

Analysis of views on social networks: An application model to measure the level of sports intolerance in Twitter

Mohammed R. Alkuzai

Department of Information Science,
King Abdulaziz University, Saudi Arabia

Hassan A. Alsereihy

Professor, Department of Information Science,
King Abdulaziz University, Saudi Arabia

This study attempted to measure the level of sports intolerance through the analysis of opinions on Twitter using an electronic tool to measure opinions developed and applied. A tool was developed to analyze opinions and feelings and use it in this study, which is an experimental in this part of the study and based on the analysis of content for the analysis of tweets and in the practical part of the study. The study found that writing in the vernacular and local dialect form a major challenge for modeling analytical methods where algorithms standard software is lost, which is an important factor in code formulation. In addition, the large number of misspellings and the use of graphic symbols may prevent the possibility of sorting the tweets through the electronic platform. Moreover, it provides the possibility to develop this model, and improve the software algorithms of the platform (Meul) to obtain more accurate results, and cover more areas.

تحليل الآراء على شبكات التواصل الاجتماعي : نموذج تطبيقي لقياس مستوى التعصب الرياضي في تويتر

محمد رده الخزاعي

قسم علم المعلومات، جامعة الملك عبدالعزيز، السعودية

أ. د. حسن عواد السريحي

قسم علم المعلومات، جامعة الملك عبدالعزيز، السعودية

المستخلص

قامت هذه الدراسة بمحاولة قياس مستوى التعصب الرياضي من خلال تحليل الآراء على شبكة تويتر باستخدام أداة إلكترونية لقياس الآراء تم تطويرها وتطبيقها . وذلك بهدف التعرف على طرق وأساليب تحليل الآراء على شبكات التواصل الاجتماعي ومعرفة العوائق والتحديات التي تواجه التحليل الآلي للآراء على الشبكات الاجتماعية. وقد تم تطوير أداة لتحليل الآراء والمشاعر واستخدامها في هذه الدراسة التي تعد تجريبية في هذا الجزء منها واعتمدت على أسلوب تحليل المحتوى لتحليل التغريدات وفي الجزء التطبيقي من الدراسة سنعرض الاجراءات التي تمت في الدراسة. وقد توصلت الدراسة الى ان الكتابة باللغة العامية وباللهجة المحلية تشكل تحدياً كبيراً لنمذجة أساليب التحليل، حيث تفقد الخوارزميات البرمجية المعيارية، وهي عامل مهم في صياغة كود البرمجة. اضافة الى ان كثرة الأخطاء الإملائية، واستخدام الرموز الرسومية، قد يحول دون إمكانية تصنيف التغريدة من خلال المنصة الإلكترونية. اضافة لذلك فان توفر الإمكانيات لتطوير هذا النموذج، وتجويد الخوارزميات البرمجية لمنصة (ميول) للحصول على نتائج أكثر دقة، وتغطية مزيد من المجالات.

أصبحت قنوات التواصل الاجتماعي هي الوسيلة المفضلة لقطاع كبير من الجمهور للتعبير عن آرائهم، وبالتالي فإنها من أهم وسائل معرفة اتجاهات الرأي العام نحو مختلف القضايا، وفي جانب الأعمال فإن قنوات التواصل الاجتماعي من المصادر المهمة لقياس رضى المستهلكين أو عدم رضاهم تجاه منتج أو خدمة معينة.

وقد اتجهت مؤخراً مؤسسات الأنظمة الإلكترونية إلى إنتاج أدوات لإدارة وتحليل الآراء على شبكات التواصل الاجتماعي بشكل آلي وبالأخص المحتوى المنشور على شبكة تويتر، كما ظهرت عدد من البحوث التطبيقية في هذا المجال.

وسوف نتناول في هذه الدراسة مفهوم تحليل الآراء Sentiment Analysis ، وإقتراح نموذج لقياس مستوى التعصب الرياضي من خلال التحليل الآلي للمشاركات المنشورة على شبكة تويتر.

مشكلة الدراسة :

تمثل شبكة تويتر في مجتمعنا العربي والخليجي على وجه الخصوص منصة رئيسة للحوار والتعبير عن الرأي في مختلف المجالات ومن مختلف الأوساط، فكثيراً ما نشأت وسوم #tags لمناقشة حدث أو قضية معينة، ومن بينها قضايا التعصب الرياضي، ولا شك أن المشاركات الكثيفة في هذه النقاشات تمثل رأياً عاماً يصعب معرفة اتجاهه في كثير من الأحيان نتيجة تباين الآراء وغازة المشاركات.

وفي هذه الدراسة نحاول قياس مستوى التعصب الرياضي من خلال تحليل الآراء على شبكة تويتر باستخدام أداة إلكترونية لقياس الآراء.

أهمية الدراسة:

تبرز أهمية هذه الدراسة من الجانب النظري، بأنها من الدراسات القلائل باللغة العربية، التي تتناول مفهوم تحليل الآراء على شبكة تويتر، وتعد إضافة مهمة للإنتاج الفكري العربي في هذا المجال. أما على المستوى التطبيقي، فإنها تطرح نموذجاً إلكترونياً لتحليل الآراء لأبعاد متعددة، بخلاف كثير من الأبحاث التي تعالج تحليل الآراء من

خلال تصنيفها إلى سلبية أو إيجابية أو محايدة، وتقع في غالبها في المجال السياسي والتسويقي. أما هذه الدراسة فتتناول تصنيف مستوى التعصب في الآراء الرياضية على شبكة تويتر.

أهداف الدراسة:

تهدف هذه الدراسة إلى :

التعرف على طرق وأساليب تحليل الآراء على شبكات التواصل الاجتماعي.

معرفة العوائق والتحديات التي تواجه التحليل الآلي للآراء على الشبكات الاجتماعية.

اقترح نموذج إلكتروني تطبيقي لتحليل الآراء على شبكة تويتر.

تقديم بعض التوصيات لرفع مستوى الدقة لقياس اتجاهات الرأي عبر أدوات تحليل محتوى شبكات التواصل.

منهج الدراسة :

تم في هذه الدراسة استخدام المنهج التجريبي Experimental وذلك عبر استخدام اداة تم تطويرها لتحليل تغريدات باللغة العربية والتعرف على خصائصها عبر تحليل المحتوى Content Analysis الممنهج عبر البرنامج لتحليل التغريدات واستخراج المؤشرات وذلك من خلال استرجاع كمية من المعلومات المنشورة على شبكة التواصل الاجتماعي تويتر، ومن ثم تحليلها لتحديد اتجاهات الآراء لكل تغريدة. وقد تم تطبيق الدراسة على المغردين الرياضيين عبر وسم #أكشن_يا_دوري، من خلال استخراج عينة مكونة من ١٠٠٠٠ تغريدة على فترات متفاوتة.

وقد تم تطوير أداة لتحليل الآراء والمشاعر واستخدامها في هذه الدراسة التي تعد تجريبية في هذا الجزء منها واعتمدت على أسلوب تحليل المحتوى لتحليل التغريدات وفي الجزء التطبيقي من الدراسة سنعرض الاجراءات التي تمت في الدراسة.

الدراسات السابقة :

الدراسة الأولى :

العتيبي، صالحة، خان، محمد (٢٠١٧) بعنوان تحديات تحليل الآراء في اللغة العربية العامية.

Alotaibi, Salihah, Khan, Muhammad (2017) “Sentiment Analysis Challenges of Informal Arabic Language”._ *IJACSA*, Vol. 8, No. 2, 2017.

والغرض من هذه الدراسة كان تسليط الضوء على تحليل الآراء المكتوبة باللغة العربية، ومناقشة التحديات والعقبات التي تواجه تحليل الآراء للغة العربية غير الرسمية (العامية) على شبكات التواصل الاجتماعية. وقد توصل الباحثان إلى أن اللغة العربية العامية مشوشة، وضعيفة الهيكل، بعكس اللغة الرسمية. وأن التغريدات العربية ومراجعات اليوتيوب تحوي الكثير من الأخطاء الإملائية، مما سبب العديد من المشاكل لعملية تحليل الآراء.

الدراسة الثانية :

خاتون، شاهين (٢٠١٧) بعنوان تحليل بيانات تويتر لحظياً لشركات الاتصالات لتحسين إدارة علاقات العملاء.

Khatoon, Shaheen (2017) “Real-time Twitter Data Analysis of Saudi Telecom Companies for Enhanced Customer Relationship Management”._ *IJACSA*, Vol. 17, No. 2, Feb. 2017.

وهدف من هذه الدراسة إلى تقديم حل آلي لمساعدة شركات الاتصالات للوصول إلى رؤية أفضل عن العملاء من خلال الاستفادة من البيانات الآنية المعروضة في تويتر، من أجل تطوير نظام إدارة عملاء متقدم، في سوق عالية التنافسية، ومن أجل معرفة رأي العملاء تجاه الخدمات التي تطرحها شركة ما والشركات المنافسة لها من خلال تحليل الآراء على تويتر مع إدخال البعد المكاني في عملية التحليل.

وتوصلت الدراسة إلى أن قيام شركات الاتصالات بمراقبة تحليل الآراء آنياً بصفة يومية على تويتر سوف يساعدها على معرفة اتجاهات رغبات العملاء، ومعرفة احتياجات كل منطقة، وبالتالي تطوير نظام لإدارة علاقات العملاء يراعي تلك الاتجاهات.

الدراسة الثالثة :

العويدي، سناء وآخرون (٢٠١٧) بعنوان تحليل الآراء الدلالي للنصوص العربية.

Alowaidi, Sana and others (2017) “Semantic Sentiment Analysis of Arabic Texts”._ *IJACSA*, Vol. 8, No. 2, 2017.

في هذه الدراسة تم تطوير نموذج للتحليل الدلالي لمحتوى تويتر العربي ATSA، اعتماداً على آلة التعلم المراقب Supervised machine learning (ML) والتحليل الدلالي، وقد تم تقييم هذا النموذج مقارنةً بنماذج أخرى، وقد تفوق نموذج من حيث الأداء على ماكينة الدعم الموجه Support Vector Machine (SVM) بنسبة ٤٨,٤٨٪، وعلى بايز الساذج naïve Bayes (NB) بنسبة ٥٨,٧٨٪.

الدراسة الرابعة :

الحربي، وعد، إمام، أحمد (٢٠١٥) بعنوان أثر اللهجة السعودية على التهيئة لتحليل الآراء في اللغة العربية.

Al-Harbi, Waad, Emam, Ahmed (2015) “EFFECT OF SAUDI DIALECT PREPROCESSING ON ARABIC SENTIMENT ANALYSIS”._ *IJACT*, Vol. 4, No. 6, Dec. 2015.

وكان الغرض هذه الورقة هو دراسة أثر اللهجة السعودية على التهيئة لتحليل الآراء على شبكات التواصل الاجتماعي باستخدام نظام المنقب السريع Rapidminer، ومن خلال ثلاثة طرق لتحليل الآراء هي : Naïve ، Support Vector Machine (SVM) and K-Nearest Neighbor Classifier ، Bayes (NB). وقد أظهرت الطرق الثلاث دقة لا بأس بها بين ٥٧,٨٩٪ و ٧٣,٣٣٪، عندما تم استبدال الكلمات باللهجة السعودية بالكلمات المقابلة لها باللغة العربية الفصحى.

الدراسة الخامسة :

إبراهيم، مناهل أحمد، سالم، نعومي (٢٠١٦) بعنوان تحليل الآراء للتغريدات العربية مع مرجع خاص لتغريدات المطاعم.

Ibrahim, Mnahel Ahmed, Salim, Naomie (2016) “Sentiment Analysis of Arabic Tweets: With Special Reference Restaurant Tweets”._ *IJCST*, Vol. 4, Iss. 3, 2016.

وتم في هذا البحث تقسيم عملية تحليل الآراء على مرحلتين: الأولى هي معالجة النصوص ومن ذلك حذف كلمات التوقف، وإعادة الكلمات إلى جذرها، أما المرحلة الثانية فهي عملية تصنيف التغريدات من خلال NB, SVM and KNN، وتمت المعالجة والتصنيف على ٩٠٠ تغريدة تتعلق بمجال المطاعم.

وخلصت الدراسة إلى أن عملية معالجة النصوص كالتجذير الخفيف مضافاً إلى حذف كلمات التوقف، كان لها أثر إيجابي في تحسين أداء عملية التصنيف، وأن أفضل طرق التصنيف كان باستخدام SVM بالتكامل مع n-gram إضافة إلى حذف كلمات التوقف stop words.

الدراسة السادسة :

ترتير، سمير، عبدالنبي، إبراهيم (٢٠١٦) بعنوان تحليل الآراء الدلالي في الاعلام الاجتماعي العربي.

Tartir, Samir, Abdul-Nabi, Ibrahim (2016) “Semantic Sentiment Analysis in Arabic Social Media”._ *Journal of King Saud Univ.* 2016.

وتهدف هذه الدراسة إلى إيجاد نموذج لتحليل الآراء على شبكة تويتر من خلال بناء أنطولوجيا عربية للآراء (ASO)، مع معالجة للنصوص العربية المكونة للتغريدات من حيث إزالة الروابط وحذف مد الحروف وكلمات التوقف. وقد خلصت الدراسة إلى أن تحليل الآراء على الإعلام الاجتماعي مهم جداً لقطاع الأعمال لفهم التوجه المستقبلي، ولصنع القرارات.

الدراسة السابعة :

كور، إقبال ديب وآخرون (٢٠١٧) بعنوان النموذج المرن لتحليل الآراء من الإعلام الاجتماعي.

Kaur, Iqbaldeep and others (2017) “Agile Model Based Sentiment Analysis from Social Media”._ *IJACSA*, Vol. 15, No. 1, Jan. 2017.

تهدف هذه الدراسة إلى تنفيذ خوارزمية برمجية التحليل الآلي لنصوص التغريدات المستقاة من شبكة تويتر، وتصنيفها إلى تغريدة إيجابية، سلبية أو محايدة. وذلك بهدف معرفة مشاعر وأفكار المشاركين نحو موضوع ما. وتمثيل تلك الآراء في شكل رسوم بيانية تعكس نسب التصنيفات الثلاث، مستخدمين مكتبة معالجة اللغة الطبيعية من ستانفورد Stanford NLP Library.

الدراسة الثامنة:

نجوين، هونج، جونج، جاسون (٢٠١٥) بعنوان المنهج الاحصائي لتحليل الآراء للغة التشبيهية على تويتر.

Nguyen, Hoang long, Jung, Jason (2015) “A Statistical Approach on Figurative Language Sentiment Analysis in Twitter”._ *SemEval*, Colorado, Jan 4-5, 2015._ pp. 679-683.

وتعرض هذه الورقة طريقة إحصائية لتحليل الآراء للغة التشبيهية أو التصويرية للنصوص القصيرة المجمعة من تويتر، وهي قائمة على تصنيف التغريدات إلى ثلاث فئات : ساخرة، تهكمية و مزاجية، وفقاً لخاصيتين رئيسيتين هما المحتوى و النسق العاطفي.

وقد تم التطبيق على مجموعتين من البيانات، بيانات تجريبية مكونة من ١٠٠٠ تغريدة، وبيانات الاختبار مكونة من ٤٠٠٠ تغريدة.

الدراسة التاسعة :

كور، جاسلين وآخرون (٢٠١٦) بعنوان تحليل الآراء من بيانات تويتر باستخدام طريقة هجينة من آلة الدعم الموجه ومثالية مستعمرة النمل.

Kaur, Jasleen and others (2016) “Sentiment Analysis of Twitter data using Hybrid Method of Support Vector Machine and Ant Colony Optimization”._ *IJCSIS*, Vol. 14, No. 7, July 2016.

واعتمدت هذه الدراسة على تصنيف الآراء إلى نوعين، سلبي وإيجابي بناءً على تحليل مجموعة من التغريدات باستخدام طريقة هجينة من آلة الدعم الموجه (SVM) ومثالية مستعمرة النمل Ant Colony Optimization (ACO)، وقد أظهرت النتائج تحسن في عملية التصنيف بنسبة ٧٥,٥٤٪ إلى ٨٦,٧٤٪.

الدراسة العاشرة :

كيلكار، أنوب جاينانت، موخاد، أنيل (٢٠١٦) بعنوان تنبؤ مؤشر نيفتي للسيارات المبني على آراء التغريدات .

Kelkar, Anup, Mokhade, Anil (2016) “Nifty Auto Index prediction on the basis of tweets sentiments”._ *GSTF Journal on Computing*, Vol. 5, No. 1, 2016.

في هذه الورقة قام الباحثان بفحص العلاقة المركبة بين آراء التغريدات وقيم سوق نيفتي لأسهم شركات السيارات، من خلال تحليل الآراء لمليون تغريدة تتعلق بسوق الأسهم لأكثر من عام كامل، وقد أظهرت النتائج علاقة قوية بين أسعار الأسهم وآراء التغريدات. حيث يستخدم ذلك عبر تحليل السببية لغرانجر Granger causality analysis لتحقيق من صحة الفرضية : أن حالة المزاج العام تنبئ بالتغيرات في قيم إغلاق مؤشر السوق.

الإطار النظري للدراسة :

مفهوم تحليل الآراء : Sentiment Analysis

تحليل الآراء أو المشاعر جانب متنامي، وهو أساساً يقوم على توظيف البرمجة اللغوية العصبية NLP، والمعالجة اللغوية، والتنقيب في النصوص، وهو يستطيع استنباط معلومات صنع القرار من محتويات المصادر، وتحليل الآراء يعمل على تحليل الأفكار والعواطف لغرض تصنيف المحتوى، وهو ليس مصطلحاً جديداً، إذ كان الناس قديماً يستنبطون الآراء من حديث الأشخاص بهدف صياغة القرارات، كما أن استقصاءات الرأي حول منتج أو خدمة معينة، هو أسلوب أحدث لتحليل الآراء، والآن فإن مواقع التواصل الاجتماعي وفرت منصات

يمكن استقراء الآراء منها من خلال أساليب التحليل، وهي تمثل قيمة ثمينة لآراء المستخدمين (Ibrahim and Salim, 2016).

وتحليل الآراء هو : قياس مستوى توافق وجهات النظر بين الناس حول موضوع معين، أو منتج، أو خدمة، أو حتى انتخاباتٍ ما (Anber and others, 2016).

وعرفه حمدي وآخرون (Hamdi and others, 2016) بأنه : التقيب عن الآراء، من خلال عملية التعرف الآلي على الآراء الواردة في نص معين حول موضوع معين.

أما جيانكينغ وكريستاني (Jianqiang and Crestani, 2016) فعرفا تحليل الآراء بأنه : عملية الكشف الآلي عن ما إذا كان النص يحوي معنى عاطفي أو رأي معين، وإمكانية تحديد قطبية هذا الرأي من حيث كونه إيجابي أو سلبي.

وكذلك فإن تحليل الآراء، أو ما يعرف بالتقيب عن الآراء، هو دراسة حسابية لآراء ومشاعر الناس حول الموضوعات والكيانات والأشخاص والأحداث، التي يتم التعبير عنها في النصوص. وهو يهدف إلى تحديد المشاعر الواردة في نصوص الإنترنت حول كونها سلبية أو إيجابية أو محايدة، ويلعب تحليل الآراء دوراً هاماً في العديد من المجالات كالسويق والسياسة والجوانب الاجتماعية (Alowaidi and Others, 2017).

وبالنظر إلى مجمل التعريفات الواردة آنفاً، يمكن القول بأن تحليل الآراء يعني : تحليل محتوى النصوص بالوسائل التقنية لتحديد اتجاه الرأي الذي تحمله هذه النصوص تجاه الموضوع الذي تعالجه، أو تتحدث حوله.

تعريف شبكات التواصل الاجتماعي Social Network :

تعرف شبكات التواصل الاجتماعي بأنها : " مجموعة من المواقع على شبكة الإنترنت، ظهرت مع الجيل الثاني WEB 2.0، تتيح التواصل بين الأفراد في بنية مجتمع افتراضي، يجمع بين أفرادها اهتمام مشترك أو شبه انتماء (بلد - مدرسة - جامعة... الخ) يتم التواصل من خلال الرسائل أو الاطلاع على الملفات الشخصية...، وهي وسيلة فعالة في التواصل الاجتماعي بين الأفراد " (الزهيري وكاظم، ٢٠١٤: ٢٨).

وذكر (عبد الحميد وآخرون، ٢٠١٣) تعريفاً لشبكات التواصل الاجتماعي بانها: مجموعة من وسائل الاعلام الاجتماعي، تتيح لمستخدميها تكوين وتنظيم وتحرير محتوى الويب، وقد أصبحت أكثر قنوات الاتصال شعبية على الإنترنت، مثل : Facebook, MySpace, and Linkedin.

والحقيقة أن شبكات التواصل الاجتماعي لم تعد مجموعات من الأفراد يجمع بينهم عامل مشترك، يقومون بتصميم صفحاتهم الشخصية على هذه الشبكات على سبيل الهواية والتسلية، وإجراء المحادثات الجماعية، بل إن شبكات التواصل حالياً أصبحت منصات شبه رسمية للتعبير عن الرأي، واستقاء الأخبار، وتبادل المعرفة في مجالات مختلفة، وأصبحت محكومة ببعض السياسات القانونية.

ويمكننا القول بأن شبكات التواصل الاجتماعي : هي فضاء افتراضي لمجتمع ضخم ومتنوع، يتواصل أفرادها عبر قنوات متعددة الأغراض والوظائف، كالفيس بوك، وتويتر، ويوتيوب، وإنستغرام، وغيرها، ويتبادلون خلالها الآراء والأفكار والمعرفة والمنافع الشخصية، بحسب اهتماماتهم، وبقدر كبير من الحرية.

طرق وأساليب تحليل الآراء :

تتعدد الطرق والأساليب لتحليل الآراء التي تتضمنها نصوص منصات التواصل الاجتماعي، ومن أبرزها منصة تويتر، وتعتمد هذه الأساليب على خوارزميات وأنظمة إلكترونية، وهناك بعض الفروقات والتفاوت بين أساليب تحليل الآراء من حيث دقة النتائج.

وذكرت عنبر وآخرون (Anber and others, 2016) منهجين لتحليل الآراء، هما:

منهج معالجة اللغة الطبيعية : معالجة اللغة الطبيعية Natural Language Processing NLP هي التفاعل بين أجهزة الحواسيب ولغات البشر، فعند تحليل آراء المستخدمين آنياً، سيما تويتر، فإنه يجب استخدام توصيفات فعالة، حيث غالبية الدراسات تستخدم توصيفات الآراء الشائعة، وهي : الإيجابية، السلبية، والمحايدة، وربما بعض الآراء تكون مختلطة (إيجابية وسلبية) في ذات الوقت.

التعلم الآلي Machine Learning : وهو المنهج العلمي الذي يستكشف البنية والدروس للخوارزميات التي يمكن تعلمها من البيانات، ويتم فيها تطبيق تقنيات قائمة على قواعد، أو موجهة، أو شبه موجهة، ثم جمع تغريدات تويتر حول موضوع معين، ثم قياس آراء الناس حول ذلك الموضوع، وتصنيفها مع إدخال المفهوم الدلالي للتغريدات.

وذكرت أيضاً جيانكينغ وكرستاني (Jianqiang and Crestani, 2016) أربعة أنواع لطرق تحليل الآراء على تويتر، على النحو التالي :

طريقة التعلّم الآلي LM :

حيث أن معظم طرق تحليل الآراء تعتمد على أدوات تصنيف في مجال التعلّم الآلي LM، ومن جملتها : بساطة بايز Naïve Bayes، التروبيات العليا (Maximum Entropy (MaxEnt، آلة الدعم الموجه Support Vector Machine (SVM)، والانحدار اللوجستي Logistic Regression، وغيرها، وجميع هذه الوسائل تعمل على تصنيف التغريدات إلى سلبية أو إيجابية.

طريقة التحليل المعتمدة على المعاجم Lexicon-Based :

هي عبارة عن قائمة بالكلمات موصوفة بالقطبية (سلبية - إيجابية) أو درجة القطبية، من تحديد النتيجة العامة لتوجهات الرأي في نص معين. وهذه الطريقة مطبقة على نطاق واسع للنصوص التقليدية، كالمدونات، والمننديات، ومراجعات المنتجات. وهي أقل ظهوراً في تحليل آراء تويتر من طريقة التعلّم الآلي، والسبب وراء ذلك هو أن نصوص تويتر تنفرد ببعض الخصائص النصية والتعبيرات العامة، ذات الطبيعة الديناميكية، كالتعبيرات المتجددة من وقت لآخر، والوسوم (# tags) المتنوعة. وهذه الطريقة تعتمد على ثلاث مراحل لتحليل الآراء، هي : تهيئة النصوص، تحديد القطبية، ثم إظهار التصنيف حسب الشروط الموضوعية.

الطريقة الهجينة Hybrid :

وهي عملية دمج للطريقتين السابقة، التعلّم الآلي، وطريقة المعاجم، وهي تقترح خوارزميات مشروطة، تستخدم المصطلحات الذاتية، في الأحكام المقارنة، وتعابير النفي، وتلك التي يحتمل أن تغير اتجاه العبارة.

طريقة الرسوم البيانية Graph-Based :

على الرغم من أن طريقة التعلّم الآلي Learning Machine تحقق أداءً جيداً لتحليل الآراء على تويتر، إلا أنها تحتاج قدر كبير من البيانات التوضيحية التي تشرح النتائج، لذا كان الحل في الرسوم البيانية أو الاجتماعية لتويتر Twitter Social Graph، فيتم توزيع التصنيفات على نقاط الالتقاء Nodes، والربط بينها بواسطة الرسوم البيانية، حيث تمت الاستفادة من الرسم البياني لعنصر المتابعين في تويتر، على اعتبار أن المشتركين يوثرون على بعضهم البعض، وتم استخدام مكونات تويتر، كالمشاركين والتغريدات والوسوم والرموز، كنقاط Nodes لبناء الرسم البياني.

مستويات وأنواع تحليل الآراء :

يقسم كور وآخرون (Kaur and others, 2017) تحليل الآراء إلى ثلاث مستويات :

التحليل على مستوى الوثيقة *Document Level Analysis* : وهو يحدد اتجاه الآراء للوثائق، من خلال تحديد قطبية الوثيقة إلى إيجابية، سلبية، أو محايدة.

التحليل على مستوى الجملة *Sentence Level Analysis* : وهو يحدد اتجاه الآراء للجملة، من خلال تحديد قطبية الجملة إلى إيجابية، سلبية، أو محايدة.

التحليل على مستوى المظهر أو الميزة *Feature/Aspect Level Analysis* : يهدف إلى مظاهر الكيانات التي تستخدم في التعبير عن الآراء.

أما إبراهيم وسالم (Ibrahim and Salim, 2016) فذكروا أربعة مستويات لتحليل الآراء، وكل مستوى يستند على الآخر :

تصنيف الآراء للوثيقة *Document sentiment classification* :

هذه الطريقة تعالج الوثيقة كوحدة واحدة، وتعكس القرار كإيجابي أو سلبي أو غير حاسم، وفيه خصائص تحليل الآراء مهمة، حيث أن مصطلح الإلغاء " ليس " يمكن ان تعكس معنى الرأي، من إيجابي إلى سلبي، أو شيء آخر.

موضوع الجملة وتصنيف الآراء *Sentence subjectivity and sentiment classification* :

في هذا المستوى، يحاول الأسلوب "التكنيك" استخراج المعاني الذاتية للجملة الموضوعية، ثم يتوقف لإعطاء قرار إيجابي أو سلبي، وهذا الأسلوب يعمل بشكل جيد للجملة البسيطة، ولكن لا يزال هناك إشكاليات عديدة لجملة المقارنات، مثل : "أنا أحب المشي لكن المكان بعيد جداً".

تحليل الآراء المظهري *Aspect-Based sentiment Analysis* :

الأسلوب لهذا المستوى هذا المستوى يقوم بتحليل الوثيقة جنباً إلى جنب مع مستوى تحليل الجملة، وهذا الأسلوب له بعض الفوائد الجانبية لا تستطيع إعطاء قرارات إيجابية أو سلبية بشكل مباشر، وبالتالي غير قادر على تشكيل رأي حاسم، وتعتبر الكلمات المستهدفة ككيان، والكلمات التفسيرية تسمى مميزات.

ملخص الرأي المظهري *Aspect-Based opinion summarization* :

الأساليب لهذا المستوى توظف جميع الآراء على أساس المظاهر المتعددة من عدة وثائق، وتعطي القرار الإيجابي أو السلبي.

تحديات وعوائق تحليل الآراء :

يواجه تحليل الآراء الواردة في النصوص باستخدام النظم الإلكترونية الكثير من التحديات والاشكاليات، وبالخصوص عندما يتعلق الأمر بتحليل الآراء لمحتوى شبكات التواصل الاجتماعي، أو الإعلام الاجتماعي، من هذه التحديات ما يتعلق اللغة الواردة في هذا المحتوى، والتي لا تخضع للضوابط المتعارف عليها، ومنها ما يتعلق بطبيعة منصات الاعلام الاجتماعي التي تتميز بالأسلوب الحر في الكتابة والتعبير، والتي يغلب عليها استخدام اللهجات المحلية للكتابة.

وفي هذا الاطار سوف نتناول بعض التحديات التي تواجه تحليل الآراء في نصوص شبكات التواصل بوجه عام، وكذلك المصاعب التي تواجه تحليل الآراء في اللغة العربية على وجه الخصوص.

ومن التحديات العامة الشائعة، فيما يتعلق بتحليل الآراء في تويتر، ما ذكره جياكانو وكريستاني (Giachanou and Crestani, 2016)، على النحو التالي :

طول النص : وهو من المزايا الفريدة لتويتر حيث لا تتجاوز التغريدة ١٤٠ حرفاً - للغة الإنجليزية - وهو ما يجعل تصنيف التغريدة إلى إيجابية أو سلبية أمراً صعباً.

الصلة بالموضوع : وهو ما يجعل عملية تحليل الآراء لتويتر تتم بعيداً عن سياق الموضوع.

اللغة غير السليمة : إن استخدام اللغة غير الرسمية، إضافةً إلى محدودية عدد الأحرف، يجعل المغردين يكتبون عبارات غير صحيحة لغوياً.

البيانات المتفرقة : تغريدات تويتر تحمل الكثير من التشويش، بسبب الرموز # ، @ ، والروابط url ، مما يشتت النص.

المعاني العكسية : وجود كلمات أو حروف النفي يؤثر بشكل كبير في تحديد قطبية الآراء (سلبية او إيجابية) للتغريدة. حيث أن إكتشاف معنى النفي في العبارة يسهم في قلب قطبية العبارة.

كلمات التوقف : وهي الحروف (إن، في، لكن، وغيرها)، وهي عادة ما تستبعد عند معالجة النصوص، وهو ما يحدث خلل في عملية التحليل، فكلمة " إعجاب " أي " Like " في اللغة الإنجليزية من الكلمات المستبعدة stop words، إلا أنها ذات أثر أساسي في تحليل الآراء على تويتر.

الترميز Tokenization أو التشفير، والتي يستخدمها المغرد بهدف الاختصار، أو بهدف السرية، أو إيصال مغزى مختلف، كل ذلك يجعل اكتشاف الرأي الصحيح للتغريدة أمراً صعباً.

تعدد اللغات في تويتر، فربما تجد تغريدات بأكثر من لغة في وسم واحد.

تعدد الوسائط، فالتغريدة يمكن أن تحوي على مقاطع فيديو أو صور.

عوائق تحليل الآراء في اللغة العربية :

يعتبر تحليل الآراء على شبكات التواصل الاجتماعي توجه حديث ويتواصل تطويره على الرغم من توفر الأدوات الخاصة بإدارة الشبكات لكن أدوات تحليل المحتوى أو الآراء أو المشاعر هي الأكثر صعوبة والحاحاً في التطوير خاصة بما يخص اللغة العربية التي تتصف بسمات لغوية خاصة، فما يخص المحتوى العربي فإن هذا التوجه لا يحظى بقدر كبير من دراسات والأبحاث، مع عدم توفر أدوات مخصصة لتحليل الآراء للنصوص العربية. وفي هذا الإطار فإن هناك بعض العوائق والصعوبات التي تخص اللغة العربية بحد ذاتها.

وقد أورد حمدي وآخرون (Hamada and others, 2016) عدداً من العوائق التي تواجه التحليل العربي للآراء، على النحو الآتي :

صعوبة اللغة العربية ذاتها : حيث أن اللغة العربية تلك بنية صرفية (نحوية) معقدة، تقوم على مجموعة أنماط من جذور الكلمات، كما أنها تتدرج بين اللغة الكلاسيكية (الفصحى) وهي لغة القرآن الكريم، واللغة العربية المعاصرة، المستخدمة في كتابة الأخبار والدروس التعليمية والحوارات الرسمية، كما أن هناك اللغة العامية - اللهجات - المستخدمة في التواصل اليومي بين الناس.

التعقيدات الصرفية (النحوية) : حيث أنها من اللغات الغنية في تركيبها، والتي تشكل تحدياً كبيراً لنظم البرمجة اللغوية العصبية NLP بشكل عام، كما أنها غنية بالاشتقاقات والمترادفات.

اللهجات العربية : وهذا يخلق مزيد من المصاعب لتحليل الآراء، حيث تستخدم للكتابة في شبكات التواصل الاجتماعي الكثير من اللهجات العربية المختلفة.

محدودية المصادر : حيث أن المصادر البحثية باللغة العربية في مجال تحليل الآراء شحيحة جداً.

ضبابية القطبية : معظم طرق تحليل الآراء تستخدم القطبية (سلبى، إيجابى)، وفي كثير من الأحيان يصعب تحديد قطبية النص في اللغة العربية، حيث كل شخص يرى توجه النص من وجهة نظره هو، وقد يختلف مع الآخر.

الاعتمادية على المجال : حيث تحديد توجهات الرأي في النصوص يعتمد على مجال الموضوع، فتحديد التصنيف في عبارة معينة قد يكون دقيقاً في مجال معين، وغير دقيق في مجال آخر، لنفس العبارة.

الآراء الضمنية : كل عبارة قد تحمل آراء ومشاعر مخفية تختلف عن معنى ظاهر العبارة، وهنا قد تكون نتائج تحليل الآراء بخلاف مقصود صاحبها.

كما بينت الحربي وإمام (Al-Harbi and Emam, 2015) بعض التحديات التي تخص اللغة العربية، كما يلي :

أن اللغة العربية هي واحدة من اللغات السامية، وهي لغة قديمة، تكتب من اليمين إلى اليسار، بطريقة مخطوطة (متصلة).

هي لغة مكونة من ٢٨ حرفاً ساكناً، ولا تحوي حروفاً كبيرة كتلك الموجودة في بداية الجمل في اللغة الإنجليزية.

اللغة العربية تعتمد على قوانين الصرف (والنحو) المعقدة، بخلاف اللغة الإنجليزية.

اللغة العربية متغيرة للغاية (يتغير شكل الكلمات بحسب قوانين الصرف)، وتحتوي على الكثير من الاشتقاقات، مما يجعل عملية التحليل أمراً صعباً ومعقداً.

الآراء في اللغة العربية تعتمد على سياق الموضوع وعلى الطابع الشخصي، فقد تواجه قطبيات مختلفة لنفس العبارات في سياقات مختلفة.

استخدام العرب عند الكتابة في الإنترنت اللهجات الخاصة، بدلاً من اللغة الفصحى، وهذا الأخطاء الإملائية (والنحوية مرتفعة)، مما يشكل تحدياً إضافياً.

الجزء التطبيقي للدراسة :

أولاً - منهجية وطريقة البحث :

الحقيقة أننا لم نتبع أسلوب تصنيف القطبية للتغريدات كباقي الدراسات السابقة التي تم تناولها، حيث قامت على تصنيف التغريدات إلى (إيجابية أو سلبية) من خلال الخوارزميات المخصصة لهذا الأسلوب، بل قمنا بتصنيف الآراء بشكل ديناميكي إلى ثلاثة تصنيفات، وهي (متعصبة، عنصرية، ومحايدة) وتعد كلا الصنفين المتعصبة والعنصرية آراء سلبية، رغم ما يحمله التعصب من ميول إيجابي أو سلبي تجاه نادي معين.

ولتوضيح ذلك، فإن التعصب يعد سلوكاً سلبياً في حد ذاته، ولكنه قد يبدو رأياً إيجابياً تجاه نادٍ معين، كأن يكون مغالياً في حب النادي أو الدفاع عنه، وقد يكون رأياً سلبياً تجاه نادٍ آخر كأن يحمل شتيمة أو انتقاصاً متعصباً تجاه ذاك النادي. وفي نموذج تحليل الآراء المقدم في هذه الدراسة فإن معالج تحليل الآراء يصنف كلا الحالتين في خانة الآراء المتعصبة.

وأما المعيارية المتبعة لتصنيف الآراء إلى ثلاثة اتجاهات (متعصبة، عنصرية، ومحايدة) فإنها تتلاءم مع طبيعة الدراسة الذي تهدف إلى قياس مستوى التعصب الرياضي، ويمكن إيضاحها على النحو الآتي:

الاتجاه المتعصب : هو أن تحمل التغريدة كلاماً متعصباً وغير موضوعي من خلال العبارات أو الكلمات الواردة فيها، فتصنف "متعصبة" بغض النظر عن كونها مع أو ضد نادٍ معين، فالتعصب هو التعصب.

الاتجاه العنصري : فإذا احتوت التغريدات على كلمات أو عبارات عنصرية، تتناول العرق أو الدين على سبيل المثال، فإن معالج الآراء يصنفها "عنصرية".

الاتجاه المحايد : أما التغريدات التي تحمل كلاماً موضوعياً من خلال عبارات أو كلمات معتدلة فإننا نصنفها محايدة، أي أنها ليست "متعصبة" ولا "عنصرية".

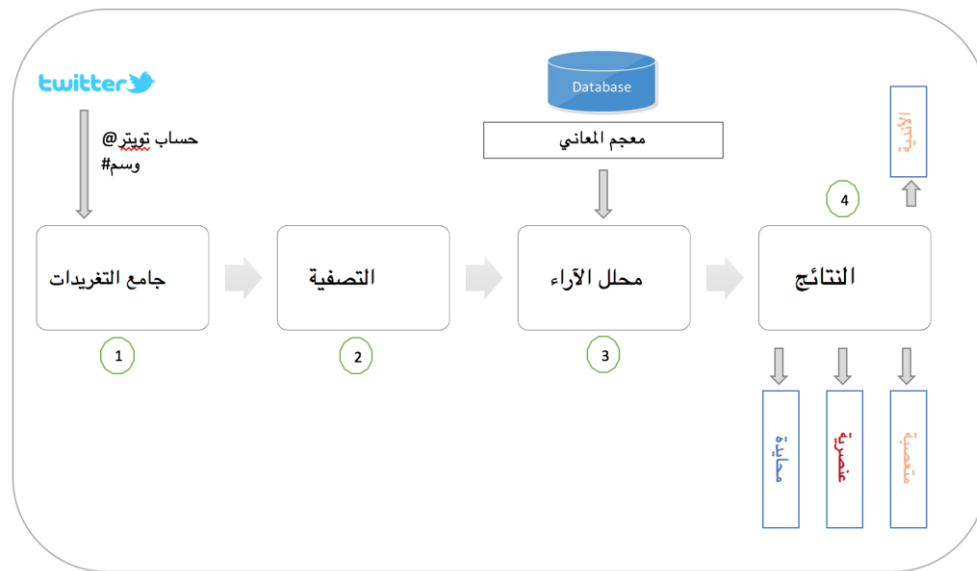
وقد تم استخدام المنهج التجريبي وتم تصميم برنامج تطبيقي لهذا الغرض، من خلال تطوير منصة إلكترونية على الويب تحاكي النموذج المقترح، بهدف تحليل الآراء الرياضية على تويتر، وتم اختبار هذه المنصة من خلال تكوين قاعدة بيانات من ١٠,٠٠٠ تغريدة، من وسم #أكشن_يا_دوري، مع قابلية النموذج والمنصة الإلكترونية للتوسع في المزيد من التصنيفات.

#أكشن_يا_دوري هو : وسم يخص أحد البرامج الرياضية الجماهيرية، الذي يناقش الدوري السعودي لكرة القدم، ويبث على قناة تابعة لإحدى الشبكات التلفزيونية العربية.

ولأجل تحديد تصنيفات الآراء تم استخدام طريقة المعجم Lexicon Method، من خلال بناء معجم كبير من الكلمات والتعبيرات الرياضية باللهجة السعودية، وتخزينها في قاعدة بيانات المنصة الإلكترونية.

أما آلية تكوين معجم الآراء، فقد تم من خلال إدخال عدد كبير من الكلمات والعبارات بواسطة مصمم التطبيق الإلكتروني، بالاستفادة من تلك المتداولة في الأحاديث والنقاشات الرياضية، ثم منح كل كلمة أو عبارة اتجاهًا محددًا (متعصب، عنصرية، أو محايدة)، وفي ذات الوقت فإن التطبيق الإلكتروني يتيح إمكانية إعادة تصنيف الكلمات والعبارات الملتبسة، بواسطة المستخدم أو المستخدم، حيث يقوم التطبيق بعرض جميع الكلمات التي تعذر عليه تصنيفها في اتجاه معين، نتيجة محدودية قاعدة بيانات المعجم، ليتمكن المستخدم من إضافتها آلياً لقاعدة معجم الكلمات، بعد تصنيفها في الاتجاه المناسب.

النموذج التطبيقي :



- وصف النموذج :

يوضح النموذج أعلاه، آلية تحليل الآراء الرياضية، من خلال ما ينشر على تويتر كأحد قنوات التواصل الاجتماعي الأكثر رواجاً بين أفراد المجتمع السعودي، ويعتمد النموذج على برمجيات الحاسب الآلي المستخدمة في تطبيقات الويب، من أجل تشغيل التطبيق الإلكتروني على منصة الإنترنت، ويعمل النموذج وفق التسلسل الآتي :

جامع التغريدات Tweets Collector : ويعمل هذا الجزء من نموذج النظام على الاتصال بشبكة تويتر، عبر ما يعرف تقنياً بواجهة التطبيق البرمجية Application Programming Interface - API، ويلزم في هذه الحالة استخدام حساب تويتر نشط. ويعمل النموذج على استخلاص التغريدات وفقاً لحساب معين أو وسم #tag محدد.

ويتم حفظ جميع التغريدات في قواعد بيانات التطبيق الإلكتروني، من أجل عرضها إلى جانب النتائج، أو الرجوع إليها لإعادة التحليل لأي سبب.

تصفية التغريدات Tweets filtering : وفي هذه المرحلة يتم تصفية التغريدات من أي وسوم أو حسابات، ضمنها المغرد تغريداته، ومن تلك الرموز على سبيل المثال (، ، - _ ' " : # ... ؛)، بالإضافة إلى توحيد شكل الحروف لتقريب الكلمات قدر الإمكان إلى الأشكال الواردة في المعجم، مثلاً الحروف (ا أ إ) تصبح جميعها " ا " .

محلل الآراء Sentiment Analyzer : ويعد هذا الجزء من النموذج هو المعالج الفعلي لإظهار النتائج النهائية، وهنا يتم تنفيذ اتجاه الرأي الرياضي، معتمداً على قاعدة بيانات تحوي عدد كبير من الكلمات والعبارات ذات الصلة بالاتجاهات الرياضية.

فكما أسلفنا، يتم عرض الكلمات والعبارات الواردة في التغريدة على معجم الكلمات في قواعد البيانات، ومن هناك يتم تحديد تصنيف الكلمة أو العبارة، ثم يتم أخذ مجمل التصنيفات لتحديد اتجاه الرأي الوارد في التغريدة، إلى "متعصب"، "عنصري"، أو "محايد".

النتائج Result : وعند هذه النقطة يقوم النظام بتصنيف اتجاه كل تغريدة إلى كونها متعصبة، أو عنصرية تجاه نادي أو لاعب معين، أو محايدة. وتظهر النتائج عبر واجهة النظام في شكل جداول ورسوم بيانية.

كما يتم تخزين نتيجة التحليل في قواعد البيانات، ليتم الرجوع إليها في عند الحاجة، دون الحاجة إلى إعادة عملية المعالجة.

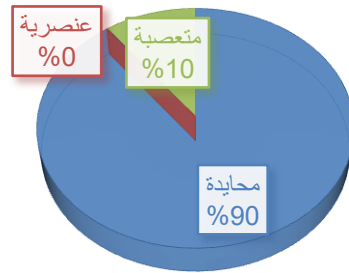
ثانياً - تشغيل النموذج واستقراء نتائج التحليل :

تم تشغيل البرنامج الإلكتروني (ميول) والحصول على القراءات الآتية :

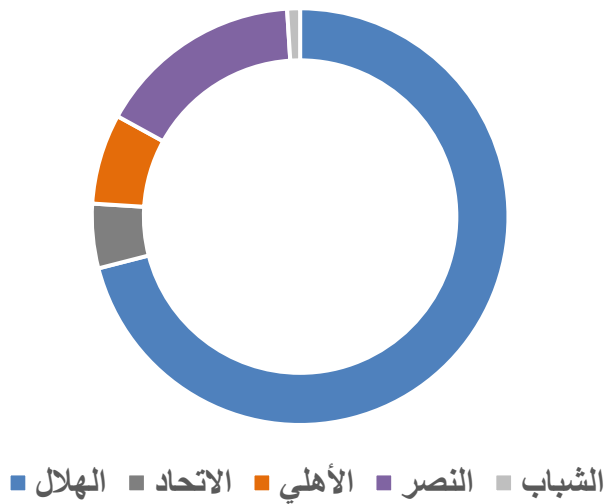
الأعداد والنسب لتصنيفات الآراء الرياضية :

التصنيف	العدد	النسبة %
متعصبة	١٠١٩	%١٥
عنصرية	١٠	%٥

محايده	٨٨٨١	%٨٠
--------	------	-----

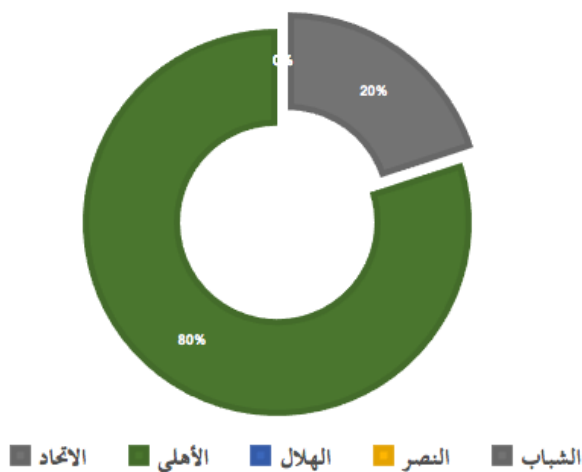


ترتيب الفرق الرياضية من حيث التعصب (وقت جمع البيانات) :



النسبة %	النادي
٧١%	الهلال
١٦%	النصر
٧%	الأهلي
٥%	الاتحاد
١%	الشباب

الفرق الأكثر تعرضاً للسخرية والعنصرية (وقت جمع البيانات) :



النسبة %	النادي
٨٠%	الأهلي
٢٠%	الاتحاد
٠%	الهلال

النصر	%٠
الشباب	%٠

النتائج

كان للدراسة الحالية هدف تجريب منصة تم تطويرها لتحليل محتوى تغريدات عربية عبر شبكة تويتر لوسم جماهيري معروف. و مما سبق يمكننا استخلاص النتائج كالاتي :

- ان التعصب الرياضي لا يعد ظاهرة، رغم بشاعته، فقد أظهرت الدراسة أن غالبية التغريدات والتعليقات في وسم #أكشن_يا_دوري هي محايدة، وقد يكون مرد ذلك إلى عدم ممارسة أو إظهار التعصب الرياضي عبر الوسائل الرسمية، أو شبه الرسمية كقناة تويتر، إبتعاداً عن الحرج، وتجنباً للملاحقات القضائية. في حين أن مظاهر التعصب كثيراً ما تتواجد في هتافات الجمهور، وقد تصل في بعض الأحيان إلى العنصرية. كما ان التحليل مرتبط بالموضوعات التي طرحت في فترة جمع البيانات وتم الحوار حولها.
- ان الكتابة باللغة العامية وباللهجة المحلية تشكل تحدياً كبيراً لنمذجة أساليب التحليل، حيث تفقد الخوارزميات البرمجية المعيارية، وهي عامل مهم في صياغة كود البرمجة.
- كثرة الأخطاء الإملائية، واستخدام الرموز الرسومية، قد يحول دون إمكانية تصنيف التغريدة من خلال المنصة الإلكترونية.
- ظهرت نسبة خطأ (لا يستهان بها) في نتائج التحليل، نرجعها إلى الأسباب المذكورة في النقاط السابقة.
- توفر إمكانية لتطوير هذا النموذج، وتجويد الخوارزميات البرمجية لمنصة (ميول) للحصول على نتائج أكثر دقة، وتغطية مزيد من المجالات.

التوصيات

- يوصي الباحثان بأهمية وضرورة تكثيف البحوث العربية في مجال تحليل الآراء، كونه أحد التوجهات المعاصرة الحديثة ذات المنفعة لتطوير الأدوات في مجال تحليل محتوى أدوات التواصل الاجتماعية.
- نحث مؤسسات البرمجة على تطوير أدوات وأنظمة برمجة لتحليل الآراء للمحتوى العربي، يتوفر الدقة والكفاءة العالية. وأن تقام مشاريع بحثية تشارك فيها فرق بحثية تقنية لغوية وليس مشاريع فردية.
- ندعو كذلك مؤسسات المعرفة لتبني منصات إلكترونية لتحليل الآراء للنصوص العربية، بعدة طرق وأساليب، للأغراض التعليمية والتدريبية.

قائمة المراجع

- الزهيري، طلال ناظم، كاظم، هشام جواد (٢٠١٤)، الشبكات الاجتماعية ودورها في الاتصال غير الرسمي : المختصين في مجال المعلومات والمكتبات أنموذجاً، المجلة العراقية لتكنولوجيا المعلومات، مج ٦، ع ٤، ص ص ٢٤-٤١.
- عبدالحاميد، طلعت أسعد وآخرون (٢٠١٣)، أثر الاتصال عبر مواقع الشبكات الاجتماعية على المشاركة السياسية للشباب : دراسة تطبيقية انتخابات رئاسة الجمهورية في مصر، المجلة المصرية للدراسات التجارية، مج ٣٧، ع ٢، ص ص ٢٩٩-٣٢٣.
- Al-Harbi, Waad, Emam, Ahmed (2015) “EFFECT OF SAUDI DIALECT PREPROCESSING ON ARABIC SENTIMENT ANALYSIS”. *IJACT*, Vol. 4, No. 6, Dec. 2015, pp. 91-99.
- Alowaidi, Sana and others (2017) “Semantic Sentiment Analysis of Arabic Texts”. *IJACSA*, Vol. 8, No. 2, 2017, pp. 256-262.
- Anber, Hana and Others (2015) “A Literature Review on Twitter Data Analysis”. *International Journal of Computer and Electrical Engineering* , Vol. 8, Iss. 3, June 2016, pp. 241-249.
- Giachanou, Anastasia, Crestani, Fabio (2016) “Like It or Not: A Survey of Twitter Sentiment Analysis Methods”. *ACM Comput. Surv.* , Vol. 49, Issu. 2, Article 28, June 2016, p. 41.
- Hamdi, Ali and Others (2016) “A Review on Challenging Issues in Arabic Sentiment Analysis”. *Journal of Computer Sciences*, Vol. 12, No. 9, pp. 471-481.

- Ibrahim, Mnahel Ahmed, Salim, Naomie (2016) “Sentiment Analysis of Arabic Tweets: With Special Reference Restaurant Tweets”. *IJCST*, Vol. 4, Iss. 3, May-Jun 2016, pp. 173-179.
- Jianqiang, Zhao, Xiaolin, Gui (2017) “Comparison Research on Text Pre-processing Methods on Twitter Sentiment Analysis”. *IEEE* , Vol. 5, pp. 2870-2879.
- Kaur, Iqbaldeep and others (2017) “Agile Model Based Sentiment Analysis from Social Media”. *IJACSA*, Vol. 15, No. 1, Jan. 2017, pp. 433-454.