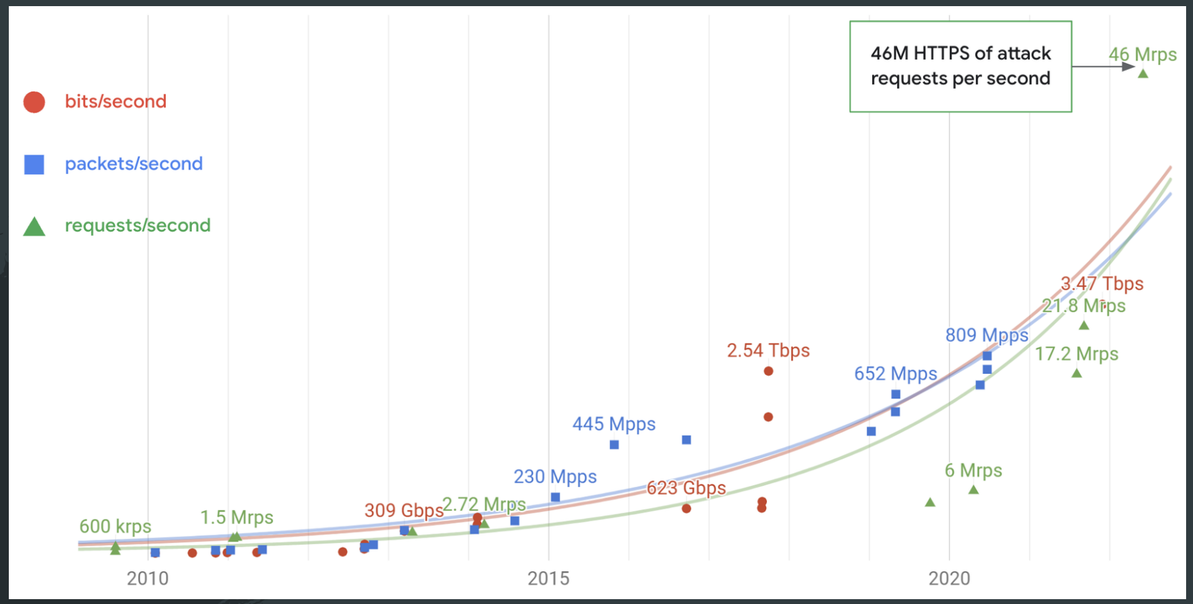
以特徵選擇與集成學習為核心之入侵偵測互動分析系統

CONTENTS

SECTION 1

Motivation

1-1 研究動機



https://blog.cloudflare.com/zh-tw/bigger-and-badder-how-ddos-attack-sizes-have-evolved-over-the-last-decade/

2010-2022 年已知最大的DDoS 攻擊。

隨著網際網路技術的迅速發展與連網裝置的普及，各類網路攻擊事件的頻率與複雜度不斷攀升。
傳統入侵偵測系統（IDS）多依賴靜態規則比對，對於新型態或零日攻擊的辨識效果有限。
因此，如何結合資料分析與機器學習技術，建構能「自我學習、即時反應」的智慧型入侵偵測框架，成為資安防護領域的重要課題。

1-2研究挑戰

然而，目前的入侵偵測資料集普遍具有以下問題：

- 特徵維度龐大、冗餘度高，造成運算負擔與過度擬合；
- 類別分佈不均，導致模型難以辨識少數攻擊樣本(正常流量遠大於異常流量)；
- 缺乏即時視覺化工具，使分析人員難以即時理解攻擊行為。

這些挑戰限制了機器學習在實際入侵偵測的應用效能。

1-3 研究目標

- 提升入侵偵測準確率與穩定性
- 降低特徵維度與訓練時間
- 建立可視化互動式分析介面
- 架構可持續更新的智慧IDS雛型

1-4 研究背景與效益

- 攻擊多樣化 → 需自動化偵測
- ML潛力大但受限於資料品質
- 可解釋性 = 實務部署關鍵

SECTION 2

Related Works

2-1文獻回顧

規則式 IDS

Koral Ilgun, Richard A. Kemmerer, and Phillip A. Porras, "State Transition Analysis: A Rule-Based Intrusion Detection Approach," IEEE Transactions on Software Engineering, Vol. 21, No. 3, March 1995

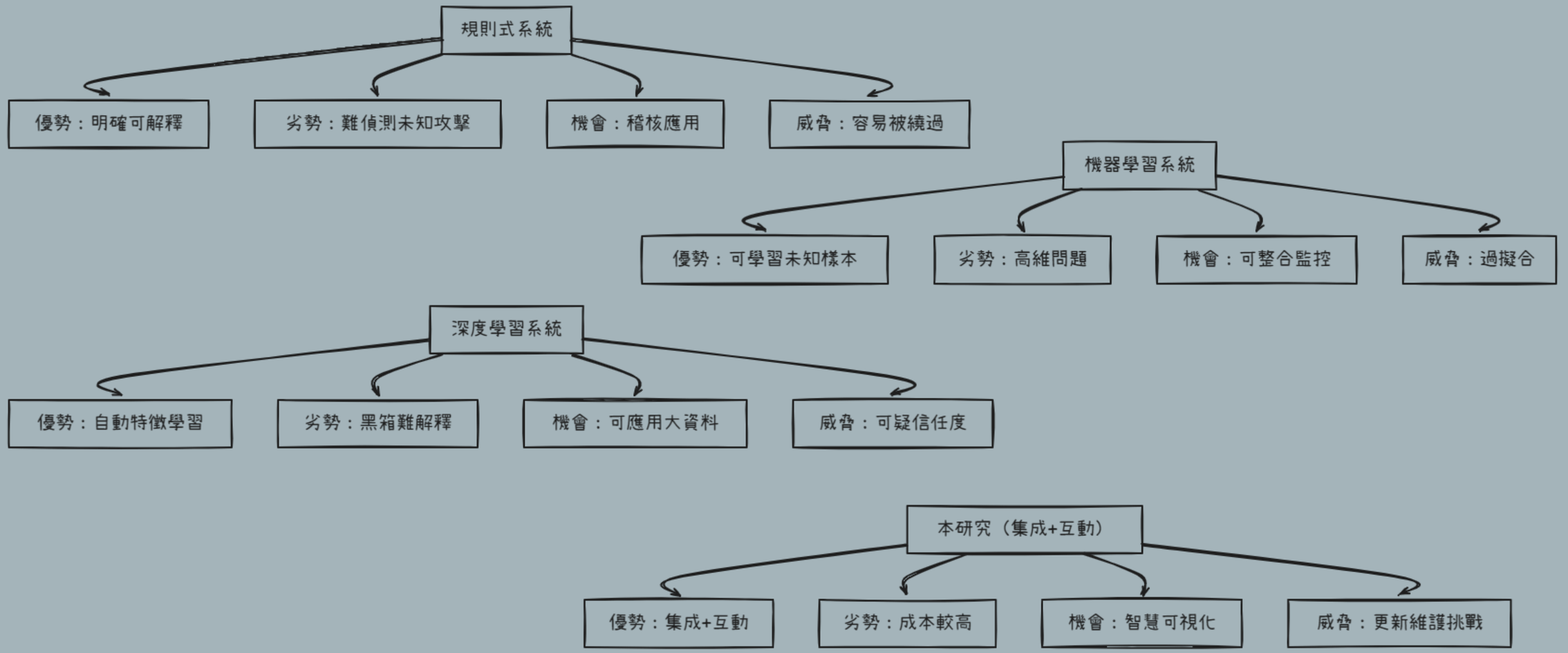
傳統ML（SVM, RF）

深度學習（CNN, LSTM）

集成學習 + 可視化平台

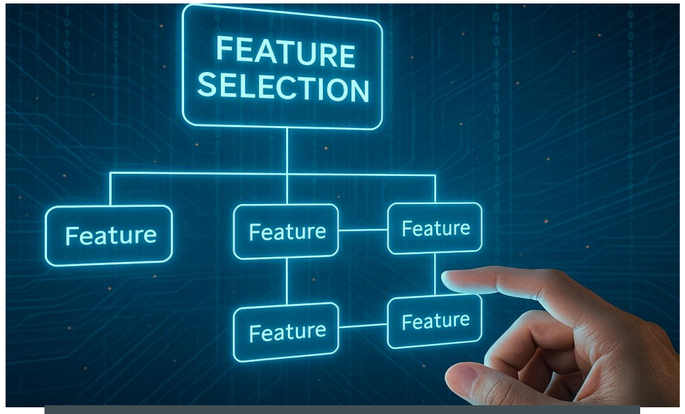
https://ndltd.ncl.edu.tw/cgi-bin/gs32/gsweb.cgi?randomimg=fc7cCL_1761616033&validpath=%2Ftmp%2F%5Enclcdr__doschk%2Ffc7cCL_1761616033__MTU0NTc2&validinput=154576&check=%E7%A2%BA%E5%AE%9A

2-2文獻回顧比較


規則式系統
優勢：明確可解釋
劣勢：難偵測未知攻擊
機會：稽核應用
威脅：容易被繞過
機器學習系統
優勢：可學習未知樣本
劣勢：高維問題
機會：可整合監控
威脅：過擬合
深度學習系統
優勢：自動特徵學習
劣勢：黑箱難解釋
機會：可應用大資料
威脅：可疑信任度
本研究（集成+互動）
優勢：集成+互動
劣勢：成本較高
機會：智慧可視化
威脅：更新維護挑戰

2-3問題與解法

Q：高維資料、類別不平衡、低解釋度



GA進行特徵選擇

RF / XGBoost 進行集成學習

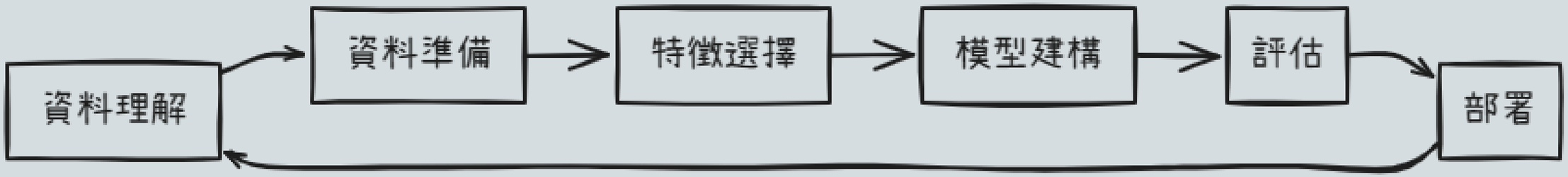
Streamlit

Streamlit 建立互動可視化平台

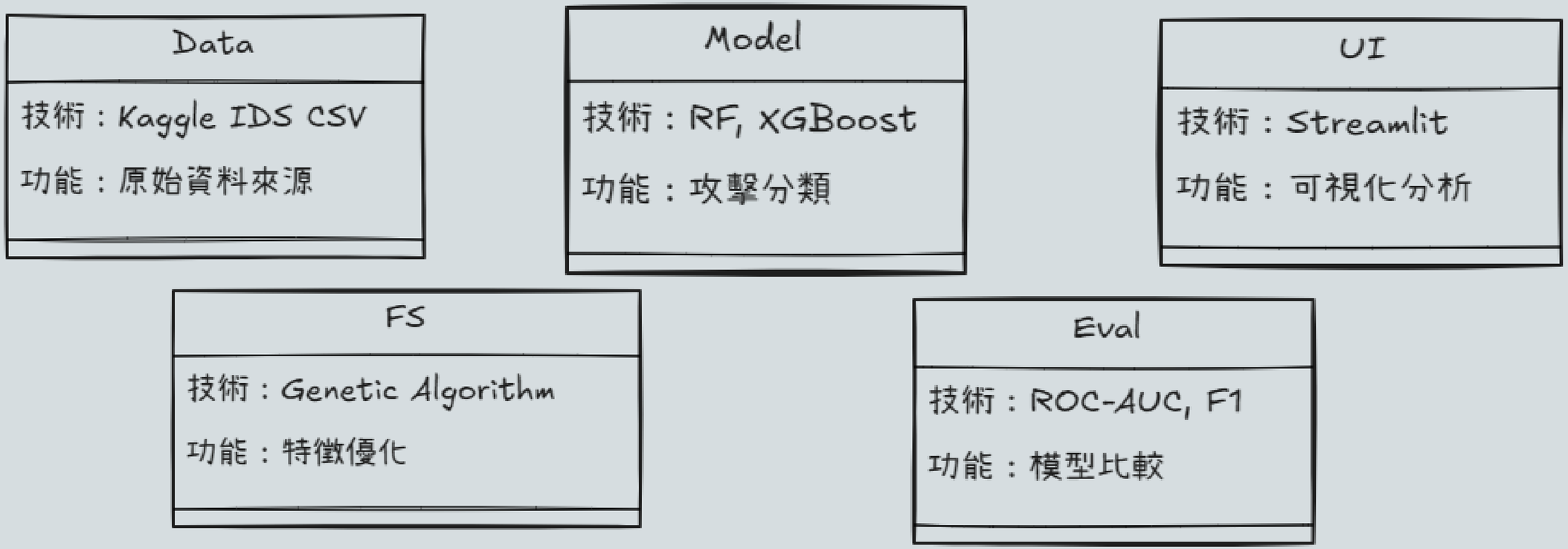
SECTION 3

Proposed Method

3-1研究設計

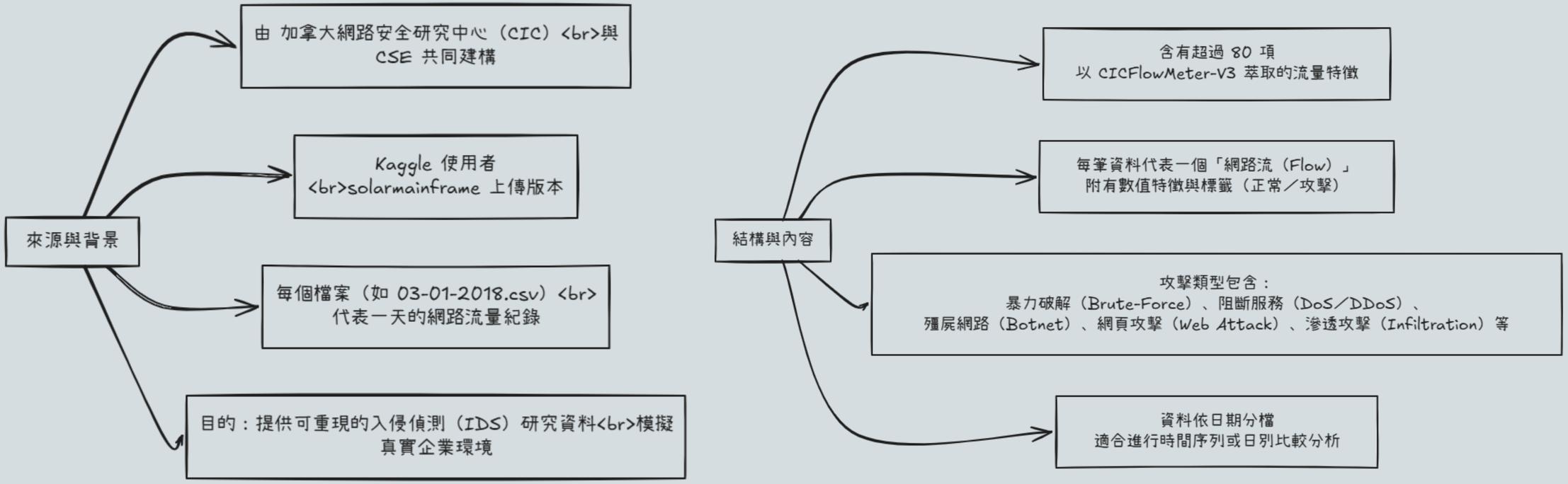


3-2研究方法與元素


Data
技術：Kaggle IDS CSV
功能：原始資料來源
Model
技術：RF, XGBoost
功能：攻擊分類
UI
技術：Streamlit
功能：可視化分析
FS
技術：Genetic Algorithm
功能：特徵優化
Eval
技術：ROC-AUC, F1
功能：模型比較

Data
技術：Kaggle IDS CSV
功能：原始資料來源

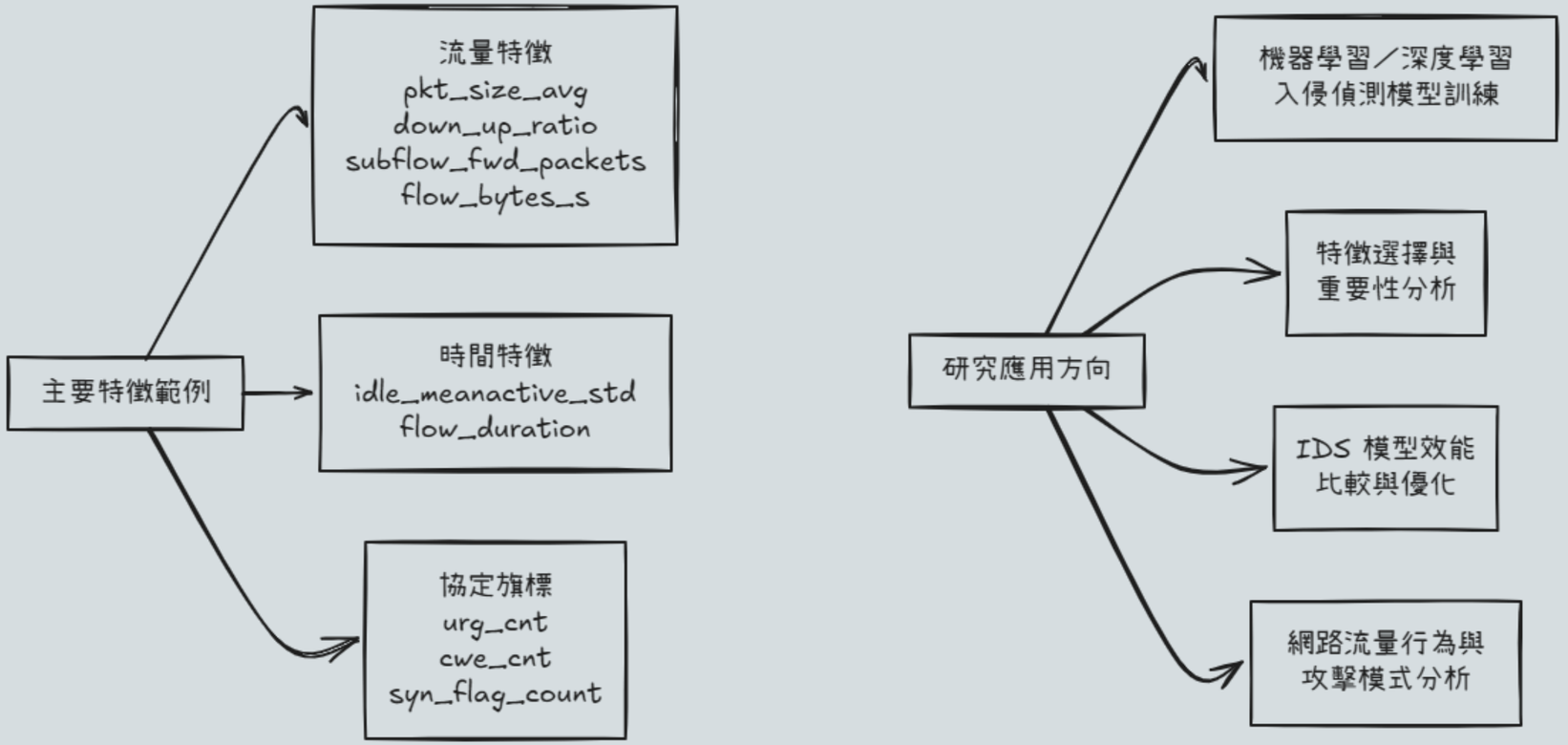
IDS 2018 Intrusion CSVs (CSE-CIC-IDS2018) - 03-01-2018.csv



https://www.kaggle.com/datasets/solarmainframe/ids-intrusion-csv?select=03-01-2018.csv

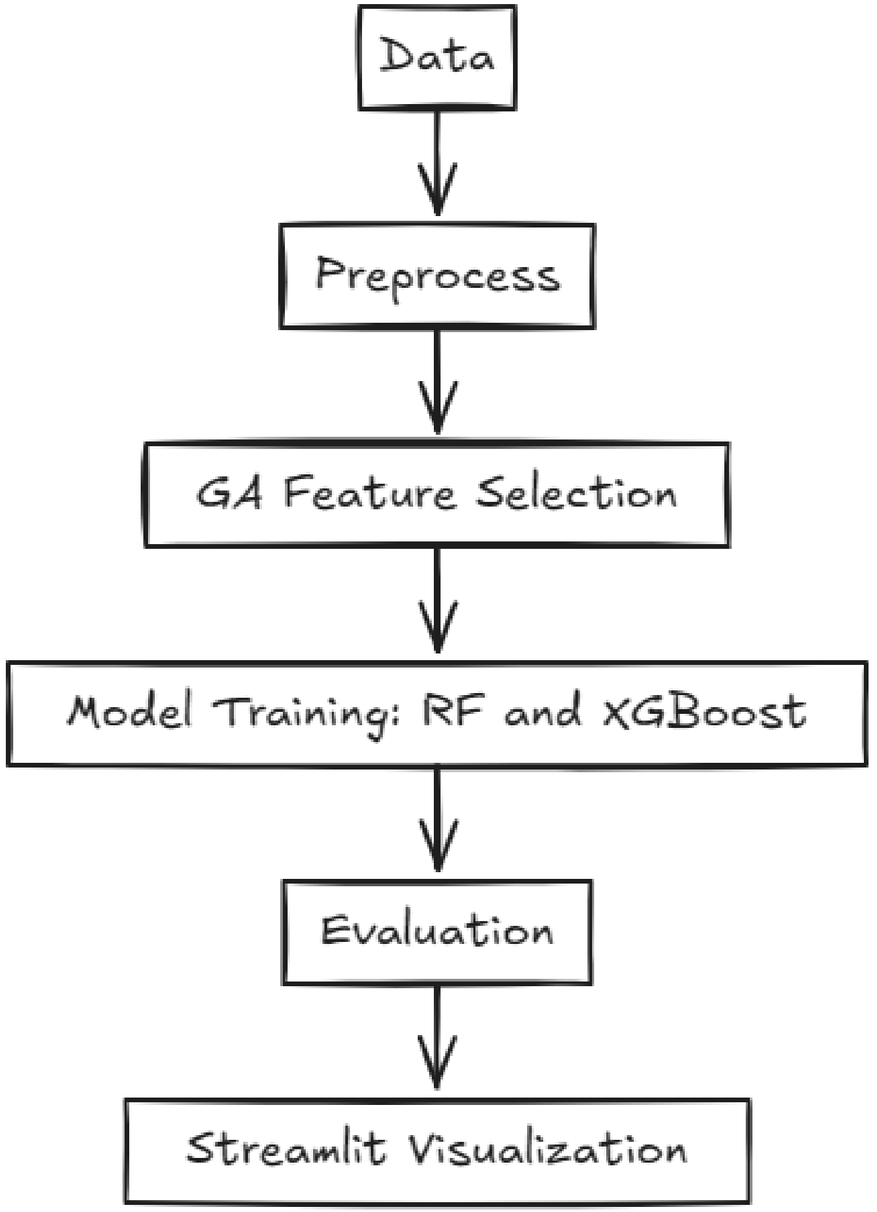
Data
技術：Kaggle IDS CSV
功能：原始資料來源

IDS 2018 Intrusion CSVs (CSE-CIC-IDS2018) - 03-01-2018.csv



https://www.kaggle.com/datasets/solarmainframe/ids-intrusion-csv?select=03-01-2018.csv

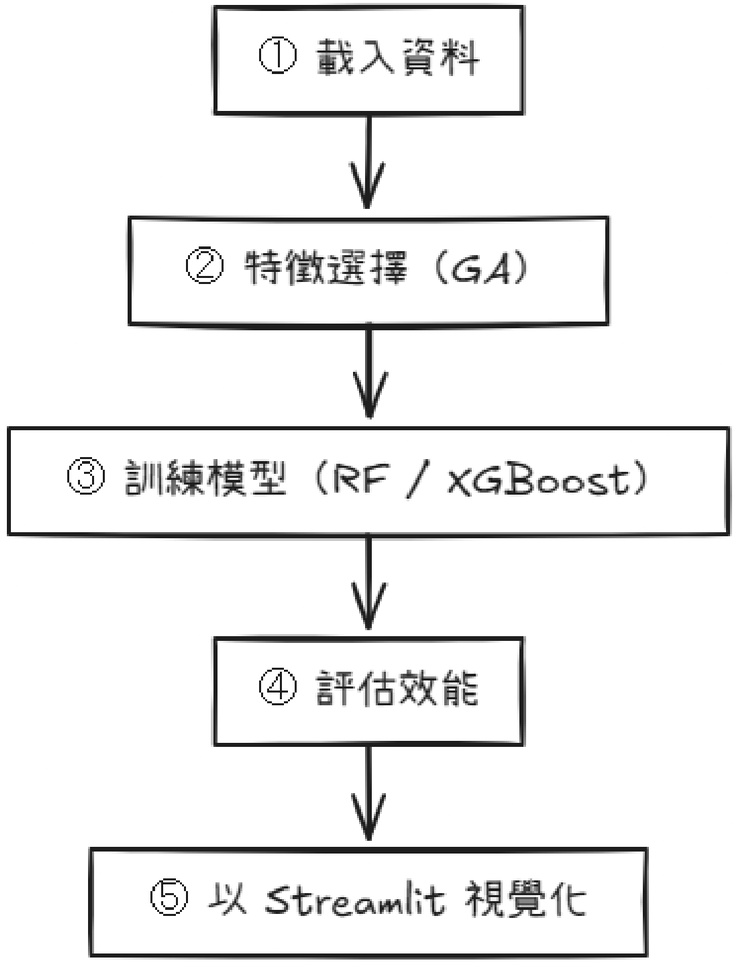
3-3 系統架構


Data
Preprocess
GA Feature Selection
Model Training: RF and XGBoost
Evaluation
Streamlit Visualization

3-4 實現方式

- 自動特徵選取避免偏差
- 集成學習提升穩定度
- 視覺化強化可理解性
- 即時互動介面縮短決策時間

3-5實作與流程



THANK YOU!