Project Report

1.Project Title :

A Lightweight Metadata Extractor for Ethical Penetration Testing

---

2. Intern Name:

Om Ganechari

---

3. Internship Role:

Penetration Testing Intern

---

4. Project Domain:

Cybersecurity, Digital Forensics, Ethical Hacking

---

5. Objective of the Project:

The objective of **MetaSniff** is to develop a lightweight, Python-based tool capable of extracting hidden metadata from commonly used file formats such as images and documents. This project aims to enhance awareness about **information leakage** via metadata and equip cybersecurity professionals with a tool for **forensic audits** and **compliance validation**.

---

6. Problem Statement:

Many users unknowingly share files (images, PDFs, Word documents) that contain sensitive metadata such as GPS locations, author names, timestamps, and device details. This can pose privacy risks or aid in malicious activities. There is a need for an open-source, beginner-friendly metadata extraction tool usable in academic, forensic, or ethical hacking contexts.

---

7. Features and Functionalities:

- Extracts **EXIF data** from image files (.jpg, .jpeg, .png)
- Extracts **core metadata** from PDF files (title, author, creation date)
- Extracts **document properties** from Word files (.docx)
- Simple CLI/Notebook interface, ideal for labs and education
- Lightweight & open-source (fully Python-based)

---

8. Tools & Technologies Used:

- **Programming Language:** Python
- **Libraries Used:**
  - `Pillow` (for image handling)
  - `PyPDF2` (for PDF metadata)
  - `python-docx` (for Word document metadata)
- **Platform:** Google Colab (for execution and testing)
- **Version Control:** Git & GitHub
- **Documentation Tools:** Markdown (README), MIT License

---

9. Use Cases:

Digital Forensics – Detect hidden metadata in legal cases
Penetration Testing – Check for data leakage in shared files
Compliance Audits – Ensure sensitive metadata is scrubbed before upload
Education & Training – Teach forensic analysis and ethical hacking basics

---

10. Ethical Considerations:

This tool is strictly intended for **educational** and **ethical research** purposes. Unauthorized use to extract metadata from someone else's files without consent is prohibited and may violate privacy laws.

---

11. Results & Output Screenshots:

(Attach screenshots from Google Colab or terminal output showing successful extraction of metadata from an image, PDF, and DOCX.)

---

12. GitHub Repository:

https://github.com/om-ganechari/metada-extractor-.git

---

13. Project Demo Video:

https://drive.google.com/file/d/1BiYhHKzXBe8P4lt2dUAMIvwcaCa-oO7-/view?usp=drive_link

---

14. Deployment :

Deployed on streamlit
https://34kyvyc8gecrxzsngdry6u.streamlit.app/

---

15. Future Enhancements:

Add metadata redaction/sanitization features
Web UI for drag-n-drop files
Integration with threat intelligence platforms
Support for more file types like `.pptx`, `.xlsx`, audio/video metadata

---

16. Conclusion:

The MetaSniff project provided real-world experience in Python development, ethical hacking concepts, and forensic analysis. This tool enhances file investigation and metadata visibility, empowering cybersecurity professionals and learners with a compact, open-source solution.