حيث تختبر هذه الدراسة نماذج التعلم الآلي المراقبة بما في ذلك شبكة التحليل التكراري (RNN)، حيث تعمل بشكل نسبي أفضل من النموذج غير المراقب ونموذج التصنيف السرجي القائم على الاحتمالات التقليدية. يناسب النهج العميق للتعلم الآلي اللغة الكورية بشكل أفضل من النماذج غير المراقبة المعتمدة على المعاجم لعدة أسباب، وتحاول الدراسة تطبيق تحليل المشاعر على السياسة الكورية، مما يشير إلى أن البيانات اللغوية للنموذج غير المراقب يجب أن تتناسب مع السياق السياسي في كوريا الجنوبية، ثانياً التحليل الصرفي صعب للغة الكورية لأنها لغة اشتقاقية (حيث يمكن أن يتغير معنى الكلمة اعتمادًا على الكلمة المضافة المجاورة). ويعتبر التضمين عملية لتحويل الكلمات إلى موجهات يمكن للكمبيوتر فهمها، ويشير حجم التضمين إلى بُعدية متجهات التضمين وهو عامل مرتبط بدقة تعقيد اللغة الطبيعية، وتستخدم هذه الدراسة أحجام تضمين قدرها 100 و200 و300 (BERT هو نموذج مدرب مسبقًا يتضمن نوعًا فريدًا من التضمين وبُعدًا مُعينًا)، ويعد نموذج BERT هو منهجية مدربة مُسبقًا، وتطبق هذه الدراسة عملية تحويل التعلم على منهجية kobert لعمل تصنيف المشاعر السياسية الخاصة بالنصوص الكورية على تويتر، وهذا يعني أن الشبكة تميز إما بين إيجابي وسلبي أو بين إيجابي ومحايد وسلبي. الخلاصة هناك ثلاثة أنواع من التضمين بثلاثة أحجام بُعدية وثلاث شبكات عصبية بخياري تصنيف، وفيما يلي تؤدي نوعين من التضمين، وثلاثة أبعاد من التضمين وثلاث شبكات عصبية وفئتي تصنيف إلى مجموع 26 من التجميعات للاختبار (تستخدم الدراسة Python لمعالجة بيانات تويتر مسبقًا وPyTorch لبناء الشبكات العصبية، يطبق نتيجة الاختبار التسلسلي الأنسب للجهاز ويحصل على المشاعر اليومية من جميع التغريدات المجمعة في عام 2019 واختبار، بينما تُستخدم المجموعات الأخرى للتحقق والاختبار والتي تقيس جدوى النظام أثناء التدريب وفي نهايته، وبالتالي لا يتغير حتى انتهاء تدريب المنهجية، ووفقًا لدرجات الدقة في الجدول 2، أي تقليل فئة التصنيف يحسن الدقة في جميع التجميعات، 7 لتحديد ما إذا كانت التغريدة تكشف عن المشاعر، 92٪ على مجموعة الاختبار، 08٪ من نموذج RNN. على سبيل المثال، Word2Vec ،FastText هو توافق أفضل مع CNN، بينما يعطي FastTrack نتائج أفضل بشكل عام مع RNN، وبشكل عام لا توجد فروع متميزة من حيث الدقة التي تتغير اعتمادًا على تركيبات العناصر المختلفة. وتحلل هذه الدراسة مشاعر جميع التغريدات المجمعة في عام 2019 باستخدام نموذج BERT المخصص للتصنيف بين فئتين. ومن بين الفروع التي تم اختبارها، يُحقِّق BERT نتائج دقة عالية لجميع المهام المعطاة مع أداء مستقر نسبيًا، ويعتبر BERT هو أكثر شبكة عصبية متقدمة بين الأنظمة التي تم اختبارها في التحليل ومصمم للتعامل مع البيانات التسلسلية المعقدة مثل اللغات الطبيعية، فهو النظام الأكثر بساطة في التنفيذ في مهمة تصنيف المشاعر مقارنة بالآخرين، وبشكل عام تتضمن الدراسة عملية تسلسلية لاستخراج الرأي العام عبر الإنترنت، أولاً يقوم الجهاز تلقائيًا بجمع النصوص التي ينشئها المستخدمين عبر تويتر في هذه الدراسة حول موضوع معين لفترة زمنية معينة، وتحدد الأغلبية القاعدة النهائية لتصنيف المشاعر للنص، ثالثًا تدرب البيانات المشفرة نموذج تعلم عميق لبناء تصنيف المشاعر، وأخيرًا يتم تجميع النصوص ذات الرأي الواحد المجمع،

lakhasly.com © تم تلخيص النص بواسطة موقع لخصلي