# Open Source Intelligence and its Applications in Next Generation Cyber Security - A Literature Review

**Yogish Pai U.[1], & Krishna Prasad K.[2]**

[1]Research Scholar, College of Computer Science and Information Science, Srinivas University, Mangalore, India

ORCID-ID: 0000-0002-4266-2809; Email: yogish77pai@gmail.com

[2]College of Computer Science and Information Science, Srinivas University, Mangalore, India

ORCID-ID: 0000-0001-5282-9038; E-mail: krishnaprasadkcci@srinivasuniversity.edu.in

**Subject Area:** Computer Science.
**Type of the Paper:** Review based Analysis.
**Type of Review:** Peer Reviewed as per |C|O|P|E| guidance.
**Indexed In:** OpenAIRE.
**DOI:** http://doi.org/10.5281/zenodo.5171580
**Google Scholar Citation:** IJAEML

**How to Cite this Paper:**

Yogish Pai, U. & Krishna Prasad, K. (2021). Open Source Intelligence and its Applications in Next Generation Cyber Security - A Literature Review. *International Journal of Applied Engineering and Management Letters (IJAEML)*, *5*(2), 1-25. DOI: http://doi.org/10.5281/zenodo.5171580

**International Journal of Applied Engineering and Management Letters (IJAEML)**
A Refereed International Journal of Srinivas University, India.

Crossref DOI : https://doi.org/10.47992/IJAEML.2581.7000.0100

# Open Source Intelligence and its Applications in Next Generation Cyber Security - A Literature Review

**Yogish Pai U.[1], & Krishna Prasad K.[2]**

[1]Research Scholar, College of Computer Science and Information Science, Srinivas University, Mangalore, India

ORCID-ID: 0000-0002-4266-2809; Email: yogish77pai@gmail.com

[2]College of Computer Science and Information Science, Srinivas University, Mangalore, India

ORCID-ID: 0000-0001-5282-9038; E-mail: krishnaprasadkcci@srinivasuniversity.edu.in

## ABSTRACT

**Purpose:** *Research serves as a springboard for new ideas, and every scholarly research begins with a review of the literature. This literature review to familiarize oneself with the domain of research and to establish the credibility of the work. It also aids in the integration and summarization of the subject.*

**Methodology**: *The necessary literature on the chosen topic have been gathered from multiple secondary data sources such as journals, conference proceedings, books, research papers published in various reputable publications, and then shortlisted the literature which are relevant for the work. The shortlisted literatures were carefully evaluated by reading each paper and taking notes as needed. The information gathered is then analyzed in order to identify the problem areas that may exist in the chosen topic.*

**Findings/Result**: *It has been observed that the chosen topic, Opensource Intelligence (OSINT) practice requires more robust and intelligent solutions from AI and its subfields. The capability of OSINT for intelligent analysis strengthens tightly integrating machine learning and automated reasoning techniques. To avoid human errors, the dependency on humans in decision-making ought to reduce. To eradicate any incorrect information, a truth discovery process is mandatory. OSINT is able to discover new knowledge by correlating intelligence from other OSINT sources. Even though Artificial Intelligence has entered the OSINT field, there is still a long way to go before OSINT fully prepares for the much-anticipated Web 3.0.*

**Originality:** *A literature review have had been carried out using secondary data gathered from various online sources, and new knowledge in the form of findings was derived in order to construct a theoretical framework and methodology for future research. It has been ensured that no judgments or decisions are made with a biased mindset or under the influence of any predetermined mentality. A concerted effort has been made to identify a research topic for further investigation.*

**Paper Type:** *Literature Review.*

**Keywords:** OSINT, Artificial Intelligence, NLP, Cyber Security, Machine learning

## 1. INTRODUCTION :

The twenty-first century witnessed a lot of changes in cyberspace as a result of technological advancement and the development of a slew of creative applications. Blogging sites, social media, photo sharing, and video sharing are just a few examples of creative applications that were launched in the early twenty-first century and quickly gained widespread popularity. Using these platforms, end users were able to publish and share material with other users. The data uploaded was not restricted to photos and videos; it was customary to express one's thoughts, opinions, and feelings on an online platform, resulting in a massive amount of data accumulating in web space (Husse et al., 2021) [1] (Dashtipour et al., 2016) [2]. Because the contents were openly available on the internet, they were accessible and readable by everyone. The introduction of smartphones and the availability of 3G internet connections greatly popularized these platforms, resulting in a record-high user base for these applications. People's

life has been inextricably linked to social media. Meanwhile, the evolution of cloud computing has prompted business communities to adopt this service. By this time, the most conventional services such as newspapers, libraries, bill payment, government, and education sector have had digitized and were storing all of their data on the web space.

Many organizations considered the data that accumulated on web space as a result of digital transformation to be a gold mine because it could have been transformed into knowledge and intelligence. As a result, Opensource Intelligence (OSINT), a more than half-century-old practice used to extract meaningful intelligence from publicly available data and which was previously used by defense personnel, regained widespread popularity (Charalambous et al., 2016) [3]. OSINT gathers publicly available data, processes it, and then transforms it into knowledge that OSINT practitioners and other business sectors demand. Online communities that are quite similar to the social networks can be browsed. OSINT evaluate**s** the postings and topics of these forums since they produce fascinating interactions (Pastrana et al., 2018) [4]. OSINT, like any other intelligence system, has its own systematic methodology for collecting data, cleansing, analyzing, and disseminating it. As the number of open source data **is** available on the public domain rose, the necessity of artificial intelligence in the OSINT process became evident. The methods employed in OSINT were developed to solve broad issues. The topics explain the problem that are being solved as well as how the solver employs them (Ponder-Sutton 2016) [5]. It was hard to discover, extract, and analyses the required information from the unstructured variety of data that people were uploading. Text mining, pattern matching, entity extraction, natural language processing, and machine learning have all acquired a lot of traction and made life easier for OSINT practitioners up to a point. Complex networks, Cloud computing, and big data have all made significant contributions to modern OSINT. Precision, reliability, promptness, and, most importantly, gaining a competitive advantage are the criteria for evaluating intelligence, including OSINT (Benes 2013) [6]. OSINT has expanded its reach to a wide number of fields in recent years, making it indispensable for many applications. The demand for OSINT is growing every day, as is the adoption rate, because it is less expensive, has no risks, and is based on publicly available data. OSINT has the potential to generate unique and novel data and insights, but it also has technical, political, and ethical issues and obstacles that must be properly addressed (Layton & Watters 2016) [7].

Corporate businesses rely on OSINT for various purposes like market forecasting, competitor analysis, and customer sentiment analysis (Santarcangelo et al., 2015) [8]. Whereas law enforcement agencies are tasked with profiling criminals and extremists, forensic analysis of criminal activity, and in the cyber security domain (Hassan & Hijazi 2018) [9], OSINT has been used to discover vulnerabilities in the IT infrastructure (Azevedo et al., 2019) [10]. Even after technological advancement and ~~th~~e implementation of AI in OSINT, there was still a need for human intervention at certain stages, such as decision-making. This highlights a potential technological gap in the current OSINT process. This review of the literature is based on secondary data gathered from a large number of papers published in journals and online sources. The topics covered in the paper are as follows: (i) Stages of open source intelligence, (ii) Present status of opensource intelligence system, (iii) Applications of OSINT in cyber security, (iv) Summary of related work, (v) Discussion & future work, (vi) Research gap, (vii) Research agenda, (viii) Analysis of research agenda, (ix) Research proposal, (x) SLOC analysis of research proposal, and (xi) Conclusion.

## 2. RESEARCH OBJECTIVE AND METHODOLOGY :

The use of Artificial Intelligence in OSINT not only improved the reliability but also speed of the process. Some AI sub functions have been employed in OSINT tools, however the degree of implementation and optimal use of AI capability must be investigated. The purpose of the literature review is to gain a better understanding of OSINT's current state and to identify the applications that rely on it. The integration of artificial intelligence with OSINT, as well as its role in cyber security, were also investigated. The goal of the project was to see how prepared OSINT is for Web 3.0 which will allow for the creation of a global data warehouse in which any data format may be shared and interpreted irrespective of any device over any network (Bruwer & Rudman 2015) [11]. As an outcome of the study, the following questions were formulated.

1. What types of applications make use of opensource intelligence?
2. On what do they rely on OSINT?

3. What services does OSINT provide to other sectors?
4. What role does artificial intelligence play in OSINT?
5. What are the numerous OSINT and Cybersecurity projects that have been undertaken?
6. What are the various ways for improving the OSINT framework's efficiency?

## 3. OVERVIEW OF OPEN SOURCE INTELLIGENCE :

Open source intelligence (OSINT) is intelligence gathered from publicly available data sources such as academic publications, journals, social media sites, online communities, and newspapers, among others. OSINT was developed for espionage purposes during the Second World War (Glassman & Kang 2012) [12]. OSINT has gained popularity in the modern era as a result of the internet revolution, which has resulted in the accumulation of massive amounts of data on the Internet. Social media posts, blogs, journals, published articles, newspapers, video-audio files, online forums, discussion groups, company websites, government documents, maps (Klaus et al., 2020) [13] (John et al., 2007) [14], and so on are all sources of data for modern-day OSINT. Governments and intelligence services are increasingly relying on OSINT to conduct investigations and combat cybercrime (Nouh et al., 2019) [15]. Open source intelligence not only finds and gathers information, but ~~it~~ also searches, chooses, and extracts relevant material from microblogging sites, and then analyses the information to provide an intelligent report (Koops et al., 2013) [16]. OSINT employs a systematic methodology to extract valuable intelligence from raw data and present it in the form of a usable intelligent report. While credibility is an important attribute for retrieved results, it does not guarantee the accuracy of any returned information (Weir 2016) [17]. Because the Cyberspace is constantly evolving, any study's conclusions are only relevant for the time period when it has been conducted, and the efficacy of the tool being used also has a significant role, which are frequently beyond the reach of practitioner (Bar-Ilan 2001) [18]. The entire OSINT activity can be divided into four parts: data collection, data processing, data exploitation, and data production or extraction. Data from websites can be collected using search engines such as Duckduckgo, Shodan, and others. The data processing stage primarily ensures that unwanted information is removed and raw data is converted to meaningful data. It's critical that data be gathered for the appropriate reasons so that no more data is gathered than is really essential (Gibson 2016) [35]. The data exploitation phase, also known as the analysis phase, is in charge of verifying the authenticity and credibility of the data processed. During the analysis and interpretation phase, open source information and data is transformed into open source intelligence (Gibson et al., 2016) [19]. Analysis is intended to achieve three goals: spatial awareness, situational awareness, and some tentative prediction (Gibson 2014) [20]. Some of the methods used for data analysis include lexical analysis, semantic analysis, geospatial analysis, and social media analysis. The intelligence derived from the OSINT process is delivered at the data extraction stage. Opensource intelligence practitioners use tools such as Maltego, Foca, Shodan, and Spiderfoot (Pastor-Galindo et al., 2020) [21]. Unstructured data is information that lacks a predefined data model and thus cannot be processed by conventional computer programs (Qureshi et al., 2011) [22]. OSINT can be used in any domain, including cyber security, forensic analysis, risk assessment, and so on. Unstructured data, false information available on the internet, and legal issues (Hribar et al., 2014) [23] are some of the challenges faced by OSINT. The publicly available information on the web space is fundamentally unorganized. This indicates that OSINT's source data is so diversified that it's tough to categories it. (Bello-Orgaz et al., 2016) [24]. Thanks to big data and cloud computing that are capable of handling such huge data. Open source Intelligence has a number of advantages including being less risky, inexpensive, and easy to obtain (Hassan 2019) [25]. The incorporation of Artificial Intelligence technology into various stages of OSINT improved its accuracy and performance.

## 4. LITERATURE REVIEW :

Since last two decades, a large number of researchers have published papers in the Opensource Intelligence domain. With the advancement of technology, opensource intelligence also evolved. The massive data that has accumulated in the public domain because of the evolution of social media platform has enticed many actors including business corporations, antisocial elements, government agencies, law enforcement agencies etc., to integrate opensource intelligence for their benefit. Thanks to Internet Accessibility that people can now readily find and post any type of information (Edwards et

al., 2017) [26].

Lee & Shon (2016) [27] proposed a new framework for cyber security threat inspection of critical infrastructure based on an Open source intelligence. This framework included four steps: developing an opensource intelligence plan, preparing opensource intelligence, gathering information from open source platforms, and producing security intelligence.

Hayes & Cappa (2018) [28] have demonstrated that OSINT may be used to do risk assessments for the company in order to prevent potential cyber-attacks on its critical infrastructure, which was part of the US electrical grid. A vulnerability assessment and various such open-source intelligence analysis procedures were carried out in order to profile the company's network, applications, devices, and critical IT resources.

Similar method for exploring website vulnerabilities was proposed by Wiradarma & Sasmita (2019) [29]. During the information gathering phases of penetration testing, OSINT tools like Maltego and others are used to get data on the victim a from open source. A system improvement recommendation was created by combining information from OSINT, penetration testing, and the ISO 31000 risk assessment standard.

Vacas et al., (2018) [30] outlined a method for using threat intelligence data acquired from open source intelligence feeds to improve the accuracy and capabilities of intrusion detection systems. The entire process starting from processing of OSINT data to aggregation and correlation of the data to create IoAs are automated. Blacklists and IDS rules are created using these IoAs and then subsequently imported into the IDS. The IDSoSint system was constructed by the author, and the approach was tested with production traffic from a few UL links. The findings suggest that OSINT data can be used to produce novel ways of portraying threat intelligence knowledge and can be used in defense systems.

Unreliable information that may get derived from unstructured data is a biggest challenge of Opensource intelligence. Johnsen & Franke (2019) [31] discussed their research on text preprocessing needs and document formation for the Latent Dirichlet Allocation (LDA) a popular topic model algorithm, in which they propose repeated preprocessing processes, such as removing common terms until the result comprises cohesive and clear subjects. Data cleansing at prepossessing stage would help opensource intelligence to analyses and produce a reliable result.

Herrera-Cubides et al., (2020) [32] conducted a study with an aim to investigate the evolution of production of research and study material in OSINT platform. This analysis looks at two of the material sources of OSINT such as research knowledge distribution databases and repositories pertaining to educational resources. This study provides academics with a roadmap to the current level of OSINT research and teaching, as well as a valuable metadata description in order to make resources more accessible and reusable in the educational ecosystem.

Fleisher (2008) [33] presented a conceptual paper on how the growing popularity of open-source data and information affects competitive and marketing intelligence. This is a descriptive conceptual article that builds arguments from a review of three uncategorized collections of material in competitive and marketing intelligence, processing of intelligence, and analysis of market.

The problems they confront in exploiting this data are described in this article, as well as the effective strategies that certain firms have exhibited in incorporating and integrating open sources in the analysis processes of competitive and marketing intelligence field. It can be seen that the study was conducted from the standpoint of a marketing analyst and the usefulness of intelligence derived from OSINT for the benefit of improving marketing efforts, rather than from the viewpoint of the individual who specialize in collecting the said data.

Magalhães & Magalhães (2019) [73] suggested TExtractor, an OSINT tool that will make gathering details concerning cyber threats easier. TExtractor is a tool that extracts text from video/audio in public sources and searches for keywords associated with harmful actors' activities. The findings are provided in the study, and they reveal that a tool such TExtractor can discover allusions to cyberattacks on audio/video sources in real time with such an accuracy of 60% to 70%.

TExtractor could also be used to keep track of a brand or automate the clipping process, which involves finding brand or product references in audio or video channels.

Kanta et al., (2020) [74] studied the possibility for Open Source Intelligence (OSINT) to be used for greater effective password cracking is investigated in this paper. A detailed review of the literature on strong passwords, cracking of password, and OSINT is presented, as well as the legal issues that these topics raise. A study of password complexity as well as demographic characteristics that influence

password choosing is also offered. Finally, the impact of OSINT by a law enforcement on password cracking is explored.

Kang (2020) [75] in order to quantify cyber threats, the authors offer the assessment variables for cyber threats among cyber-attack databases and analyze the priority of those elements. As evaluation variables for cyber threats, he choose the objective of the assault, attack type, target, convenience of attack, attack durability, frequency of OSINT database, and elements of the lowest layer of each component. The priority of each element is assessed only by using the analytic hierarchy process after it has been chosen.

### 4.1 Stages Of Open Source Intelligence:

In order to provide useful intelligence about a target, opensource intelligence adheres to a well-defined and precise methodology. The CIA's intelligence cycle and the book of Intelligence Studies describe this process in slightly different ways, but both have collection, processing, analysis, production, and dissemination in common, with the latter adding classification and the former adding planning and direction as additional steps. In this paper as shown in Fig.1 below four major steps, collection, processing, exploitation, and production, have been considered and explored for the purpose of a literature review (Williams 2018) [34].

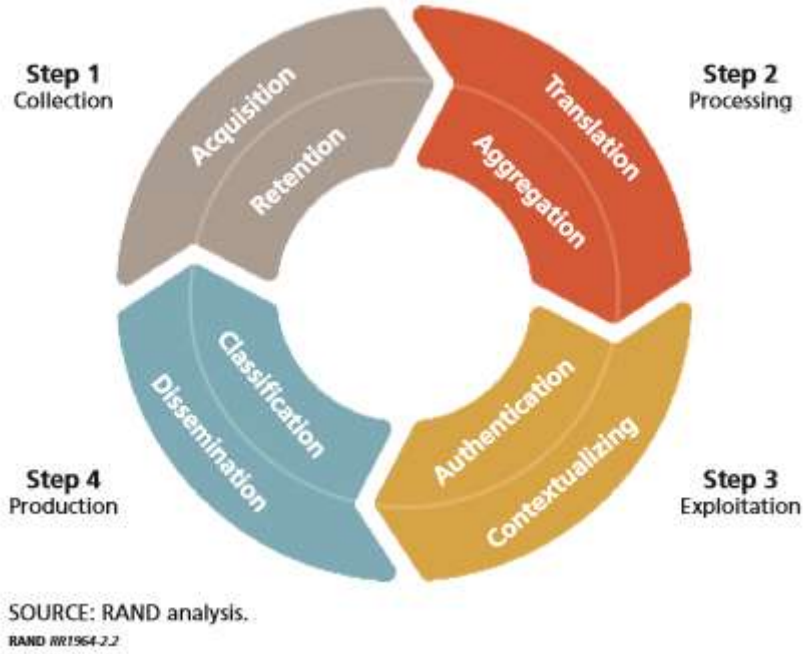


**Fig. 1:** The OSINT Operations Cycle (Williams 2018) [34]

#### 4.1.1 Collection Of Data

Data is an important asset for carrying out any kind of intelligence activity. Just like any other intelligence method, OSINT heavily depends on the data, which are extracted from publicly available sources. In this phase one has to identify the source of data and type of data to be collected. (Gibson 2016) [35] presented some of the strategies that a practitioner might utilize to get information from public sources, as well as the rationale for doing so. In this chapter, the author focuses on the collection of publicly available data that is consistent with the intelligence-led strategy since he feels that automated data gathering methods can deliver the most benefit at this moment. However, the authors assert that for the sole investigator, automated approaches-can accelerate their manual search that would be extremely beneficial to them, and will work to emphasize areas where manual procedures can transform to automated ones. The data types and data sources used for OSINT are also discussed in this study. Structured data are data stored in any relational database, and unstructured data are web pages, journals, pictures, video, and audio. Both these data forms are addressed in detail.

Quick et al., (2016) [36] proposed a data volume reduction approach that focuses on imaging a selection

of important files and data, including the system registry, data files, sheets, email, browser history, conversations, logs, photographs, videos, and other essential file types. When applied to test scenarios, the actual media quantity was reduced by a factor of a hundred. The Digital Forensic Data Reduction procedure can be used as a screening tool to quickly understand data and determine the media or machine may contain possible evidence that should be investigated first. If the essential information is identified during the first examination of a subset, the data reduction method may eliminate the need for full analysis.

Text mining is a technique for extracting useful information, wisdom, or patterns from unstructured data from various sources. It turns unstructured data's words and phrases into numerical values that may be mapped to structured data in a database and examined using old data mining techniques [37]. The effectiveness of some of the open source tokenization technologies is examined in this research report. As a result of this research, some tools have only read text documents and taken into account a certain number of characters. When all of the factors are considered, Nlpdotnet tokenizer produces the greatest results when compared to other tools.

### 4.1.2 Processing Of Data

The processing step primarily deals with verifying and removing noises from the raw data received during the data collection phase in order to make it usable for analysis. Filtering out irrelevant data, translating texts from another language into English, converting photos, audio, and video files into useful data, and so on are all tasks carried out during the processing phase. The massive amount of data acquired from open source makes it difficult to interpret and draw useful insights, necessitating increased processing power, such as cloud storage and big data computing capabilities. Ji et al., (2012) [38] outlined a methodical flow of research on big data processing in the field of cloud computing and addressed significant concerns such as cloud storage and computing architecture, major parallel processing frameworks, key applications, and MapReduce optimization. Among the most prominent parallel processing models, such as MPI, General Purpose GPU (GPGPU), MapReduce, and MapReduce-like, a work has studied MapReduce to learn how to improve its performance when processing large amounts of data. It is also covered how to employ algorithms and parallelization techniques to improve scalability and performance when processing big data.

Milne & Witten (2012) [39] presented a multilingual, highly effective toolkit for mining Wikipedia's large amount of semantic knowledge. This open source toolkit enables developers and researchers to utilize Wikipedia's massive information asset into their projects. In addition to that, it generates database containing condensed versions of Wikipedia's structure and content, as well as a Java API for accessing them. Pages, categories, and redirects on Wikipedia are represented as classes, which may be searched, viewed, and repeated over quickly. Parallelized processing of Wikipedia dumps, XML-based web services, annotation features and semantic related measures is developed using machine learning which are among the advanced features. Wikipedia Miner is meant to be a place where data mining skills can be shared.

Gong et al., (2018) [40] proposed a new model for analyzing data reliability and validity, which uses the comparative analysis of the cyber threat intelligence data and offers a set of criteria for assessing the reliability of feeds that provide cyber threat intelligence data. A new model has been proposed to analyze the data reliability using Opensource intelligence Cyber threat intelligence feeds.

### 4.1.3 Exploitation Of Data

Exploitation also known as analysis phase is in charge of determining whether the material processed in the preceding phase is what it claims to be and how valuable it is to the Intelligence community. Exploitation phase comprise of three steps such as authentication, credibility evaluation and contextualization. Verification of authenticity and credibility of information plays a key role in developing a trustworthy knowledge. Contextualizing entail assembling several open source information pieces from any source into an output that gives a comprehensive understanding of a subject (Williams 2018) [34]. Most commonly used methods for analysis are lexical analysis, semantic analysis, geospatial analysis, and social media analysis.

**(1) Lexical analysis**

Lexical Analysis is a program that collects and analyses enormous amounts of text from the internet. Identifying frequently searched phrases on Google is straightforward application of lexical analysis. More advanced systems try to deduce information about persons who use social media, such as demographic factors.

Baldini (2007) [41] discussed about the Multilingual Text Mining platform for Open Source Intelligence which was chosen by The Joint Intelligence and EW Training Centre to equip armed forces and civilian employees of the Italian Defense in the OSINT discipline. Multilanguage lexical analysis allows for the automatic indexing, easy navigation, and categorization of papers, regardless of their language or the source from which they were gathered. This method allows intelligence analysts to search, analyze, and classify large amounts of heterogeneous material, assisting them in cutting through the information maze.

Denecke (2008) [42] presented two main methodologies: Multilingual Sentiment based on SentiWordNet and Sentiment Analysis focusing on SentiWordNet (a lexical resource for sentiment analysis). The first demonstrates that when deployed to a multilingual scenario, Opinion Mining poses unique issues. This strategy is inadequate. Indeed, the authors state that statistical techniques necessitate training courses that are typically distributed in multiple languages or are even absent. Lexical approaches, in fact, entail language-specific lexical and linguistic resources. Creating these resources takes a long time and often requires manual labour. The latter method is based on SentiWordNet. This paper proposes a method for assessing text polarity in a multilingual framework. The lexical resources that are available in English language are utilized in this method in order to conduct sentiment analysis. First, regular translation software is used to convert a text written in a language other than English into English. The translated material is then sorted into one of two categories based on its sentiment: "positive" or "negative." A document is scanned for sentiment-bearing terms such as adjectives in order to classify sentiment. SentiWordNet is used to calculate positivity and negative ratings for these words. The polarity of the paper is then determined by an interpretation of the ratings. The author has put this strategy to the test using reviews from Amazon for a German film and has found it to be effective. The method is contrasted to a statistical polarity classifier based on n-grams and evaluated using German movie reviews collected from Amazon. According to the finding working with the combination of sentiment analysis methodologies that already exists and standard technology is a convenient approach to sentiment analysis in a multilingual context.

**(2) Semantic analysis**

Semantics is a subset of linguistics that studies the meaning of language which is concerned with the meaning of words and phrases. Within the context of natural language processing, semantic analysis evaluates and reflects human language and analyses words written in English as well as other natural languages with human-like interpretations.

Golestan et al., (2015) [43] shown an information fusion approach that can combine data from human-based sources with information from physical sensors. This model is built on the same author's prior work, which was a fuzzy extension to the Multi-Entity Bayesian Network language and semantic analysis.

(Hassan et al., 2016) [44] The lack of comprehensive semantic information for optimal semantic understanding was addressed in this paper. The method entails first determining the relevance of software requirement representations, followed by determining the impact of semantic analysis. This was achieved by employing a semantic structure that was created expressly to analyze and disambiguate the data. A strategy was given to handle the challenge of semantic information being unavailable for enhanced semantic analysis. The system's architecture is built on semantic technology, which may be incorporated into the documentation and execution of the program. By adapting current necessary ecosystem concepts and using them to design conducting experiments and procedures that benefit in knowledge management to the software platform, the research review indicates that intelligence gathered from already available software essential documents and knowledge derived from existing applications are combined in the used framework, which is centered on semantic technology. The provided technique demonstrates that extant Ontologies may be adapted and combined to assist knowledge management, as well as designing systems and conducting tests on real-world software requirements.

Wang (2017) [45] presented an approach for analyzing semantic content extraction methods in depth. A new algorithm for website de-duplication which makes use of tf-idf and word vector distance has been proposed by the author. A model for generating a semantic cloud is also proposed in this thesis.

Sleimi et al., (2018) [46] offered a semantic legal metadata technique that allows the user to know and grasp legal clauses. The discovery of consistent legal requirements requires the use of metadata. There is a scarcity of information on how to assess the integrity of metadata which is essential for formal

specifications analysis. There is a paucity of literature on ascertaining metadata consistency for formal specifications analysis. Besides that, semantic legal metadata extraction process automation capability itself is underutilized. It doesn't make complete use of the Natural Language Processing.

Gupta et al., (2017) [47] proposed a methodology for constructing an Intelligent Querying Framework that allows the end users to build their own initial questionnaire. The system includes a module that translates English phrases into SQL-like queries that may be used to respond to customer requests. As a result, it reduces the quantity of efforts required by making their research easier.

**(3) Geospatial analysis**

In environmental research, geospatial analysis refers to the use of geographic data to find ecologically important information that is both geographically and temporally referenced. Environmental threats identification, tracking pollutant spread over time, study of patterns in environmental factors such as temperature and acidification of the sea over time, and associating different environmental characteristics with locations are some of the basic functions of Geospatial analysis. Geospatial analysis, in particular, is the transformation of data depending on geography. Geographic information systems, remote sensing, global positioning systems, metadata, remote sensing and georeferencing are some of the technologies used in this type of analysis. Geospatial Analytics is used widely in assessing Weather related risks, urban planning and development, during covert operations, while exploring natural resources etc. The data being used for Geospatial analysis comes from variety of sources such as images uploaded to the social media, mobile device data, and details from GPS, information from location sensors which is used to build a meaningful intelligence.

Thakur et al., (2015) [48] have presented Planet Sense, a revolutionary distributed processing system that offers end-to-end capabilities for geospatial intelligence, from collecting raw data to delivering actionable insights in real-time. This design is expandable to allow for progressive information gathering and integration with diverse and aging data sources. The platform consists of four key components, such as GeoData Cloud – an architecture for hoarding and managing diverse datasets, Real-time streaming data harvesting method, Data and analytics platform and demonstration and display via website and RESTful APIs.

Yue et al., (2013) [49] using geospatial semantics and services, this study presents a workflow-based technique for discovering complicated geospatial features. The basic features derived from pictures are stored in Web Feature Services that are accessible through a catalogue service. The findings reveal that complex feature placement can be determined by locating one of its constituent features that follows precise spatial correlations with other features. The workflow method aids in the formalization of computing procedures for the discovery of various types of complicated features. Transparency, dynamics, and interoperability are all advantages of using service technologies.

Triglav et al., (2010) [72] used spatio-temporal evaluation matrices and index of spatio-temporal anticipations matrices is presented in this study as a tool for evaluating spatio-temporal quality. Geospatial data producers can utilize these two easy tools to methodically categorize and show the precision of their spatio-temporal data, and consumers can use business intelligence concepts and a Web 2.0 strategy to convey their needs in the same way. The study presents the basic principles and various instances, as well as some prospective future applied research endeavors.

**(4) Social media analysis**

The practice of gathering the most essential information from people's social media platforms and deriving practical conclusions is known as social media analysis. The information being analyzed comes from people's previous postings, conversations with their followers, and earlier social media initiatives, among other things. The purpose of social media analysis is to obtain a valuable information of individual attitudes and preferences. The majority of users use social media to convey their emotions such as happiness, anger, agreement, disagreement, and annoyance etc. through text messages or postings. When individuals mention or talk about a business or product on social media, sentiment analysis method can be used to determine the feelings or sentiment behind the phrases they use. The detailed analysis of the word's individuals uses to express themselves about a scenario, event, product, brand, company, or other topic will provide public opinion on the subject under consideration. Organizations can utilize social media analysis to uncover commonalities in consumer preferences and complaints, as well as online talk about a certain individual, business, or event, if they have the correct tool.

D'Avanzo& Pilato (2015) [51] used a practical methodology for addressing disconnect between

customers' expectations and product/user reviews. Author used a popular collaborative learning model to achieve this, simulating a scenario wherein two or more individuals' study or make an effort to learn together something. In this way, online shoppers took advantage between one another's capabilities and talents by using the Bayesian social sentiment analysis tool proposed and collectively soliciting one another's opinions in order to make their purchases. The new technique collects user feedback from specific types of markets and visually summarizes it in order to relieve customer overload and speed up their purchasing experience. This strategy was used on the Facebook sites of mobile phones and fashion retailers.

T.K et al., (2021) [52] present a thorough overview of different applications of Social Media analysis utilizing powerful machine learning techniques in this study. The author began by providing a summary of machine learning algorithms that are employed in social media analysis following a thorough overview of machine learning methods to social media analysis. It also discusses the obstacles and benefits of using Machine Learning in social media analysis. Finally, they discussed unresolved difficulties and implications in social media analysis in preparation for future research.

#### 4.1.4 Production - Extracting of Knowledge

The final stage of open source intelligence is the delivery of a meaningful intelligence report to the consumers. Because the report will be comprehensive and of high priority, it can be shared directly with the judiciary, law enforcement agencies, and other relevant parties. An OSINT product's classification level is also assigned during the production stage. The specifics of gathering, analyzing, and exploiting the data may necessitate a higher classification level. Distribution is an important part of the production stage. The most common way to share open-source analysis is through a formal report. Products, on the other hand, can be in the form of verbal instructions or visual representations. Al-khateeb & Agarwal 2019 [53] used many community edition tools, such as the Maltego tool, provide data gathering and graphing features that make reviewing the data easier. Maltego is one of the few helpful applications that can gather data from numerous sources and present it in a convenient style. Some of the other such tools are Gephi, Spider foot, Lampyre, etc.

### 4.2 Tools Used for Open Source Intelligence:

The OSINT tools are used to obtain intelligence about their potential target during the research. When an analyst uses the correct OSNIT tool, he or she may give a more accurate intelligence report. The OSINT technologies employ artificial intelligence to locate confidential data on the internet. The OSNIT tools fulfil three functions, however each one focuses on a different aspect. Firstly, locating assets that are visible to the public, then collecting sensitive details from outside the organization and finally turning it into meaningful intelligence. OSINT tools are classified as those, that query numerous search engines at the same time like social media search engines, domain and person search engines, and so on, and those designed for big data analytics platforms. Chauhan & Panda (2015) [54] discussed some of the automated tools and online applications that are often used by experts in many intelligence-related sectors particularly information security to conduct investigation. The author has covered every aspect of the OSINT tools, from deployment to comprehending their interface, as well as their functioning and use. Some of the programs evaluated have a graphical user interface, while others are command line only. The author has discussed the tools listed below.

**(1) Creepy**

Creepy is a Python program that uses EXIF information in photographs to retrieve geolocation and display it on a map.

**(2) TheHarvester**

An open source intelligence application capable of extracting valuable information from public sources such as a person's name, e-mail addresses, ports that have been kept open, computer and network device banners, organization subdomains, and other important information.

**(3) Shodan**

A search engine that allows users to do device searches over the internet and offers filters to assist users narrow down their results.

**(4) Search Diggity**

A tool that offers a vast database of queries for several search engines that can be used to obtain incriminating information about the target and has a wide range of options.

**(5) Recon-ng**

This framework assists all OSINT fanatics in practicing different phases of reconnaissance in an automated manner. It primarily focuses on web based open source reconnaissance and offers its consumers distinct transforming modules to execute deep and swiftly reconnaissance.

### 4.3 Present Status of Opensource Intelligence System:

Working with OSINT would have been challenging due to the massive amount of data available unless tools supported by Artificial Intelligence were not available. Machine reasoning, automatic speech recognition (ASR) or speech to text, Machine perception, and translators are just a few of the activities supported by AI in the field of OSINT. Machine Learning, Pattern Recognition and Natural Language Processing are AI subfields that play a significant role in OSINT activity. We can see some of the recent work done in the field of Opensource Intelligence with the help of sub-areas of artificial intelligence.

Evangelista et al., (2020) [55] in their paper, author conducted a comprehensive Studies In order to look into the use of Open Source Intelligence in conjunction to Artificial Intelligence in this research. This research reveals significant patterns in the application of OSINT to AI. The distribution of OSINT publications featuring AI was analysed. It began in 2015, with the recommendation of using OSINT in conjunction with AI. Following that, I gathered information from books and articles available on the Internet. OSINT and AI were first applied in the domains of languages and translations, military applications, and social media, with promising results, before reaching its pinnacle: cybersecurity. Security-related OSINT papers account for 41% of all AI-related OSINT papers, or nearly half of all articles. When we consider that the expansion of these media began to pick up momentum in 2016, this appears to be an attractive value. In recent years, there has been a trend toward the publication of Opensource intelligence linked with artificial intelligence for the subject of cybersecurity, which has the highest share of applications.

Sagnika et al., (2020) [56] provided a thorough examination of sentiment analysis methods used on languages other than English. The tools used, the benefits and drawbacks of each method, and their efficiency are all discussed. The associated difficulties are also discussed. The paper discusses both methods for analyzing translated data and methods for analyzing data available in the target language. Lexicon-based approaches and Machine Learning techniques were the two main techniques used in this paper. The available research in multilingual sentiment analysis was examined in this study, and the key languages that have been addressed or for which a corpus has been generated were identified, as well as the techniques utilized and their contributions, as well as their accuracy rates.

Big Data presents significant potential and difficulties for information focused activities in both the public and corporate sectors. As the online data ecosystem has evolved, software-based techniques have become a critical component of the OSINT effort. Analysts might simply be unable to manage with the avalanche of information circulating about the web without the assistance of automated systems. However, while OSINT aficionados must engage with algorithmic solutions, it is critical that they do so without being overloaded. The trained analyst possesses a level of competence, knowledge, and discretion that cannot be codified. We must combine the expertise of analysts and algorithms in order to effectively leverage open-source data, while maintaining the distinction between them (Eldridge et al., 2017) [71].

### 4.4 Applications Of Osint in Cyber Security:

Because of the consequences of cybercrime, intelligence department along with law enforcement team all across the globe are working to combat cyber threats. All industries are grappling with the same problem of how to best combat cybercrime and successfully promote security to individuals and companies. Mining public records to develop a full profile of specific targets to derive unique and high-value intelligence is quickly becoming a valuable tool for intelligence agencies. Countering cybercrime is becoming increasingly reliant on innovative software tools and strategies to gather and handle data in an effective and efficient way as the volume of open-sources available grows. The existing initiatives to use open source information for cyber-criminal investigations were examined in this chapter, and an integrated OSINT Cybercrime Investigation Framework was developed (Tabatabaei& Wells 2016)

[57].

Layton et al., (2013) [58] studied the investigation of the use of authorship analysis to determine when malicious profiles were created by the same person. The approach achieved a base accuracy of 0.8400 in the proof of concept, and 0.9050 accuracies when given a threshold value utilizing the second match for successful predictions. By employing this threshold number, the author was able to demonstrate that wrong guesses do not appear to be right. The current technique revealed 11 definite matches between 132 potential malicious Twitter accounts in this application to malicious profiles. There were also nine possible matches for which the expert lacked sufficient evidence in one direction or the other. This finding might not have been feasible without the use of an automated method, and it demonstrates that automated open source intelligence has a lot of promise for linking profiles together indirectly. The author intends to improve on these findings by employing ensembles to develop more robust authorship algorithms.

Yeboah-Ofori (2017) [59] systematically reviewed uncovered amazing facts and contradictory thoughts. It also exposed the underlying research issues that have an impact on Open Source Intelligence. Because of the indestructibility of social media technologies, they are used for business, social and intelligence gathering purposes. However, in order to ensure effective and advanced mitigating circumstances, more research is needed to gain situational awareness, respond to threats, and establish adequate defensive measures.

Shere (2020) [60] published their research results to provide a comprehensive overview of OSINT professionals' perspectives on the GDPR's impact on their competencies in the United Kingdom one year after it was introduced. According to this report, the GDPR has coincided with a shift in social networking site behaviors as well as the availability of OSINT tools. The threat of legal action by social media companies against OSINT tools used to access their social media channels has exacerbated the suffocation of OSINT abilities.

Quick et al., (2018) [61] with a large amount of data obtainable through forensic assessment, there is a great deal of data that may be enhanced with publicly available data to allow for greater knowledge of events or people, as well as better decision-making opportunities. The expansion of the media landscape, and the time it takes to conduct searches and evaluate information, is one challenge affecting the quick and timely processing of vast collection of forensic information. In the suggested process for open source analysis forensic, data from a wide range of systems and data storage is supplemented with OSINT. Methodology used to reduce the volume of forensic data and the analysis intelligence that deals with forensic activities plays a major role in building a framework to process OsINT and forensic intelligence. The topic of external source data was discussed in this chapter. The method allows for the merging of open source with closed and confidential data sources. The swift processing of diverse case data using semi-automated data mining tools and external source gathering software improved the data mining and intelligence capabilities of digital forensic data holdings.

Casanovas (2016) [62] in his EU project CAPER created an OSINT tool to tackle organized crime. This chapter shows how to incorporate the EU's General Data Reform Package and ethical concerns into information security and surveillance platforms. This study also discusses an analytical proposal for a Meta-rule of law, which is completely compatible with the idea of establishing a Global Ethics to deal with cybercrime, cyberterrorism, and eventually cyberwarfare.

Quick et al., (2018) [63] have proposed a framework for better digital forensic data analysis. This entails reducing the amount of data to only what is required to carry out the analysis's goal. Semi-automated computing of the data to discover entity and related information, as well as automated searching of the entity information with other data sources, including OSINT resources, to increase the database's value.

González-Granadillo et al., (2021) [64] described Enriched threat intelligence. ETIP is a threat intelligence platform that enhances existing threat intelligence platforms by adding extended import, quality evaluation processes, and information sharing capabilities. The study also includes a comparison of ETIP's components to aggregated IoCs and single IoCs that build collated ones, as well as an evaluation of ETIP in a real-world use-case scenario. Derbyshire et al., (2021) [65] reviewed of current risk assessment material and industrial practice reveals that adversaries' understanding of cost is a significant gap. This paper establishes the associated costs witnessed by an adversary as Time, Finance, and Risk, which is backed by a functional study conducted with appropriate security professionals. A methodology is proposed and constructed based on these parameters to facilitate the probabilistic assessment of an adversary's cost. The paper also uses a case study to demonstrate this paradigm, which

is a significant addition to existing cyber security risk assessments. A small group of users looked into threats in greater depth using one of two methods: threat intelligence or breaking the opponent down into its constituent parts. Threat intelligence will provide more detailed information on which enemies are likely to target a particular client. This data is gathered from a variety of sources throughout the participants, including open source intelligence (OSINT), government documents, and even privately run cyber security operation centers.

Martinez Monterrubio et al., (2021) [66] conducted a study in the field of OSINT and MedOSINT. The goal of this study is to design and prototype a tool for doing open source intelligence (OSINT), especially on official medical bulletins, in order to detect false news. MedOSINT is a modular system that may be configured to process data from a variety of medical official bulletins. Intelligence is generated for decision-making from the analyzed data, confirming the accuracy of the COVID-19 news. When evaluating official bulletins, the tool is compared to other possibilities, and it is proven that MedOSINT outperforms the present options. It is also supplemented by a Case-Based Reasoning (CBR) system, which provides an expert explanation. This has shown to be an excellent complement since it can locate explanatory cases for justification by example.

Lande & Shnurko-Tabakova (2019) [67] made a study into the establishment of fundamental and applied concepts for assessing information flows across global computer networks while performing open source intelligence is represented in this paper. The parameters of today's informational environment and available theoretical and technological remedies are demonstrated particularly in terms of cyber-security, which reveals the importance of this problem. Around the world, software and technology solutions for OSINT are now being developed and applied.

Mittal et al., (2016) [68] explained the CyberTwitter architecture, which provides end users with cybersecurity intelligence warnings based on publicly available Twitter data. The authors use the Security Vulnerability Concept Extractor (SVCE) to extract terms linked to security vulnerabilities. The extracted intelligence will be stored as Resource Description Framework triples in a cyber security knowledge base. On the basis of a ‘user system profile,' SWRL rules will be used to generate alerts for security analysts. User system profile contains information about the operating system installed, other software installed, and version numbers, among other things. To assess temporal cybersecurity events and ensure that the issued warnings are current and relevant, the author constructs an "intelligence" classification. The user can then use these alerts to keep the organization's system up to date and safe.

Ziolkowska, A. (2018) [69] discusses how military intelligence may benefit from opensource intelligence, thereby safeguarding citizens' lives and the country's security. Military intelligence relies heavily on the armed forces' social support and operations. Open Source Intelligence can be utilized to gain these kinds of insights. Nowadays, it is critical to make appropriate use of open source data (OSINT). Much of the data provided could be useful information for intelligence services after adequate processing, verification, and analysis. As a result, public information is included in military reconnaissance. International organizations, such as the NATO Alliance and the European Union, which have adequate intelligence divisions, also use such information searches. The information available on the open internet, such as diplomatic negotiations and geopolitical plans, is critical for operational activities, which can be obtained through technologies like OSINT. Internal, financial, environmental, scientific, and technological policies, as well as demographic challenges, are all included in this area.

Hernandez et al., (2018) [70] found from a study of several OSINT technologies and how they can be applied to a nation's cyber intelligence activities. The authors offer a set of changes tailored to the Colombian context that were applied and donated to the society, enabling law enforcement authorities to construct data gathering processes using Colombian open sources. The true value of the data is provided by the implementation of three machine learning models that do sentiment analysis on it in order to determine the adversary's perspective on a given topic, understand their motive, and, as a result, design effective cyber defense plans. The sentiment analysis approaches used for text processing are explained in detail, including keyword location, lexical affinity, statistical methodologies, and concept level.

Lee et al., (2020) [50] proposed a method for measuring cyber threats and recommend a cyber threat prediction model depending on artificial neural networks in this study. Cyber-attacks have become increasingly sophisticated in recent years. One of the most effective countermeasures to this advanced cyber threat is to foresee cyber-attacks ahead of time. Predicting cyber threats necessitates a significant

amount of information and effort. People can easily determine cyber threats if they use Open Source Intelligence, which is at the heart of recent information acquisition. To indicate cyber threats using OSINT, a database for cyber-attacks from opensource intelligence must be created, and elements that can assess cyber threats from the formed DB must be selected. Based on past research, the author constructed a cyber-attack database using data mining and analyzed the significance of core factors among cumulative DG factors using the AHP technique.

### 4.5 Summary of Related Work:

**Table1:** Summary of findings from 2007-2021 presented by various authors.

| Sl. No. | Authors | Year | Inventions/Findings/Results |
|---|---|---|---|
| 1 | Baldini [41] | 2007 | Discusses a Multilingual Text Mining system for Open Source Intelligence, which will be used to educate military and civilian employees in the OSINT discipline in Italy. |
| 2 | Fleisher [33] | 2008 | A detailed, conceptual work that draws on and develops ideas from three uncategorized areas of literature such as intelligence related to marketing, competitive and processing intelligence to establish an argument. |
| 3 | Denecke [42] | 2008 | The approach makes use of English lexical resources for sentiment analysis. |
| 4 | Triglav et al. [72] | 2010 | The usage of spatio-temporal assessment matrices is presented as a tool for evaluating spatio-temporal quality. |
| 5 | Ji et al. [38] | 2012 | Several techniques for enormous amounts of data to be processed, both from a system and application standpoint are studied here. |
| 6 | Milne & Witten [39] | 2012 | Wikipedia Miner is a toolkit that builds databases with summarized versions of Wikipedia's content, as well as a Java API for accessing them. Wikipedia Miner is meant to be a place where data mining skills can be shared. |
| 7 | Layton et al. [58] | 2013 | Explores the concept of associating online accounts automatically for open source information gathering. |
| 8 | Yue et al. [49] | 2013 | Uses geospatial semantics and services to present a work flow-based approach for discovering complex geographical features. |
| 9 | Golestan et al. [43] | 2015 | Fuzzy Multi-Entity Bayesian Networks Fuzzy-MEBN and semantic analysis are used in a novel technique for soft data association. To find its related entity class, context, and state, the soft data is evaluated using different semantic analysis algorithms. |
| 10 | Thakur et al. [48] | 2015 | PlanetSense is a unique distributed processing system that offers end-to-end capabilities for geospatial intelligence, from gathering raw data to delivering actionable insights in real time. |

| | | | |
|---|---|---|---|
| 11 | Chauhan & Panda [54] | 2015 | Professionals from many intelligence-related sectors, particularly information security, routinely employ automated technologies and web-based services to conduct reconnaissance. |
| 12 | D'Avanzo & Pilato [51] | 2015 | Introduces a cognitive-based process for extracting users' opinions from specific types of markets and visually summarizing them to reduce buyer overload. |
| 13 | Lee & Shon [27] | 2016 | The recommended framework's application is discussed. The framework could be used to complement and leverage earlier threat detection methods as well as inspect cyber threats to critical infrastructure. It has been debated how to complement signature-based threat detection methods and how to effectively use anomaly detection methods. |
| 14 | Tabatabaei& Wells [57] | 2016 | The present efforts to use open source data for cyber-criminal investigations are discussed in this paper, as well as the development of an integrative OSINT Cybercrime Investigation Framework. |
| 15 | Casanovas [62] | 2016 | Discussing how to incorporate the legal and ethical concerns posed by the European Union's General Data Reform Package into security and surveillance platforms The CAPER Workflow, PbD Strategies, and PbD and Security are all explored. |
| 16 | Gibson [35] | 2016 | Explains how an investigator might collect data from free sources and how to convert that data into useable formats for future study. When accessing, examining, and exploiting open-source data, it is important to follow some privacy, legal, and ethical best practices. |
| 17 | Quick et al. [63] | 2016 | This research presented a methodology for improved digital forensic data analysis. This comprises lowering the volume of data to the minimum required for analysis. |
| 18 | Quick et al. [36] | 2016 | The suggested Digital Forensic Data Reduction technique detailed in this study helps to significantly reduce forensic analysis time and storage requirements. |
| 19 | Vijayarani, S., & Janani, R. [37] | 2016 | A comparative analysis of the performance of the seven open-source tokenization technologies is conducted. |
| 20 | Hassan et al. [44] | 2016 | An approach was presented to address the problem of non-availability of semantic information required for better semantic analysis |

| | | | |
|---|---|---|---|
| 21 | Mittal et al. [68] | 2016 | The author created the 'intelligence' ontology to analyze temporal cybersecurity incidents and to ensure that the produced notifications are fresh and accurate. The user can then use these notifications to maintain the system up to date and secure for the organization. |
| 22 | Yeboah-Ofori [59] | 2017 | A detailed evaluation and synthesis of findings from existing empirical research on cyber intelligence and OSINT profiling in order to identify and mitigate risks and vulnerabilities on online social networks. |
| 23 | Eldridge et al. [71] | 2017 | The paper will pave the way for future research in this field, as there is a lot of need for more research, especially empirical research. Around OSINT process models in various industries and organizational settings with diverse limitations |
| 24 | Edwards et al. [26] | 2017 | Offered a set of safeguards, including an automated social engineering vulnerability scanner that businesses may use to assess their risk of social engineering assaults based on open source intelligence. |
| 25 | Wang [45] | 2017 | TF-IDF and word vector distance are used to suggest a new webpage de-duplication algorithm. |
| 26 | Gupta et al. [47] | 2017 | Method for creating an Intelligent Querying System IQS that allows a user to fire inquiries in his or her own natural language. |
| 27 | Gong et al. [40] | 2018 | The first instance to work with OSINT CTI data reliability analysis. For optimal usage of CTI systems, this suggested model includes data reliability and validity criteria. |
| 28 | Quick et al. [61] | 2018 | The process of Data Reduction by Selective Imaging is investigated, as well as Quick Analysis and the DFINT+OSINT framework. |
| 29 | Hayes & Cappa [28] | 2018 | The usefulness of OSINT technologies in conducting risk assessments to avert cyber assaults is highlighted in this article. |
| 30 | Williams [34] | 2018 | The project is aimed at intelligence professionals who want to learn more about open-source analysis and tools. |
| 31 | Vacas et al. [30] | 2018 | Covers the entire process from the collecting of OSINT data feeds through the implementation of new rules and blacklists. The technique was tested with 49 OSINT feeds and production traffic as part of the IDSoSint system. |
| 32 | Sleimi et al. [46] | 2018 | The authors offer a standardized conceptual model for semantic meta data types relevant to legal requirements analysis, as well as automated extraction procedures based on NLP for various metadata types. |

| | | | |
|---|---|---|---|
| 33 | ZIÓŁKOWSKA [69] | 2018 | Various aspects of open source intelligence in the military field has been discussed |
| 34 | Hernandez et al. [70] | 2018 | Various OSINT tools, methods, sentiment analysis techniques etc. have been reviewed |
| 35 | Johnsen & Franke [31] | 2019 | According to research, automated algorithms like LDA must adhere to a set of guidelines in order to minimize vocabulary size and increase quality. It suggests a series of preprocessing actions that should be repeated. |
| 36 | Wiradarma & Sasmita [29] | 2019 | The author explains how the ISO 31000 framework was used to conduct an IT risk assessment based on the findings of penetration testing with the OWASP methodology. Finding the finest and most successful technique to create IT risk management guidelines using the OWASP and ISO 31000 frameworks is the significance and value of this research. |
| 37 | Magalhães & Magalhães [73] | 2019 | The author suggests an OSINT tool TExtractor to make gathering information about cyber risks easier. TExtractor is a tool that extracts text from video/audio in open sources and searches for keywords associated with harmful actors' activities. |
| 38 | Al-khateeb & Agarwal [53] | 2019 | SCF is a term that has been defined. We present Maltego, a tool that may be used to execute SCF. Two case studies were featured that used the aforementioned approaches. |
| 39 | Lande & Shnurko-Tabakova [67] | 2019 | The findings of a study into the establishment of fundamental and applied concepts for assessing information flows across global computer networks while conducting open source intelligence are presented in this paper. |
| 40 | Evangelista et al. [55] | 2020 | To look into the applications of OSINT with AI, a literature review was conducted. Analyzing the 244 publications discovered reveals which are the most OSINT-related article bases. |
| 41 | Shere [60] | 2020 | The Paper discusses how new laws have and public perceptions of the UK surveillance state influenced OSINT investigations? |
| 42 | Herrera-Cubides et al. [32] | 2020 | Examines how the creation of research and educational materials in OSINT has changed throughout time. This study provides academics with a roadmap to the current level of OSINT research and teaching, as well as a valuable metadata description to make resources more accessible and reusable in the educational setting. |

| | | | |
|---|---|---|---|
| 43 | Sagnika et al. [56] | 2020 | a thorough examination of sentiment analysis methods used in languages other than English The tools employed, their benefits and drawbacks, and the effectiveness of all techniques, as well as the issues they pose |
| 44 | Kanta et al. [74] | 2020 | A detailed review of the literature on strong passwords, password cracking, and OSINT is provided, as is the law enforcement challenges associated with these topics. |
| 45 | Kang [75] | 2020 | To quantify cyber threats, the authors offer cyber threat assessment variables from cyber-attack databases and analyze the priority of those elements. |
| 46 | Lee et al. [50] | 2020 | A cyber threat prediction model based on artificial neural networks was used. |
| 47 | González-Granadillo et al. [64] | 2021 | In comparison to current TIPs, this study shows ETIP with expanded capabilities in terms of visualization, import, quality assessment processes and information exchange. |
| 48 | T.K et al. [52] | 2021 | Multiple applications of social media analysis utilizing machine learning algorithms are examined in this survey. Machine learning algorithms for social media analysis have been discussed. |
| 49 | Derbyshire et al. [65] | 2021 | A methodology is proposed and built to aid in the probabilistic estimation of an adversary's cost. |
| 50 | Martinez Monterrubio et al. [66] | 2021 | About Designing and prototyping a tool to perform intelligence on open sources, intelligence is generated for decision making, and the veracity of the COVID-19 news is validated. |

## 5. DISCUSSION & FUTURE WORK :

This review paper provides an overview of open source intelligence and its use in cyber security. Each stage of open source intelligence is described using some of the most extensively utilized methodologies and tools available on the market. With the use of the appropriate technology, skillset, and analytical competence of the user, it has been demonstrated that Opensource intelligence is capable of forecasting events well in advance of their occurrence. It can be utilized not only for early prediction but it may also be utilized for root cause analysis or forensic investigation of an incident such as a civil disturbance that has already occurred. OSINT has been utilized in a wide range of everyday applications, including risk assessment, sentiment analysis, marketing campaigns, social media analysis, investigative journalism, and, most crucially, cyber-security. According to peer-reviewed papers and journal publications, a lot of innovative research is going on in the field of OSINT, but the huge volume of constantly growing unstructured data, along with fake news, data reliability, and legal aspects of conducting OSINT, is still a challenge for the OSINT community. With the increased accessibility and use of the internet, the amount of data being added to web space has exploded. At the same time, the availability of high-speed computing capacity has improved data processing and analysis significantly. Artificial intelligence and its sub-areas increased the processing and analytical power of the complete opensource intelligence activity, from data collecting to cleansing, analysis, and dissemination of the information obtained. The transition to Web 3.0, which is based on three key technological innovations such as artificial intelligence, decentralized data networks, and edge computing, will drastically alter OSINT. The reliance on machine learning and automated reasoning in Web 3.0 would benefit OSINT.

At the same time, widespread adoption of encryption by the general public would impede data collection and raise some legal issues based on local law.

## 6. RESEARCH GAP :

After the literature review, it is observed that the current methodology, models, and publications appear to be lagging in addressing the benefits and challenges of the much-anticipated Web 3.0 era. Artificial Intelligence, Ubiquity and Semantic Web will reinvent how data is collected, processed, and distributed in Opensource intelligence activities. Of course, OSINT techniques make the most of artificial intelligence capabilities like natural language processing, machine learning, and network analysis, among others. There is a need to improve the ability to classify and manipulate data in order to enable machines to understand the meanings and the phrases used to describe data, as well as the ability to acquire data from a larger and more diverse set of sources, as well as the capacity to create and distribute all kinds of data across all network types.

During the review, it was discovered that some areas of OSINT do not make extensive use of AI and its subsets.

Some of the research gaps discovered are as follows:

**Research Gap 1**: OSINT lacks **an** intelligent analysis mechanism.

OSINT analysis does not currently use intelligent mechanisms. The tools employed are populating the gathered data and its clear and direct connections.

Incorporating semantic analysis, pattern analysis, and correlation with other events into the analysis phase would reduce reliance on humans.

**Research Gap 2**: OSINT will need to have machine learning and automated reasoning skills.

The primary motive for automated reasoning adoption is to create an OSINT that uses logical reasoning to solve a wide range of problems, including open questions.

**Research Gap 3**:

Incorporation of the Truth Discovery Process Because of the massive amount of data available from numerous sources, the outcomes will be confusing and contradictory. Such cases would be reduced by an automated truth discovery procedure.

**Research Gap 4**: Rather than relying on human intervention, identifying misleading information or false processes should be done through an automated process.

The Internet is open to interpretation by nature, and the majority of the content has no assurance of being accurate and formal, which would taint the results.

## 7. RESEARCH AGENDA :

1. What model can be recommended to make the Opensource Intelligence System Web 3.0 ready?
2. What new framework can be proposed to integrate OSINT with artificial intelligence subsets in order to improve cyber security?
3. What Machine Learning Technology can help improve the OSINT process?
4. Which are the best Machine Learning Algorithms for enhancing OSINT's intelligent analysis mechanism?
5. What AI subset can be used to enhance the reasoning abilities of the OSINT process?
6. What algorithm is capable of automating the Truth Discovery Process effectively?

## 8. ANALYSIS OF RESEARCH AGENDA :

Understanding Web 3.0 and its features in-depth will aid in the development of a conceptual OSINT model that is compatible with next-generation technology. The algorithms that are currently utilized in OSINT were developed to solve broad problems. The review paper explains the problems that are being solved as well as how the solver approaches them. Natural language processing and adaptive resonance theory are examples of semi-supervised learning topics that cover a wide range of algorithms and learning approaches. K-means, expectation, BIRCH, maximization and hierarchical are examples of unsupervised learning. Decision trees, and ensemble methods such as bagging, boosting, and random forests, as well as K means nearest neighbors, are all examples of supervised learning. There are many clustering algorithms differing by what measure **is** used to cluster the data. In OSINT, these are well represented.

## 9. RESEARCH PROPOSAL :

Regardless of whether they are used by law enforcement agencies, security professionals, or criminal hackers, Opensource Intelligence operations use advanced techniques to dig through the large quantity of visible data for the knowledge they need to achieve their objectives. Discovering public-facing assets, evaluating the data acquired, and extracting knowledge from the studied data are all key tasks that must be complete as part of the OSINT practice. The tools utilized to accomplish these tasks must be extremely capable of meeting the requirements and ensuring precision, dependability, and speed. To meet modern world necessities and to make them future ready, it is inevitable to use Artificial Intelligence and its subsets at each stage of the OSINT activity, regardless of the applications they are used for. A preliminary model will be developed first, based on the literature review, identified gaps, and future requirements, followed by improvement of the model against a wider field of OSINT techniques. Third, final product testing to ensure the effectiveness of the evolved model.

## 10. SLOC ANALYSIS OF RESEARCH PROPOSAL :

SLOC analysis is primarily concerned with exploring strengths, limitations, opportunities, and challenges in a systematic manner. We conducted a SLOC analysis of the Opensource Intelligence field to assess its effectiveness [76-80].

**Table. 2**: SLOC Analysis

| Strengths | Limitation |
|---|---|
| • Availability of the open source data<br>• Technological advancements<br>• Availability of the tools<br>• Computing capacity | • Data handling complexity<br>• Reliability of the data<br>• Limited use of AI capabilities<br>• Dependency on the human factor |
| **Opportunities** | **Challenges** |
| • Wide demand from corporate companies for marketing and sales purpose<br>• Data driven approach<br>• New business opportunities in other areas such as Cyber Security, Political analysis etc. | • Legal issues such as data protection, privacy law etc.<br>• Unstructured data will be difficult analyze effectively<br>• Misinformation such as fake news, biased opinions<br>• Information overload |

## 11. CONCLUSION :

Evolution is the rule of nature; the world of open source intelligence will not remain stagnant; advancement in other technologies will pose a challenge to OSINT practice due to changes in the nature of data and the manner in which it is accumulated; and the same advancement in technology will improve OSINT practice's ability to deal with such challenges effectively.

Since the inception of OSINT by the US military in the late 1980s, a lot has changed along with the advancement of technology. The revolution of Internet technology resulted in a paradigm shift. The widespread popularity of social media has enriched the accumulation of open source content on the web.OSINT is now being used in a variety of many other fields such as marketing, cyber security, political strategy analysis, and so on. The flood of data during this era was beneficial not only to law enforcement agencies and professional practitioners, but also to threat actors. OSINT-based cyber security began to gain popularity, and corporate businesses began to use it for self-assessment in order to identify any potential security loopholes. By incorporating AI technology, OSINT has become more powerful and precise. With the introduction of 3G internet services on smartphones, users were able to express their opinions and share a large amount of information about current affairs related to the surrounding area and their respective country to the rest of the world via social media applications that were freely accessible to the rest of the world. This has accelerated the growth of open source data.

In this paper, we reviewed the current state of OSINT as well as the various stages involved in producing Opensource intelligence. In addition, some OSINT techniques for basic searches were presented, as well as the most sophisticated OSINT tools available today for advanced investigations. The study revealed that the degree of AI implementation still needs to be improved in order to achieve a fully automated solution with no human intervention in the decision-making stage. The goal is to effectively implement Artificial Intelligence in OSINT practice in order to improve overall performance and minimize misinterpretations that may occur as a result of the limitations that we reviewed. Our future work should focus on implementing AI in Next Generation Cyber Security through the use of OSINT.

## REFERENCES :


[1] Hussen Maulud, D., Zeebaree, S. R., Jacksi, K., Mohammed Sadeeq, M. A., & Hussein Sharif, K. (2021). State of art for semantic analysis of natural language processing. *Qubahan Academic Journal*, *1*(2), 21-28.

[2] Dashtipour, K., Poria, S., Hussain, A., Cambria, E., Hawalah, A. Y., Gelbukh, A., & Zhou, Q. (2016). Multilingual sentiment analysis: State of the art and independent comparison of techniques. *Cognitive Computation*, *8*(4), 757–771.

[3] Charalambous, E., Kavallieros, D., Brewster, B., Leventakis, G., & Koutras, N. (2016). Combatting Cybercrime and Sexual Exploitation of Children: An Open Source Toolkit. In *Open source intelligence investigation: From strategy to implementation* (pp. 233–249). essay, Springer.

[4] Pastrana, S., Hutchings, A., Caines, A., & Buttery, P. (2018). International Symposium on Research in Attacks, Intrusions, and Defenses. In *Research in attacks, intrusions, and Defenses: 21ST International Symposium, RAID 2018, Heraklion, CRETE, Greece, September 10-12, 2018, proceedings* (Vol. 11050, pp. 207–227). Cham, Switzerland; Springer.

[5] Ponder-Sutton, A. M. (2016). The Automating of Open Source Intelligence. In *Automating open source intelligence: Algorithms FOR OSINT* (pp. 1–20). essay, Elsevier/Syngress.

[6] Benes, L. (2013). OSINT, New Technologies, Education: Expanding Opportunities and Threats. A New Paradigm. *Journal of Strategic Security*, *6*(3), 22–37.

[7] Layton, R., & Watters, P. A. (2016). The Automating of Open Source Intelligence. In *Automating open source intelligence algorithms FOR OSINT* (pp. 1–17). essay, Syngress.

[8] Santarcangelo, V., Oddo, G., Pilato, M., Valenti, F., & Fornaro, C. (n.d.). Social Opinion Mining: An Approach for Italian Language. In *Future internet of things and Cloud (FICLOUD), 2015 3rd International conference on* (pp. 693–697). Rome, Italy.

[9] Hassan, N. A., & Hijazi, R. (2018). The evolution of open SourCe intelligenCe. In *Open source intelligence methods and tools a practical guide to online intelligence* (pp. 11–11). essay, APRESS.

[10] Azevedo, R., Medeiros, I., & Bessani, A. (2019). PURE: Generating Quality Threat Intelligence by Clustering and Correlating OSINT. In *2019 18th IEEE International Conference on Trust, Security and Privacy In Computing And Communications/13th IEEE International Conference On Big Data Science And Engineering (TrustCom/BigDataSE)* (pp. 483–490).

[11] Bruwer, R. (H., & Rudman, R. (2015). Web 3.0: Governance, risks and safeguards. *Journal of Applied Business Research (JABR)*, *31*(3), 1037.

[12] Glassman, M., & Kang, M. J. (2012). Intelligence in the internet age: The emergence and evolution of open source Intelligence (OSINT). *Computers in Human Behavior*, *28*(2), 673–682.

[13] Klaus, S., Franziska, S., & Reiner, C. (2020). Conception and implementation of professional laboratory exercises in the field of open source intelligence (OSINT). *Society for Imaging Science and Technology*, *2020*(3), 1-99.

[14] John, D. S. M., Goodchild, M. F., & Longley, P. (2007). In *Geospatial analysis: A comprehensive guide to principles, techniques and software tools* (pp. 39–39). essay, Matador.

[15] Nouh, M., Nurse, J. R. C., Webb, H., & Goldsmith, M. (2019). Cybercrime investigators are users too! understanding the socio-technical challenges faced by law enforcement. *Proceedings 2019 Workshop on Usable Security*, 1-11.

[16] Koops, B.-J., Hoepman, J.-H., & Leenes, R. (2013). Open-source intelligence and privacy by design. *Computer Law & Security Review*, *29*(6), 676–688.

[17] Layton, R., & Watters, P. A. (2016). The limitations of automating OSINT: understanding the question, not the answer. In *Automating open source intelligence algorithms FOR OSINT* (pp. 159–169). essay, Syngress.

[18] Bar-Ilan, J. (2001). Data collection methods on the Web for infometric purposes — A review and analysis. *Scientometrics*, *50*(1), 7–32.

[19] Gibson, H., Ramwell, S. S., & Day, T. (2016). Analysis, Interpretation and Validation of Open Source Data. In *Open source intelligence investigation from strategy to implementation* (pp. 95–110). essay, Springer-Verlag.

[20] Gibson, S. D. (2014). Exploring the Role and Value of Open Source Intelligence. In *Open source intelligence in the twenty-first century: New approaches and* (pp. 9–23). essay, Palgrave Macmillan.

[21] Pastor-Galindo, J., Nespoli, P., Gomez Marmol, F., & Martinez Perez, G. (2020). The not yet exploited goldmine of osint: Opportunities, open challenges and future trends. *IEEE Access*, *8*(1), 10282–10304.

[22] Qureshi, P. A. R., Memon, N., & Wiil, U. K. (2011). LanguageNet: A novel framework for processing unstructured text information. In *2011 IEEE International conference on intelligence and Security Informatics (ISI)* (pp. 95–100). IEEE / Institute of Electrical and Electronics Engineers Incorporated.

[23] Hribar, G., Podbregar, I., & Ivanuša, T. (2014). OSINT: A "GREY ZONE"? *International Journal of Intelligence and Counter Intelligence*, *27*(3), 529–549.

[24] Bello-Orgaz, G., Jung, J. J., & Camacho, D. (2016). Social big data: Recent achievements and new challenges. *Information Fusion*, *28*(1), 45–59.

[25] Hassan, N. A. (2019). Gathering Evidence from OSINT Sources. In *Digital forensics basics: A practical guide using Windows OS* (pp. 311–322). essay, Apress.

[26] Edwards, M., Larson, R., Green, B., Rashid, A., & Baron, A. (2017). Panning for gold: Automatically analysing online social engineering attack surfaces. *Computers & Security*, *69*(8), 18–34.

[27] Lee, S., & Shon, T. (2016). Open source intelligence base cyber threat inspection framework for critical infrastructures. *2016 Future Technologies Conference (FTC)*, 4(1), 1375-1384.

[28] Hayes, D. R., & Cappa, F. (2018). Open-source intelligence for risk assessment. *Business Horizons*, *61*(5), 689–697.

[29] Wiradarma, A. A., & Sasmita, G. M. (2019). IT risk management based on Iso 31000 and OWASP framework using OSINT at the information gathering Stage (Case Study: X Company). *International Journal of Computer Network and Information Security*, *11*(12), 17–29.

[30] Vacas, I., Medeiros, I., & Neves, N. (2018). Detecting Network Threats using OSINT Knowledge-Based IDS. In *2018 14th EUROPEAN Dependable Computing CONFERENCE: 10-14 SEPTEMBER 2018, Iasi, Romania* (pp. 128–135). Piscataway, NJ; Institute of Electrical and Electronics Engineers.

[31] Johnsen, J. W., & Franke, K. (2019). The impact of preprocessing in natural language for open source intelligence and criminal investigation. In *2019 IEEE International Conference on Big Data (Big Data)* (pp. 4248–4254). Los Angeles,, CA; IEEE.

[32] Herrera-Cubides, J. F., Gaona-García, P. A., & Sánchez-Alonso, S. (2020). Open-source intelligence educational resources: A visual perspective analysis. *Applied Sciences*, *10*(21), 7617.

[33] Fleisher, C. S. (2008). Using open source data in developing competitive and marketing intelligence. *European Journal of Marketing*, *42*(7/8), 852–866.

[34] Williams, H. J. (2018). In *Defining second generation open source Intelligence (osint) for the defense enterprise* (pp. 1–42). essay, RAND | National Defense Research Institute.

[35] Akhgar, B., Bayerl, P. S., Sampson, F., & Helen Gibson. (2016). Acquisition and Preparation of Data for OSINT Investigations. In *Open source intelligence investigation: From strategy to implementation* (pp. 69–93). essay, Springer.

[36] Quick, D., & Choo, K.-K. R. (2016). Big forensic data reduction: Digital forensic images and electronic evidence. *Cluster Computing*, *19*(2), 723–740.

[37] Vijayarani, S., & Janani, R. (2016). Text mining: Open source tokenization tools – an analysis. *Advanced Computational Intelligence: An International Journal (ACII)*, *3*(1), 37–47.

[38] Ji, C. J., Li, Y., Qiu, W., Awada, U., & Li, K. (2012). Big Data Processing in Cloud Computing Environments. In *2012 12th International Symposium on Pervasive systems, algorithms, and NETWORKS (ispan 2012) San Marcos, Texas, USA, 13-15 December 2012* (pp. 17–23). Piscataway, NJ; IEEE.

[39] Milne, D., & Witten, I. H. (2013). An open-source toolkit for mining Wikipedia. *Artificial Intelligence*, *194*(1), 222–239.

[40] Gong, S., Cho, J., & Lee, C. (2018). A reliability comparison method FOR OSINT Validity Analysis. *IEEE Transactions on Industrial Informatics*, *14*(12), 5428–5435.

[41] Baldini, N., Neri, F., & Pettoni, M. (2007). A multilanguage platform for open source intelligence. *Data Mining VIII: Data, Text and Web Mining and Their Business Applications*, *38*(1), 325-334.

[42] Denecke, K. (2008). Using SentiWordNet for multilingual sentiment analysis. In *2008 IEEE 24th International conference on data engineering workshop* (pp. 507–512). Cancun,, Mexico; I E E E.

[43] Golestan, K., Karray, F., & Kamel, M. S. (2015). An integrated approach for Fuzzy Multi-entity Bayesian Networks and semantic analysis for soft and hard data fusion. In *2015 IEEE International conference on Fuzzy Systems (FUZZ-IEEE)* (pp. 1–8). IEEE / Institute of Electrical and Electronics Engineers Incorporated.

[44] Hassan, T., Hassan, S., Yar, M. A., & Younas, W. (n.d.). Semantic analysis of natural language software requirement. In *2016 sixth international conference on Innovative computing Technology (intech)* (pp. 459–463). IEEE.

[45] Wang, S.-Z., Zhang, Q.-C., & Zhang, L. (2017). Natural language semantic corpus construction based on cloud service platform. In *2017 international conference on machine learning and Cybernetics (ICMLC)* (pp. 670–674). Ningbo; IEEE.

[46] Sleimi, A., Sannier, N., Sabetzadeh, M., Briand, L., & Dann, J. (2018). Automated extraction of Semantic Legal metadata using natural language processing. *2018 IEEE 26th International Requirements Engineering Conference (RE)*, 124-135.

[47] Gupta, P., Goswami, A., Koul, S., & Sartape, K. (2017). Iqs-intelligent querying system using natural language processing. *2017 International Conference of Electronics, Communication and Aerospace Technology (ICECA)*, 410-413.

[48] Thakur, G. S., Bhaduri, B. L., Piburn , J. O., Sims, K. M., Stewart , R. N., & Urban, M. L. (2015). PlanetSense: a real-time streaming and spatio-temporal analytics platform for gathering geo-spatial intelligence from open source data. In *Proceedings of the 23rd Sigspatial International conference on advances in geographic information systems* (pp. 1–4). New York, NY; ACM.

[49] Yue, P., Di, L., Wei, Y., & Han, W. (2013). Intelligent services for discovery of complex geospatial features from remote sensing imagery. *ISPRS Journal of Photogrammetry and Remote Sensing*, *83*(1), 151–164.

[50] Lee, J., Moon, M., Shin, K., & Kang, S. (2020). Cyber threats prediction model based on artificial neural networks using quantification of open source Intelligence (OSINT). *Journal of Information and Security*, *20*(3), 115–123.

[51] D'Avanzo, E., & Pilato, G. (2015). Mining social network users opinions' to aid buyers' shopping decisions. *Computers in Human Behavior*, *51*(10), 1284–1294.

[52] Balaji, T. K., Annavarapu, C. S., & Bablani, A. (2021). Machine learning algorithms for social media analysis: A survey. *Computer Science Review*, *40*(5), 1-32.

[53] Al-khateeb, S., & Agarwal, N. (2019). Social cyber forensics: Leveraging open source information and social network analysis to advance cyber security informatics. *Computational and Mathematical Organization Theory*, *26*(4), 412–430.

[54] Chauhan, S., & Panda, N. K. (2015). OSINT Tools and Techniques. In *Hacking web intelligence: Open source intelligence and web reconnaissance concepts and techniques* (pp. 101–131). essay, Syngress.

[55] Evangelista, J. R., Sassi, R. J., Romero, M., & Napolitano, D. (2020). Systematic literature review to investigate the application of open source Intelligence (OSINT) with artificial intelligence. *Journal of Applied Security Research*, *16*(3), 345–369.

[56] Sagnika, S., Pattanaik, A., Shankar Prasad Mishra, B., & Meher, S. K. (2020). A review on Multi-Lingual sentiment analysis by machine learning methods. *Journal of Engineering Science and Technology Review*, *13*(2), 154–166.

[57] Akhgar, B., Bayerl, P. S., Sampson, F., Tabatabaei, F., & Douglas Wells. (2018). OSINT in the Context of Cyber-Security. In *Open-source intelligence investigation from strategy to implementation* (pp. 213–231). essay, Springer International Publishing.

[58] Layton, R., Perez, C., Birregah, B., Watters, P., & Lemercier, M. (2013). Pacific-Asia Conference on Knowledge Discovery and Data Mining. In *Trends and applications in knowledge discovery and data mining revised selected papers* (pp. 36–46). Heidelberg; Springer.

[59] Yeboah-Ofori, A. (2018). Cyber intelligence and osint: Developing mitigation techniques Against Cybercrime threats on social media. *International Journal of Cyber-Security and Digital Forensics*, *7*(1), 87–98.

[60] Shere, A. R. (2020). Now you [don't] see me: How have new legislation and changing public awareness of the UK surveillance state impacted osint investigations? *Journal of Cyber Policy*, *5*(3), 429–448.

[61] Quick, D., & Choo, K.-K. R. (2018). In *Big digital Forensic Data: Volume 2: Quick analysis for evidence and intelligence* (pp. 67–81). essay, Springer Singapore.

[62] Taddeo, M., Glorioso, L., & Casanovas, P. (2017). Cyber Warfare and Organised Crime. A Regulatory Model and Meta-Model for Open-Source Intelligence (OSINT). In *Ethics and policies for CYBER operations: A NATO Cooperative cyber Defence Centre of Excellence Initiative* (Vol. 124, pp. 139–167). essay, Springer.

[63] Quick, D., & Choo, K.-K. R. (2018). Digital forensic intelligence: Data subsets and open-source Intelligence (dfint+osint): A timely and Cohesive mix. *Future Generation Computer Systems*, *78*(2), 558–567.

[64] González-Granadillo, G., Faiella, M., Medeiros, I., Azevedo, R., & González-Zarzosa, S. (2021). ETIP: An enriched threat intelligence platform for Improving osint Correlation, analysis, visualization and sharing capabilities. *Journal of Information Security and Applications*, *58*(5), 1-15.

[65] Derbyshire, R., Green, B., & Hutchison, D. (2021). "Talking a different language": Anticipating adversary attack cost for cyber risk assessment. *Computers & Security*, *103*(4), 1-24.

[66] Martinez Monterrubio, S. M., Noain-Sánchez, A., Verdú Pérez, E., & González Crespo, R. (2021). Coronavirus fake news detection VIA Medosint check in health Care OFFICIAL bulletins with CBR explanation: The way to find the real information source through OSINT, the verifier tool for official journals. *Information Sciences*, *574*(10), 210–237.

[67] Lande, D., & Shnurko-Tabakova, E. (2019). OSINT as a part of cyber defense system. *Theoretical and Applied Cybersecurity*, *1*(1), 103-108.

[68] Mittal, S., Das, P. K., Mulwad, V., Joshi, A., & Finin, T. (n.d.). CyberTwitter: Using Twitter to generate alerts for cybersecurity threats and vulnerabilities. In *2016 IEEE/ACM International conference on advances in social Networks analysis and Mining (ASONAM)* (pp. 860–867). San Francisco, CA; IEEE.

[69] Ziolkowska, A. (2018). Open source Intelligence (osint) as an element of Military recon. *Security and Defence Quarterly*, *19*(2), 65–77.

[70] Hernandez Mediná, M. J., Pinzón Hernández, C. C., Díaz López, D. O., Garcia Ruiz, J. C., & Pinto Rico, R. A. (2018). Open source Intelligence (OSINT) in a Colombian context and sentiment analysis. *Revista Vínculos*, *15*(2), 195–214.

[71] Eldridge, C., Hobbs, C., & Moran, M. (2017). Fusing algorithms and analysts: Open-source intelligence in the age of 'big data.' *Intelligence and National Security*, *33*(3), 391–406.

[72] Triglav, J., Petrovič, D., & Stopar, B. (2011). Spatio-temporal evaluation matrices for geospatial data. *International Journal of Applied Earth Observation and Geoinformation*, *13*(1), 100–109.

[73] Machado, A. M., & Magalhães, J. P. (2019). TExtractor: An OSINT Tool to Extract and Analyse Audio/Video Content. In *Innovation, engineering and entrepreneurship* (Vol. 505, pp. 3–9). Cham; Springer International Publishing.

[74] Kanta, A., Coisel, I., & Scanlon, M. (2020). A survey exploring open source intelligence for smarter password cracking. *Forensic Science International: Digital Investigation*, *35*(12), 1-11.

[75] Kang, S., Moon, M., Shin, K., & Lee, J. (2020). A study on Priority analysis of Evaluation factors for cyber threats using open source Intelligence (OSINT). *Journal of Information and Security*, *20*(1), 49–57.

[76] Aithal, P. S., & Kumar, P. M. (2015). Applying SWOC analysis to an institution of higher education. *International Journal of Management, IT and Engineering*, *5*(7), 231-247.

[77] Anantha Murthy, & Nethravathi P. S. (2021). The Evolution of the E-Vehicle Industry and its Path Towards Setting up Dominance in Automobile Industry - A Case Study. *International Journal of Case Studies in Business, IT and Education (IJCSBE)*, *5*(2), 38–49.

[78] Yogish Pai U, & Nandha Kumar K.G. (2021). Operational Resilience of the Indian IT-BPM Industry during the COVID-19 Pandemic – A Case Study. *International Journal of Case Studies in Business, IT and Education (IJCSBE)*, *5*(1), 1–13.

[79] Laveena C. Crasta, & Shailashri V. T. (2021). A Comprehensive Study of Talent Management Process adopted by Tata Consultancy Services (TCS). *International Journal of Case Studies in Business, IT and Education (IJCSBE)*, *5*(1), 267–281.

[80] Yogish Pai U, & Nandha Kumar K.G. (2021). E-Commerce to Multinational Conglomerate: Journey of Alibaba Group – A Case Study. *International Journal of Case Studies in Business, IT and Education (IJCSBE)*, *5*(1), 25–35.

**********