

# MULTILINGUAL SMS SPAM DETECTION USING NATURAL LANGUAGE PROCESSING AND MACHINE LEARNING TECHNIQUES

SAMRAT KUNDU<sup>1\*</sup>, DEEPANJAN SEN<sup>2</sup>, SUBHASH MUKHERJEE<sup>3</sup>  
AND SUBHO BISWAS<sup>1</sup>

---

*Mobile users unduly suffer from spam SMS messages containing unsolicited advertisements, fraudulent schemes, lottery scams and phishing attempts that usually emanate from unknown numbers. To overcome this issue, in this project, an automatic spam detection system is developed by using Natural Language Processing and Machine Learning. Text messages are pre-processed by lowercasing, removal of stop words, tokenisation, and stemming. TF-IDF works to convert text to numbers for training the model. Several machine learning algorithms, such as Naive Bayes, Logistic Regression, Decision trees, Random forest, K-Nearest Neighbour, and decision tree algorithms, are evaluated. Among them, the Multinomial Naive Bayes model obtained the best performance with approximately 97- 98% accuracy and high precision, which provides a practical and effective solution to SMS spam filtering.*

---