PROJECT REPORT ON

META DATA EXTRACTOR AND CORRELATION TOOL

Submitted by

ANKOLU CHANDRA SEKHAR and
Supraja Technologies employee id: (ST#IS#8422)

Under the Supervision of

UPENDRA

Senior Security Analyst

Krishna

Security Analyst

Registered And Head Office

D.NO: 11-9-18, 1st Floor,

Majjivari Street, Kothapeta,

Vijayawada - 520001.

+91 9550055338 / +91 7901336873

contact@suprajatechnologies.com

Company Identification Number (CIN Number) : U72900AP2019PTC111909
PAN Number : AAICC1770E, TAN Number : HYDC10078F
ISO Certificate Number : 031122019107
MSME Registration Certificate No : UDYAM-AP-06-0035821
Supraja Technologies Trademark Application ID No for Class 41 : 4146139
Supraja Technologies Trademark Application ID No for Class 42 : 4146144

COMPANY INTRODUCTION:

Supraja Technologies is a leading Knowledge and Technical Solutions Provider and pioneer leader in IT industry, is operating based out of Vijayawada.

R&D at Supraja

With a 24X7 work in Research & Development, experts at Supraja Technologies work under our :

- **Supraja Technologies Cyber Security Cell**

About Supraja Technologies:

Supraja Technologies (a unit of CHSMRLSS Technologies Pvt. Ltd.) with its foundation pillars as Innovation, Information and Intelligence is exploring indefinitely as a **Technology Service Provider (Corporate Consulting)** and as a **Training Organization (Ed-Tech)** as well.

You may visit us at :
www.suprajatechnologies.com

The multi domains of trainings which Supraja Technologies operate include the following :

- **Workshops & Hackathons**
  - o Engineering Colleges
  - o Corporate Companies (Startups & MNC's)
  - o Government Organizations

Company Identification Number (CIN Number) : U72900AP2019PTC111909
PAN Number : AAICC1770E, TAN Number : HYDC10078F
ISO Certificate Number : 031122019107
MSME Registration Certificate No : UDYAM-AP-06-0035821
Supraja Technologies Trademark Application ID No for Class 41 : 4146139
Supraja Technologies Trademark Application ID No for Class 42 : 4146144

- **Classroom Trainings Cum Certification Courses**
  - o Summer Training (30-45 Days)
  - o Winter Training (10 - 15 Days)
  - o Weekend Training (2 Days)
  - o 1 Month / 3 Months / 6 Months Courses

- **On-site Trainings**
  - o Value Added Courses / Two Credit Courses for Colleges
  - o Faculty Development Programs
  - o College Summer Training (15 Days, 30 Days, 45 Days & 60 Days)
  - o Govt Agencies, Police Academies, Corporates etc

- **Cloud Campus**
  - o (Distance Learning Program) *Coming Soon

- **Internships**
  - o Internship for Engineering Students (30 Days, 45 Days & 60 Days)
  - o Internship for Graduates (6 Months)

- **Lab Setup**
  - o Cyber Lab

- **CoE**
  - o Cyber Security Centre of Excellence

- **Supraja Technologies Security Assessment Product**
  - o (Our SaS Product is currently under development) *Coming Soon

SUPRAJA TECHNOLOGIES
(a unit of CHSMRLSS Technologies Private Limited)
An MSME & ISO 9001:2015 Certified Company
Regd. & Head Office : Door No. 11-9-18, 1st Floor, Majjivari Street, Kothapeta, Vijayawada - 520001.
Branch Office : Prabhu Villas, Plot No. 124, 1st Floor, Church Road, Ayyappa Nagar, Vijayawada - 520007.
contact@suprajatechnologies.com | www.suprajatechnologies.com | +91 - 9550055338

Why Supraja Technologies:

Be it Training or a workshop, the course content is always from R&D Cell of Supraja.

- A proven track record of delivering quality services.
- **68,500+** Students trained by our trainers till date.
- Training Partners of recognized institutions.
- Trainers with excellent research and teaching pedagogy illustrate their findings through corporate standard **practical demonstrations** during their sessions.
- Easy to learn and **hands-on sessions** are given, with additional benefits of Study Material, Tool kit and immediate query handling.
- Self-Prepared **Cyber Security Cell.**
- Supraja Technologies has the best, experienced and highly **skilled bunch of R&D Engineers, Security Analysts, Security Consultants & Trainers.**
- We provide training in Innovating and Trending Technologies to Govt. Officials, Corporate Houses and Colleges.

✓ Something we are proud of :

1. Supraja Technologies CEO Mr.Santosh Chaluvadi is an Alumni of Potti Sriramulu Chalavadi Mallikharjuna Rao College of Engineering and Technology, Vijayawada.

   With our CEO this college conducted/organised a 50 hours Nonstop Marathon Training Workshop on Ethical Hacking & Cyber Security for which this respective college and our CEO both holds their name in **“LIMCA BOOK OF RECORDS 2017”**

2. We are very happy to inform you all that our company, Supraja Technologies has been shortlisted for **"Top 50 Tech Companies" award 2019**, conferred at InterCon - Dubai, UAE.

   Supraja Technologies is one out of thousands companies that were initially screened by InterCon team of 45+ research analysts over a period of three months and the final shortlist includes 150+ firms and we are very proud to inform you all that our company

Supraja Technologies also happens to be a part of the same.

✓ Life changing solution/service :

After working on R&D for around 2 years, finally in the mid 2019 we have successfully developed a service/solution of various techniques and strategies for the Film Industry through which he can kill piracy of any film in online up to 35% right now. This betaservice is being appreciated & adopted by various Tollywood Film Industry Producers & Hero's to safeguard their film from piracy in online and to gain more profits.

By the end of 2033 our vision is to rollout a complete full packed service/solution where we can kill piracy entirely 100% everywhere in online for sure.

Appreciation :

Received a great appreciation from our 1st Tollywood Film Industry client Mr.Saptagiri for providing our Anti-Piracy betaservice for his film VAJRA KAVACHADARA GOVINDA

✓ Achievements:

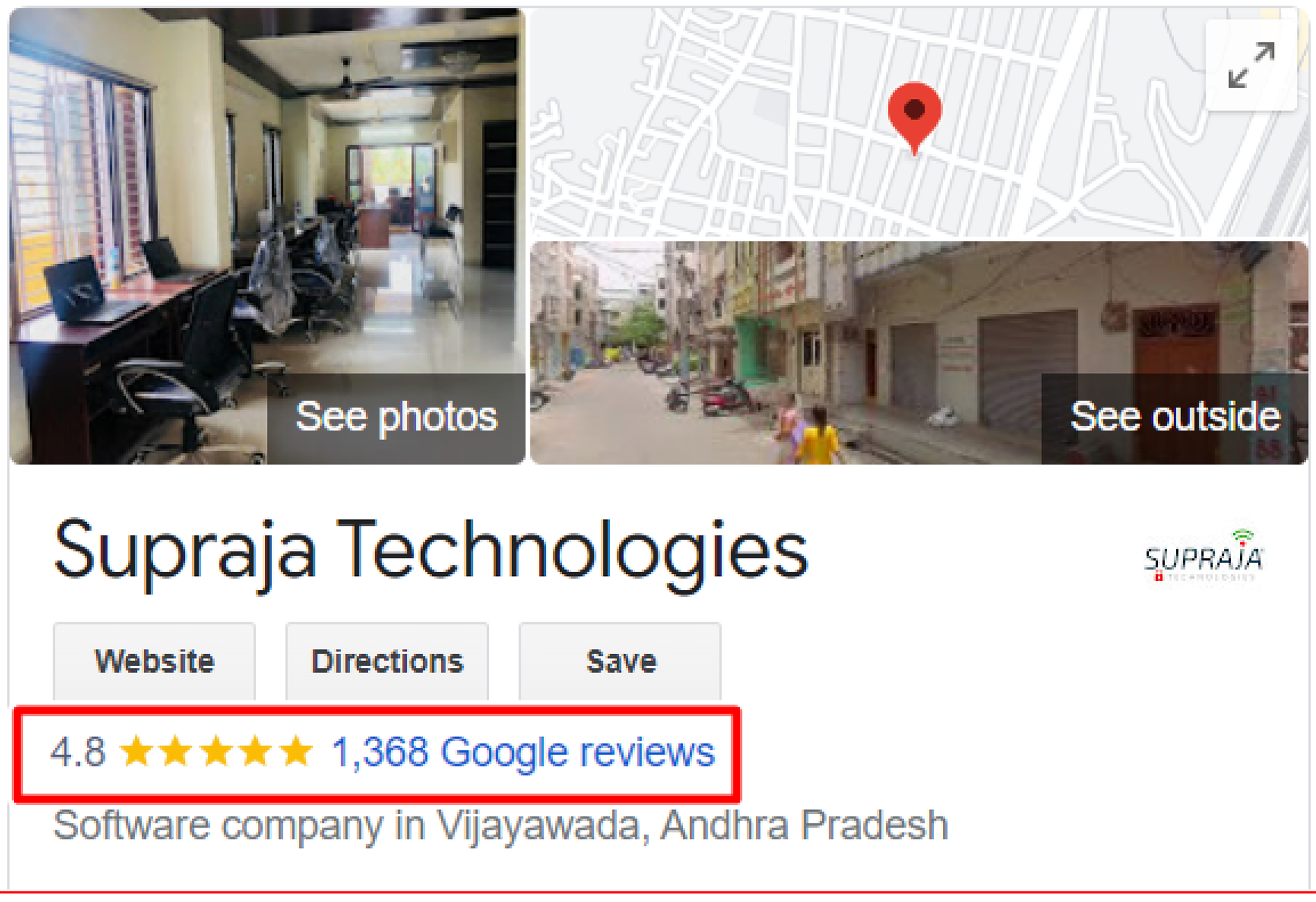
- On 17th August 2024, We (Supraja Technologies) launched our company's **Centre of Excellence in Cyber Security** (CoE) at Ramco Institute of Technology, Rajapalayam

- On 18th September 2024, We (Supraja Technologies) launched our company's **Centre of Excellence in Cyber Security** (CoE) at SRM University (Ramapuram Campus), Chennai

- On 20th November 2024, We (Supraja Technologies) launched our company's **Centre of Excellence in Cyber Security** (CoE) at St.Joseph's Institute of Technology, Chennai

Company Identification Number (CIN Number) : U72900AP2019PTC111909
PAN Number : AAICC1770E, TAN Number : HYDC10078F
ISO Certificate Number : 031122019107
MSME Registration Certificate No : UDYAM-AP-06-0035821
Supraja Technologies Trademark Application ID No for Class 41 : 4146139
Supraja Technologies Trademark Application ID No for Class 42 : 4146144

Our Company Supraja Technologies is one of the emerging startup in Andhra Pradesh with 4.8 Google Ratings



Link : https://bit.ly/SuprajaGoogle

AND

also check our Company CEO Instagram Profile for our recent more success stories:

SUPRAJA TECHNOLOGIES
(a unit of CHSMRLSS Technologies Private Limited)
An MSME & ISO 9001:2015 Certified Company
Regd. & Head Office : Door No. 11-9-18, 1st Floor, Majjivari Street, Kothapeta, Vijayawada - 520001.
Branch Office : Prabhu Villas, Plot No. 124, 1st Floor, Church Road, Ayyappa Nagar, Vijayawada - 520007.
contact@suprajatechnologies.com | www.suprajatechnologies.com | +91 - 9550055338

https://www.instagram.com/chaluvadisantosh/

Santosh Chaluvadi

Founder & CEO
Supraja Technologies

He is a 28-year-old entrepreneur, one of the India's efficient Cyber Security Analyst and also he is an expert Digital Marketer as well. He is a digital marketer by profession and security enthusiast by passion. He primarily focuses on content building, testing and monetization of blogs. He has successfully developed many websites and done the security testing himself to ensure that the user's data is in safe hands and their privacy is protected. He is very active on social media and shares lot of tech stuff with his followers. The young student hacker has solved many issues with the vulnerabilities present in various websites and databases, given a solution in clearing the loopholes in order to protect the data to be leaked from the databases. Besides Ethical Hacking & Cyber Security, he also has a passion in Blogging & Digital Marketing.

While pursuing his engineering itself, he has trained many young generation people/students of more than 3500+ from various parts across Andhra Pradesh through his workshops, seminars, courses in Cyber Security and this makes him one of the youngest student trainers in India.

At the age of 20 he conducted his first workshop on Blogging & Ethical Hacking which was the beginning to his success in this field and right now he has handful of workshops to train students, government and corporate organizations as well in Andhra Pradesh & Telangana. He is the only student trainer who started conducting workshop for his peers, professors and for corporates.

The year 2016 gave me the conviction and confidence to notch up whatever I was doing. I'd organized a 50-hour marathon event on Ethical Hacking and Cyber Security in PSCMR College, Vijayawada that went on to bag to achieve Limca Book of Records for non-stop longest duration workshop. The impact we were making was clearly visible by now. Diverse people from Jammu & Kashmir in the North to Kanyakumari in the South had attended the event. Some of the Government of India officials took notice of this record and invited our team for couple of conferences at New Delhi. The country's esteemed institutions were reaching out to me to conduct training, workshops, and events on a myriad of subjects related to online security. Multi-National Corporations (MNC's) had begun to consult me on the Cyber Security of their systems.

❖ Life changing solution/service :

After working on R&D for around 2 years, finally in the mid 2019 we have successfully developed a service/solution of various techniques and strategies for the Film Industry through which he can kill piracy of any film in online up to 35% right now. This betaservice is being appreciated & adopted by various Tollywood Film Industry Producers & Hero's to safeguard their film from piracy in online and to gain more profits.

By the end of 2033 our vision is to rollout a complete full packed service/solution where we can kill piracy entirely 100% everywhere in online for sure.

Appreciation:

Received a great appreciation from our 1st Tollywood Film Industry client Mr.Saptagiri for providing our Anti-Piracy betaservice for his film VAJRA KAVACHADARA GOVINDA

❖ Our Company Achievements:

- On 17th August 2024, We (Supraja Technologies) launched our company's **Centre of Excellence in Cyber Security** (CoE) at Ramco Institute of Technology, Rajapalayam
- On 18th September 2024, We (Supraja Technologies) launched our company's **Centre of Excellence in Cyber Security** (CoE) at SRM University (Ramapuram Campus), Chennai
- On 20th November 2024, We (Supraja Technologies) launched our company's **Centre of Excellence in Cyber Security** (CoE) at St.Joseph's Institute of Technology, Chennai

❖ Records, Appreciations, Awards & Recognitions etc at a glance :

1. Holds a National Record in “Limca Book of Records – 2017”
2. Ex-Associate Member for National Cyber Safety and Security Standards (NCSSS)
3. Steering Committee Member for United Conference on Cyber Space (UNITEDCON 2020)
4. Judge for the Grand Finale of SIH (Smart India Hackathon 2024) Software Edition for the Nodal Centre at Sri Sai Ram Engineering College, Chennai which is an initiative by Ministry of Education (Govt. of India) and AICTE
5. Received Appreciation from Mr.Amit Narayan, Executive Director at Rajahmundry Asset of Oil and Natural Gas Corporation Limited (ONGC) on 16th December, 2022 for training their employees on Cyber Security
6. Awarded as a "Karmaveer Chakra - 2019", on 12th October 2019 at IIT Delhi, which was instituted by iCONGO in partnership with the United Nations
7. Received Appreciation from Mr.Sandeep Rathore, Commissioner of Police, Chennai on 2nd March 2024 for the exceptional commitment and invaluable contribution as a Member of JURY at Greater Chennai Police Cyber Hackathon 3.0
8. Evaluator for the “Innovative Bharat” which is organized by AICTE and Ministry of Education held on 6th January 2024
9. Judge for the Grand Finale of SIH (Smart India Hackathon 2023) Senior Software Edition for the Nodal Centre at PVPSIT, Vijayawada which is an initiative by Ministry of Education (Govt. of India) and AICTE
10. Appointed as a Member for Board of Studies on 19th April 2025 for the Departments of Cyber Security, Data Science and AIML at Bapatla Engineering College, Bapatla
11. Appointed as a Member for Board of Studies on 12th March 2025 for the Department of Cyber Security at St. Joseph’s Institute of Technology, Chennai
12. Appointed as one of the Industry Academia Advisory Council Member on 30th September 2023 for the Department of Cyber Security at VNR Vignana Jyothi Institute of Engineering & Technology, Hyderabad
13. Appointed as a Member for Board of Studies on 4th May 2023 for the Department of Cyber Security at Madanapalle Institute of Technology & Sciences, Madanapalle
14. Appointed as a Member for Board of Studies for the Department of MCA at NRI Institute of Technology, Perecherla (Guntur)
15. Awarded as a “Social Media Influencer - 2019”, on 30th June 2019 by Jignasa in association with Government of Andhra Pradesh
16. Nominated for “INDIA 500 CEO AWARD 2019”
17. Invited & Interviewed by ETV Andhra Pradesh news channel on 27th July, 2019 for a Special Story Interview on “Spy Apps”
18. Appreciated by Mr.Sridhar, Sub-Inspector of Police at Central Crime Branch, Vijayawada on 23rd October, 2018 for exclusively training him on Special Investigation Course, which will help him to solve the cyber crime cases easily
19. Received a great appreciation from our 1st Tollywood Film Industry client Mr.Saptagiri, for providing Anti-Piracy betaservice for his movie VAJRA KAVACHADARA GOVINDA

❖ Something we are proud of :

We are very happy to inform you all that our company, Supraja Technologies has been shortlisted for **"Top 50 Tech Companies" award 2019**, conferred at InterCon - Dubai, UAE.

Supraja Technologies is one out of thousands of startup companies that were initially screened by InterCon team of 45+ research analysts over a period of three months and the final shortlist includes 150+ firms and we are very proud to inform you all that our company Supraja Technologies also happens to be a part of the same.

Some Glimpses of our Journey

Mr.Santosh Chaluvadi – CEO, Supraja Technologies
Giving hands-on Cyber Security training workshop to the CSE students
at IIT Kharagpur

Mr.Santosh Chaluvadi – CEO, Supraja Technologies was
Invited & Interviewed by ETV Andhra Pradesh news channel on 27th July, 2019
for a Special Story Interview on “Spy Apps”

Mr.Santosh Chaluvadi – CEO, Supraja Technologies was awarded as a
"**Social Media Influencer 2019**" in recognition of his remarkable achievements in the social media as a part of First International Social Media Festival on 30th June 2019 by Jignasa in association with Government of Andhra Pradesh

On 23rd October 2018 Mr.Santosh Chaluvadi, CEO - Supraja Technologies and Mr.Krishna Chaitanya, CTO - Supraja Technologies had successfully completed delivering Special Investigation Course training in Cyber Security to Mr.Sridhar Garu, Sub-Inspector of Police at Central Crime Branch, Vijayawada which will help him to solve the cyber crime cases easily

ETV Andhra Pradesh News Channel interviewed Mr.Santosh Chaluvadi, CEO - Supraja Technologies for his achievements in the domain of Cyber Security & Digital Marketing

Supraja Technologies was invited by Indian Air Force (Air Wing NCC) to deliver a session on Latest Cyber Crimes & Awareness for the NCC cadets, staff and officers
on 4th July, 2019

Supraja Technologies – CEO, CTO & CMO with
Indian Air Force (Air Wing NCC) Group Captain Sandeep Gupta.
We thank Mr.Sandeep Gupta for inviting us on 4th July, 2019 to deliver a session on Latest Cyber Crimes & Awareness for the Indian Air Force (Air Wing NCC) cadets, staff & officers

TABLE OF CONTENTS

1. INTRODUCTION

1.1 OVERALL DESCRIPTION

The Metadata Extractor Tool is designed to extract meaningful metadata from structured data sources and organize it in a systematic and understandable format. Metadata refers to descriptive information about data such as source, type, format, creation details, and other attributes that help in better data understanding and management. This tool focuses on identifying and separating such metadata from raw data to improve data usability.

The system follows a modular approach where input data is first collected and pre-processed, after which relevant metadata attributes are extracted. The extracted metadata is then organized and stored for further analysis and reference. The tool emphasizes simplicity, accuracy, and scalability, making it suitable for academic, research, and basic data analysis applications.

The Metadata Extractor Tool reduces manual effort involved in data analysis and helps users gain insights into data characteristics efficiently. It can be extended in the future to support additional data formats, advanced correlation techniques, and visualization features, making it a flexible and reusable solution for metadata management.

2 . EXISTING SYSTEM

In the existing system, data analysis is mainly performed on raw data while metadata is either ignored or handled manually. Traditional methods do not provide an automated way to extract metadata in a structured format, which makes the process time-consuming and prone to errors. Metadata is often scattered across files or datasets and lacks proper organization, making it difficult to understand data characteristics and relationships. Additionally, existing systems do not support effective correlation between different metadata attributes, which limits deeper data analysis. As the volume of data increases, these manual and non-automated approaches become inefficient, less reliable, and difficult to scale, highlighting the need for an improved metadata extraction solution.

3 . PROPOSED SYSTEM

The proposed system introduces an automated Metadata Extractor Tool that efficiently extracts metadata from structured data sources and organizes it in a systematic manner. Unlike the existing system, this approach eliminates manual effort by automatically identifying relevant metadata attributes such as data source, type, format, and creation details. The extracted metadata is processed and stored in a structured format, enabling easy understanding and analysis of data characteristics. The system also supports basic correlation of metadata attributes to identify relationships within the data. This automated and modular approach improves accuracy, reduces processing time, and enhances scalability, making the proposed system more efficient and suitable for academic, research, and data analysis applications.

4. SYSTEM DESIGN

4.1 FEASIBILITY STUDY

The feasibility study is conducted to evaluate the practicality and effectiveness of developing the Metadata Extractor Tool. It analyzes whether the proposed system can be implemented successfully within the given constraints such as cost, technology, and user acceptance. This study helps in determining the viability of the project before full-scale implementation. The feasibility of the system is analyzed under three major aspects: economical feasibility, technical feasibility, and social feasibility.

4.1.1 ECONOMICAL FEASIBILITY

The Metadata Extractor Tool is economically feasible as it is developed using open-source technologies and does not require any licensed software. The system can be implemented on existing hardware without the need for additional infrastructure. Since development, deployment, and maintenance costs are minimal, the project is cost-effective and suitable for academic institutions and small-scale users.

4.1.2 TECHNICAL FEASIBILITY

The proposed system is technically feasible because it uses well-known and widely supported technologies such as Python and basic data handling techniques. The required tools are easy to install and operate on standard computer systems. The system design is simple and modular, which allows smooth development, testing, and future enhancements without technical complexity.

4.1.3 SOCIAL FEASIBILITY

The system is socially feasible as it is user-friendly and easy to understand. It reduces manual effort in handling and analyzing metadata, thereby improving efficiency and productivity. The tool does not disrupt existing workflows and can be easily adopted by students, researchers, and professionals for data organization and analysis purposes.

4.2 INPUT AND OUTPUT DESIGN

4.2.1 INPUT DESIGN

Input design is an important phase of the system as it defines how data is provided to the Metadata Extractor Tool for processing. The primary objective of input design is to ensure that the data entered into the system is accurate, complete, and in a suitable format for metadata extraction. A well-designed input mechanism reduces errors and improves overall system efficiency.

In the proposed system, the input consists of structured data files such as CSV or text files that contain datasets from which metadata can be extracted. These input files may include attributes like data

source, file type, creation date, author, category, or other descriptive information. The system validates the input data to ensure correctness and consistency before processing.

The input design is kept simple and user-friendly so that users can easily provide datasets without requiring technical expertise. By accepting standardized file formats and performing basic validation checks, the system ensures smooth data processing and reliable metadata extraction.

4.2.2 OBJECTIVES

The main objective of the Metadata Extractor Tool is to extract meaningful metadata from structured data sources and organize it in a systematic manner. The project aims to simplify the process of metadata identification and reduce manual effort involved in data analysis. By extracting metadata automatically, the system helps users better understand the characteristics of data.

The specific objectives of the project are:

- To collect input data from structured file formats such as CSV or text files
- To extract relevant metadata attributes from the input data
- To organize extracted metadata in a structured and readable format
- To correlate metadata attributes to identify basic relationships within the data
- To improve data usability and support efficient data analysis

The project focuses on providing a simple, scalable, and user-friendly solution for metadata management that can be applied in academic and basic data analysis environments.

4.2.3 OUTPUT DESIGN

Output design focuses on presenting the processed metadata in a clear, structured, and understandable format. The main objective of output design is to ensure that the extracted metadata is easy to interpret and useful for further analysis. A well-designed output improves user understanding and supports effective decision-making.

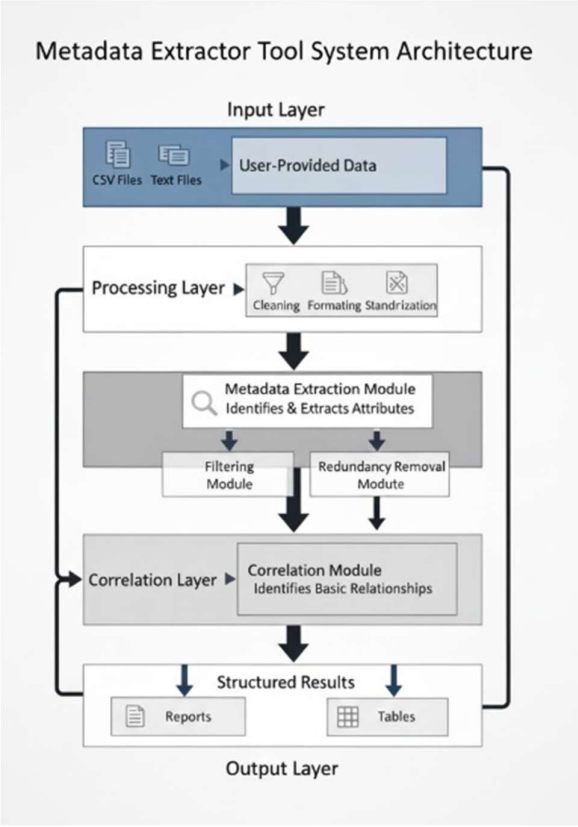
In the proposed system, the output consists of organized metadata presented in the form of structured tables, reports, or files. The extracted metadata attributes such as data source, type, format, and other descriptive details are displayed in a systematic manner. The output may be stored in formats such as CSV or text files, making it easy to view, share, and reuse.

The output design emphasizes simplicity, accuracy, and readability. By providing well-structured and meaningful output, the Metadata Extractor Tool helps users quickly understand the characteristics of the input data and utilize the metadata efficiently for academic and data analysis purposes.

5. IMPLEMENTATION

The implementation of the Metadata Extractor Tool follows a modular and structured approach to efficiently process input data and extract meaningful metadata. The system begins with data collection and pre-processing, where input datasets are cleaned, formatted, and standardized to ensure consistency. The pre-processed data is then passed to the metadata extraction process, which identifies relevant descriptive attributes such as source, type, format, and other characteristics. Filtering is applied to remove unnecessary or duplicate metadata entries, thereby improving accuracy and efficiency. The spamming module handles redundancy by eliminating repeated metadata records and retaining only unique information. The mailing module ensures smooth communication and data transfer between different modules of the system. Finally, the processed and correlated metadata is organized and generated as structured output in the form of reports or tables. This modular implementation reduces manual effort, improves data organization, and provides a scalable and easy-to-understand solution for metadata management in academic and basic data analysis environments.

5.1 SYSTEM ARCHITECTURE

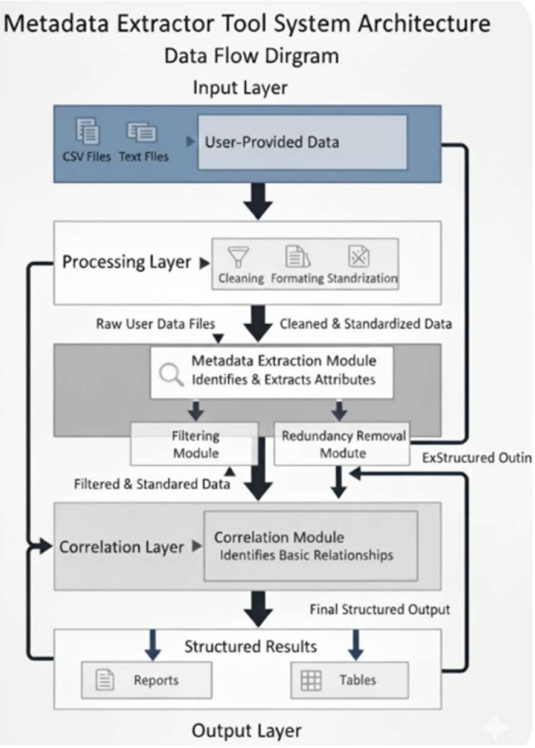


6. ALGORITHMIC IMPLEMENTATION

The algorithmic implementation of the Metadata Extractor Tool follows a systematic and sequential approach to extract and organize metadata from structured input data. The process begins by accepting input data files such as CSV or text files from the user. The input data is then pre-processed to remove inconsistencies, handle missing values, and standardize formats to ensure data quality. After pre-processing, relevant metadata attributes such as data source, type, format, and descriptive details are extracted from the data. Filtering is applied to remove irrelevant metadata, and duplicate entries are eliminated to avoid redundancy. The extracted metadata attributes are then correlated to identify basic relationships within the data. Finally, the processed metadata is organized into a structured format such as tables or reports and generated as output, providing a clear and efficient representation of metadata for analysis and reference.

7. SYSTEM DESIGN

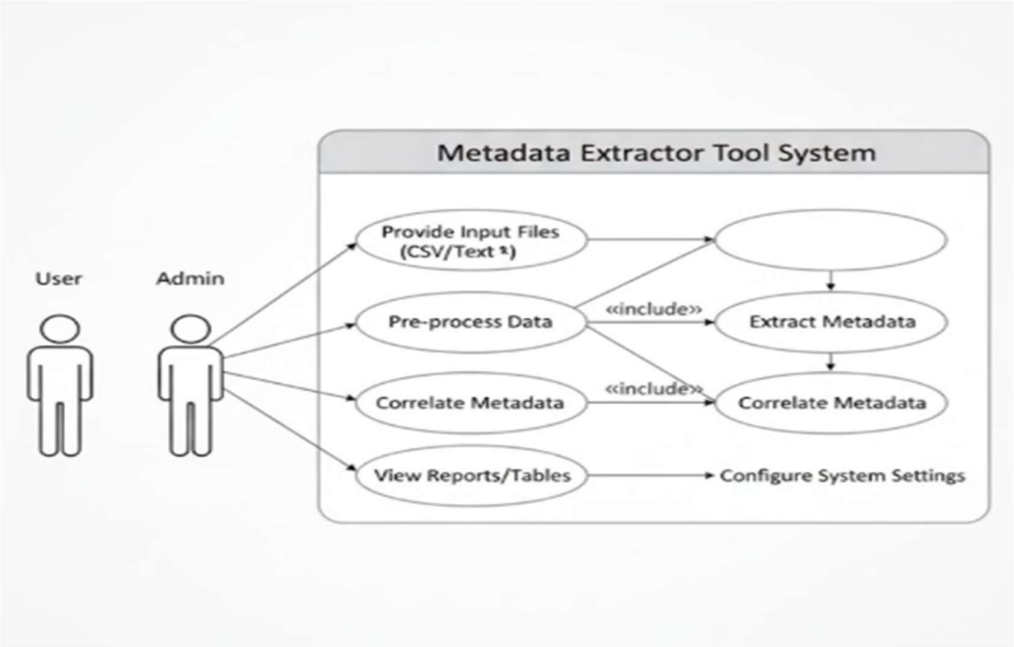
7.1 DATA FLOW DIAGRAM



The Data Flow Diagram (DFD) represents the flow of data through the Metadata Extractor Tool and illustrates how input data is transformed into structured output. The process begins at the Input Layer, where user-provided data such as CSV files and text files are supplied to the system. This raw data is then passed to the Processing Layer, where it undergoes cleaning, formatting, and standardization to ensure consistency and accuracy. After pre-processing, the cleaned data is forwarded to the Metadata Extraction

Module, which identifies and extracts relevant metadata attributes from the input data. The extracted metadata is then processed by the Filtering Module to remove unnecessary information and by the Redundancy Removal Module to eliminate duplicate metadata entries. The refined metadata is passed to the Correlation Layer, where basic relationships among metadata attributes are identified. Finally, the processed and correlated metadata is delivered to the Output Layer, where it is presented as structured results in the form of reports and tables. This data flow ensures smooth processing, organized metadata management, and clear output generation**.**

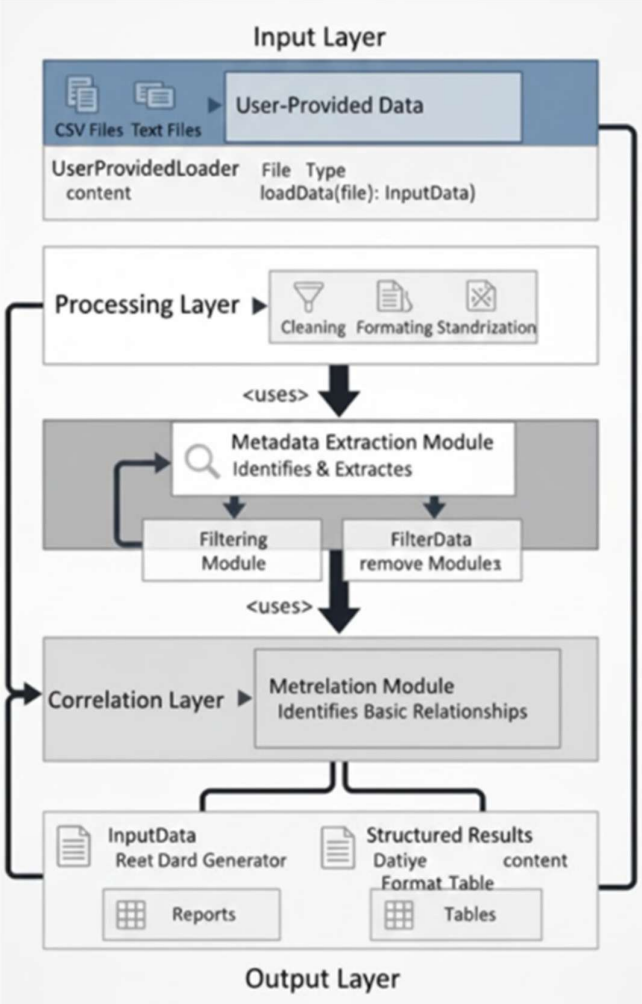
7.2 USE CASE DIAGRAM



The Use Case Diagram represents the functional interactions between the users and the Metadata Extractor Tool system. It illustrates how different actors interact with the system to perform various operations related to metadata extraction and analysis. The primary actors involved in the system are the User and the Admin.The User interacts with the system by providing input data files such as CSV or text files. After providing the input, the user can initiate the pre-processing of data, which includes cleaning, formatting, and standardization. The user can then request the system to extract metadata from the processed data and perform correlation of metadata attributes to identify relationships. Finally, the user can view the generated outputs in the form of reports and tables.The Admin has additional privileges and is responsible for managing and configuring system settings. The admin can monitor system operations, manage input handling, and ensure proper functioning of the metadata extraction and correlation processes. The use case diagram also shows relationships such as «include», indicating that certain operations like metadata extraction are dependent on pre-processing,and correlation depends on successful metadata extraction.Overall, the Use Case Diagram provides a clear understanding of system

functionality, user roles, and interactions, ensuring that all required operations of the Metadata Extractor Tool are well-defined and structured**.**

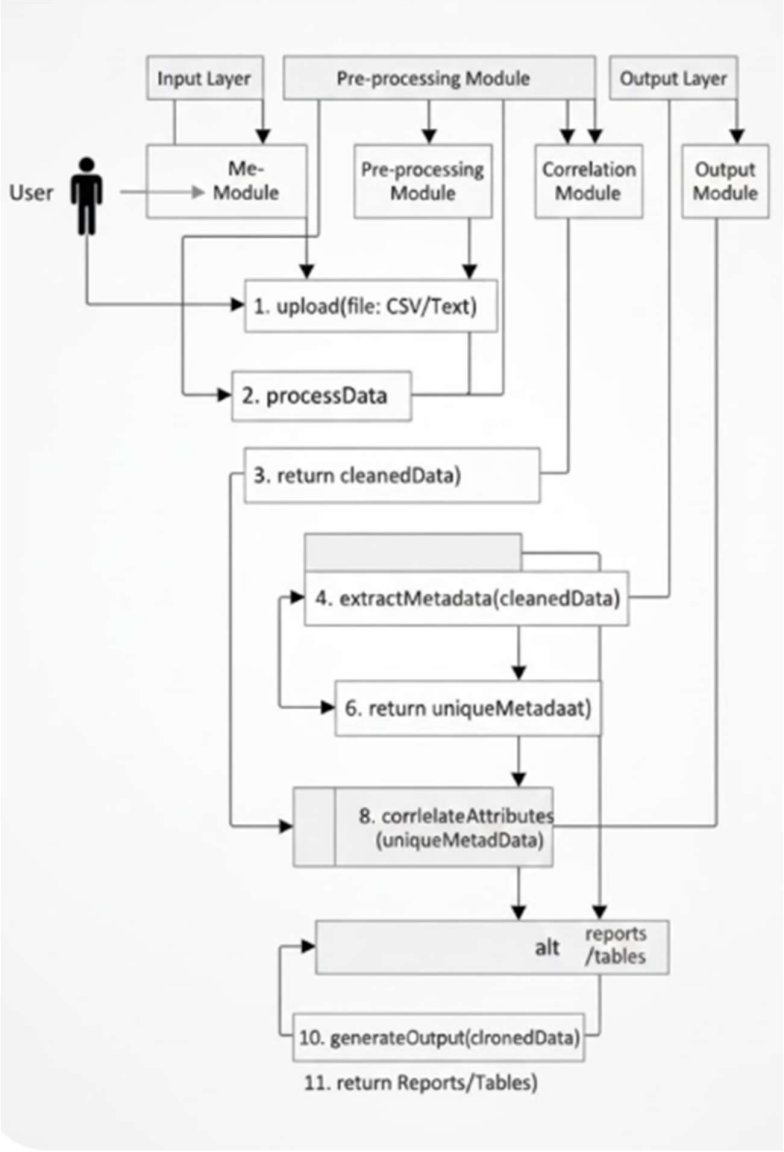
7.3 CLASS DIAGRAM



The Class Diagram represents the static structure of the Metadata Extractor Tool by illustrating the system's classes, their attributes, methods, and relationships. The system is designed using a layered class structure to ensure modularity and ease of understanding. The UserProvidedLoader class is responsible for accepting user-provided input files such as CSV and text files and loading the data into the system. This class contains methods to identify the file type and load input data appropriately. The loaded data is passed to the ProcessingLayer class, which performs data cleaning, formatting, and standardization to ensure consistency. The MetadataExtractionModule class identifies and extracts relevant metadata attributes from the processed data. It interacts with the FilteringModule, which removes irrelevant metadata, and the FilterData class, which eliminates duplicate or redundant entries. The refined metadata is then passed to the CorrelationModule, which identifies basic relationships among metadata attributes.

Finally, the OutputLayer class generates structured results in the form of reports and tables using classes such as ReportGenerator and TableGenerator. The class diagram clearly defines responsibilities for each class and shows how classes interact to transform raw input data into structured metadata output, ensuring clarity, reusability, and scalability of the system.

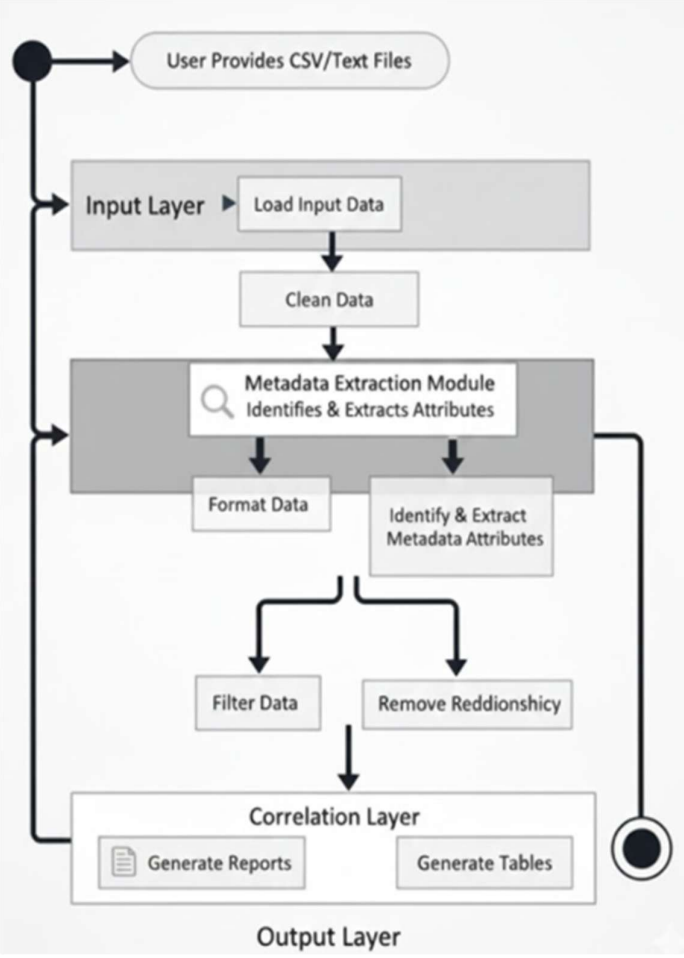
7.4 SEQUENCE DIAGRAM



The Sequence Diagram illustrates the step-by-step interaction between the user and the different modules of the Metadata Extractor Tool over time. The sequence begins when the User uploads an input file in CSV or text format through the Input Layer. The uploaded file is received by the Input Module, which

forwards the data to the Pre-processing Module. In the pre-processing stage, the system performs data cleaning, formatting, and standardization, after which the cleaned data is returned. The cleaned data is then passed to the Metadata Extraction Module, where relevant metadata attributes are extracted. After extraction, duplicate or redundant metadata entries are removed, and unique metadata is returned to the system. The unique metadata is then sent to the Correlation Module, which correlates metadata attributes to identify basic relationships. Based on the correlation results, the system proceeds to the Output Module, where structured output is generated in the form of reports or tables. Finally, the generated reports and tables are returned to the user. This sequence diagram clearly represents the flow of control and data exchange among system components, ensuring proper coordination between modules and smooth execution of the metadata extraction and correlation process

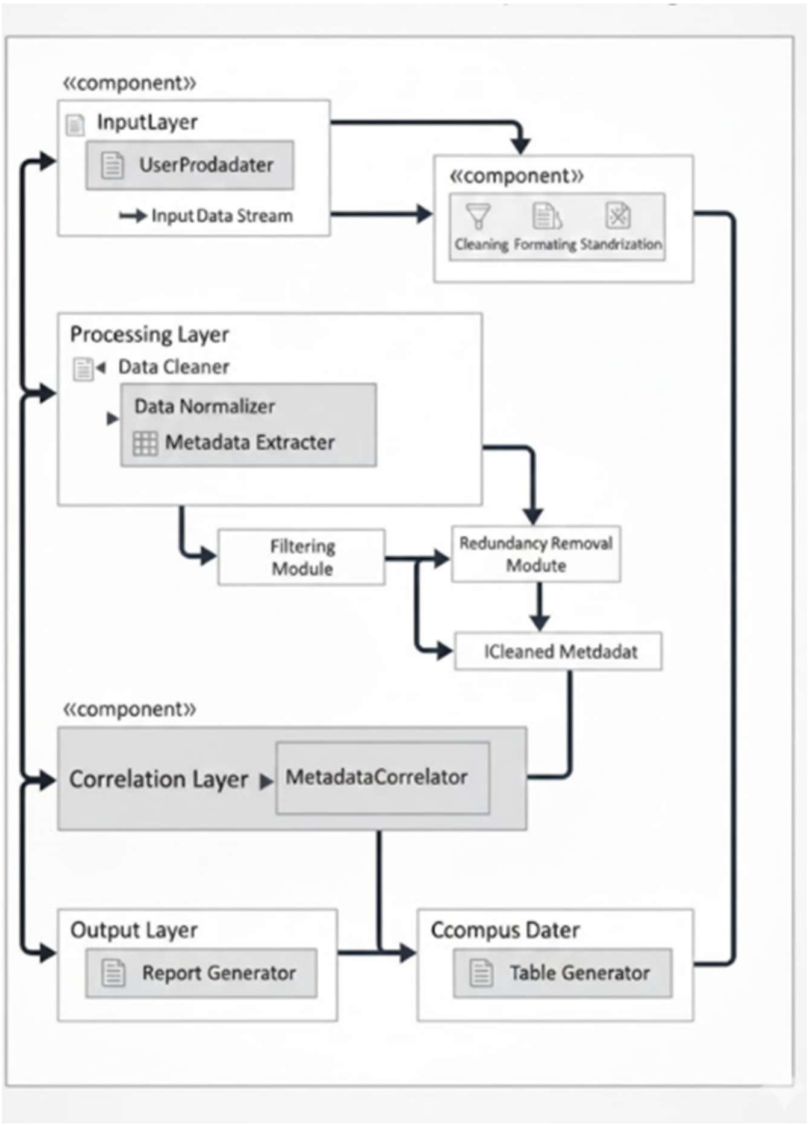
7.5 ACTIVITY DIAGRAM



The Activity Diagram represents the overall workflow of the Metadata Extractor Tool by showing the sequence of activities performed from start to end. The process begins when the user provides input data

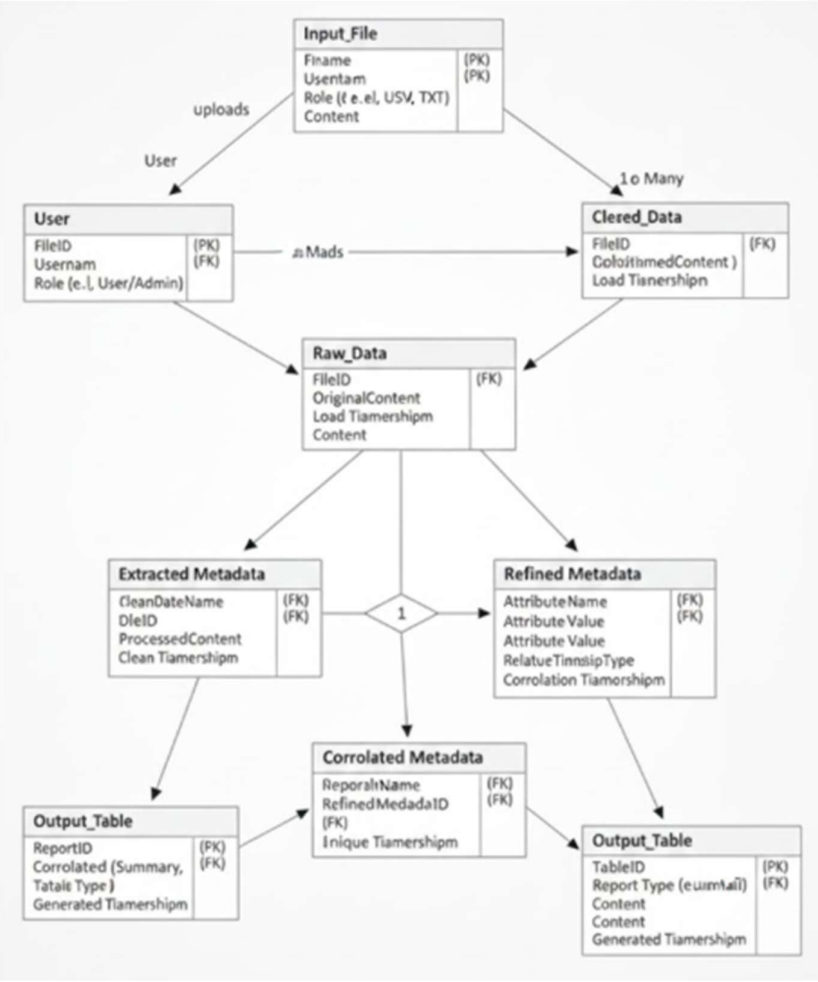
in the form of CSV or text files. The system then moves to the input layer, where the input data is loaded into the system. After loading, the data undergoes a cleaning process to remove inconsistencies and errors. Once the data is cleaned, it is passed to the metadata extraction module, where relevant metadata attributes are identified and extracted. The extracted metadata is then formatted properly to maintain uniformity. Following this, filtering is performed to remove unnecessary metadata, and redundancy removal is applied to eliminate duplicate entries. The refined metadata is then sent to the correlation layer, where basic relationships among metadata attributes are established. Finally, the processed data reaches the output layer, where structured outputs are generated in the form of reports and tables. The activity diagram clearly illustrates the continuous flow of operations and ensures an organized and systematic execution of the metadata extraction and correlation process.

7.6 COMPONENT DIAGRAM



The Component Diagram illustrates the high-level structure of the Metadata Extractor Tool by showing its major components and the interactions between them. The system is divided into several logical components to ensure modularity and efficient data processing. The Input Layer component is responsible for receiving user-provided data in the form of CSV or text files and converting it into an input data stream. This data is forwarded to the Processing Layer, which consists of components such as Data Cleaner, Data Normalizer, and Metadata Extractor. These components perform cleaning, formatting, standardization, and metadata extraction operations. The extracted metadata is then passed through the Filtering Module and Redundancy Removal Module, which remove irrelevant and duplicate metadata entries, producing cleaned metadata. The cleaned metadata is sent to the Correlation Layer, where the Metadata Correlator component identifies basic relationships among metadata attributes. Finally, the processed data reaches the Output Layer, which includes the Report Generator and Table Generator components that generate structured outputs in the form of reports and tables. The component diagram clearly demonstrates how each component interacts with others, ensuring smooth data flow, separation of responsibilities, and scalability of the Metadata Extractor Tool.

7.7 E-R DIAGRAM



The Entity–Relationship (E–R) Diagram represents the database structure of the Metadata Extractor Tool by defining the entities, their attributes, and the relationships among them. The primary entity User stores user information such as user ID, username, and role, and is responsible for uploading input files. The Input_File entity contains details about uploaded files including file name, file type, and content, and is associated with the user through an upload relationship. The uploaded files generate Raw_Data, which stores the original content along with timestamps. This raw data is processed to produce Cleaned_Data, representing formatted and standardized content. From the cleaned data, Extracted_Metadata is derived, which stores processed metadata attributes and timestamps. The extracted metadata is further refined to form Refined_Metadata, where individual attribute names, values, and relationship types are maintained. The Correlated_Metadata entity captures relationships among refined metadata attributes and stores correlation information. Finally, the processed and correlated metadata is presented through Output_Table entities, which store generated reports and tables along with timestamps. The E–R diagram clearly illustrates how data flows from user input to final output while maintaining proper relationships, ensuring structured data storage and efficient metadata management

8. REQUIREMENTS SPECIFICATION

8.1 FUNCTIONAL REQUIREMENTS

The functional requirements define the core functionalities that the Metadata Extractor Tool must perform to meet project objectives. These requirements describe how the system should behave and what operations it must support to successfully extract and manage metadata.

The functional requirements of the system are as follows:

- The system should accept structured input data files such as CSV or text files from the user.
- The system should read and process the input data correctly without data loss.
- The system should perform pre-processing on input data to handle formatting issues and inconsistencies.
- The system should extract relevant metadata attributes from the input data automatically.
- The system should filter out irrelevant or unnecessary metadata information.
- The system should remove duplicate or redundant metadata entries to improve clarity.
- The system should correlate metadata attributes to identify basic relationships within the data.
- The system should store the extracted metadata in a structured format.
- The system should generate output in the form of reports or tables for easy understanding.
- The system should allow users to view and access the generated metadata output

8.2 SOFTWARE REQUIREMENTS

The Metadata Extractor Tool requires basic and commonly available software to ensure smooth development and execution. The system is designed using open-source technologies, which makes it easy to implement and cost-effective. The primary programming language used for developing the tool is Python, as it provides strong support for data handling and processing. Libraries such as Pandas are used for reading, processing, and organizing structured data files like CSV or text files. A standard operating system such as Windows, Linux, or macOS is sufficient to run the application. Additionally, any basic code editor or integrated development environment (IDE) such as Visual Studio Code, PyCharm, or IDLE can be used for development and testing. These software requirements ensure simplicity, flexibility, and ease of use for academic and learning purposes.

8.3 OPERATING SYSTEMS SUPPORTED

The Metadata Extractor Tool is designed to be platform-independent and can operate efficiently on commonly used operating systems. Since the system is developed using open-source and cross-platform technologies, it does not depend on any specific operating system. The tool can be executed on Microsoft Windows, Linux, and macOS environments without requiring any special modifications. These operating systems provide the necessary support for running the required software tools and libraries used in the project. As a result, the system ensures flexibility, ease of deployment, and accessibility for users across different platforms.

8.4 TECHNOLOGIES AND LANGUAGES USED TO DEVELOP

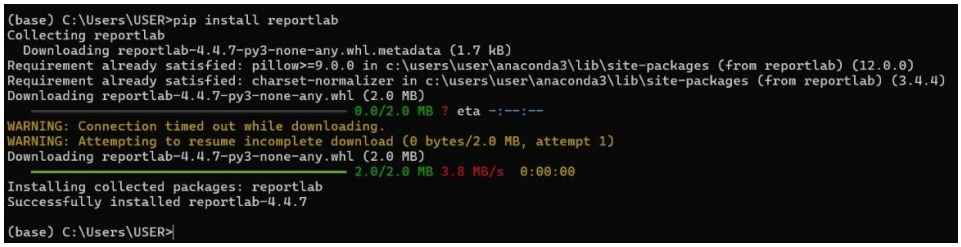
The Metadata Extractor Tool is developed using simple, reliable, and open-source technologies to ensure ease of implementation and understanding. The primary programming language used for developing the system is Python, as it provides strong support for data processing and file handling. Python libraries such as Pandas are used for reading, processing, and organizing structured data formats like CSV and text files. Basic file handling techniques are used to manage input and output data efficiently. The development process is carried out using standard development tools such as code editors or integrated development environments (IDEs). These technologies and languages are chosen to provide flexibility, scalability, and ease of use, making the system suitable for academic, research, and basic data analysis applications.

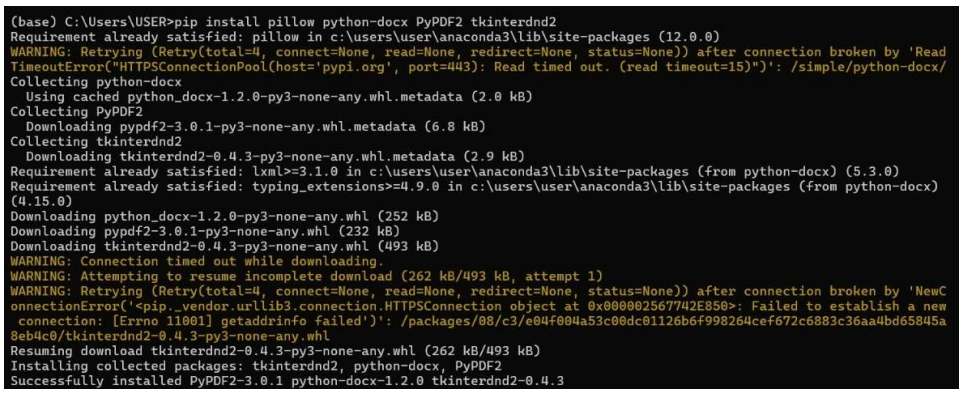
8.5 HARDWARE REQUIREMENTS

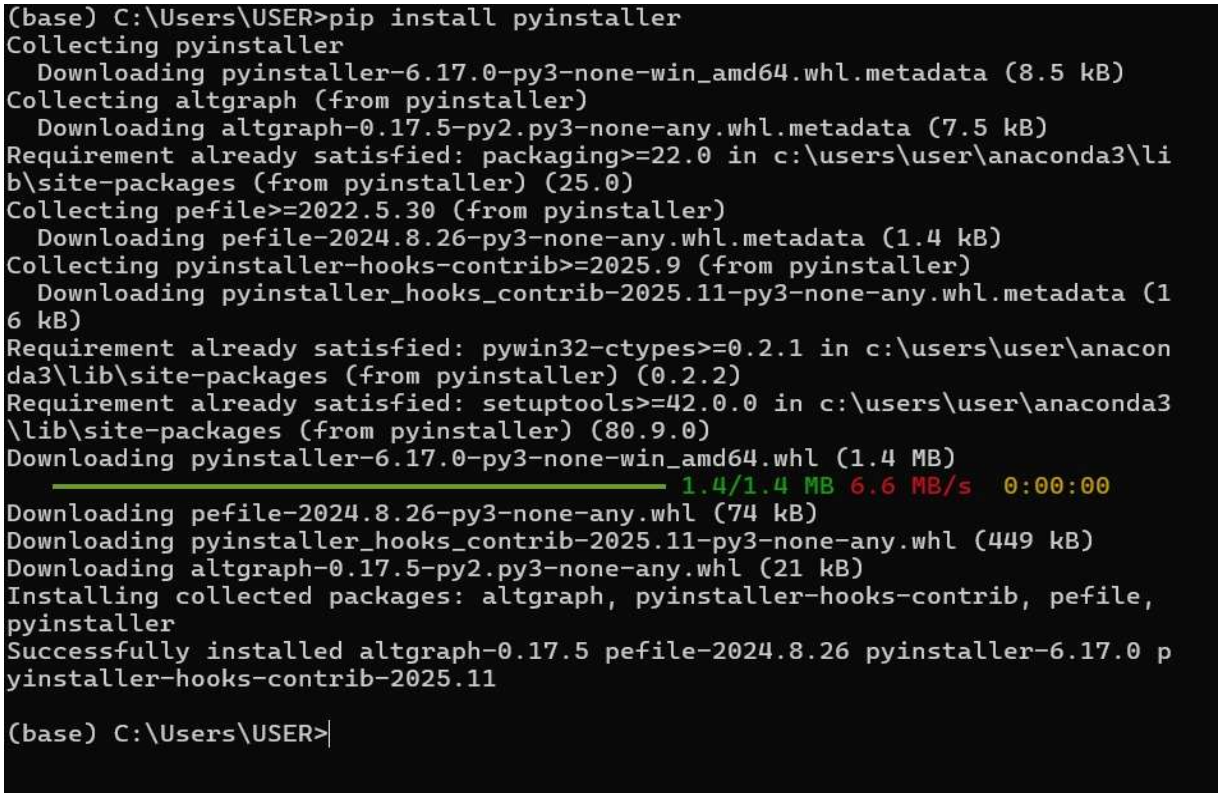
The Metadata Extractor Tool does not require high-end hardware for its operation, as it is designed to run on standard computer systems. A basic system configuration is sufficient to execute the application efficiently. The minimum hardware requirements include a processor with standard computing capability, at least 4 GB of RAM to handle data processing tasks smoothly, and adequate storage space to store input data files and generated output files. A basic keyboard, mouse, and display unit are required for user interaction. Since the system is lightweight and intended for academic and basic data analysis purposes, it can be easily executed on commonly available personal computers or laptops without any additional hardware components

9. EXECUTION

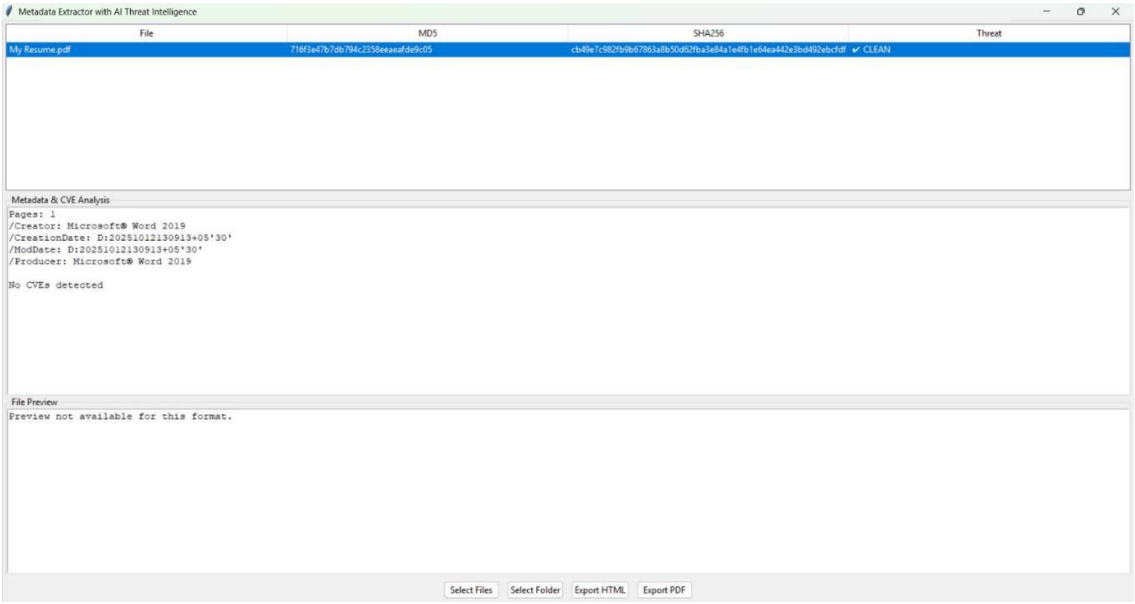
Converting the source code into an executable file (.exe file)







ToolInterface

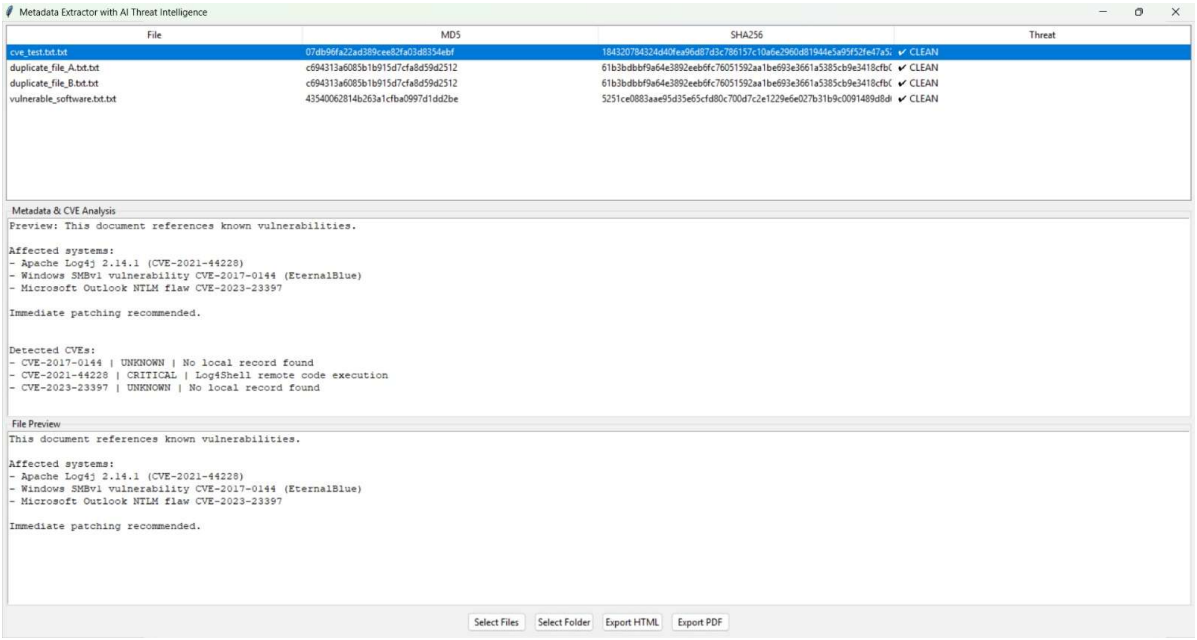


The diagram represents the working interface of the Metadata Extractor with AI Threat Intelligence tool. The top section displays the list of selected files along with their calculated MD5 and SHA-256 hash values, which uniquely identify the file and help verify its integrity. The Threat column indicates the security status of the file; in this case, the file is marked as CLEAN, meaning no known threat is detected.

The Metadata & CVE Analysis section shows extracted metadata such as the number of pages, creator application, creation date, modification date, and producer details. This information helps in understanding the origin and properties of the file. Below this, the system performs CVE analysis to check whether the software metadata is associated with known vulnerabilities, and the result shows that no CVEs were found.

The File Preview section attempts to display file content when supported. At the bottom, action buttons allow users to select files or folders and export analysis results in HTML or PDF format, making the tool suitable for documentation and security analysis.

Final Result



The final output of the Metadata Extractor with AI Threat Intelligence displays the analysis results of the input files selected by the user. Each file is processed to generate MD5 and SHA-256 hash values for integrity verification and duplicate identification. The system extracts metadata and correlates file content with known CVE vulnerability information. Detected vulnerabilities, affected systems, and recommended actions are clearly shown in the analysis section. The threat status indicates whether files are safe or suspicious. This output confirms the successful extraction, correlation, and security analysis of metadata.

10 . CONCLUSION

The Metadata Extractor with AI Threat Intelligence project successfully demonstrates an effective approach to extracting, analyzing, and correlating metadata from digital files. The system accurately generates cryptographic hash values to verify file integrity, extracts meaningful metadata, and identifies references to known vulnerabilities using CVE analysis. By combining metadata extraction with basic threat intelligence, the project enhances file analysis and security awareness. The user-friendly interface and report generation features make the tool suitable for academic, forensic, and basic security applications. Overall, the project proves that metadata-driven analysis can improve data understanding, integrity verification, and vulnerability awareness in a simple and efficient manner.