

## Use of Random Forest Classifier to Identify Counterfeited E-Commerce Listings

Gunawardhana H.M.K.T.<sup>1\*</sup>, Kumara B.T.G.S.<sup>1</sup>, Rathnayaka R.M.K.T.<sup>1</sup> and Jayaweera P.M.<sup>2</sup>

<sup>1</sup>*Department of Computing and Information Systems, Sabaragamuwa University of Sri Lanka*

<sup>2</sup>*University of Sri Jayawardanapura, Sri Lanka*

\*[kalintush@gmail.com](mailto:kalintush@gmail.com)

Online counterfeiting has become a significant threat to the e-commerce industry recently. It is becoming difficult to take countermeasures as the methods, tactics, and approaches of counterfeiting are evolving, and it is difficult to create a one-stop solution. According to the Organization for Economic Co-operation and Development (OECD), counterfeiting accounted for USD 464 billion or 2.5% of world trade in 2019. Counterfeiting generally contributes to factors such as child labour, illegal drug trafficking, and money laundering, which highlights this as a significant area for further study. This study uses 23000 e-commerce listings related to Paris Saint Germain (PSG) in thirty (30) e-commerce marketplaces such as Alibaba, Amazon, Redbubble, and Mercado Libre to train a text classifier based on title, description, seller name, and product URL. This study uses Random Forest Classifier and presents results with 95% accuracy. Also, this study focused on the provisions of an image classifier to make better decisions in anti-counterfeiting strategies in e-commerce.

**Keywords:** *Business intelligence, Counterfeiting, E-commerce, Machine learning*