ENHANCING EMAIL SECURITY: A BIOLOGICALLY INSPIRED OPTIMIZED APPROACHED ALGORITHM FOR SPAM DETECTION IN MACHINE LEARNING:

1. Primary Goal:

The project aims to develop an optimized machine learning system for spam detection using bio-inspired algorithms. It combines Genetic Algorithm (GA) for feature selection and Harris Hawk Optimization (HHO) for parameter optimization, improving accuracy and reducing false positives.

2. Challenges and Solutions

Challenges:

1. High false positives/negatives in traditional spam detection.
2. Spam techniques evolve, making detection harder.

Solutions:

1. GA optimizes feature selection for better spam detection.
2. HHO enhances classification by optimizing model parameters.
3. Combining GA and HHO creates a more adaptable and accurate spam detection system.

3.Methodology

Data Preprocessing: Tokenization and feature extraction convert emails into numerical vectors.

Bio-Inspired Algorithms: GA selects the best features and **HHO** optimizes model parameters.

Model Training: Trained on labeled spam/non-spam emails and evaluated using metrics like accuracy, precision, recall, and F1-score.

4. Technology Stack:

Frontend: HTML, CSS, JavaScript.

Backend: Python (Django/Flask), with libraries like NumPy and Pandas.

Database: Manages the storage of training and testing email datasets.

CSV files used for training/testing datasets.

5. Algorithms Used:

Existing: Naive Bayes, Rule-Based Filters, Bayesian Filters.

Proposed: GA for feature selection, HHO for optimization, Multinomial Naive Bayes for spam classification.

6.what datasets used:1.

1. SpamAssassin Public Corpus: A widely used dataset for spam detection.

2. Enron Email Dataset: Another large dataset containing both spam and non-spam emails.

3. Ling-Spam Dataset: Often used in email spam classification research.