# An Optimized Solution for Ranking Based On Data Complexity

**Sheenam Malhotra, Williamjeet Singh**

***ABSTRACT Cloud Computing is an emerging field with lot of possibilities for the maintenance at the Infrastructure Layer and Software Layer. A storage architecture is associated with two processes namely the storage and the retrieval process. The storage architecture plays a vital role in how quickly the data is retrieved. The retrieved data is presented as per the weight of the retrieved data. This paper presents a novel secure storage and the ranking mechanism for the documents for cloud. As no previous reference for any data is kept at the server, the data is encrypted based on the co-relation between the data files calculated by Cosine similarity. The ranking of the retrieved data is done through Supervised Machine learning mechanism. The evaluation of the parameters are done on the base of computation time and total number of true retrievals on multi-keyword search. Multiple dataset from Kaggle are used to perform and cross validate the proposed algorithm.***

***Keywords: Data Complexity Data Ranking Data Encryption Cosine Similarity***

## I. INTRODUCTION

Cloud computing provides a viable and reliable services to the users in terms of storing and preserving the data by enjoying the high quality services. The greater flexibility and economic savings are the motivating factors for the businesses and enterprises to manage the complex local data in the cloud(Gill and Chana (2016)). The data protection and its security are the prime concern in the cloud with data management system is an added advantage(Velumadhav Rao and Selvamani (2015)). There are various encryption methods to protect the data such as e-mail, tax documents, health records, and financial transaction etc. from the unauthorized users, enterprises encrypts the data before outsourcing to the commercial public. Although, such services obsolete the conventional data access services such as keyword searching. The huge amount of data downloading and then its de-encryption seems to be impractical as the huge bandwidth cost is the major barrier(Sookhak et al., 2017; Naeem et al., 2016). Moreover, bulk data on the cloud if not encrypted, it can be easily utilized and searched by the other party. Thus, effective privacy and search service for the encrypted data is utmost important. For instance, there is a large count of on-demand users and data files in the cloud, which is a challenging task to protect as it is difficult to meet the performance requirements as per scalability and usability level (Ji Liu et al., (2015); Gborlick et al., (2019)).

Meanwhile, the requirement of effective and reliable data retrieval in terms of large number of documents requires the ranking. However ranking criteria using parallel processing in the cloud server lags the time gap and reduces the performance of the network (Khurana et al., 2019).This is done to preserve the cloud network form undifferentiated outcomes. The efficient ranked search system helps the user to find the most relevant information in less time. In addition, this avoids the burdensome task of the cloud server, and unnecessary traffic by transferring the most relevant information to the users. Such a service 'pay as per data' is highly recommended and desirable in the cloud network. However, in case of privacy protection, such ranking criteria does not leak any confidential data such as keyword related to the information (Deshpandeand Talware, 2019). On the flip side, the accuracy of the search result has been revamped by introducing the more than two keywords rather than single keyword in the search platform(Zarezadeh et al., 2019). For example Google Scholar retrieve more information on the website, when data users provide more keywords rather than single, which often does not provide good results.In addition, coordinate matching is a paradigm in which each keyword helps to provide the result quickly such that as the number of similar matches in the web related to the data, effective are the results obtained. Such technique is widely used to retrieve the information in the plaintext community.Moreover, the similarity measure works for such system as many techniques developed in literature such as cosine similarity, soft cosine, dice coefficient etc. which group the similar data. However, providing security to the encrypted data is a difficult task due to security and inherent privacy issues, which includes strict index privacy, keyword privacy, and many other privacy obstacles. In the state of art techniques, data encryption, security and privacy techniques has been elaborated well to encrypt the data files securely and allows a search through single and multi-keywords to retrieve the data as per the requirement. Moreover, direct implementation of such approaches to secure the cloud data not confined suitable, as these techniques does not accommodate such requirements like enrich user experience, and its usability. In the past, Boolean keyword search preferred to enhance the search flexibility, but these are seldom applicable for ranking functionality (Xu et. al., 2019; Wu et al., 2019). However, designing an efficient, secure and reliable system which supports the single and multi-keyword search with its ranking stills seems to be a challenging task.In addition, there are specific risk management policies also in order to prevent the data from malicious users in the cloud.

The cloud is classified into public, private, and hybrid cloud based on the accessibility provided by the system (Wei et al., 2018; Pillai and Rao (2016)).Few security issues are described below (Chang and Ramachandran (2015); Velumadhav Rao and Selvamani (2015); Kumar et al., 2018; Aikat et al., 2017; Khan, 2016):

- Data classification issue

The cloud computing consists of three layers Software layer, a Platform layer, and Infrastructure layer. Software layer provides the user interface to the client. It helps to exploit the administrations operating on the cloud framework. The service layer is responsible for the execution of the supplied tasks, and the execution process is termed as the service.

- Data availability issue

While protecting data at a distant location which is managed by others, the data owner might face the trouble of system failure of the service provider. Moreover, if the cloud stops operating, data will not be accessed, as the information relies on a single service provider. Threats to data accessibility may cause a denial of services and Direct /Indirect (DOS) attack. Cloud computing offers on-demand service for different levels (Aikat et al., 2017). If a provided service is no longer accessible or the service quality cannot meet the Service Level Agreement (SLA), clients may lose faith in the cloud system.

- Data integrity issue

The data integrity ensures the privacy and originality of the data. It becomes vital to keep the data in its original form. A data passes several transactions in its cloud storage life span. In such a scenario, any unwanted change in the data element will ultimately result in loss of the customer. Several Data Integrity tools have been presented and analyzed till data, which monitors authorization and access of the data files and elements (Kumar et al., 2015).

- Data security issue

Security plays a vital role in any storage architecture. Cloud itself supports DAAS (Database as a Service) in which it provides space to keep the data elements. Even the PAAS and IAAS are storage dependent as an operating system which requires 10GB of space, cannot be installed over a 2 GB space.

Many security patches runs in the background who analyses the security threats in the cloud server or DAAS (Khan, 2016; Gai et al., 2016).

- Trust issue

Confidence (trust) in both current IT business, as well as cloud computing, need to be earned. One of the major problems in cloud computing is trust. Trust relates to "assurance and confidence that people, data, objects, the information will have in the cloud computation." Trust can be within, humans, machines, human to machine and machine to human. In cloud computing when any user stores their data on cloud storage, the user must have trust on the cloud provider otherwise Gmail server, yahoo server can be used as they are trusted servers (Horvath and Agarwal (2015)). Nowadays, the cloud is becoming accepted, many people use the cloud, but still, people have doubt that their data might not be safe in the cloud (user don't put their confidential information). Therefore cloud provider must deal with the trust issues so that more people may utilize the services of cloud computing.

The scientists and the researchers around the world are working hard to optimize storage performance and secured data transaction. In addition to that, relevant data retrieval is as important security of the data. Decision making and ranking makes the retrieval process more precise and accurate. Figure 1 presents the entire structure of the retrieval

The user sends the Query (Qd) to the database or computation server. The database server (db server)extracts the keywords by removing the stop words from user query, and the db server applies the matching algorithm to extract the relevant data. A ranking algorithm is implemented over retrieved data values, and it is returned to the user

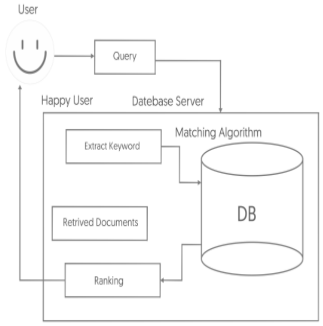


**Fig.1.User Data Process**

$\mathrm{Qd_n} = \mathrm{Qd} - \mathrm{f_{stop}}$ (1)

$\mathrm{f_{stop}} = \sum_{k=0}^{n} \mathrm{stop_{words}}$ (2)

$\mathrm{Mv_r} = \mathrm{f_x}(\mathrm{Qd_n})$ (3)

where $= \mathrm{R_x}(\mathrm{Mv_r})$ (4)

$\mathrm{Mv_r}$is the retrieved document,

$\mathrm{f_x}$is the matching algorithm and $\mathrm{R_x}$is the ranking mechanism

Researchers periodically learn and develop a new architecture for the matching and retrieval process. A lot of previous algorithms can perform the classification process with great accuracy in different domains like data classification, image processing, and speech processing. This paper presents a combinational structure of data storage and retrieval mechanism. A list of ordinal measures is as follows.

### *1.1. Data Elements*

Definition 1: Let D be a finite set of objects with $m * n$ matrix values where m is the total number of documents with n number of maximum words in the list.

$$\mathrm{D} = \sum_{i=0}^{m*n} \sum \mathrm{d_{ij}} \qquad (5)$$

### *1.2. Word To Vector*

Definition 2: The general processing of words is not easy as string matching leads to a lot of ambiguities. Word to vector is significant leap advancements for data processing. The maximum likelihood (ML) policy aims to maximize the most relevant value for a word

$\mathrm{W_i}$

$$\mathrm{W_v(h)} = \frac{\mathrm{Score(W_i)}}{\sum_{k=1}^{n} \mathrm{Word_d}} \qquad (6)$$

where h is the conversion function

### *1.3. Cosine Similarity*

If the data is to be identified based on a co-relation, then the documents must be represented in terms of the vector. This makes the data more precise and accurate for the evaluation. The proposed architecture takes cosine similarity as co-relation evaluatorfor the cloud storage and encryption selector.

Definition 3: Provided two vector set $\mathrm{V} = \{\mathrm{v_1, v_2, v_3 \ldots . v_n}\mathrm{)}$ and $\mathrm{G} = \{\, \mathrm{g_1, g_2, \ldots g_n}\}$ over equal universe U with same cardinality $|\mathrm{U}| = \mathrm{k}$then the cosine similarity is defined as follows

$$\mathrm{CosSim} = \frac{<V,G>}{||\mathrm{V}||\,||\mathrm{G}||} = \frac{<V,G>}{1^2+2^2+3^2 \ldots n^2} \qquad (7)$$

The rest of the paper is organized as follows

Section II represents the proposed solution; Section III represents the results and analysis, and Section IV concludes the paper.

## 1. Existing Work

Alsmadi and Prybutok (2018) investigates the behaviour of cloud computing considering the security and privacy as a primary concern. Research has been made to understand these perspectives and explores the effect of cloud computing services. The perceived results clear that the developed model in literature protect the confidential information. But the main limitation was that using the cloud computing services raised the accountability issues.Khan and Al-Yasiri (2016) also focussed on security and privacy issues to understand the cloud computing challenges. Therefore, interviews had been conducted and literature had been studied to determine the future and current aspects which effects the cloud computing. But, researchers wereunable to understand the threats and vulnerabilities in cloud computing. Thus, further research had been carried out to determine the security challenges.Velumadhav Rao and Selvamani (2015) highlighted the data security challenges in a cloud environment and also provided various solutions to overcome them.The data encryption techniques had been applied to enhance the security and it wasfurther verified to maintain the data integrity.The methodology presented by the scholars identifies the fraud users and securing the data processes between various data centres.But, the proposed method was inefficient in following the concrete satandards to overcome the data security challenges. In addition, there was no key for various cloud users to access the data and no access to retreive the large data files. Researchers (Chang and Ramachandran (2016); Khan, 2016)explainedthe general information, justification and components of cloud computing to safeguard information security which wasentirely based on the requirements and implementation of multi-layered security.There waslarge amount of informative data available in the data centres, which requiresreal-time protection and security. Researchers utilized the Business Process Sample Representation (BPMN) to simulate how data was used. Using the BPMN simulation allowedthem to evaluate the selected safety performances before real exercise. The results showed that the security breach could be controlled between 50 and 125 hours. This means that additional security was required to ensure that all the data is properly secured in these hours.In addition, researchers also demonstrated that cloud computing can maintain multi-layer security in real time which comprised of three layers Security: 1) firewall and access control; 2) Identity management and intervention prevention and 3) Convergent encryption according to Sookhak et al. (2017). However, studies showed that there might be a blocking percentage reduction of viruses and continuous fraud users. But the main drawback was that when the number of hours rise then security issue also surges. Meanwhile, Guo proposes a secure search technique considering the multi-keyword in the cloud environment. In the presented research, two challenges has been focussed by the scholars such as inconvenient management of the key and rank documents as per their quality. The technique is applicable for the multi-owners. Binary tree index was constructed considering the search algorithm. But, time and efficiency still poses a major barrier (Guo et al., 2018).Consequently, Fu et al. (2016) proposedan effective search scheme which wasentirely based on conventional techniques of Wang to address the keyword searching problem and data handling issues (Wang et al., 2015). A multi-keyword search scheme proposed following the past approaches to rank the data. A new method of word conversion which wasentirely based on uni-gram had been developed to ameliorate theaccuracy. In addition, the same root keywords may be questioned and solved usingstemming algorithm. The use of real world data shows that proposed scheme is practically effective and provides better results. The experimental results include the accuracy, precision, time and efficiency to conclude the results. But the main drawback of the proposed technique is it lags the search time for different query word length as the keywords increases then accuracy reduces. Therefore to overcome such alimitation, a robust ranking and searching mechanism proposed to rank the data and ease the searching procedure.

## 2. The Proposed Solution

The proposed solution is divided into two parts, namely storage and retrieval

### *2.1. Storage*

The document set D has m*n data files. To ensure that D is stored securely, D is encrypted utilizing encryption algorithm $\boldsymbol{E(E_1, E_2, E_3)}$. To select an algorithm out of $\boldsymbol{E}$, a sophisticated encryption selection algorithm is implemented. The first step is to apply the cosine similarity over $\boldsymbol{D}$.

**Algorithm 1: The Similarity Calculator**

```
1. $\mathrm{Cosine_{sim}} = \mathrm{functionCosineSimilarity(docs)}$
2. The input of this calculator is the numeral vector
   of each data file  //Cosine_doc_similarity =
   [ ]; //
```

```
3.  Empty array for the calculation of similarity index $\mathrm{Sim_{count}} = 0$; // Total number of identified similarities
    For $\mathrm{m = 0\ to\ docs.length}$ //
4.  For 1 to total number of the data files $\mathrm{Current_{doc} = docs\,(m)}$; // Doc I For $\mathrm{n = I + 1\ to\ docs.length}$ //
5.  Next data value series
    $\mathrm{L\ |Cosine\_doc(current\_doc) - cos(docs(n))|; Cosine\_doc\_similarity[similarity\_count, (}$
    current_doc // The similarity measure has three columns
6.  $\mathrm{Cosine_{doc\_similarity}[similarity_{count}, 1] = docs(n)}$;// First column is the main data file, Second Column is Connecting data file
7.  $\mathrm{Cosine_{doc\_similarity}[similarity_{count}, 2] = L}$;
8.  // The similarity value
9.  $\mathrm{Sim_{count} = Sim_{count} + 1}$; // Counter incremented by 1
10. Endfor
11. Endfor
12. Endfunction
```

Based on the cosine similarity of the data files at the time of storage, an encryption selection algorithm termed as Complexity evaluator is formed.

**Algorithm 2: The complexity Evaluator**

```
1.  The complexity Evaluator$\mathrm{(Similarity_{Indexes})}$
2.  Inputs: $\mathrm{Similarity_{indexes}}$
3.  Output: $\mathrm{Selected_{encryption}Algorithm}$
4.  $\mathrm{A_C} \dashrightarrow \mathrm{Average_{Complexity}} = \sum_{k=0}^{n} \frac{\mathrm{Similarity}}{\mathrm{n}}$
5.  Foreach $\mathrm{data_{file}}$in $\mathrm{Data_{list}}$
6.  $\mathrm{f = search(data_{file}in\ Data_{list})}$
7.  $\mathrm{s} = \frac{\mathrm{Similarity(f)}}{\mathrm{Total_{count_f}}}$
8.  $\mathrm{C_R} \dashrightarrow \mathrm{Complexity_{Range}} = [\frac{\mathrm{A_C}}{4}\frac{\mathrm{A_C}}{2}\mathrm{A_C}]$
9.  If $s > \mathrm{C_R}(1) \& s < \mathrm{C_R}(2)$
10. $\mathrm{Selected_{Encryption_{Algorithm}}} = 1$
11. ElseIf $s > \mathrm{C_R}(2) \& s < \mathrm{C_R}(3)$
12. $\mathrm{Selected_{Encryption_{Algorithm}}} = 2$
13. ElseIf $s > \mathrm{C_R}(3)$
14. $\mathrm{Selected_{Encryption_{Algorithm}}} = 3$
15. End If
16. End For
```

The proposed algorithm selects 3 encryption algorithms at the initial level. The proposed algorithm computes the average similarity index of the entire document set.If $\mathbf{D}$ is a document set having 4 documents $\mathbf{D} = \{\mathbf{d1}, \mathbf{d2}, \mathbf{d3}, \mathbf{d4}\}$, then the similarity index will be calculated in the following pattern.

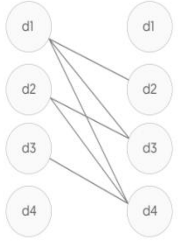


**Fig.2.Similarity Calculation**

Let $\mathbf{SS}$ be the similarity vector,

$\mathbf{SS} = \{(\mathbf{d1}, \mathbf{d2}), (\mathbf{d1}, \mathbf{d3}), (\mathbf{d1}, \mathbf{d4}), (\mathbf{d2}, \mathbf{d3}), (\mathbf{d2}, \mathbf{d4}), (\mathbf{d3}, \mathbf{d4})\}$.

Let $\mathbf{S_v}$is the vector set of the similarity value.

$$\mathbf{S_v} = \mathbf{R_v}(\mathbf{SS}) \qquad (8)$$

The similarity calculator decides which encryption algorithm should be applied over which data frame which not only secures the data element but also reduces the space complexity of the cloud storage. Redundant bit pattern may increase the space complexity of the cloud structure.

### *2.2. Retrieval and Ranking*

The retrieval process has to follow the storage process also as if the data is encrypted, and the search term should also be encrypted to support the architecture. The query which is supplied by the user follows Algorithm 1 and Algorithm 2 to search the relevant data. The retrieval of data elements is done on the base of string matching pattern, and ranking is done utilizing Artificial Neural Network. Algorithm 3 presents the architecture of ANN for the retrieval process.

**Algorithm 3: The ranking structure**

```
1.  Input: $\mathrm{Retrieved_{Document_{Set(RDS)}}}$ , $\mathrm{Orignial_{Data}}$
2.  Output: Trained Neural Network
3.  $\mathrm{Training_{Data} = [\ ]}$
4.  Foreach rt in RDS
5.  $\mathrm{Data_{value} = Original_{Data}(rt)}$
6.  $\mathrm{Weight_{value} = Weight(Data_{value})}$
7.  Foreach Wv in $\mathrm{Weight_{value}}$
8.  $\mathrm{Training_{Data}[rt\ , wv] = Weight_{value}[wv]}$
9.  $\mathrm{Target_{Matrix}[\ wv\ ] = rt}$
10. End For
11. End For
12. // Initialize Neural Network with n number of neurons
13. $\mathrm{Neural_{object}} =$
    $\mathrm{Initialize_{Neural}(Training_{Data}, Target_{Matrix}, n\ )}$
14. Train_Neural()
15. End Algorithm
```

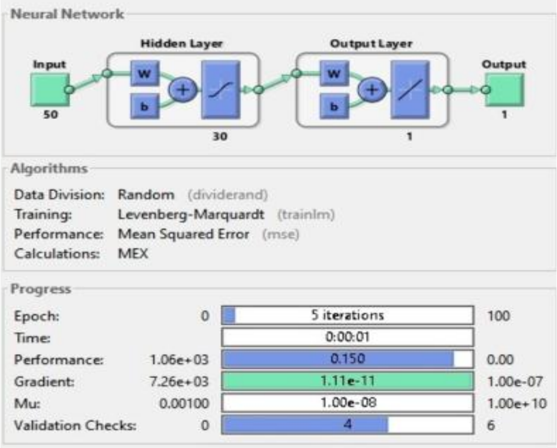
The training algorithm takes two elements as inputs namely $\mathbf{Retrieved_{Document_{Set(RDS)}}}$and the$\mathbf{Original_{Data}}$. If the retrieved document vector is $\mathbf{d1}, \mathbf{d2}, \mathbf{d3}, \mathbf{d4} \ldots \mathbf{dn}$, and each document has words in them which are relevant to the search term.

Let **d1** has weight value**3 43  33 46** , **d2**has**11 34 23 21** , **d3** has **54 55 56 44** and **d4** has **36 67 77 55** then the training matrix would be as follows [**23 43 33 46**: **11 34 23 21** : **54 55 56 44**: **36 67 77 55**] and associated group will[ **1**: **2**: **3**: **4**]. It will propagate through the Neural Architecture. A supervised machine learning of Feed Forward Back Propagation Neural Network is applied. Here all the retrieved document files are important so they must be visible in the result but as each matched file is relevant to query hence a supervised mechanism is applied here. The ordinal measures for Neural Network are as follows:

**Table 1.Propagation Details**

| **Total Passed Propagation Iteration** | **100** |
|---|---|
| Validation Parameters | Time<br>Gradient<br>Mutation |
| Cross-Validation Parameter | Mean Square Error |
| Outcast | Regression Model |
| Propagation Architecture | LevenBerg |
| Calculation Type | MEX |
| Data Distribution | Random |

The proposed work model takes the data propagation as the regression model, which is followed by the Levenberg architecture. The data distribution is random for the regression model. The design pattern for the regression model is as follows.



Algorithms

Data Division: Random (dividerand)
Training: Levenberg-Marquardt (trainlm)
Performance: Mean Squared Error (mse)
Calculations: MEX

Progress

| | | | |
|---|---|---|---|
| Epoch: | 0 | 5 iterations | 100 |
| Time: | | 0:00:01 | |
| Performance: | 1.06e+03 | 0.150 | 0.00 |
| Gradient: | 7.26e+03 | 1.11e-11 | 1.00e-07 |
| Mu: | 0.00100 | 1.00e-08 | 1.00e+10 |
| Validation Checks: | 0 | 4 | 6 |

**Fig.3.**The propagation architecture of proposed work model

As shown in Fig.3. the total supplied iterations are 100. When any of the satisfying parameters are satisfied, the propagation of the weight is stopped. It is observed that even if the hidden neurons are varied, the simulation stopped within a range of 5-15. Table 2. demonstrates the variation details of Neurons

**Table 2.**

Variation of Neuron Structure

| **Total Neurons** | **Satisfied Structure at Neuron Epoch** |
|---|---|
| 10 | 5 |
| 12 | 6 |
| 15 | 5 |
| 20 | 7 |
| 25 | 10 |
| 30 | 12 |

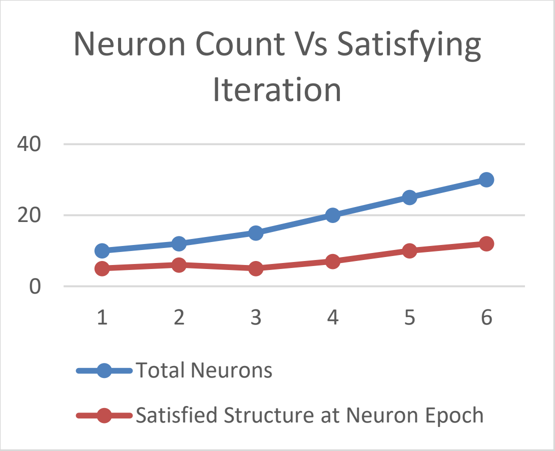


**Fig.4.**Neuron variation with Satisfying Epoch

The regression model varies the input structure by a marginal threshold by lambda.The regression model ensures that the validation is structured and the classification structure is supported for accurate results

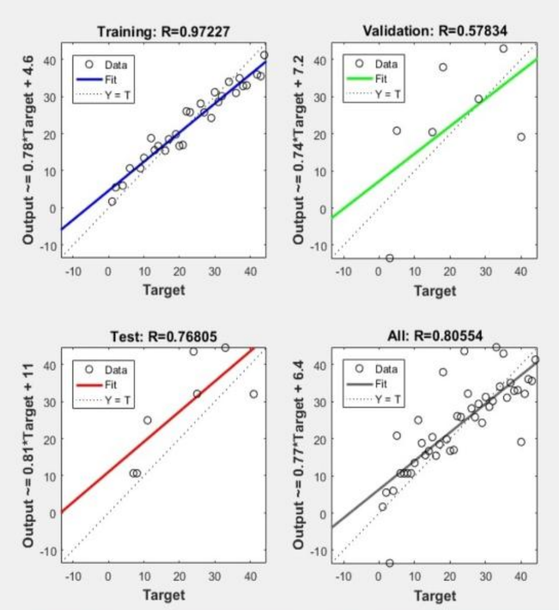


**Fig.5.**Regression Model

The regression value R is said to be as good as it is close to 1. So R lies between 0 to 1 i.e0 ≤ R ≤ 1. The R-value of the proposed architecture has an initial value ~.97and ends up with ~.80 on an average.

## 3. Dataset

Three datasets have been considered for the analysis of the proposed work. Their utilization concerning the research is defined below:

### *3.1. FIFA 18 dataset*

i. Content

Dataset for those who love data science and have grown up playing FIFA.

ii. Content

1. Every player featuring in FIFA 18
2. 70+ attributes
3. Player and Flag Images
4. Playing Position Data
5. Attributes based on actual data of the latest EA's FIFA 18 game
6. Attributes include on all player style statistics like Dribbling, Aggression, and GK Skills, etc.
7. Player personal data like Nationality, Photo, Club, Age, Wage, Salary, etc.

The data is scraped from the website https://sofifa.com by extracting the Player personal data and Player Ids and then the playing and style statistics

### *3.2. ERON Email Dataset*

The Eron email dataset contains approximately 500,000 emails generated by employees of the Enron Corporation. The Federal Energy Regulatory Commission obtained it during its investigation of Enron's collapse.This is the May 7, 2015 Version of the dataset, as published at https://www.cs.cmu.edu/~./enron/

### *3.3. Financial Dataset for Fraud Detection*

There is a lack of publicly available datasets on financial services and especially in the emerging mobile money transactions domain. Financial datasets are essential to many researchers and in particular to us performing research in the domain of fraud detection. The data set is available at https://www.kaggle.com/ntnu-testimon/paysim1.

## 4. Results and Discussion

The evaluation of the proposed work has been divided into two sections (a) Results Analysis (b) Comparative Study. The result section is based on the following parameters.

### *5.1 Results Analysis*

Ranking Time vs. Total Result Values: It is the time to rank the data elements which are classified against the provided query. The evaluation is done in seconds.

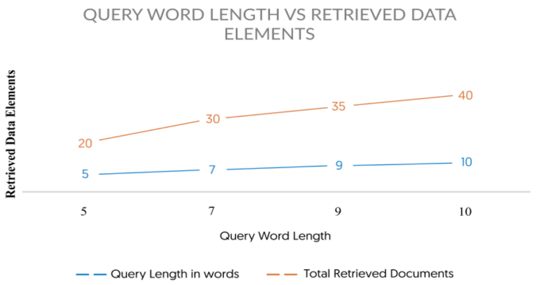


**Fig.6.**Total Query Word Length vs. Retrieved Data Elements

Fig.6. shows that a maximum word length of 10 words is passed as the query line. 10wordscontain the stop words as well. A maximum of 40 data elements is retrieved for 10 words. Similarly, for 9 words, the retrieved data elements are 35. Fig. 7 demonstrates the ranking time for the retrieved data values.

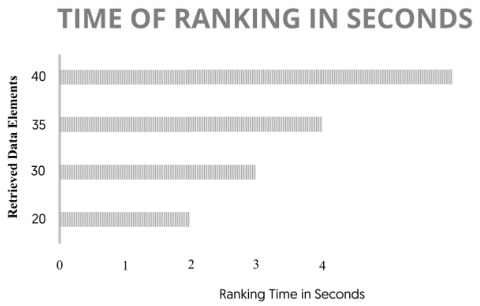


**Fig.7.**Total Consumed Time in Seconds vs passed query

When there are more keywords in the query, obviously there would be more related data elements. This truth is no different in this case. For a query word length of value 10, a total of 40 data elements are extracted from the repository, and a total of 6 seconds is consumed to rank the elements. Minimum passed word length is 20, and the total completion retrieval time is 2 seconds. Fig.8. and Fig.9 represented the overall performance of the proposed algorithm, whereas we have also evaluated the parameter even for each data set individually also. However, ranking time entirely based on the search time of the file. The search time has been computed considering the query word length.

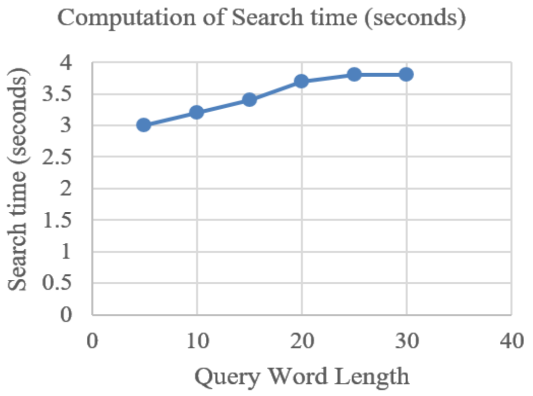


**Fig.8.**Total Search time in Seconds vs number of keywords

The given Fig. 8 clearly demonstrates the search time for number of different keywords. When the keyword length is 10 then search time is almost 3 seconds. When the number of keywords increases then search time also surges to 3.7 seconds. But after 20 words it again rises to 3.5 seconds. Thus, it is clear that as the number of keywords increases then search time also increases.Table 3. represents the values of retrieval and ranking details of all the three data sets.

**Table 3. bRetrieval and Ranking Details**

| Data Set | Query Length in Words | Retrieved Data Elements | Time to Rank in seconds |
|---|---|---|---|
| FIFA 18 | 5 | 25 | 2.85 |
| FIFA 18 | 7 | 37 | 2.63 |
| FIFA 18 | 10 | 41 | 9.36 |
| ERON EMAIL | 5 | 18 | 2.1 |
| ERON EMAIL | 7 | 24 | 3.56 |

| | | | |
|---|---|---|---|
| ERON EMAIL | 10 | 34 | 7.14 |
| FINANCIAL FFD | 5 | 23 | 2.14 |
| FINANCIAL FFD | 7 | 31 | 3.65 |
| FINANCIAL FFD | 10 | 42 | 6.12 |

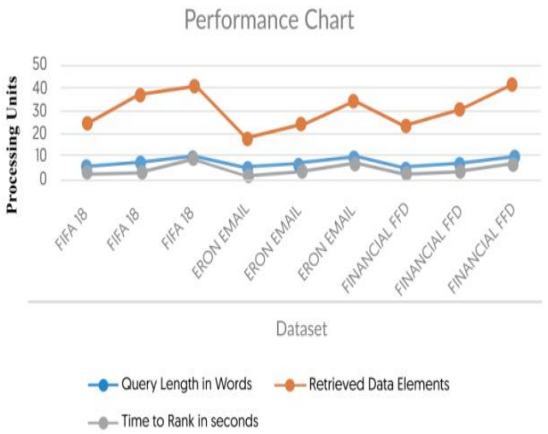


**Fig.9.**Performance Chart of Individual Dataset

Fig.9 is the pictorial representation of the proposed work model. The best performance out of all the three data sets is evaluated for Financial FFD, which stands a quick 6.12 seconds against the ranking of 42 evaluated document set. The proposed algorithm is not flaw-full, but the data of this dataset is little precise to the topic, and hence, the evaluation is quick. The worst performance is noted to be for FIFA 18 dataset, which is due to its massive amount of attributes and intricate architecture.

### *5.2 Comparative Analysis of the proposed work with the past studies*

In this section, a comparative analysis has been presented with the existing work to determine the effectiveness of the proposed work.

Fig. 10 clearly shows that the search time of the proposed work is less than the existing work. The average time taken by the existing work (Fu et al. (2016)) is 4.51seconds. However, proposed work take search time of 3.48 seconds. The average computed search time of both systems lucid that performance of the proposed work is better than the existing work. The overall improvement in search time is $\frac{4.51-3.48}{4.51} \times 100 = 23\%$.

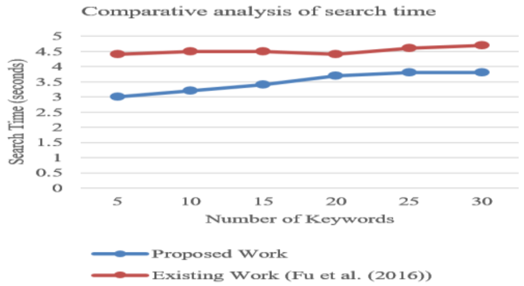


**Fig. 10.**Comparison of Search time with the existing work

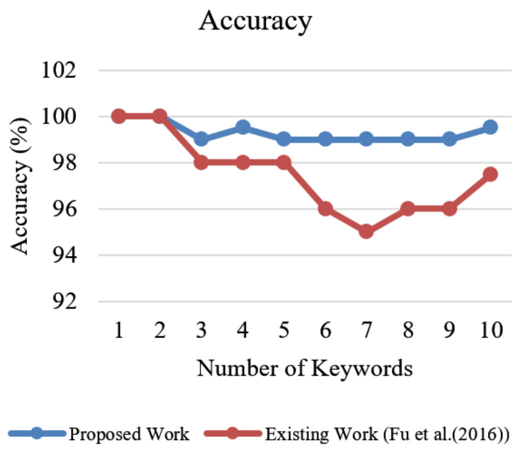


**Fig. 11**. Comparison of Accuracywith the existing work

Figure 11 compares the proposed work in terms of accuracy for searching the keywords with the existing work (Fu et al. (2016)). The given results clear that the proposed method outperforms the existing technique. The average accuracy obtained by the proposed method is 99.3% while that of past technique, it is 97.4%. Thus, the overall accuracy has been improved by 2%. In addition, efficient results have been obtained which secures the data and reduces the search time.

## 5. Conclusion

The proposed algorithm develops a new ranking algorithm based on Feed Forward Back Propagation Neural Network. The proposed architecture utilizes Cosine similarity to evaluate the co-relations between the documents. The correlation values are passed to the training architecture of Feed-Forward and propagate back based on the satisfaction parameters. A total of 100 simulation iterations were supplied at the training layer, and it is observed that the training gets completed within 20 simulation iterations. The classified architecture matches the output label to the target label, and the matched values stand a high precision as compared to the other available data elements. A cross-layer regression is also performed to ensure the best results for the ranking structure. A total of 3 different datasets, namely FIFA 18, ERON EMAIL, and FINANCIAL FFD is utilized for the evaluation of the proposed algorithm. The evaluation of the performance of the proposed algorithm is based on time to rank the retrieved documents. The performance of the proposed algorithm was best for FINANCIAL FFD in which the proposed algorithm ranked 42 data elements in about 6.12 seconds whereas the proposed algorithm showed a little slower response for FIFA 18 dataset in which it ranked 41 data elements in about 9.38 seconds. In addition, the proposed work compared with the state of art technique in terms of search time. The results have been improved by 23% in comparison to past techniques. The proposed work model leaves a lot of futuristic approaches. It would be interesting to see any structural algorithm at the classification of the retrieval part of this proposed algorithm. Algorithms like SVM, ANN can be utilized.

## REFERENCES

1. Alsmadi, D. and Prybutok, V., 2018. Sharing and storage behavior via cloud computing: Security and privacy in research and practice. Computers in Human Behavior, 85,218-226.
2. Aikat, J., Akella, A., Chase, J.S., Juels, A., Reiter, M., Ristenpart, T., Sekar, V. and Swift, M., 2017. Rethinking security in the era of cloud computing. IEEE Security & Privacy.
3. BursellM.. U.S. Patent No. 9,251,115. Washington, DC: U.S. Patent and Trademark Office,2016.
4. Chang, V. and Ramachandran, M., 2015. Towards achieving data security with the cloud computing adoption framework. IEEE Transactions on Services Computing. 9(1), 138-151.
5. Fu, Z., Wu, X., Guan, C., Sun, X. and Ren, K., 2016. Toward efficient multi-keyword fuzzy search over encrypted outsourced data with accuracy improvement. IEEE Transactions on Information Forensics and Security. 11(12), 2706-2716.
6. Gai, K., Qiu, M., Sun, X. and Zhao, H., 2016. Security and privacy issues: A survey on FinTech. In International Conference on Smart Computing and Communication. Springer, Cham, 236-247.
7. Gborlick,M., Gupta, L. M., R. G.Hathorn and K. A. Nielsen. 2019. U.S. Patent Application No. 10/171,585-612.
8. Horvath, A. S., and AgrawalR., 2015. Trust in cloud computing. SoutheastCon. Fort Lauderdale. FL, 1-8.
9. Kumar, V., Reddy, N.C.S. and Reddy B.S.,2015. Preserving Data Privacy, Security Models and Cryptographic Algorithms in Cloud Computing. International Journal of Computer Engineering and Applications.7(1), 71-82.
10. Khan, N. and Al-Yasiri, A., 2016. Identifying cloud security threats to strengthen cloud computing adoption framework. Procedia Computer Science, 94, 485-490.
11. Khan,M. A., 2016. A survey of security issues for cloud computing. Journal of network and computer applications. 71, 11-29.
12. Kumar, P.R., Raj, P.H. and Jelciana, P., 2018. Exploring data security issues and solutions in cloud computing. Procedia Computer Science, 125,691-697.
13. Liu, J., Pacitti, E., Valduriez, P. and Mattoso, M., 2015. A survey of data-intensive scientific workflow management. Journal of Grid Computing, 13(4),457-493.
14. Liu, J., Wang, S., Zhou, A., Kumar, S.A., Yang, F. and Buyya, R., 2016. Using proactive fault-tolerance approach to enhance cloud service reliability. IEEE Transactions on Cloud Computing, 6(4), 1191-1202.
15. Naeem M. M., Mahar H., Memon F. and Siddique M., 2016. Overview of virtualization in cloud computing, In Colossal Data Analysis and Networking (CDAN), Symposium, IEEE, 1-4.
16. Pillai, P.S. and Rao, S., 2016. Resource allocation in cloud computing using the uncertainty principle of game theory. IEEE Systems Journal, 10(2), 637-648.
17. Singh, S. and Chana, I., 2016. QoS-aware autonomic resource management in cloud computing: a systematic review. ACM Computing Surveys (CSUR), 48(3), 42-51.
18. Sookhak, M., Abdullah G., Khan K. and Buyya R. 2017. Dynamic remote data auditing for securing big data storage in cloud computing, Information Sciences, 380,101-116.
19. Velumadhav Rao, R. and Selvamani K., 2015. Data security challenges and its solutions in cloud computing, Procedia Computer Science, 48, 204-209.
20. Wei, W., Fan, X., Song, H., Fan, X. and Yang, J., 2016. Imperfect information dynamic stackelberg game based resource allocation using hidden Markov for cloud computing. IEEE Transactions on Services Computing, 11(1), 78-89.
21. Xu, P., Tang, S., Xu, P., Wu, Q., Hu, H. and Susilo, W., 2019. Practical Multi-Keyword and Boolean Search over Encrypted E-mail in Cloud Server. IEEE Transactions on Services Computing, 11(2), 88-98.
22. Wang, J., Chen, X., Huang, X., You, I. and Xiang, Y., 2015. Verifiable auditing for outsourced database in cloud computing. IEEE transactions on computers, 64(11),3293-3303.
23. Deshpande, P.V. and Talware, U.L., 2019. Secure Ranked Keyword Search Method with Conditional Random Fields over Encrypted Cloud Data. In Innovations in Computer Science and Engineering. Springer, Singapore,455-462.
24. Wu, Z., Li, K., Li, K. and Wang, J., 2019. Fast Boolean Queries with Minimized Leakage for Encrypted Databases in Cloud Computing. IEEE Access, 7, 49418-49431.
25. Khurana, A., Krishna, C.R. and Kaur, N., 2019. Improved Ranking for Search Over Encrypted Cloud Data Using Parallel Index. In International Conference on Advanced Computing Networking and Informatics, Springer, Singapore,97-107.
26. Zarezadeh, M., Mala, H. and Ashouri-Talouki, M., 2019. Multi-keyword ranked searchable encryption scheme with access control for cloud storage. Peer-to-Peer Networking and Applications,1-12.
27. Guo, Z., Zhang, H., Sun, C., Wen, Q. and Li, W., 2018. Secure multi-keyword ranked search over encrypted cloud data for multiple data owners. Journal of Systems and Software, 137, 380-395.