
- يعتبر التعلم العميق استلهامًا يحاكي طريقة الدماغ البشري في تبسيط المشكلات إلى وحدات أكثر تجريدًا

عن طريق الشبكات العصبية المكونة من ثلاثة أنواع من الطبقات: طبقات المدخلات وطبقاتٍ داخلية، وطبقة المخرجات التي تحوي التصنيف النهائي للنص بين الفئات المعرفة . تتدرج الطبقات في الشبكة العصبية في مستوى التعقيد وفي التدفق؛ حيث إن كل طبقة تتلقى مخرجات الطبقة السابقة لها. تتغير المعلومات المتدفقة عن طريق الأوزان المرتبطة بكل عصب في الشبكة العصبية، أي أن مُخرج الشبكة العصبية هو نتيجة جمع حاصل ضرب الأوزان بطبقة المدخلات ثم الطبقات المخفية. تمثل هذه الأوزان منتج التعلم، حيث إن



صورة ٥: توضيح لآلية عمل آلة الشبكات العصبية العميقة

٣- الخوارزميات الهجينة

تدريب الشبكة العصبية يهدف إلى الوصول إلى أفضل الأوزان القادرة على استنتاج سمات للنصوص قد لا تكون منطقية للعقل البشري، ولكنها تجسد خصائص مشتركة بين نصوص ذات الفئات المتشابهة.

التقنيات المتوافرة لتحليل المشاعر

١ - Smart Reader

أحد إصدارات ParallelDots الذي يستخدم خوارزميات الذاكرة قصيرة المدى (LSTM) التصنيف مشاعر النص إلى مشاعر إيجابية أو سلبية أو محايدة. يتميز بسهولة الاستخدام وسرعة الأداء ودعمه للعديد من اللغات كالإنجليزية والإسبانية والروسية، وكثير غيرها. كما تتضمن واجهة برمجة التطبيقات خصوصية البيانات والنصوص المستخدمة، حيث إنها تدعم عمليات النشر السحابية الخاصة عن طريق حاويات Docker. إضافة إلى ذلك فإنه يدعم تحليل الفقرات بل النصوص الكاملة، ويأخذ في الحسبان علامات الترقيم والأيقونات المعبرة ويُظهر نسب التأكد من كل فئة. وهو غير مجاني، ولكنه يتيح فترة تجربة مجانية بإمكانات مخفضة.

٢ – مشاعر

أحد إصدارات "أدوات عربي" ويتخصص في تحليل المشاعر في النصوص العربية، ويعطي نتيجة إيجابية، أو سلبية، أو محايدة بالاعتماد على تحليل النص المدخل. يتميز بسهولة الاستخدام وسرعة الأداء إضافةً إلى أنه مجاني بالكامل، لكنه لا يدعم تحليل الفقرات أو النصوص الكاملة، ولا يأخذ في الحسبان علامات الترقيم والأيقونات المعبرة، ولا يُظهر نسب التأكد إلا عند التأكد التام.

٣– Amazom Comprehend

إحدى الخدمات التي تقدمها Amazon Web Services (AWS) التي تستخدم تعلم الآلة لكي تستطيع استيعاب النص والعلاقات التي يتضمنها بشكل دقيق. يعتبر الإصدار الأكثر عمقًا على الإطلاق في تحليل المشاعر باللغة العربية، فهو يستطيع استخراج العبارات الرئيسية، وتحليل المشاعر، والتعرف على الكيانات، ونماذج الموضوعات، كما أنه يتميز بسهولة الاستخدام وسرعة الأداء إضافةً إلى أنه يتيح استخدامًا مجانيًا لغير الشركات. يدعم تحليل الفقرات والنصوص الكاملة ويأخذ في الحسبان كل مكونات الجملة، وأيضًا يظهر نسب التأكد، كما يدعم العدد الأكبر من اللغات مقارنةً بأي تقنية أخرى.

أخيرًا، يضيف احتمالية رابعة للنصوص المختلطة.

يحتاج لحسابٍ عامل ذي بطاقة ائتمان عاملة ويضع حدًّا لنسبة الْCPU/GP المستخدمة في الحسابات المجانية. لا يحدد العواطف المسيطرة على النصوص المختلطة أو الاستنتاج النهائي.

تقييم التقنيات المستخدمة لتحليل المشاعر

بسبب طبيعة المخرج في معظم محللي المشاعر وكونه على هيئة فئة، فإن المعيار الأساسي للتقييم هو متوسط عدد الفئات الصحيحة لكل تقنية من بين الفئات الصحيحة لثماني جمل استرشادية تتدرج في عدم وضوح القطبية، حيث تم اختبار كل تقنية باستخدام ثماني الجُمَل نفسها، وحساب متوسط التصنيف الصحيح لكل تقنية تبعًا لهذه الصيغة:

متوسط معدل التصنيف الصحيح = (حاصل جمع التصنيفات الصحيحة /عدد الجمل) × ١٠٠

في الجدول التالي، سوف نستعرض نتائج التقنيات المستخدمة في تحليل المشاعر

باستخدام الجمل الاسترشادية ونلاحظ أن بعض محللي المشاعر تنتج نسبة تأكد من التصنيف بالإضافة إلى التصنيف، ولكن لعدم وجود هذه النسبة في كل التقنيات فقد تم ذكرها كملاحظة فقط .

تعد المعدلات الناتجة متطرفة قليلًا. حيث إنها نتاج تجربة على ثماني جملٍ فقط؛ في حين أنه عادةً ما يتم اختبار هذه التقنيات على آلافٍ إن لم يكن مئات الآلاف منها الجمل.

الجملة	الفئة الصحيحة	Smart Reader	مشاعر	Amazon Comprehend
"دعم العملاء الخاص بكم كارثي. لقد بقيت على الانتظار أكثر من نصف ساعة!"	سلبي	سلبي بنسبة: ٩١,٨٪	محايد	سلبي بنسبة: ٩٦ ٪
"استيقظت اليوم وأنا أشعر بسعادةٍ بالغة ولم أستطع فهم مصدرها"	إيجابي	إيجابي بنسبة: ٦٧٪	إيجابي	إيجابي بنسبة: ٩٧ ٪
"هذا المنتج لا بأس به، يستطيع أن يؤدي الغرض"	محايد	محايد بنسبة: ٥٢,٣ ٪	محايد	إيجابي بنسبة: ٩٧ ٪
"أنا حقاً أقدر هذه الخدمة الفعالة للعملاء "	إيجابي	إيجابي بنسبة: ٨٤,٧٪	محايد	إيجابي بنسبة: ٩٩ %
"مع أنه يصعب تركيبه قليلاً إلا أنني راضٍ عن هذا المنتج"	إيجابي	إيجابي بنسبة:٧٣ ٪	سلبي	مختلط بنسبة: ٩٩ ٪
"جودة عالية ولكن بسعر مبالغ فيه"	سلبي	سلبي بنسبة: ٦٩,٨ ٪	سلبي	مختلط بنسبة: ٩٩ ٪
"كيف تتوقعون من العملاء فهم دليل الاستخدام"	سلبي	محايد بنسبة: ٦٠ ٪	محايد	محايد بنسبة: ٦٤ ٪
"إن كنت تنوي شراء هذا العطر، ضعه في البيت ولا تسمح لشخصٍ آخر أن يشمه"	سلبي	سلبي بنسبة: ٦٩,٨ ٪	محايد	إيجابي بنسبة: ٤٤٪
متوسط معدل التصنيف الصحيح	٩٠ ٪	٣٠ ٪	٣٠ ٪	

ومن خلال تأمل النتائج في الجدول السابق، لاحظنا أنه في حالة استخدام Amazon Comprehend تم احتساب الجمل مصنفة كمختلطة ذات تصنيفٍ خاطئ حتى وإن كانت فعلًا مختلطة، حيث إن التقنية لم توضح ما هي مجموعة المشاعر الموجودة أو الاستنتاج النهائي من الجملة.

اقتراحات للباحثين لتحسين دقة أداء تقنيات تحليل المشاعر

إن تعرف الإنسان على المشاعر يشبه إلى حد كبير طريقة الآلة، حيث لا يمكن اقتصار الفرد على الدلالة المنفصلة للكلمات المذكورة ، بل يجب الاهتمام بطبيعة السياق الذي يحكم الموقف الذي بصدده ذُكرت الجملُ، وطَبَقَة الصوت التي تستخدم في التعبير عن تلك المشاعر وغيرها من العوامل التي تؤثر في طبيعة فهم الإنسان لفهم المشاعر. وبالتالي، فإن الوصول بالآلة لمستويات دقيقة من القدرة على تحليل المشاعر يمكن أن يتقدم عن طريق:

- تدريب النماذج على نصوص أكبر تحتوي على مشاعر ضمنية توظف التحديات السابق ذكرها.
- توظيف المحولات مزدوجة الاتجاه وطبقات الانتباه لأخذ السياق في الاعتبار بطريقة أفضل.
- محاولة تخمين الفئة التي لم يستطع النموذج التعرف السليم عليها عن طريق أخذ بقية الجملة في الاعتبار وحساب احتماليات الفئات الممكنة.
- تطوير واجهة التقنيات لإظهار تفاصيل التصنيف كما كان يجب في حالة تصنيف Amazon Comprehend للجمل المختلطة.

المصادر

[1] "كيفية تحليل المشاعر في النص باستخدام Amazon Comprehend | AWS", *Amazon Web Services, Inc.*, 2021. [Online]. Available: https://aws.amazon.com/ar/getting-started/hands-on/analyze-sentiment-comprehend/. [Accessed: 19- Apr- 2021].

[2]"Sentiment Analysis | ParallelDots AI APIs", *Paralleldots.com*, 2021. [Online]. Available: https://www.paralleldots.com/sentiment-analysis. [Accessed: 19- Apr- 2021].

[3] "أدوات عربي - نظام تحليل المشاعر"، *Arabitools.com*, 2021. [Online]. Available: https://www.arabitools.com/sentest.html. [Accessed: 19- Apr- 2021].

[4]M. Al-Ayyoub, A. Khamaiseh, Y. Jararweh and M. Al-Kabi, "A comprehensive survey of arabic sentiment analysis", *Information Processing & Management*, vol. 56, no. 2, pp. 320-342, 2019. Available: 10.1016/j.ipm.2018.07.006.

[5]A. Alnawas and N. Arıcı, "The Corpus Based Approach to Sentiment Analysis in Modern Standard Arabic and Arabic Dialects: A Literature Review", *Journal of Polytechnic*, 2018. Available: 10.2339/politeknik.403975.

[6]M. Al-Ayyoub, A. Nuseir, G. Kanaan and R. Al-Shalabi, "Hierarchical Classifiers for Multi-Way Sentiment Analysis of Arabic Reviews", *International Journal of Advanced Computer Science and Applications*, vol. 7, no. 2, 2016. Available: 10.14569/ijacsa.2016.070269.

[7] "تحليل الشعور: تقريبًا كل ما تحتاج إلى معرفته (٢) - المبرمج العربي"، *Arabicprogrammer.com*, 2021. [Online]. Available: https://arabicprogrammer.com/article/5523134020/. [Accessed: 19- Apr- 2021].