نموذج محسن للكشف عن التغريدات العربية الغير مرغوب فيها في تويتر

مرام محمد دوغان

اسم المشرف على الرسالة

د/ محمد بشيري

د/ منال كلكتاوي

المستخلص

تعد الرسائل الغير مرغوب فيها من الأنشطة التي تؤثر سلبا على تجربة المستخدمين للإنترنت. أحد أشهر تطبيقات التواصل الاجتماعي هو تويتر، حيث أن المستخدمين يقومون بتبادل رسائل قصيرة حول الأخبار، السياسة، وتجارب الحياة اليومية. وهذا ما أدى لانتشار واسع للرسائل الغير مرغوب فيها في منصة تويتر وتضم تلك الرسائل ما يلي: نشر إعلانات الغير مدفوعة، نشر محتوى ضار أو غير ذي صلة وتعتبر هذه الرسائل من المشاكل الأمنية الحديثة وأيضا هدراً للموارد. ظهرت مؤخرا عدة طرق وأساليب للتعرف وتحديد هوية مرسلو هذه الرسائل المزعجة. على الرغم من أن هذه الأساليب قدمت مساهمات أساسية في مجال اللغة الإنجليزية. إلا أن القليل منها حتى الآن مغطى باللغة العربية. هناك تحديات عدة تواجه الأساليب الحالية للكشف عن الرسائل الغير مرغوب فيها باللغة العربية، وبالأخص التعامل مع طبيعة اللغة العربية واستخراج الميزات. لذلك نحن نركز على تطوير نظام قوي للكشف عن الرسائل الغير مرغوب فيها يمكنه اكتشاف الإعلانات الموجودة في علامات التصنيف الشائعة بالمملكة العربية السعودية. وبناءً على ما سبق، تقترح هذه الرسالة نموذج التعلم العميق القائم على ذاكرة طويلة قصيرة الأمد (LSTM)نوع من بنية الشبكة العصبونية التكرارية (RNN) تعتمد هذه الآلية على تضمين الكلمات المدربة مسبقا باعتبارها هندسة ميزات حديثة للكشف عن الرسائل الغير مرغوب فيها المنشورة باللغة العربية. يحقق هذا النموذج المقدم درجة F1 بنسبة ٩٩.٢٨٪. والتي تتفوق على خوارزميات التعلم الآلي الأخرى المستخدمة للكشف عن الرسائل الغير مرغوب فيها باللغة العربية.

An Enhanced Spam Detection Model For Arabic Tweets

Meram Mehmet Mahmut Dogan

Supervised By
Dr. Mohammed Basheri
Dr. Manal Kalkatawi

ABSTRACT

Spam is an activity that impacts the experience of users on the internet. One of the most popular social networks is twitter, where people exchange short text messages about news, politics, life experiences, etc. Twitter has led to an increase in the spread of spam which is used for advertisements, spread malicious, or just irrelevant content which introduces new security issues and waste of resources. Recently, several approaches have been identified in research for identifying spammers. Even though these approaches presented essential contributions to the field in the English language, few until now covered in the Arabic language. Several challenges are facing existing approaches for Arabic spam detection, especially, handling the morphological nature of Arabic and feature extraction. Therefore, we focus on having a robust spam detection model that could detect the advertisements in trending hashtags in Saudi Arabia. Accordingly, this thesis proposes a deep learning model based on Long Short Term Memory (LSTM) an artificial Recurrent Neural Network (RNN) architecture supported by pre-trained word embedding as a modern feature engineering to detect Arabic spam tweets. Our model achieves an F1-score of 99.28%, which outperforms other machine learning algorithms used in Arabic spam detection.