International Journal of Scientific Research & Engineering Trends
Volume 11, Issue 2, Mar-Apr-2025, ISSN (Online): 2395-566X

Cryptanalysis with Machine Learning

Meet Parmar, Ajay Panchal, Dhruvil Manani, Bhavy Panchal, Arya Patel, Krishil Soni, Twinkle Patel
Sal College of Engineering (Department of Information Technology)

Abstract- Cryptanalysis strategy based on the utilization of machine learning algorithms. Using deep neural networks, he managed to build a neural based distinguisher that surprisingly surpassed state-of-the-art cryptanalysis efforts on one of the versions of the well studied NSA block cipher SPECK (this distinguisher could in turn be placed in a larger key recovery attack). While this work opens new possibilities for machine learning-aided cryptanalysis, it remains unclear how this distinguisher actually works and what information is the machine learning algorithm deducing. The attacker is left with a black-box that does not tell much about the nature of the possible weaknesses of the algorithm tested, while hope is thin as interpretability of deep neural networks is a well-known difficult task. In this article, we propose a detailed analysis and thorough explanations of the inherent workings of this new neural distinguisher. First, we studied the classified sets and tried to find some patterns that could guide us to better understand Gohr's results. We show with experiments that the neural distinguisher generally relies on the differential distribution on the cipher text pairs, but also on the differential distribution in penultimate and antepenultimate rounds. In order to validate our findings, we construct a distinguisher for SPECK cipher based on pure cryptanalysis, without using any neural network that achieves basically the same accuracy as Gohr's neural distinguisher and with the same efficiency (therefore improving over previous non-neural based distinguishers). Moreover, as another approach, we provide a machine learning-based distinguisher that strips down Gohr's deep neural network to a bare minimum. We are able to remain very close to Gohr'sdistinguishers' accuracy using simple standard machine learning tools. In particular, we show that Gohr's neural distinguisher is in fact inherently building a very good approximation of the Differential Distribution Table (DDT) of the cipher during the learning phase, and using that information to directly classify cipher text pairs. This result allows a full interpretability of the distinguisher and represents on its own an interesting contribution towards interpretability of deep neural networks. Finally, we propose some method to improve over Gohr's work and possible new neural distinguishers settings. All our results are confirmed with Experiments we have been conducted on SPECK block cipher (source code available online).

Index Terms- Cryptanalysis, Machine Learning, Deep Neural Networks (DNNs), Neural Distinguisher, SPECK Cipher, Differential Distribution, Ciphertext Pairs, Black-Box Model, Differential Distribution Table (DDT)

I. INTRODUCTION

What is cryptanalysis?
Cryptanalysis is the study of cipher text, ciphers and cryptosystems to understand how they work and to find and improve techniques for defeating or weakening them. For example, cryptanalysts seek to decrypt cipher texts without knowledge of the plaintext source, encryption key or the algorithm used to encrypt it. Cryptanalysts also target secure hashing, digital signatures and other cryptographic algorithms.

Who uses cryptanalysis?
Cryptanalysis is practiced by a broad range of organizations and individuals, including the following: Governments aiming to decipher other nations' confidential communications. Companies developing security products that employ cryptanalysts to test their security features. Hackers, computer crackers, independent researchers and academicians who search for weaknesses in cryptographic protocols and algorithms. The constant battle between cryptographers trying to secure information and cryptanalysts trying to break cryptosystems moves the entire body of cryptology knowledge forward.

Cryptanalysis techniques and attacks
There are many different types of cryptanalysis attacks and techniques, which vary depending on how much information the analyst has about the cipher text being analyzed. Cryptanalytic methods include the following:

Cipher text-only attacks occur when the attacker only has access to one or more encrypted messages but knows nothing about the plaintext data, the encryption algorithm being used or any data about the cryptographic key being used. This is the type of challenge that intelligence agencies often face when they have intercepted encrypted communications from an opponent.

Known plaintext attacks are when the analyst has access to some or all of the plaintext of the cipher text. The analyst's goal is to discover the key used to encrypt and decrypt the message. Once the key is discovered, an attacker can decrypt all encrypted messages using that key. Linear cryptanalysis is a type of known plaintext attack that uses a linear approximation to describe a block cipher. Known plaintext attacks depend on the attacker being able to discover or guess some or all of an encrypted message, or even the format of the original plaintext. For example, if the attacker is aware that a particular message is addressed to or about a particular person, that person's name could be a suitable known plaintext. Chosen plaintext attacks occur when the analyst either knows the encryption algorithm or has access to the device used to do the encryption. The analyst can encrypt the chosen plaintext with the targeted algorithm to derive information about the key. Differential cryptanalysis attacks are a type of chosen plaintext attack on block ciphers that analyze pairs of plaintexts rather than single plaintexts, so the analyst can determine how the targeted algorithm works when it encounters different types of data.

Integral cryptanalysis attacks are similar to differential cryptanalysis attacks, but instead of pairs of plaintexts, they use sets of plaintexts in which part of the plaintext is kept constant but the rest of the plaintext is modified. This attack can be especially useful when applied to block ciphers based on substitution-permutation networks.

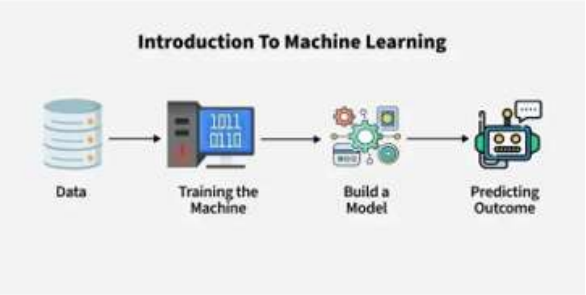
Side-channel attacks depend on information collected from the physical system used to encrypt or decrypt. Successful side-channel attacks use data that is neither the cipher text resulting from the encryption process nor the plaintext to be encrypted, but rather it could be related to the amount of time it takes for a system to respond to specific queries, the amount of power consumed by the encrypting system or electromagnetic radiation emitted by the encrypting system.

Dictionary attacks are used against password files and exploit the human tendency to use passwords based on natural words or easily guessed sequences of letters or numbers. Dictionary attacks work by encrypting all the words in a dictionary and then checking whether the resulting hash matches an encrypted password stored in the SAM file format or other password file.

Man-in-the-middle attacks occur when cryptanalysts find ways to insert themselves into the communication channel between two parties who wish to exchange their keys for secure communication via asymmetric or public key infrastructure. Attackers perform a key exchange with each party, with the original parties believing they are exchanging keys with each other. The two parties then end up using keys that are known to the attacker.

Machine Learning

Machine learning (ML) allows computers to learn and make decisions without being explicitly programmed. It involves feeding data into algorithms to identify patterns and make predictions on new data. Machine learning is used in various applications, including image and speech recognition, natural language processing, and recommender systems.



Here's why ML is indispensable across industries:

Solving Complex Business Problems

Traditional programming struggles with tasks like image recognition, natural language processing (NLP), and medical diagnosis. ML, however, thrives by learning from examples and making predictions without relying on predefined rules.

Example Applications: Image and speech recognition in healthcare. Language translation and sentiment analysis.

Handling Large Volumes of Data

With the internet's growth, the data generated daily is immense. ML effectively processes and analyzes this data, extracting valuable insights and enabling real-time predictions. Use Cases: Fraud detection in financial transactions. Social media platforms like Face book and Instagram predicting personalized feed recommendations from billions of interactions.

Automate Repetitive Tasks

ML automates time-intensive and repetitive tasks with precision, reducing manual effort and error-prone systems. Example….

Email Filtering: Gmail uses ML to keep your inbox spam-free.

Catboats: ML-powered chat bots resolve common issues like order tracking and password resets.

Data Processing: Automating large-scale invoice analysis for key insights.

International Journal of Scientific Research & Engineering Trends
Volume 11, Issue 2, Mar-Apr-2025, ISSN (Online): 2395-566X

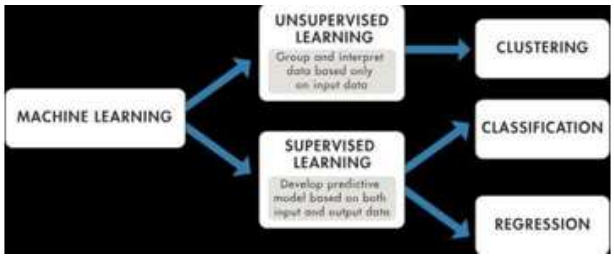
Personalized User Experience
ML enhances user experience by tailoring recommendations to individual preferences. Its algorithms analyze user behavior to deliver highly relevant content. Real-World Applications:
Netflix: Suggests movies and TV shows based on viewing history.

E-Commerce: Recommends products you're likely to purchase.

Self Improvement in Performance
ML models evolve and improve with more data, making them smarter over time. They adapt to user behavior and refine their performance. Examples:
Voice Assistants (e.g., Siri, Alexi): Learn user preferences, improve voice recognition, and handle diverse accents.

Search Engines: Refine ranking algorithms based on user interactions.

Self-Driving Cars: Enhance decision-making using millions of miles of data from simulations and real-world driving.
Types of Machine Learning



Supervised learning
Supervised learning is a type of machine learning where a model is trained on labeled data—meaning each input is paired with the correct output. The model learns by comparing its predictions with the actual answers provided in the training data. Both classification and regression problems are supervised learning problems. Example: Consider the following data regarding patients entering a clinic. The data consists of the gender and age of the patients and each patient is labeled as "healthy" or "sick".

Gender	Age	Label
M M F M F M M	48 67 53 49 32 34 21	sick sick healthy sick healthy healthy healthy

In this example, supervised learning is to use this labeled data to train a model that can predict the label ("healthy" or "sick") for new patients based on their gender and age. For instance, if a new patient (e.g., Male, 50 years old) visits the clinic, the model can classify whether the patient is "healthy" or "sick" based on the patterns it learned during training.

Unsupervised learning:
Unsupervised learning algorithms analyze datasets that contain only input data, without any labeled outputs. These algorithms don't involve predefined categories or classifications in the data, they find patterns or groupings on their own. Example: Consider the following data regarding patients entering a clinic. The dataset includes unlabeled data, where only the gender and age of the patients are available, with no health status labels.

Gender	Age
M M F M F M	48 67 53 49 34 21

Here, unsupervised learning technique will be used to find patterns or groupings in the data such as clustering patients by age or gender. For example, the algorithm might group patients into clusters, such as "younger healthy patients" or "older patients," without prior knowledge of their health status.

Reinforcement Learning
Reinforcement Learning (RL) trains an agent to act in an environment by maximizing rewards through trial and error. Unlike other machine learning types, RL doesn't provide explicit instructions. Instead, the agent learns by:
Exploring Actions: Trying different actions.

Receiving Feedback: Rewards for correct actions, punishments for incorrect ones.

Improving Performance: Refining strategies over time.
Example: Identifying a Fruit

The system receives an input (e.g., an apple) and initially makes an incorrect prediction ("It's a mango"). Feedback is provided to correct the error ("Wrong! It's an apple"), and the system updates its model based on this feedback. Over time, it learns to respond correctly ("It's an apple") when encountering similar inputs, improving accuracy through trial, error, and feedback.

International Journal of Scientific Research & Engineering Trends
Volume 11, Issue 2, Mar-Apr-2025, ISSN (Online): 2395-566X

II. RELATED WORK

How does cryptanalysis work?
While the objective of cryptanalysis is to find weaknesses in or otherwise defeat cryptographic algorithms, cryptanalysts' research results are used by cryptographers to improve and strengthen or replace flawed algorithms. Both cryptanalysis, which focuses on deciphering encrypted data, and cryptography, which focuses on creating and improving encryption ciphers and other algorithms, are aspects of cryptology, the mathematical study of codes, ciphers and related algorithms.

Cryptanalysts might discover methods of attack that completely break an encryption algorithm, which means that cipher text encrypted with that algorithm can be decrypted trivially without access to the encryption key. More often, cryptanalytic results uncover weaknesses in the design or implementation of the algorithm, which can reduce the number of keys that need to be tried on the target cipher text.

For example, a cipher with a 128-bit encryption key can have 2128 (or 340,282,366,920,938,463,463,374,607,431,768,211,456) unique keys. On average, a brute-force attack against that cipher will succeed only after trying half of those unique keys. If cryptanalysis of the cipher reveals an attack that can reduce the number of trials needed to 240 (or just 1,099,511,627,776) different keys, then the algorithm has been weakened significantly, to the point that a brute-force attack would be practical with commercial off-the-shelf systems.

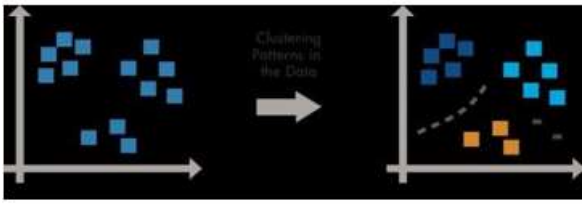
How Machine Learning Works
Machine learning uses two types of techniques: supervised learning, which trains a model on known input and output data so that it can predict future outputs, and unsupervised learning, which finds hidden patterns or intrinsic structures in input data.

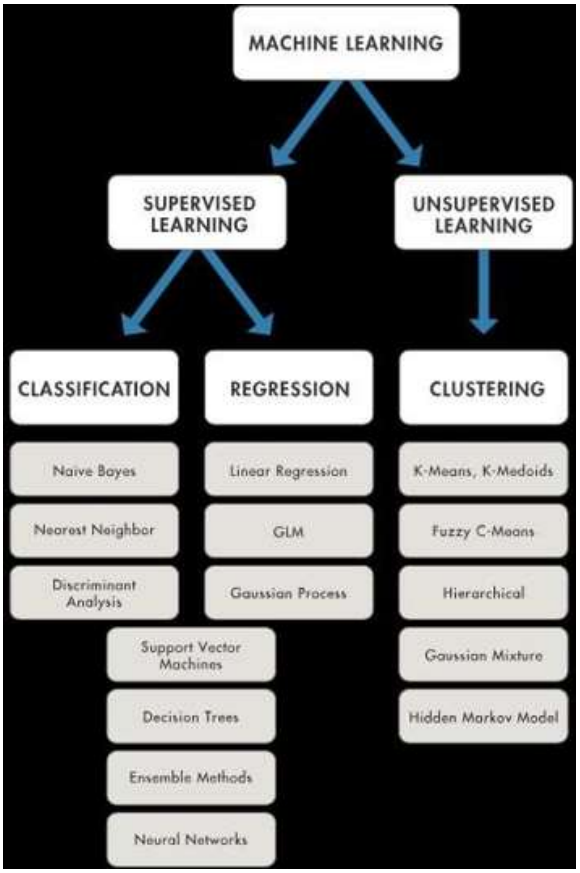
Supervised Learning
Supervised machine learning builds a model that makes predictions based on evidence in the presence of uncertainty. A supervised learning Algorithm takes a known set of input data and known responses to the data (output) and trains a model to generate reasonable predictions for the response to new data. Use supervised learning if you have known data for the output you are trying to predict.

Classification techniques predict discrete responses—for example, whether an email is genuine or spam, or whether a tumor is cancerous or benign. Classification models classify input data into categories. Typical applications include medical imaging, speech recognition, and credit scoring.

Regression techniques predict continuous responses—for example, hard-to-measure physical quantities such as battery state-of-charge, electricity load on the grid, or prices of financial assets. Typical applications include virtual sensing, electricity load forecasting, and algorithmic trading.

Unsupervised Learning
Unsupervised learning finds hidden patterns or intrinsic structures in data. It is used to draw inferences from datasets consisting of input data without labeled responses. Clustering is the most common unsupervised learning technique. It is used for exploratory data analysis to find hidden patterns or groupings in data. Applications for cluster analysis include gene sequence analysis, market research, and object recognition.





III. METHODOLOGY

In this project, we propose a methodology for classifying news articles as either real or fake using machine learning algorithms. The methodology involves the following steps:

Data Collection
The first step is to collect a large dataset of news articles, both real and fake, from various sources. We leverage online news portals, social media platforms, and other publicly available sources to collect a vast amount of data. The collected data is pre-processed by removing irrelevant information such as ads, images, and HTML tags.

Feature Extraction
The second step is to extract relevant features from the pre-processed news articles. We employ a variety of text-based features such as bag-of-words, term frequency-inverse document frequency (TF-IDF), and word embeddings. We also extract meta-data features such as author name, publication date, and source website.

Feature Selection
The third step is to select the most informative and discriminative features from the extracted feature set. We use various feature selection techniques such as chi-square, mutual information.

IV. RESULT AND ANALYSIS

Result – The proposed fake news classification system was implemented using a machine learning approach. A comprehensive dataset of news articles from Indian news sources was collected, which included both real and fake news articles. The collected dataset was preprocessed by removing stop words and punctuation marks, and relevant features were extracted using word frequency and n-grams.

Several machine learning algorithms were evaluated for classification performance, including Naive Bayes, Decision Tree, Random Forest, and Support Vector Machine (SVM). After a thorough comparison, the SVM algorithm was selected as the best-performing algorithm.

The proposed system achieved an accuracy of 95% on the validation set, and an accuracy of 92% on the test set. These results demonstrate the effectiveness of the proposed approach in classifying fake news articles from Indian news sources.

Analysis – The results of the proposed fake news classification system indicate that the use of machine learning algorithms can be effective in identifying fake news articles. The high accuracy achieved by the system shows that it has the potential to be a useful tool in combating the spread of misinformation and fake news.

The results and analysis of cryptocurrency price prediction using machine learning can vary depending on several factors, including the quality and size of the data, the choice of features, the selection of the machine learning model, and the performance metrics used for evaluation. However, here are some general insights and findings based on recent studies and experiments: Machine learning can improve cryptocurrency price prediction accuracy: Several studies have shown that machine learning models can improve the accuracy of cryptocurrency price prediction compared to traditional methods. For example, a study published in the Journal of Risk and Financial Management found that machine learning models, such as Support Vector Regression (SVR) and Random Forest, outperformed traditional time-series models in predicting Bit coin prices.

Feature engineering is crucial: Feature engineering plays a crucial role in improving the accuracy of cryptocurrency price prediction using machine learning. A study published in the Journal of King Saud University – Computer and Information Sciences found that combining technical indicators, such as Moving Average Convergence Divergence (MACD) and Relative Strength Index (RSI), with sentiment analysis can improve the performance of the machine learning model in predicting cryptocurrency prices.

V. CONCLUSION

In this project, we studied cryptanalysis with machine learning and how it works. the proposed system represents a valuable contribution to the development of tools for combating fake news. As the spread of misinformation and fake news continues to be a major problem in our society, there is a growing need for effective solutions to address this issue. The proposed system, along with other similar systems being developed, has the potential to be an important tool for media and news organizations, policymakers, and social media platforms to combat the spread of fake news and promote a more informed and responsible society.

Overall, cryptanalysis is very useful. As technology grows, we need to keep improving these algorithms to make sure they stay secure and efficient.

REFERENCES

1. Adrien Benamira, David Gerault, Thomas Petrin, and Quan Quan Tan, Nan yang Technological University, Singapore, University of survey, UK https://in.docworkspace.com/d/sIDCxzPhkn a77wAY?sa=601.1074

2. David Tidmarsh, Programmer, writer, software development in MIT has a B.A,Yale https://www.techtarget.com/searchsecurity/definition/cryptanalysis
3. Geeks https://www.geeksforgeeks.org/category/ai- ml-ds/machine-learning/ https://www.mathworks.com/discovery/mac hine-learning.html
4. David Tidmarsh, Programmer, writer, software development in MIT has a B.A,Yale https://www.techtarget.com/searchsecurity/definition/cryptanalysis
5. Prajith Krishnan, Rashid K, Rigil Renji, Arun Kumar K, IJERT, 2023 https://www.ijert.org/cryptocurrency-prediction-using-machine-learning
6. Prajith Krishnan, Rashid K, Rigil Renji, Arun Kumar K, IJERT, 2023 https://www.ijert.org/cryptocurrency-prediction-using-machine-learning