الكشف عن نغمة الاكتئاب التي تؤدي إلى الانتحار في وسائل التواصل الاجتماعي العربية باستخدام تقنيات التعلم الآلي والتعلم العميق

اعداد

عدي مأمون صدقة

المشرف

الدكتور طارق غازي كنعان

المشرف المساعد

الدكتور محمد البس

الملخص

تهدف هذه الدراسة إلى اقتراح نظام للكشف عن الأكتئاب الذي يؤدي إلى الانتحار من المحتوى العربي على وسائل التواصل الاجتماعي ، على وجه الخصوص ، Facebook و Twitter ، باستخدام طريقتين شهيرتين ، الطريقة الأولى خوارزميات التعلم الآلي ، والطريقة الثانية خوارزميات التعلم العميق.

هناك العديد من الدراسات والبحوث التي تركز على هذا الموضوع باللغة الإنجليزية ، لكنها غير موجودة باللغة العربية. لذلك ، يركز بحثنا على اللغة العربية ، على الرغم من الصعوبات التي نواجهها في اللغة العربية ، قمنا بتبسيط مفاهيم اللغة العربية باستخدام ثلاثة من أشهر معالجات اللغة الطبيعية العربية ، والتي تشمل الجذع والتطبيع ووقف إزالة الكلمات .

لذلك نجمع مجموعة البيانات من مواقع التواصل الاجتماعي فيسبوك وتويتر ، وقد تم جمع العدد الإجمالي لحوالي 10000 منشور. نستخدم تطبيق Netvizz لاستخراج البيانات من منصة Facebook وتطبيق TAGS لاستخراج البيانات من منصة Twitter.

بعد ذلك ، قمنا بتطبيق خمسة مصنفات في تقنيات التعلم الآلي ، والتي تشمل Naive Bayes (NB) ، وآلة ناقلات الدعم (SVM) ، و K-Nearest Neighbor (KNN) ، وشجرة القرار (J48) ، والغابة العشوائية (RF). بالإضافة إلى ذلك ، من أجل تقييم أداء المصنفات ، قمنا بحساب قياس F1 ، واستدعاء ، وقياس الدقة.

أظهرت النتائج تفوق المصنف RF على القيمة الأعلى لمقياس F1 بنسبة (99.1٪). من ناحية أخرى ، تم تحقيق أفضل نتيجة مع جميع مجموعات البيانات دون تطبيق تقنيات المعالجة المسبقة ، من خلال SVM مع قياس F1 (93.4٪).

بالإضافة إلى ذلك ، عند فصل مجموعة البيانات في مجموعة بيانات twitter ومجموعة بيانات Facebook ، أظهرت النتيجة تفوقًا لمصنف التردد اللاسلكي على أعلى قيمة لقياس F1 بنسبة (99.3٪) و (99.1٪) على التوالي مع جميع مجموعات البيانات وتطبيق تقنيات المعالجة المسبقة.

من ناحية أخرى ، نطبق تقنيات التعلم العميق ، والتي تشمل الشبكات العصبية التلافيفية (CNN) والشبكات العصبية المتكررة (RNN) والشبكات العصبية العميقة (DNN). نستخدم لغة برمجة Python ، على وجه الخصوص ، مكتبة Keras لتطبيق DL. بالإضافة إلى ذلك ، نطبق طريقة استخراج الميزة ، والتي تتضمن تضمين الكلمات وحقيبة الكلمات (BOW) في مجموعة البيانات.

أظهرت النتائج تفوق خوارزمية DNN على أعلى قيمة لقياس F1 بنسبة (96٪) ، تليها CNN بنسبة (94٪) ، و RNN بنسبة (62٪) على التوالي مع جميع مجموعات البيانات.

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

.