Hate Speech Sentiment Analysis for Arabic Language Using Support Vector Machine (SVM) and Convolutional Neural Network

Prepared By

Worod Hawamdeh

Supervisor BY

Dr. Ghaith Jaradat

Abstract

This thesis presents a comprehensive study on hate speech sentiment analysis in the context of Arabic language. The objective is to develop effective machine learning and deep learning models with genetic algorithm that can accurately classify hate speech sentiments and contribute to mitigating the negative impact of online hate speech. The research explores various techniques for data preprocessing, feature selection, and model training, including Genetic Support Vector Machines, Genetic Naive Bayes, Genetic Convolutional Neural Networks. Performance evaluation is conducted using metrics such as precision, recall, F1-score, and accuracy.

When comparing the previous classifiers, it was observed that the support vector machine algorithm was superior with an accuracy of 0.93 and an F1-score of 0.96, followed by genetic deep learning networks with

an accuracy of 0.91 and an F1-score of 0.92, and finally the genetic Naïve Bayes algorithm with an accuracy of 0.87 and an F1-score of 0.92.The research presented a comparison between its results and the results of previous studies conducted on the same dataset, and it was noted that the SVM algorithm was superior with an accuracy of 0.93 over the study that used AraBERT (Husain, 2021), which gave an accuracy of 0.80, and over the study that used SVM (Mulki, 2019) which gave an accuracy of 0.87. The experiments investigate the impact of dataset size, cross-domain generalization, robustness to adversarial attacks, model interpretability, and comparison with human annotation. Insights gained from these experiments help improve the models' accuracy, scalability, generalization capabilities, and robustness in real-world scenarios. The findings highlight the strengths and limitations of the proposed hate speech sentiment analysis technique, providing valuable insights for future research and development. Overall, this thesis contributes to the advancement of hate speech detection and supports efforts towards a safer and more inclusive online environment

تحليل مشاعر الكراهية في اللغة العربية باستخدام آلة المتجهات الداعمة (SVM) والشبكة العصبية الالتفافية

إعداد

ورود فراس حوامدة

إشراف

الدكتور غيث جرادات

الملـخص

تقدّم هذه الرسالة دراسة شاملة حول تحليل مشاعر خطاب الكراهية في سياق اللغة العربية. الهدف هو تطوير نماذج فعّالة تعتمد على التعلم الآلي والتعلم العميق باستخدام الخوارزمية الجينية لتصنيف مشاعر خطاب الكراهية بدقة والمساهمة في التخفيف من التأثير السلبي لخطاب الكراهية على الإنترنت. ويستكشف البحث تقنيات مختلفة لمعالجة البيانات واختيار السمات وتدريب النماذج بما في ذلك خوارزمية آلة المتّجه الداعم والاحتمالات الجينية البسيطة، وشبكات التعلم العميق الجينية. تم إجراء تقييم الأداء باستخدام مقاييس الدقة والاسترجاع و **F1-score** والدقة **Precision**.

عند مقارنة المصنفات السابقة تم ملاحظة تفوق خوارزمية آلة المتجه الداعم بدقة 0.93 و**F1-score** بقيمة 0.96 تليها شبكات التعلم العميق الجينية بدقة 0.91 و **F1-score** بقيمة 0.92 وأخيراً خوارزمية الاحتمالات الجينية البسيطة بدقة 0.87 و **F1-score** بقيمة 0.92. قدم البحث مقارنة بين نتائجه ونتائج الدراسات السابقة التي تمت على نفس مجموعة البيانات وقد لُوحظ تفوق خوارزمية **SVM** بدقة 0.93 على الدراسة التي استخدمت **AraBERT (Husain,2021)**

والتي أعطت دقة قدرها 0.85 وعلى الدراسة التي استخدمت(**Mulki,2019**) **SVM** والتي أعطت دقة قدرها 0.83 .وتستكشف التجارب تأثير حجم مجموعة البيانات، والعمومية بين المجالات والقوة ضد الهجمات العدائية وتفسير النموذج والمقارنة مع التفسير البشري. تساهم الرؤى المكتسبة من هذه التجارب في تحسين دقة النماذج وقابليتها للتوسع والتعميم في السيناريوهات الواقعية، تُسلط النتائج الضوء على نقاط القوة والقيود في تقنية تحليل مشاعر خطاب الكراهية المقترحة وتقدّم رؤى قيمة للبحوث والتطوير المستقبلي.

بشكل عام، تساهم هذه الرسالة في تقدم اكتشاف خطاب الكراهية وتدعم الجهود نحو بيئة أكثر أمانا وشمولية على الإنترنت.