Abstract

In the realm of cybersecurity, the detection of malware in executable files represents a pressing challenge. Conventional signature-based methods often struggle to keep pace with evolving threats, necessitating innovative solutions. This research investigates the application of advanced machine learning techniques, specifically Long Short-Term Memory (LSTM) and Bidirectional LSTM (BiLSTM) architectures, augmented by word embedding methodologies, for robust malware detection.

The research initiates with a systematic investigation of fundamental machine learning principles and rigorous data processing methodologies, forming a robust foundation for subsequent phases. Leveraging this acquired knowledge, the study embarks on the creation and refinement of a specialized deep learning model intricately designed for the accurate detection of concealed malware within executable files.

Every aspect of model construction receives meticulous attention, encompassing data collection, preprocessing, rigorous experimentation, and the finetuning of hyperparameters through hyperparameter optimization (HPO).

The HPO process systematically explores and refines various model configurations. The results of this optimization process unveil the most effective model configurations, with thorough analysis of performance metrics. The evaluation of the final model provides a comprehensive assessment of its capabilities in malware detection.

In summary, this research presents an adaptive and robust deep learning model for malware detection, strengthened by LSTM and BiLSTM architectures and enriched by word embedding techniques. It offers a comprehensive account of the research process, encompassing data collection, preprocessing, hyperparameter optimization, and model evaluation. This work contributes valuable insights to the dynamic field of cybersecurity and underscores the potential of machine learning in fortifying the security of digital systems.