Using Machine Learning to Optimize Key-Length Prediction for Polyalphabetically Encrypted Text

Pingali, Shriya (School: West High School)

With the continuous rise of machine learning technology, today's cybersecurity algorithms are faced with a new challenge; their vulnerabilities are ever-furthered by continually more advanced decryption programs that utilize machine learning. This project seeks to examine the efficiency of algorithms that employ machine learning on predicting the key length used to encrypt text via the Vigenere cipher, a method of text encryption which, due to its polyalphabetic nature, serves as the basis for more complex security algorithms used today. The Kasiki method is a generally unreliable means of determining key-length for the Vigenere cipher. It makes use of character combination frequency in language, and it is so far the only conceived method. This experiment examines the accuracy to which the Kasiki method and machine learning can individually and jointly predict key-length. The Kasiki method alone yielded an accuracy rate of 55%, but because it utilizes an implicit greatest-common-divisor algorithm, the Kasiki method often inaccurately predicted key-lengths outside of the valid range that were exact multiples of their respective true values. Using this information to train a linear regression machine learning model yielded a 75% accuracy rate in predicting key length. Full decryption of Vigenere-encrypted text is greatly facilitated upon extracting key length, and the significance of this accuracy rate exposes the vulnerability pertaining to the Vigenere cipher and consequently that of today's cybersecurity algorithms, but prediction accuracy could potentially be further optimized in a future investigation through the use of feature selection within a machine learning model.

Awards Won:

Drexel University: Full tuition scholarship $250,000

National Security Agency Research Directorate : Honorable Mention "Science of Security"