**Title:**

**PDF Malware Detection: Toward Machine Learning Modeling With Explainability Analysis**

### **Abstract**

The Portable Document Format (PDF) is a widely used file type that has become a target for fraudsters who embed harmful code to compromise users' systems. Traditional detection techniques often fall short in effectively identifying PDF malware due to its versatile nature and reliance on a limited set of features. This work aims to enhance PDF malware detection through the development of a comprehensive dataset consisting of 15,958 PDF samples, encompassing benign, malicious, and evasive behaviors. We utilize three established PDF analysis tools—PDFiD, PDFINFO, and PDF-PARSER—to extract significant characteristics from these samples. Additionally, we derive various features proven effective in classifying PDF malware. An efficient and interpretable feature set is constructed through rigorous empirical analysis of the extracted and derived features. We evaluate several baseline machine learning classifiers, achieving a notable accuracy improvement of approximately 2% with the Random Forest classifier using the selected feature set. Furthermore, we enhance model explainability by generating a decision tree that provides rules for human interpretation, showcasing the effectiveness of Support Vector Classifier, K-Nearest Neighbors, Logistic Regression, and SVM with Optimal Hyperparameters in the context of PDF malware detection.

### **Keywords**

PDF malware detection, machine learning, dataset creation, feature extraction, Random Forest classifier, model explainability