

<sup>1</sup> Apurva  
Khandekar\*  
Dr. Prathipati  
Ratna Kumar<sup>2</sup>

# Trusted Personalised Marketing Communications with Big Data Analytics for Product Offerings Using Encryption Algorithm



**Abstract:** - Digital marketing is becoming popular and every organisation is trying to promote their business through digital advertising platforms. The service industry is known for personalisation where services are personalised as per the requirement of customers. The service ensures customer satisfaction and helps in retaining existing customers. In this research, a secure IoT environment with big data analytics is developed to achieve a Smart and secure environment. Initially, a Hadoop and Top-k query processing algorithm is proposed in this work to handle data acquisition, which reduces the redundancy in the collected data. Continuously, a Shannon-Fano algorithm is presented for data compression. To protect the confidentiality of big data, the system usually encrypted the big data before uploading them to the cloud. Accordingly, a Fully Homomorphic encryption algorithm is introduced to encrypt the data to enhance cloud storage security. Subsequently, to preserve improved privacy, secure data transportation, and data access management, the Secure and Robust Data Access Management (SRDAM) Algorithm is presented. Customers will be able to manage their consent, modify their profile, and fully control their data subject rights with the help of the proper consumer identity and access management solution. The proposed work is evaluated in the Matlab software and the performance metrics are accuracy, precision, recall, F1 score, and encryption time. The accuracy of the proposed method is approximately 3% higher than the existing Dnn4C, 6% than the RNN LM, 7.5% higher than the DNN LM, 8% higher than the SLAMC, and 9% higher than the N-gram methods. Accordingly, these results reveal that the proposed method has the best performance and it produces secure data transportation, and data access management, respectively.

**Keywords:** Personalized Marketing, Data Acquisition, Hadoop, Top-K Query, Shannon-Fano Algorithm, Data Compression, Secure And Robust Data Access Management, Secure Data Transportation, And Data Access Management.

## 1. INTRODUCTION

Broadcasters have entered the market competition from the conventional monopoly industry, and the operating mode has gradually shifted from reliance on licensing resources to personalization and refining, relying on data with the gradual merger of the three networks [1]. Traditional content distribution based on licenses is positive, but it lacks market and customer awareness, and users are in a passive state of content receipt. Furthermore, in many scenarios, marketing strategies are crude and inaccurate, failing to account for the differentiated characteristics of individual users and thus leading to a slew of negative outcomes such as decreasing marketing revenue, increasing marketing costs, deteriorating marketing efficiency, and user experience. Furthermore, big data technologies can be used to make customized recommendations [2-3]. Content-based recommendation, collaborative filtering, association rule-based recommendation, neural network-based model, and hybrid recommendation are the most common recommendation algorithms now in use [4]. The technique of regulating a website to a specific user's features or interests is known as personalization. To improve customer service and e-commerce sales, use this tool. The webpage is tailored to each customer [5].

### Personalized Communication

Personalization entails more successfully and efficiently addressing consumers' requirements, making contacts quicker and easier and as a result, enhancing customer gratification and the likelihood of future visits. Personalization software packages such as Monetate, OptinMonster and Broadvision and others are available. Retargeting operations are a type of advertising that can adapt in real-time based on customer behaviour and data insights [6]. Over the baseline, personalization increases click to the top place by 3.5 percent and decrease the average mistake in the rank of a click by 9.43 percent. According to a survey of 200 marketing executives conducted by "Forbes Insights" and "Arm Treasure Data," personalization is yielding positive outcomes [7].

Despite the increasing collaboration of quantitative and qualitative marketing and consumer researchers to address these developing problems, healthcare and medical decision-making endure understudied practical fields. In healthcare and medical decision-making from both consumer research and marketing science perspectives, to

<sup>1</sup>\*Research Scholar, Koneru Lakshmaiah Education Foundation, Hyderabad, Email: khandekarapurva@gmail.com

<sup>2</sup>Assistant Professor, Koneru Lakshmaiah Education Foundation, Hyderabad, Email: rk30111972@klh.edu.in

progress the acceptance of firm, consumer and regulatory elections and their interplay on relevant public policy [8-9]. Contextual, touch-point, omniscient, and integrated marketing, along with neuro-marketing are among the current problems of digital marketing. They usually presume a synergetic mix of digital marketing technologies, such as location, social, mobile, and web-based marketing [10]. Security in health is the knowledge that one's health is stable; if it isn't, there are resources available to help one return to a healthy state. The purpose is to provide a baseline level of protection against diseases and unhealthy behaviours.

### Secure Data Analytics

Health security encompasses operations and procedures that extend beyond national borders to ensure people's health. The cost of connectivity facilities through flat and secure connections among customers, clinics, and particular health care systems is a vital goal of any health care service. IP-based networking, high-speed and low-cost internet, rising data consumption, miniaturization, cloud computing, and improved data analytics are just a few of the technologies and market developments that are enabling the interconnection of small devices [11-12]. Millions of users visit multiple websites these days, and if they want to access something, they must use the id and password concept in various text boxes to safeguard the application. However, this approach just requires one textbox for both the id and the password [13]. Subsequently, the data is stored in a string and multiple programmes in various languages are built to direct the data into the database. Concurrently, just one column is essential in the database. However, the gathered data can be sensitive or personal to the user, and it is frequently used in marketing analytics [14]. In cloud architecture, communication among smart devices and the cloud is frequently required, which can be insecure if proper precautions are not followed. Artificial Intelligence (AI) on the edge can help address some of these issues. When diverse sources of data are combined, consumer digital profiles can be created, which can be a valuable source of information for marketing intelligence and decision-making. In an IoT/IoE setting, smart devices track many more activities of a consumer, these sorts of tracking and data collecting are only projected to proliferate in the future. In digital marketing, AI is utilised to incorporate machine learning into the context [15]. It's also worth noting that digital marketing isn't just for regular shopping or advertising, but also for essential areas like food and health marketing. The rest of the paper is organized as follows, section 2 portrays the literature of the study, and section 3 reveals the problem definition and motivation. Section 4 reveals the proposed methodology, section 5 demonstrates the experimentation and result in a discussion, and section 6 reveals the conclusion of the work.

## 2. LITERATURE SURVEY

The main advantages of science and technology mean for example artificial intelligence and big data analysis are increasingly appearing in digital marketing. Zhuo Jun Li *et al* [16] investigated digital marketing efficiency and presented an intelligent algorithm based on data analysis to improve marketing communication effectiveness. It has the potential to increase the effectiveness of digital marketing communication as well as the social impact of digital marketing. Advertising platforms with a bad reputation could experience negative effects from personalization disclosures because the motivations for posting such disclosures are viewed as phoney. Ooijen *et al* [17] suggest that when a personalised disclosure is offered as a warning that is published alongside an advertisement by the unreliable platform itself, the advertising efficacy is decreased in contrast to when the disclosure was supplied by a more trusted third party.

Because of privacy concerns, the collection of multidimensional crowdsourcing data has sparked public outrage. To overcome this, Shen *et al* [18] propose using local differential privacy (LDP) to protect crowd-sourced data without sacrificing usability. Despite their usefulness for multidimensional crowd-sourced data, conventional LDP methods ignore users' privacy concerns. By developing the concept of personalized LDP (PLDP), recognized the personality of data owners in the preservation and usage of their multidimensional data (PLDP). To perturb data owners' data, we design personalized multiple optimized unary encoding (PMOUE), which meets total-PLDP. Current techniques still place a premium on user identity, mutual privacy, and session-specific key agreements between content owners, trusted clients, and cloud service providers. Therefore, Ezhilarasan *et al* [19] introduced a Safe and Robust Data Access Management (SRDAM) algorithm to preserve better privacy, and secure data transit. For documents, pictures, and videos, the proposed SRDAM algorithm decreases data downloading time, data uploading time, and communication overhead compared to traditional approaches.

Personal data has numerous values, including dignity and independence of the individual, economic use, and public administration. In the meantime, the stakeholders who have access to personal data have grown increasingly diversified, resulting in an increased demand for sharing and exploiting personal data. In this situation, the Rational Expectation rule emerges as a new alternative for personal data protection in the age of big data in YongLin *et al* [20]. Moreover, it covers the criteria of risk control under the rational expectation rule by measuring the risk of personal information sharing in application contexts using the matrix approach. Yan *et al* [21] conduct important research and analysis based on theory and practice. Simultaneously, relative analysis was utilized to focus on the state of commercial banks, regarding the current state of the Internet. The proposed scheme mainly focused on the issues that have arisen as a result of contemporary Internet development in many forms, such as the bank's sales position and client products. The current research attempts to provide a complete understanding of the main difficulties linked to user privacy that hinder data-driven innovation DDI in Saura *et al* [22]. The study finishes with a discussion of the importance of user privacy in DDI in digital markets.

A single case study was conducted by Jones-Smith *et al* [23] which was based on relationship management theory, to look at how three executive leaders of a nonprofit organization in the north-eastern United States employed digital marketing to grow and sustain their donor base. The possibility of building a regular support stream from stakeholders, which allows executive non-profit executives to extend services supplied to communities and customers, has positive social change implications. Zarouali *et al* [24] discovered that the perceived security, privacy, and WhatsApp as a platform are positively related to credulous brands on that messaging network, using data from a nationwide typical survey. As a result, consumers' intent to share information with brands on WhatsApp is positively influenced by brand trust. For eras, AI was used in the healthcare industry for a variety of purposes. However, building a strong AI model is tough, and building a generalized prediction model is difficult due to the fragmented structure of patient data across the healthcare system. Aich *et al* [25] proposed a solution based on blockchain and AI technology to address the aforementioned issues. The blockchain will safeguard data access and AI-based federated learning, allowing for the development of a strong model for global and real-time use.

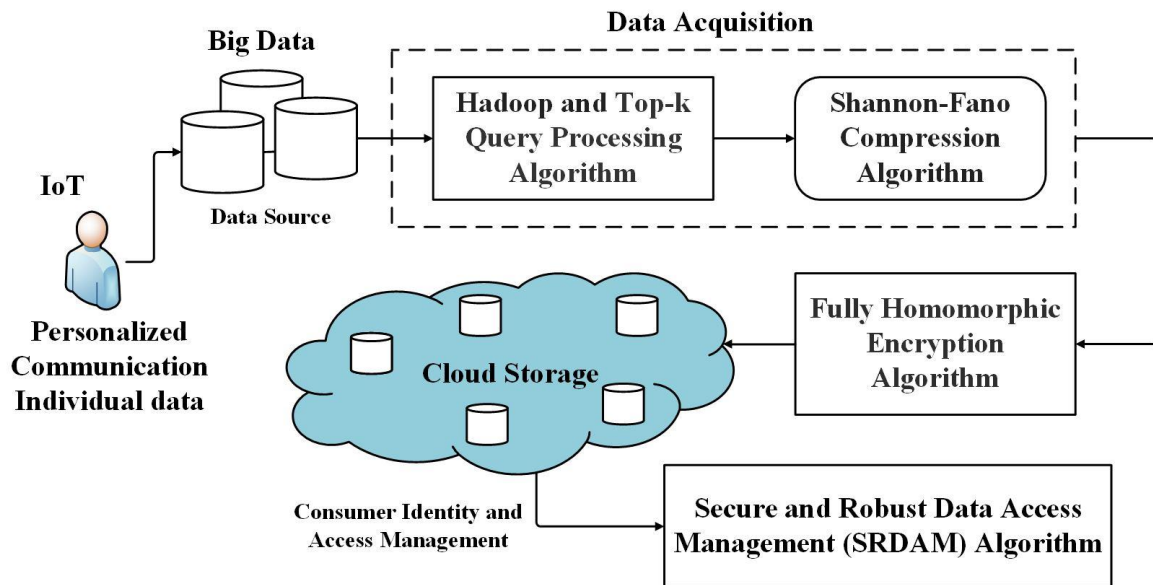
### 3. RESEARCH PROBLEM DEFINITION AND MOTIVATION

The Internet of Things (IoT) is an emerging area, this technology is made to connect any object around us to the Internet with a unique IP, and these connected objects can be communicated with each other remotely as per the user's convenience. Accordingly, it has applications in all most all the fields like industries, factories, environment, agriculture, transport, education, healthcare, energy, and retail. IoT leads to new technologies like big data and cyber-physical systems. Connecting any object, from anywhere at any time, is not simple. Privacy, discovery, software complexity, security, scalability, interoperability, and fault tolerance are the various challenges faced by this work. Accordingly, security is one of the key issues. The weak links used to connect things to the Internet lead to many security issues at different levels of the IoT. This research presents the novel security architecture for IoT-enabled personalized communication and it also presents the various security issues. Fitness, social life, health care, education, entertainment, energy conservation, home automation, environmental monitoring, and transportation systems are the different sectors, that IoT identifies various applications in these sectors. IoT technology has dramatically improved the quality of life and reduced human effort.

Actually, in modern times where processes such as personalized marketing and filter bubble are present, many people fear that their privacy is at risk. It is important to note that a large part of insights associated with big data includes predictions that are made regarding the details of consumers. Most of the time, these details are quite personal in terms of their nature, which is one of the reasons why even the chance or possibility of them falling into the hands of the wrong people, is enough to eliminate any possible trust that consumers have in different firms and organizations. Realizing the importance of privacy and the value that is possessed by sensitive information, it is necessary for the survival of a firm that they consider different measures for preventing the obstruction of the privacy of consumers. Digital media is a fast and widely growing source of information, entertainment, news, shopping and social interaction. Consumers want brands which they can trust, companies they are familiar with, and interaction that is personalized and relevant, and offer customized to their needs and preferences conveniently. This motivates the research to propose an IoT-based artificial intelligent system for secure consumer data for personalized communication in personalized marketing.

#### 4. PROPOSED RESEARCH METHODOLOGY

Personalization in marketing is a strategy companies use to connect with customers on a one-to-one basis. In order to produce individualised content, companies gather and analyse information on income level, preferences, customers' demographics, occupations, and purchasing habits. Marketing automation delivers this content to customers through social media, email, blogs, and other means. While this approach can be easy to implement, prospects aren't necessarily qualified, and their "intent to buy" signal is weak. This study examines IoT applications in marketing research and offers details on how organizations use big data technologies like social media, machine learning (ML), and artificial intelligence (AI) to advertise a range of products and services. The flow diagram of the research work is demonstrated in figure 1.



**Figure 1:** Block Diagram of the Research Work

Personalization has proven to be a highly effective marketing strategy. A key element of personalization success is data collection and automation, and with this comes the need for an intelligent algorithm. Consequently, the research combines the functions of the IoT with Cloud Computing, and Big Data to achieve a smart and secure environment. In big data, data acquisition was handled by Hadoop and Top-k query processing algorithm which reduces the redundancy in the collected data. Further data compression was processed by the Shannon-Fano algorithm. Before being stored in the cloud, the data encryption process was taken by a Fully Homomorphic encryption algorithm which enhances the cloud storage security. Further for consumer identity and access management, the research developed a Secure and Robust Data Access Management (SRDAM) Algorithm which maintains data access management and secure data transportation.

##### 4.1 Secure IoT Environment with Big Data Analytics

The Internet of Things is a significant advancement in the Big Data era, which supports many real-time engineering applications through enhanced services. With the quick advancement of sensor technology, for IoT, WSNs will become the key technology in which millions of sensors and devices are connected over the Internet. Big data analytics has many benefits, including the capacity to create new goods and services, reduce costs, act quickly, and make better decisions. This wireless network ensures Big Data for managing the consumer data during personalized marketing. Taking this into account, the research proposes a scenario that tries to combine the functions of the IoT with Cloud Computing, and Big Data to achieve a Smart and secure environment. Based on redundancy elimination technology, a compatible cloud storage method is presented to fulfil the demands of users in terms of storage efficiency, security, speed, and capacity, to minimise data storage space and increase data redundancy. Consequently, the research proposed a novel big data acquisition platform which is designed based on Hadoop and Top-k query processing algorithm. The collected data are classified and processed by the classifier. The classified big data are compressed by the Shannon-Fano algorithm, and the data security is improved by data encryption.

#### 4.1.1 Hadoop and Top-k Query Processing Algorithm

The open-source Hadoop framework is used to process and store a huge amount of data, but in a decentralised way, by storing data on several devices and then distributing it among these devices to speed up processing. A large amount of structured and unstructured data can be handled by Hadoop. In order to store and process the big data, the Hadoop system is based on a distributed clustering of many nodes that work together. When the files are saved under multiple names, the Hadoop may be duplicated storing the same file several times, therefore, it is still not characterized by effective storage. This results in a reduction of device processing efficiency and loss of storage space. Data deduplication through HDFS is performed to prevent this duplicate information.

Data deduplication is a storage-optimization technique that reduces data by avoiding multiple copies of redundant data and it keeps storing unique data. In general, data deduplication involves four main phases. The incoming large data is split into small files by the original file known as “Chunks”. To prevent hash collisions, the unique hash value is then calculated and allocated to each chunk. The redundant chunk will be deleted if the new chunks are already present in the database; otherwise, the new chunk will be stored with a new hash value and added to the index table.

However, data deduplication is both I/O intensive and communication-intensive, which leads to a negative impact on system performance. The best data duplication ratio and the scalable throughput are achieved by effectively utilising the computational capacity and storage capacity of several data nodes. Additionally, index data updates are necessary. However, system overhead will be caused by global index updates and duplicate data detection, particularly when there are tens of data nodes. With consistent performance, deduplication over HDFS may support a large amount of raw data storage. A top-k query processing algorithm is proposed to enhance the performance of this system.

#### 4.1.2 Top-k Query Model

The user specifies a number  $k$ , and the system should return the  $k$  most relevant answers using the top-k query. A scoring function determines the relevant degree of responses to the query. The algorithm is run over sorted lists (also called inverted lists) by using the common method for efficient top-k query processing and formally defined them.

Let  $D$  be a set of  $n$  data items, then the sorted lists are  $m$  lists  $L_1, L_2, \dots, L_m$ , such that each list  $L_i$  contains every data item  $d \in D$  in the form of a pair  $(id(d), s_i(d))$  where  $id(d)$  is the identification of  $d$  and  $s_i(d)$  is a value that denotes the local score (attribute value) of  $d$  in  $L_i$ . The data items in each list  $L_i$  are sorted in descending order of their local scores. Let  $f$  be a scoring function given by the user in the top-k query. Applying the function  $f$  to the local scores  $d$  for each data item  $d \in D$  results in the calculation of an overall score, which is represented by  $ov(d)$ . Formally,  $ov(d)$  is portrayed as follows.

$$ov(d) = f(s_1(d), s_2(d), \dots, s_m(d)) \quad (1)$$

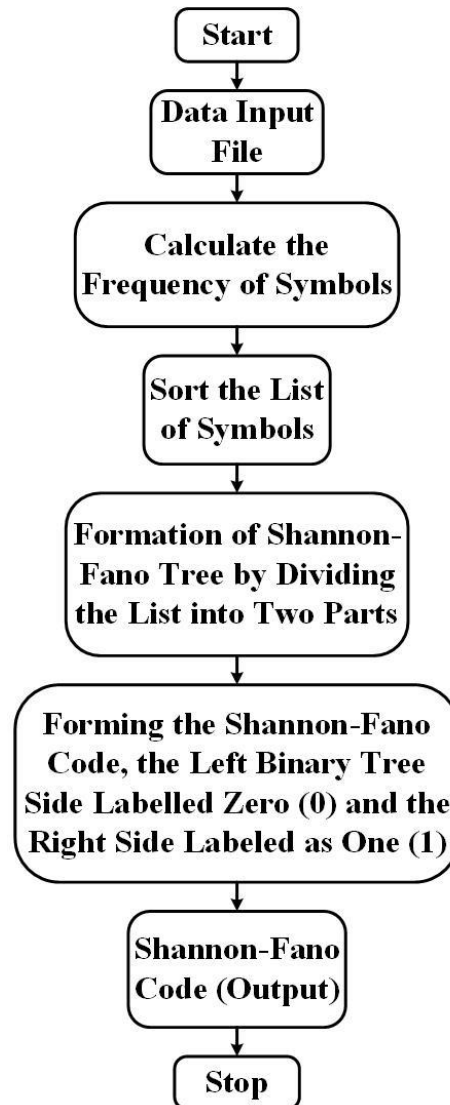
The set of  $k$  elements with the greatest overall scores among all database elements is the output of a top-k query. The scoring function in this work belongs to the category of linear functions with positive coefficients (denoted by LFPC). The storage efficiency, security, speed, and capacity are managed by this Hadoop and top-k query-based processing technique, which also decreases data redundancy and storage space.

#### 4.1.3 Shannon-Fano Algorithm

Shannon Fano compression algorithm is the type of lossless algorithm that constructs a prefix code based on the set of symbols and their probabilities or frequency. According to their frequencies, the various symbols are sorted and they are then divided into two groups based on their frequency. It is an entropy encoding algorithm. In order to create the binary tree for the Shannon Fano algorithm, which starts at the root to the node, this approach uses the occurrence frequency, it forms the binary tree that is used for Compression. By allocating a large number of

bits for information with a low frequency and a small number of bits for information with a high frequency, the Shannon Fano algorithm encodes the data based on its frequency of occurrence.

The Flow Chart of the Shannon-Fano Algorithm is portrayed in figure 2 as follows:



**Figure 2:** Flowchart of Shannon-Fano Algorithm

The following steps are used to implement the Shannon Fano compression algorithm.

- Create a list of frequency or probability counts for a given list of symbols to determine the relative frequency of occurrence for each symbol.
- According to frequency, sort the list of symbols so that the most frequent symbols are on the left and the least frequent ones are on the right.
- Split the list into two parts, keeping the frequency counts for the left part as close as feasible to the total of the right.
- Create a tree with the restriction that the binary digits 0 are assigned to the left part of the list and digit 1 is assigned to the right of the list. This means that all of the symbol codes for the first half will begin with 0, and all of the symbol codes for the second part will begin with 1.
- Apply steps 3 and 4 backwards to each of the two halves until each symbol has a corresponding code leaf on the tree. This is done by subdividing groups into smaller ones and adding bits to the codes.

#### 4.2 Data Encryption Algorithm

One of the most popular cloud computing applications is the safe storing and sharing of big data. For big data security and privacy, big data cloud storage and sharing also bring great challenges, although providing economy and convenience. To protect the confidentiality of big data, the system usually encrypted the big data before uploading them to the cloud. Consequently, the research encrypted using a Fully Homomorphic encryption algorithm to enhance cloud storage security.

#### 4.2.1 Fully Homomorphic Encryption

Homomorphism is a structure-preserving transformation between two sets, where an operation on two members in the first set is preserved in the second set on the corresponding members. Let  $P$  and  $C$  be sets with members  $p_1, p_2 \in P$ ,  $T$  is a transformation with an operation  $\oplus$ . If  $\forall (p_1, p_2) \in P, (p_1 \oplus p_2) = T'(T(p_1) - T(p_2))$  the system is homomorphic.

An FHE scheme was defined by Craig Gentry; the scheme contains sets  $N = \lambda, P = \lambda^2, Q = \lambda^5$  and the security parameter  $\lambda$ . The scheme additionally makes use of two integer parameters  $0 < \alpha < \beta$  and the following algorithms, which,  $\beta$  indicates the size of the sparse subset elements and  $\alpha$  denotes the number of elements in a sparse subset, which is necessary for the scheme's Recrypt operation.

**KeyGen**( $\lambda$ ): Generate a random  $P$ -bit odd integer,  $p$ . A set  $\vec{y} = \{y_1, y_2, \dots, y_\beta\}$  is created in a way that  $y_i \in [0, 2)$ . A sparse subset  $S \subset \vec{y}$  of these  $\alpha$  elements must exist out of these elements, such that  $\sum_{y_j \in S} (y_j) = \frac{1}{p} \text{ mod } 2$ . Set  $sk$  to be a binary encoding  $s$  of the sparse subset  $S$ , where  $s = (0, 1)^\beta$ . Set  $pk \leftarrow (p, \vec{y})$ .

**Encrypt**( $pk, m$ ): Obtain the ciphertext  $c = m' + pq$ , where  $m'$  is a random  $N$ -bit integer st.  $m = m' \text{ mod } 2$ . Generate  $\vec{z} : z_i \leftarrow c \cdot y_i \text{ mod } 2$ . Return  $c^* = (c, \vec{z})$ . In the rest of the paper, mention **Encrypt**( $pk, m$ ) as **Encrypt**.

**Decrypt**( $sk, c^*$ ): Output  $LSB(c) \oplus LSB(\lfloor \sum_t S_t z_t \rfloor)$ , where  $\lfloor \cdot \rfloor$  returns the nearest integer to the input and  $LSB(\cdot)$  returns the least significant bit of the input. Decryption is currently functional (within small precision errors)  $\sum_t S_t z_t = \sum_t c S_t y_t = \frac{c}{p} \text{ mod } 2$ .

On encrypted data, it can be argued that the above encryption allows quite wonderfully arbitrary computations. Consequently, defined operations such as  $Evaluate(f, c_1, \dots, c_t)$  where  $f$  is an arbitrary operation on the ciphertext  $c_1, \dots, c_t$ . The output of the computation is always a ciphertext,  $c$  whose decryption would be the same as the function  $f$  applied on the plaintexts corresponding to,  $c_1, \dots, c_t$ . However, the decryption can be erroneous if the noise (measured as  $c \text{ mod } p$  increases. In order to reduce the error during the computations, there is an additional operation, called Recrypt which takes the ciphertext,  $c$  and produces another ciphertext, say  $c'$  which corresponds to the same plaintext, but with a reduced noise level. The operation is done by allowing to compute of the decryption function, as the function  $f$  in the Evaluate function.

Considering an encryption scheme with an odd number  $p$  as a shared secret key:

- ❖ To encrypt a bit  $m$ , a random large  $q$  and small  $r$  are chosen and output ciphertext is computed as  $c = pq + 2r + m$ . The ciphertext is close to a multiple of  $p$  and  $m = LSB$  of distance to the nearest multiple of  $p$ .

❖ The decryption is performed as:  $m=(c \bmod p) \bmod 2$

**Homomorphism Additive Property Verification:** Considering two ciphertexts  $c_1=m_1+2r_1+pq_1$  and  $c_2=m_1+2r_2+pq_2$  under this encryption scheme, the addition and multiplication operations are defined with the following equations. The addition of the ciphertexts can be defined as:

$$c_3=c_1+c_2=(m_1+m_2)+2(r_1+r_2)+p(q_1+q_2) \tag{2}$$

If the  $(m_1+m_2)+2(r_1+r_2)$  is significantly smaller than  $p$ , then

$$c_3=(c_1+c_2) \bmod p=(m_1+m_2)+2(r_1+r_2) \tag{3}$$

The requirements for additive homomorphism conditions are met by this algorithm.

**Homomorphic Multiplicative Property Verification:** Multiplication of the ciphertext can be defined as follows.

$$c_4=c_1 \times c_2=(m_1+2r_1+pq_1) \times (m_2+2r_2+pq_2) \tag{4}$$

$$c_4=m_1 m_2 + 2(2r_1r_2 + r_1m_2 + r_2 m_1) + p[pq_1q_2 + q_2(m_1+2r_1) + q_1(m_2+2r_2)] \tag{5}$$

If the  $m_1 m_2 + 2(2r_1r_2 + r_1m_2 + r_2 m_1)$  is significantly smaller than, then

$$c_4=(c_1 \times c_2) \bmod p = m_1 m_2 + 2(2r_1r_2 + r_1 m_2 + r_2 m_1) \tag{6}$$

The multiplicative homomorphic criteria are met by this algorithm.

**Privacy Protection:** Users transmit and save their data to the cloud encrypted. Both ensure safe data storage and in the process of transmission, it ensures data security. They can't easily obtain the information in plaintext when the cloud computing service providers handle it.

**Data Processing:** Users or a trusted third party can process ciphertext data directly, instead of the original data which is enabled by a fully homomorphic encryption mechanism. To decrypt and get accurate data, users can obtain arithmetic findings. For instance, in the medical information system, electronic medical records are saved on a cloud server as ciphertext. When the health department deals with potential safety problems, they must know some areas of certain disease locations and age distribution. With specialised data processing services, they can provide secure electronic medical record data. After decryption, they can obtain the correct data.

**Ciphertext Retrieval:** Direct searches of the ciphertext data are possible with fully homomorphic encryption technology based on the ciphertext retrieval method. In addition, it is not only enhancing retrieval effectiveness and ensures query privacy, but it also allows for the addition and multiplication of retrieval data without altering the corresponding plaintext.

### 4.3 Identity and Access Management

As the commercial risk of collecting, storing, and exploiting consumer data to personalize services increases, most marketing leaders remain surprisingly unconcerned with how to manage data security and privacy. However, this causes damage to the name and brand, and while difficult to quantify, can hurt revenue and future growth even more. Data breaches have become such a common problem; customers now consider data security a factor in overall customer happiness ratings. Customers will be able to monitor their consent, edit their profile, and fully control their data subject rights with the help of the proper consumer identity and access management solution. In order to maintain improved privacy, safe data transportation, and data access management, the research suggested the SRDAM. The suggested method combines evaluating cloud service providers and then takes into account the need for both cloud users and cloud service providers' property requirements.

#### 4.3.1 Secure and Robust Data Access Management (SRDAM) Algorithm

The Secure and Robust Data Access Management (SRDAM) Algorithm, which is suggested to preserve improved privacy, secure data transportation, and data access management is described in this section. The suggested

method combines evaluating cloud service providers and then takes into account the need for both cloud users and cloud service providers' property requirements. In order to improve the data security and data contribution and retrieval process, there are implementation pre-processing steps of the proposed framework explained. The proposed method optimizes the memory and cleans un-wanted memory in a cloud server.

**Data Owner:** Data owners incorporate slight login elements and register points of interest. The module promotes the use of encryption to transfer the data by the data owner. The method guarantees the data to be shielded from the unauthorised user.

**Cloud Server Provider:** The cloud service provider deals with a cloud to give information for maintaining cloud services. Data owner encrypt their data and store them in the cloud for informing users. To retrieve the information, cloud users can download encrypted data of their interest from the cloud and afterwards decrypt them. It is responsible to approve the accountability of registered cloud users.

**Cloud User:** Here, cloud users complete the registration and login details. The module is utilized to help the user to look at the data utilization into different catchphrases ideas and get the precise outcome list in light of the cloud user query. The cloud user will choose the required data and enrol the user points of interest and get the Verification key in the email before entering the Secret Key.

**Data Ranking Search:** The proposed system assures the user to search the information that is looked at habitually utilizing the rank search. By using his or her secret key to decode the downloaded data, the suggested approach enables the client to download the record. The downloaded records and the transmitted data are enabled to view by the data owner.

**Platform Authorization:** Each System additionally has a unique character that can be utilized to create a different set of characters. Since the authority is joined to the cloud user and can't be (effortlessly) expelled, the authority can utilize these personalities to demonstrate the specific cloud user. It is not the same cloud user or even profiles and credential information validation, which just demonstrates that a specific cloud user. In worst cases that an attacker says a system, has prior information about allowed cloud users, it is workable for the system to validate itself to the challenges. That is, it can demonstrate that it is one of the known and privileged cloud users.

**Aggressor (Attacker):** With specific transmitted data to the cloud storage server, the attacker adds the malicious or malicious affected information. At that point, the unauthorised cloud user will consider an attacker. During data transmission, if, any data is not validated with a secret key, it is treated as malicious data.

The proposed algorithm is highly dedicated to avoiding key complexities. Here, trusted authorities can view the list of CSP (cloud service Providers) and Data Owner details. However, trusted authorities are unable to access content and unknown with a valid secret key. In this process, the data owner can store the data in the preferable cloud, after approval from the respective cloud server. Where owner and user can utilize the cloud-based on their subscription time without any data privacy, or hesitations. The technique uses a secret key to convert the encrypted archive to Zip data. As a result, to download the content, SK sends it to the user.

If anyone is caught with any malicious activity, then it can be smoothly detected and the user can have deactivated. The private key generation process is valid for an authorized user including the user id to generate a secret key where the user alone can access the cloud. It enables a system or cloud server to recognize the cloud users that are getting or asking to access the information for evaluating the trusts. In terms of cost, trust, and reputation of the administration of CSP, to empower cloud users to pick up the genuine CSP for reliable services and helps CSP fine reliable cloud users to generate revenue. The system has three entities namely cloud server, content owner, and data user which is denoted as  $Z$ . The system has computation abilities because some attributes partially contain users' personal identity-based information. Content owner outsources encrypted content in the cloud. The cloud server has adequate content storage capability, and does nothing but save them. New data user requests private keys from the system and is unaware of which attributes are accessed by access structures. When a data user requests their private keys from the access tree; hence, the system creates a composite private key and forwards it to respective users. Every type of data user can download any type of encrypted content; if their private keys fulfil the validation criteria of the access tree  $Ta$  and perform the operation associated with access  $a$ . The

cloud server proceeds with the content management (upload or download) operation access  $a$ , if and only if the users are verified with the access tree,  $Ta$ .

If the set  $Au$  accomplishes any access tree in the set  $\{Ta\}a \in \{0, \dots, r-1\}$ , the technique returns a message  $M$  or a verification. If, Cloud servers utilize  $VS$  to authenticate it, and successfully confirm the authentication parameter the operation request will be continued.

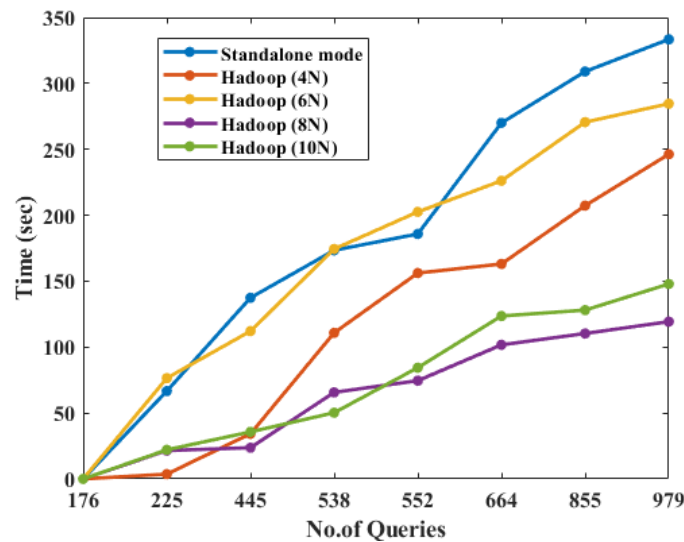
### 5. EXPERIMENTATION AND RESULT DISCUSSION

The simulation system configuration of the proposed work is portrayed in table 1. Subsequently, the proposed technique is evaluated and tested under the Matlab R2021a software. The proposed work operates under windows 10 home and its memory capacity is 6GB DDR3. Additionally, it utilizes an Intel Core i5 @ 3.5GHz processor and the simulation time of the work is 10.190 seconds.

**Table 1:** Simulation System Configuration

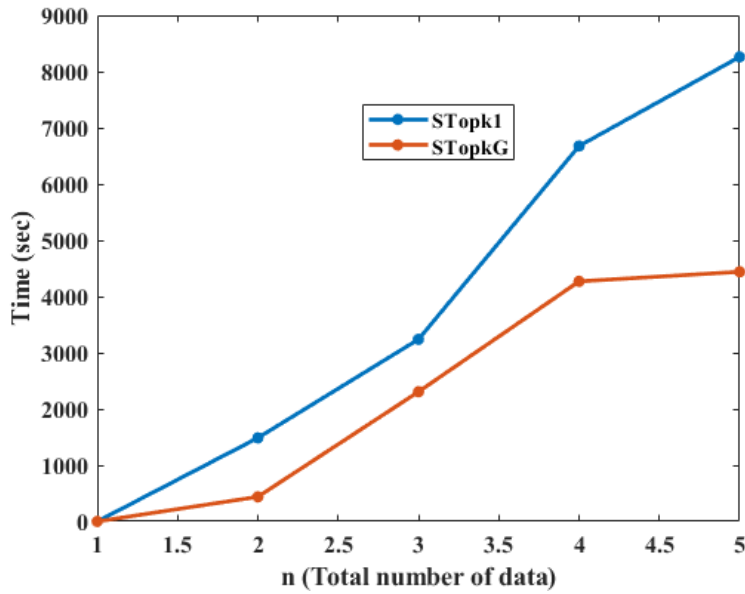
Simulation System Configuration	
MATLAB	Version R2021a
Operation System	Windows 10 Home
Memory Capacity	6GB DDR3
Processor	Intel Core i5 @ 3.5GHz
Simulation Time	10.190 seconds

The performance metrics are accuracy, precision, recall, F1 score, encryption time, and throughput. These performance metrics are evaluated as follows.



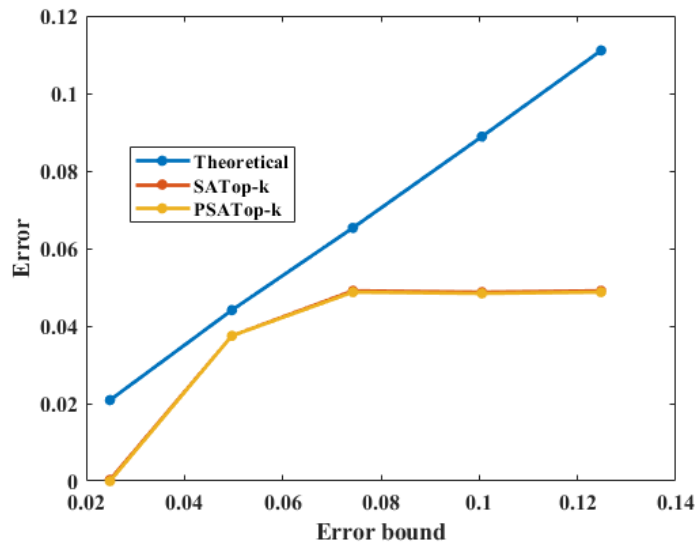
**Figure 3:** Time Graph for Number of Queries

The time graph for a different number of queries is depicted in above figure 3. It consists of standalone mode, Hadoop (4N), Hadoop (6N), Hadoop (8N), and Hadoop (10N). The number of queries ranges from 176 to 979 and the time ranges from 0 to 350 secs. The time for standalone mode ranges from 0 to 340 seconds, and the time for Hadoop (4N) ranges from 0 to 249 secs. Time for Hadoop (6N) varies from 0 to 280 secs, time for Hadoop (8N) ranges from 0 to 119 sec, and the time for Hadoop (10N) ranges from 0 to 150 secs, respectively.



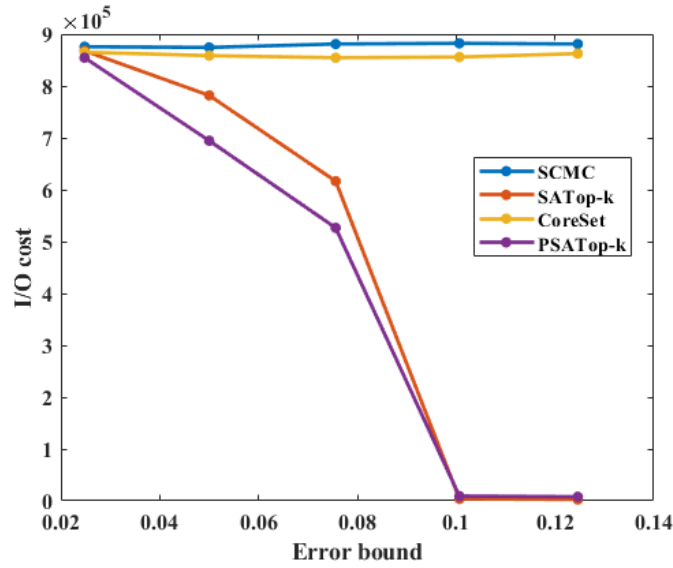
**Figure 4:** Time Graph for Total Number of Data

Figure 4 illustrates the time graph for the total number of data. It is measured for STopk1 and STopkG, for STopk1, the time ranges from 0 to 9000 sec and for STopkG, the time ranges from 0 to 5000 seconds. Accordingly, when the total number of data is 1, the time is 0, when the total number of data is 2, the time for STopk1 and STopkG is 1500 sec and 400 sec. When the number of data is 3, the measured time for STopk1 and STopkG is 2350 sec and 3100 secs, and when the number of data is 4, the time for STopk1 and STopkG is 4300 and 6800. Subsequently, the measured time for STopk1 and STopkG is 4500 and 8250 during the total number of data is 5.



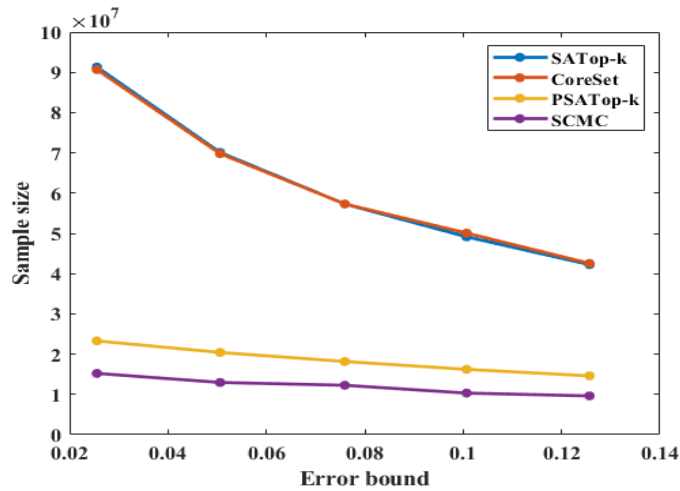
**Figure 5:** Error Graph for Top-k Querying Method

Figure 5 demonstrated the error graph for the proposed method. The error is measured based on the error bound as 0.02 to 0.14. During the error bound is 0.03, the error is measured as 0.028, when the error bound is 0.05, the error produces 0.047 for the theoretical value. Accordingly, the theoretical value for error is 0.07, 0.09, and 0.103 during the error bound is 0.08, 0.1, and 0.12.



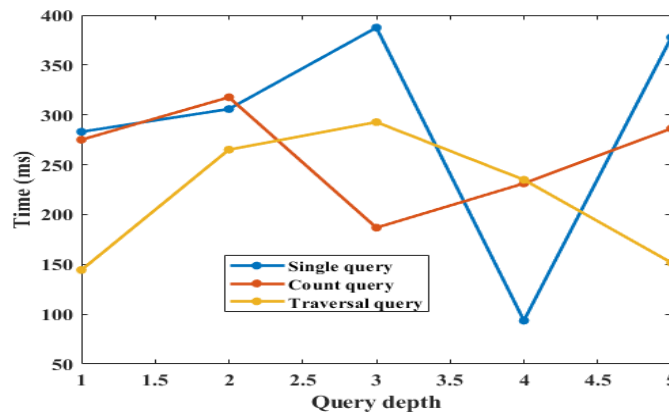
**Figure 6:** Performance Graph for I/O Cost

Figure 6 reveals the performance graph for I/O cost. The cost is measured based on the error bound and is measured for the SCMC, SATop-k, Coreset and PSATop-k. The cost is decreased when the error bound increases. The cost value for SCMC approximately ranges from  $8.8 \times 10^5$ . Then the cost value for Coreset ranges approximately  $8.7 \times 10^5$ . The cost value for SATop-k and PSATop-k initialized as  $8.8 \times 10^5$  and  $8.75 \times 10^5$ , and it reaches 0 when the error bound is 0.1.



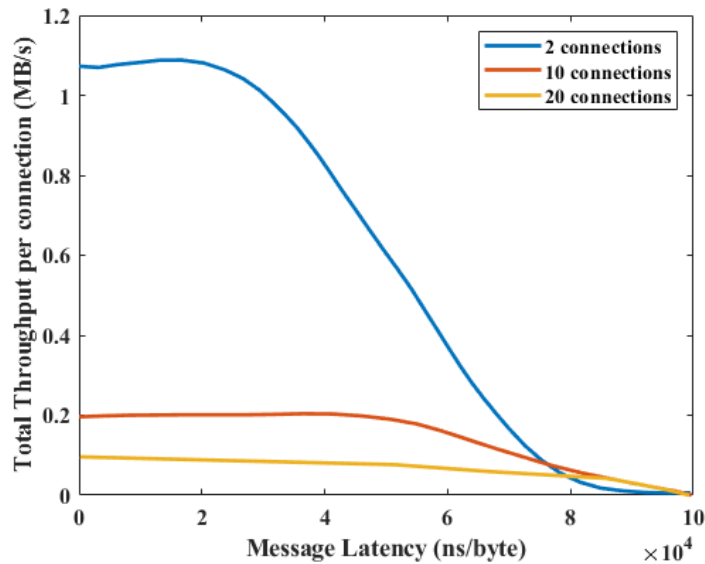
**Figure 7:** Graph for Sample Size

The sample size graph is demonstrated in figure 7; it is measured for the error bound. The sample size is revealed for SCMC, SATop-k, Coreset and PSATop-k. The SATop-k and Coreset contain the same sample size, i.e. it ranges from  $9$  to  $4.5 \times 10^7$  for the error bound of 0.02 to 0.12. The PSATop-k value produces the value of  $2.2$  to  $1.7 \times 10^7$ , and SCMC produces the sample size of  $1.5$  to  $1 \times 10^7$ .



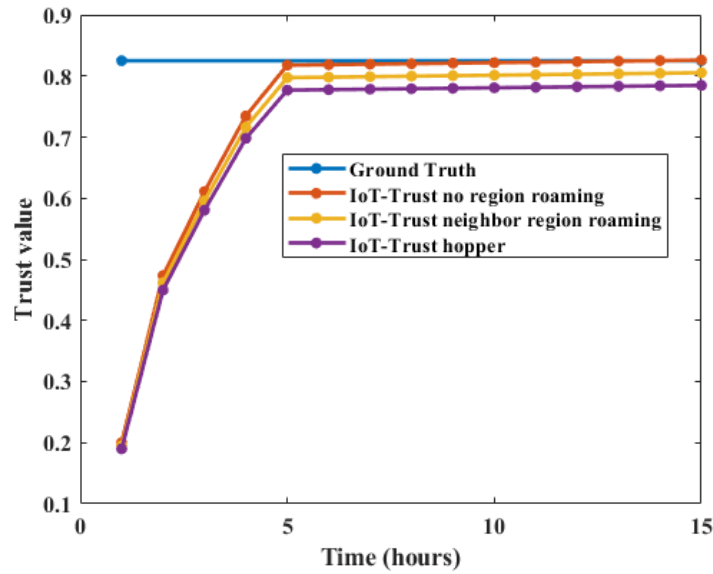
**Figure 8:** Time Graph for Query Depth

Figure 8 demonstrates the time graph for query depth, it consists of a single query, count query, and traversal query. Query depth is a predefined property that informs about the depth of a query. The time for a single query is 280 ms for the query depth is 1 and 380 ms for query depth is 5. For the count query, the time is measured as 275 ms when the query depth is 1 and when the query depth is 5, the time is 285 ms. The traversal query produces times of 146 ms and 150 ms during the query depth is 1 and 5.



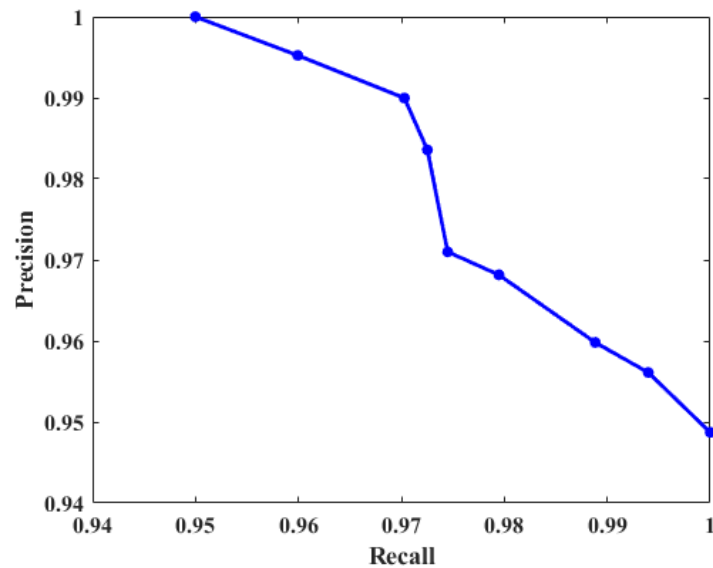
**Figure 9:** Performance Graph for Total Throughput

Figure 9 portrays the total throughput graph. The throughput is calculated as requests/units of time. From the first sample's beginning to the end of the last sample, the time is determined. As it is intended to represent the demand on the server, this includes any intervals between sampling. Here, the throughput is measured based on the message latency and it is measured for 2, 10 and 20 connections. For 2 connections, the total throughput is 1.05 to 0.03 during the latency is 0 to 100000. Accordingly, for this condition as latency is 0 to 100000, the 10 connections and 20 connections are produces 0.2 and 0.1 values.



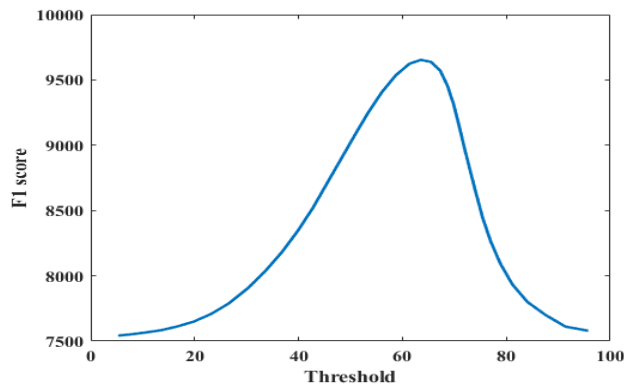
**Figure 10:** Performance Graph for Trust Values

The trust value graph for the proposed method is demonstrated in above figure 10. Here, the ground truth values maintain the constant value of 0.82. The IoT-trust no region roaming produces the trust value of 0.832 for 15 hours. The IoT-trust neighbour region roaming produces the trust value of 0.8 and the IoT-trust hopper produces the trust value of 0.78, respectively.



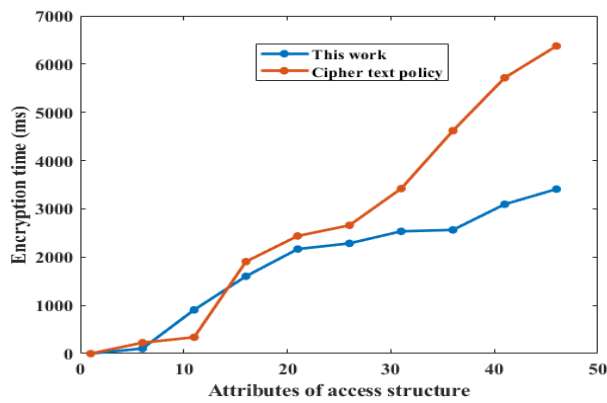
**Figure 11:** Performance Graph for Precision Vs Recall

Figure 11 demonstrates the precision vs recall graph. The precision-recall plot is a model-wide evaluation measure that is based on two basic evaluation measures - recall and precision. The recall is a performance measure of the whole positive part of a dataset, whereas precision is a performance measure of positive predictions. The figure reveals that when the recall increases, the precision of the work is decreased. The precision values change from 0.95 to 1 with the recall value, varying between 1 and 0.948.



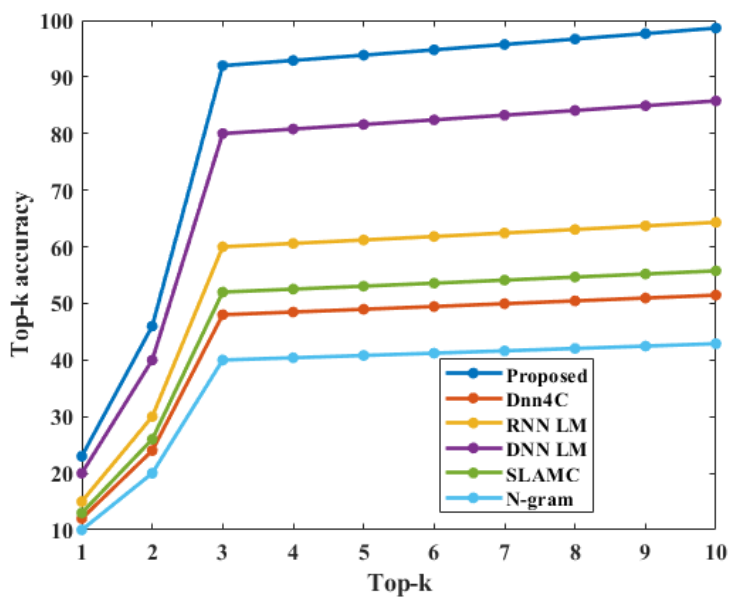
**Figure 12:** Performance Graph for F1 Score

Figure 12 reveals the performance graph for the F1 score. The F1 score takes both precision and recalls into account, which also means it accounts for both FPs and FNs. The higher the precision and recall, the higher the F1 score. Therefore, the F1 score value ranges from 7500 to 10000 during the threshold of 0 to 100.



**Figure 13:** Encryption Time Graph

Figure 13 reveals the encryption time graph for the attributes of access structure. The encryption time is used to calculate the throughput of any process of encryption, which is calculated as the total encrypted plaintext (in bytes) divided by the encryption time (in ms). The encryption time for cipher text policy ranges from 0 to 6400 ms and the encryption time for this work ranges from 0 to 3450 ms during the attribute of access structure is 0 to 50.



**Figure 14:** Top-k Accuracy Graph

The accuracy graph for the proposed Top-k method is portrayed in figure 14. The proposed method is compared with the existing Dnn4C, RNN LM, DNN LM, SLAMC, and N-gram. The accuracy of the proposed method produces 98%, which is higher than the other methods. The proposed method is approximately 3%, 6%, 7.5% 8%, and 9% higher than the existing DNN LM, RNN LM, SLAMC, Dnn4C, and N-gram methods. Therefore, the proposed method produces better performance than the other existing methods.

## 6. RESEARCH CONCLUSION

Today, there are numerous marketing tools to reach and influence customers. All the marketing tools may not be applicable for promoting every industry however, a few marketing tools may become a lifeline for a particular industry in this competitive business environment. In this research, a Secure IoT Environment with Big Data Analytics is developed. Initially, a Hadoop and Top-k query processing algorithm is introduced to the big data acquisition platform. The collected data are classified and processed by the classifier. The classified big data are compressed by the Shannon-Fano algorithm, and the data security is improved by data encryption. One of the most popular cloud computing applications is secure big data storage and sharing. Big data cloud storage and sharing, also pose significant big data security and privacy problems, while offering economy and convenience. To protect the confidentiality of big data, the system usually encrypted the big data before uploading them to the cloud. Consequently, the research encrypted using a Fully Homomorphic encryption algorithm to enhance cloud storage security. Accordingly, to maintain enhanced privacy, secure data transportation, and data access management, a Secure and Robust Data Access Management (SRDAM) Algorithm is proposed. The proposed work is implemented in the Matlab software. The evaluation metrics are precision, recall, F1 score, and accuracy. The accuracy of the proposed method is approximately 7% higher than the existing methods. Consequently, the proposed work produces better performance than the other methods, and it enhanced privacy, secure data transportation, and data access management.

## REFERENCES

- [1] Mo, L. and Yang, L., 2022. Research on Application Effective Evaluation of Artificial Intelligence Technology in Marketing Communication. *Security and Communication Networks*, 2022.
- [2] Gu, J., 2022. Research on Precision Marketing Strategy and Personalized Recommendation Method Based on Big Data Drive. *Wireless Communications and Mobile Computing*, 2022.
- [3] Nikolajeva, A. and Teilans, A., 2021. Machine Learning Technology Overview In Terms Of Digital Marketing and Personalization. *ECMS*, pp.125-130.
- [4] Zhu, M., Chakravarti, D. and Ni, J., 2022. Emerging Marketing Research on Healthcare and Medical Decision Making: Toward a Consumer-Centric and Pluralistic Methodological Perspective. *Journal of the Association for Consumer Research*, 7(2), pp.000-000.
- [5] Badica, A.L. and Mitucă, M.O., 2021. IOT-Enhanced Digital Marketing Conceptual Framework. *BRAIN. Broad Research in Artificial Intelligence and Neuroscience*, 12(4), pp.509-531.
- [6] Kumar, A., Syed, A.A. and Singh, A., 2022. Future Aspects of Digital Sustainability in Hotels: A Study on Digital Marketing Challenges with Proposed Solutions (Opportunities) during and Post COVID Era. *International Management Review*, 18, pp.79-94.
- [7] Köves, A. and Király, G., 2021. Inner drives: Is the future of marketing communications more sustainable when using backcasting? *Futures*, 130, p.102755.
- [8] Ghazal, T.M., 2021. Internet of Things with Artificial Intelligence for Health Care Security. *Arabian Journal for Science and Engineering*, pp.1-12.
- [9] Noninska, I. and Romansky, R., 2022. Organization of Technological Structures for Personal Data Protection. *International Journal on Information Technologies & Security*, 14(1).
- [10] Petrescu, M., Krishen, A. and Bui, M., 2020. The internet of everything: implications of marketing analytics from a consumer policy perspective. *Journal of Consumer Marketing*.
- [11] Sachdev, R., 2020, April. Towards security and privacy for edge AI in IoT/IoE based digital marketing environments. In *2020 Fifth International Conference on Fog and Mobile Edge Computing (FMEC)* (pp. 341-346). IEEE.
- [12] Pal, D., Arpikanondt, C. and Razzaque, M.A., 2020. Personal information disclosure via voice assistants: the personalization–privacy paradox. *SN Computer Science*, 1(5), pp.1-17.
- [13] Alkurd, R., Abualhaol, I. and Yanikomeroğlu, H., 2020. Big-data-driven and AI-based framework to enable personalization in wireless networks. *IEEE Communications Magazine*, 58(3), pp.18-24.

- [14] Kumar, K., Som, S., Tanwar, S. and Kumar, S., 2021. An IoT-Based Cryptographic Algorithm: Mapping the Credentials for Micro-application in Single Column. In *Emerging Technologies in Data Mining and Information Security* (pp. 281-287). Springer, Singapore.
- [15] Sarma, A., 2020. Healthcare Marketing in India with special reference to hospitals: Challenges, Opportunities and Strategies. *Journal of Management in Practice (Online Only)*, 5(1).
- [16] Li, Z., 2022. Accurate Digital Marketing Communication Based on Intelligent Data Analysis. *Scientific Programming*, 2022.
- [17] van Ooijen, I., 2022. When disclosures backfire: Aversive source effects for personalization disclosures on less trusted platforms. *Journal of Interactive Marketing*, p.10949968221080499.
- [18] Shen, Z., Xia, Z. and Yu, P., 2021. PLDP: Personalized Local Differential Privacy for Multidimensional Data Aggregation. *Security and Communication Networks*, 2021.
- [19] Ezhilarasan, E. and Dinakaran, M., 2021. Privacy preserving and data transpiration in multiple cloud using secure and robust data access management algorithm. *Microprocessors and Microsystems*, 82, p.103956.
- [20] Lin, Y., Shen, Z. and Teng, X., 2022. Personal Information Protection and Interest Balance Based on Rational Expectation in the Era of Big Data A Case on the Sharing of Mobile Phone Signaling Big Data in Smart City Planning. *International Review for Spatial Planning and Sustainable Development*, 10(1), pp.1-23.
- [21] Yan, C., Zhu, J., Ouyang, Y. and Zeng, X., 2021. Marketing Method and System Optimization Based on the Financial Blockchain of the Internet of Things. *Wireless Communications and Mobile Computing*, 2021.
- [22] Saura, J.R., Ribeiro-Soriano, D. and Palacios-Marqués, D., 2021. From user-generated data to data-driven innovation: A research agenda to understand user privacy in digital markets. *International Journal of Information Management*, 60, p.102331.
- [23] Jones-Smith, V., 2021. Nonprofit Leaders' Digital Marketing Strategies to Secure and Sustain Donors (Doctoral dissertation, Walden University).
- [24] Zarouali, B., Brosius, A., Helberger, N. and De Vreese, C.H., 2021. WhatsApp marketing: a study on WhatsApp brand communication and the role of trust in self-disclosure. *International Journal of Communication*, 15, p.25.
- [25] Aich, S., Sinai, N.K., Kumar, S., Ali, M., Choi, Y.R., Joo, M.I. and Kim, H.C., 2022, February. Protecting personal healthcare record using blockchain & federated learning technologies. In *2022 24th International Conference on Advanced Communication Technology (ICACT)* (pp. 109-112). IEEE.