

استخدام طرق التعلم العميق وتعلم الآلة لاستخلاص المنشورات العربية في مواقع

التواصل الاجتماعي ذات المحتوى العنصري

إعداد

أشرف عبد الحميد المهيرات

المشرف

الدكتور طارق غازي كنعان

الملخص

ازداد في الآونة الأخيرة استخدام الإنترنت ، وخاصة مواقع التواصل الاجتماعي مثل Twitter و Facebook؛ الذي يساهم في توفير مساحة مفتوحة للعالم ويشجع الناس على المشاركة في مشاعرهم ، مما يساهم في نمو العادات السلبية بين مستخدمي مواقع التواصل الاجتماعي ؛ ويُعرف أحد هذه السلوكيات بالتمييز العنصري. وتعد العنصرية من أحد أنواع خطابات الكراهية التي تستغل وسائل التواصل الاجتماعي ؛ وتهدف دراستنا إلى اقتراح نموذج لتصنيف العنصرية في محتوى الوسائط الاجتماعية العربية متعدد الاصناف (العرق ، الدين ، الإعاقة ، أخرى) من خلال طريقتين: وهي التعلم الآلي وايضا التعلم العميق ، حيث تكشف العديد من الدراسات عن ظاهرة العنصرية في اللغة الإنجليزية بينما تقتصر هذه الدراسات على اللغة العربية. حيث استخدمنا أدوات المعالجة المسبقة للتعامل مع اللغة العربية. طبقنا لثلاثة تقنيات شهيرة في البرمجة اللغوية العصبية (التطبيع ، وإيقاف إزالة الكلمات ، والجذع). وقمنا بإنشاء مجموعات من البيانات من وسائل التواصل الاجتماعي العربية من تويتر وفيسبوك بما يقارب من (10.319) منشورًا.

استخدمنا مصنفًا مشهورًا متعدد الفئات في التعلم الآلي ؛ أشهرها آلة ناقلات الدعم (SVM) و Naive Bayes (NB) و شجرة القرار (J48) و الغابة العشوائية (RF) و K-Nearest Neighbor , (kNN) وقمنا بتقييم دقة المصنف من خلال حساب قياسات F1-Measure والاستدعاء والدقة من ناحية أخرى ، واستخدمنا أساليب التعلم العميق ، والتي تحتوي على الشبكات العصبية التلافيفية (CNN) ، والشبكات العصبية المتكررة (RNN) ، والشبكات العصبية العميقة (DNN) التي تتضمن طريقة استخراج الخصائص من خلال استخدام Embedding Word (Word2Vector) .

وأظهرت النتائج أن مصنف SVM حصل على أعلى قيمة لـ F1-Measure في كل من مجموعة بيانات Twitter بنسبة (83.7%) ومجموعة بيانات Facebook بنسبة (77.4%). بينما في تقنيات التعلم العميق

، حيث حصلنا على نتائج أفضل من مصنف التعلم الآلي مثل RNN / LSTM حيث اعطت أعلى قيمة في تقنيات الشبكة العصبية الأخرى في كل من مجموعة بيانات Twitter على مقياس F1-Measure بنسبة (89.97%)، ومجموعة البيانات Facebook بنسبة (87.3%).

الكلمات المفتاحية: التعلم الآلي ، التعلم العميق ، البرمجة اللغوية العصبية ، اللغة العربية ، وسائل التواصل الاجتماعي.