**Система для аналізу цифрових доказів із використанням машинного навчання (**CyberInsight) – проникнення в суть.

**Ідея проєкту**

Розробка інструменту, який допомагає аналізувати цифрові докази у кримінальних справах. Система може:

1. Аналізувати тексти (наприклад, підозрілі електронні листи чи повідомлення) на наявність ознак шахрайства, погроз чи інших злочинів.

2. Виявляти патерни кіберзлочинної поведінки (наприклад, мережу шахрайських акаунтів).

3. Автоматизувати перевірку метаданих файлів (час створення, зміни, геолокація) для виявлення підробок.

**Ключові компоненти**

1. **Аналіз тексту**

Використовувати Python-бібліотеки (наприклад, NLTK, SpaCy) для обробки природної мови.

Розробити модель, що виявляє ознаки злочинної діяльності в текстах.

Використати попередньо навчені моделі, наприклад, GPT чи BERT, для аналізу контексту.

2. **Аналіз метаданих**

Створити скрипт для витягування метаданих з документів, зображень, відео тощо (бібліотеки: pymediainfo, exiftool).

Інтегрувати механізм перевірки автентичності файлів.

3. **Побудова мережі зв’язків**

Аналіз графів для виявлення зв’язків між підозрілими об’єктами (наприклад, акаунтами чи транзакціями).

Використовувати бібліотеки networkx або igraph.

4. **Інтерфейс**

Розробити простий веб-додаток для демонстрації роботи інструменту (можна використати Flask/Django).

**Приклади використання**

1. Допомога юристам і криміналістам у розслідуваннях кіберзлочинів.

2. Автоматизація перевірки цифрових доказів у судових процесах.

3. Підтримка правоохоронних органів у профілюванні кіберзлочинців.

**Як розпочати?**

1. **Вивчити кейси**: Ознайомлюємось із реальними випадками кіберзлочинів та методами роботи криміналістів.

2. **Налаштувати середовище**: Visual Studio Code або інше середовище.

3. Розробити базовий прототип аналізатора текстів та метаданих.

4. **Додати функціонал поступово**. Залучити алгоритми машинного навчання чи аналізу графів для ускладнення задачі.

1. **Аналіз текстів**

Розробити модель для виявлення злочинної діяльності.

Додати базову класифікацію текстів за допомогою бібліотек NLTK, SpaCy чи моделей GPT/BERT.

1. **Аналіз метаданих**

Написати скрипт для витягування метаданих із файлів (зображень, документів тощо).

Використовувати бібліотеки pymediainfo або exiftool.

1. **Аналіз графів**

Почати будувати базові зв’язки між даними, наприклад, акаунтами чи транзакціями, використовуючи networkx.

1. **Інтерфейс**

Додати базовий веб-інтерфейс на Flask/Django для демонстрації аналізу.

Розпочнемо з аналізу текстів. Перша мета – створити базову систему обробки природної мови (NLP), яка буде:

1. **Очищувати текст** від зайвих символів, пунктуації, стоп-слів тощо.
2. **Виявляти ознаки злочинної діяльності** за ключовими словами або контекстом.
3. Інтегрувати попередньо навчену модель для більш складного аналізу (наприклад, використання BERT).

Покроковий план (аналіз текстів).

1. **Попередня обробка текстів**

Токенізація.

Лематизація або стемінг.

Видалення стоп-слів.

1. **Ключові слова**

Створити список слів, пов’язаних із шахрайством, погрозами тощо.

1. **Класифікація тексту**

Використовувати модель машинного навчання (наприклад, Logistic Regression) або бібліотеки трансформерів (Hugging Face).

Почнемо зі створення базового коду для попередньої обробки тексту та визначення ключових слів.