

# **Cyberbullying Activity Detection over Social Media Using Deep Learning Techniques**

**Submitted by**

**Saif Ayman Najdawi**

**Supervisor**

**Dr. Ghaith Jaradat**

## **Abstract**

Social media's growth has brought about several problems, primarily due to a misunderstanding of the notion of freedom of expression. Cyberbullying is a noteworthy phenomenon that stands out because it has a serious global impact on both victims and society at large. It is a type of bullying that takes place on social media platforms. While numerous attempts in the literature have been made to address, prevent, or alleviate cyberbullying, their effectiveness is contingent on the victims' cooperation.

This study aims to address the growing concern of cyberbullying on social media platforms by developing an effective deep learning (DL) model to accurately identify instances of cyberbullying, enabling timely interventions and mitigating potential harm. A large dataset of labelled social media posts, distinguishing between cyberbullying and non-cyberbullying content, will be collected, and data pre-processing and

cleaning will be performed to ensure model Acc and effectiveness. Five different models (BiLSTM, CNN- BiLSTM, BiLSTM-GRU, and ANN) will be evaluated for a multi-class classification problem involving five target variable classes.

The analysis is expected to reveal that the BiLSTM model will achieve the highest scores in Acc in 91% of instances in the dataset. The BiLSTM-GRU model is also expected to perform well, with an Acc of 90%. In contrast, the CNN-BiLSTM and ANN models are anticipated to have lower performance scores. These results will demonstrate the effectiveness of the proposed model in accurately classifying instances into different classes related to cyberbullying. BiLSTM, CNN- BiLSTM, BiLSTM-GRU, and ANN models show the best performance. The developed model will potentially contribute to creating a safer online environment, protecting individuals from the harmful effects of cyberbullying, and informing the development of policies and regulations to combat this pervasive issue.

**Keywords: Cyberbullying, Social Media, Tweets, Machine Learning, Deep Learning, Classification.**

الكشف عن نشاط التمر الإلكتروني عبر وسائل التواصل الاجتماعي باستخدام

تقنيات التعلم العميق

إعداد

سيف أيمن النجداوي

إشراف

الدكتور غيث جرادات

الملخص

نمت وسائل التواصل الاجتماعي بشكل كبير، مما أدى إلى ظهور العديد من المشكلات، والتي تنشأ في المقام الأول من سوء فهم مفهوم حرية التعبير. التمر الإلكتروني هو ظاهرة تبرز بشكل لافت لأن لها تأثير عالمي خطير على الضحايا والمجتمع بشكل عام. إنه نوع من التمر يحدث على وسائل التواصل الاجتماعي. بالرغم من العديد من المحاولات في الأدبيات لمعالجة أو منع أو تخفيف التمر الإلكتروني، فإن فعاليتها تعتمد على تعاون الضحايا.

هدفت هذه الدراسة إلى معالجة القلق المتزايد بشأن التمر الإلكتروني على منصات وسائل التواصل الاجتماعي بتطوير نموذج تعلم عميق (DL) فعال لتحديد حالات التمر الإلكتروني بدقة، مما يتيح التدخلات في الوقت المناسب وتقليل الأضرار المحتملة. تم جمع مجموعة بيانات كبيرة من المنشورات المسماة على وسائل التواصل الاجتماعي، والتي تميزت بين المحتوى الذي يشمل التمر الإلكتروني والمحتوى الذي لا يشمل، وتمت معالجة البيانات وتنظيفها لضمان دقة النموذج

وفعاليتها. تم تقييم خمسة نماذج مختلفة (BiLSTM، CNN- BiLSTM، BiLSTM-GRU، و ANN) لمشكلة التصنيف متعدد الفئات التي تضمنت خمس فئات متغيرة هدف.

من المتوقع أن تكشف النتائج أن نموذج BiLSTM سيحقق أعلى الدرجات في الدقة، حيث يتوقع بشكل صحيح 91% من الحالات في مجموعة البيانات. كما من المتوقع أن يؤدي نموذج BiLSTM-GRU بشكل جيد، مع دقة تصل إلى 90%. بينما من المتوقع أن تكون درجات أداء نماذج CNN-BiLSTM و ANN أقل. من شأن هذه النتائج أن توضح فعالية النموذج المقترح في تصنيف الحالات بدقة إلى فئات مختلفة تتعلق بالانترنت الإلكتروني، حيث يظهر النماذج BiLSTM، CNN- BiLSTM، BiLSTM-GRU، و ANN أفضل أداء. سيكون للنموذج المطور القدرة على المساهمة في خلق بيئة إلكترونية أكثر أماناً، حماية الأفراد من التأثيرات الضارة للانترنت الإلكتروني، وإبلاغ تطوير السياسات واللوائح لمكافحة هذه المشكلة المتفشية.

الكلمة المفتاحية: التمر الإلكتروني، وسائل التواصل الاجتماعي، التغريدات، التعلم الآلي، التعلم العميق، التصنيف.