Automated Bengali Document Summarization By Collaborating Individual Word & Sentence Scoring

Porimol Chandro[1]
Md. Faizul Huq Arif[4]
Department of Computer Science & Engineering,
World University of Bangladesh(WUB)
{porimolchandroroy[1], arifict27[4]}@gmail.com

Md. Mahbubur Rahman[2]
Md. Saeed Siddik[3]
Mohammad Sayeedur Rahman[6]
Institute of Information Technology(IIT),
Md. Abdur Rahman[5]
Centre for Advanced Research in Sciences (CARS),
University of Dhaka
{de.fahimmf[2], mukul.arahman[5],
de.nayeem[6]}@gmail.com
saeed.siddik@iit.du.ac.bd[3]

***Abstract*— Bengali documents are increasing on the World Wide Web and it is becoming a overwhelming problem for the increasing large number of web users to reviewing and reduce the information. Many researches have been conducted in the field of Natural Language Processing for English documents and in order to serve with satisfactory accuracy. This research work proposed a simple and powerful extraction based method for summarizing of the Bengali text documents. The system could summarize a single document at a time. The ultimate objective of the proposed methodology helps readers to get summary and insight of the Bengali documents without reading revealing the in-depth details. In the proposed Bengali documents summary generation method there are four features: Preprocessing, Sentence Ranking and Summarization, Combining Parameters for Sentence Ranking, Summary Generator. The results of performance evaluation show that the average scores of Precision, Recall and final scores are 0.80, 0.67, and 0.72 respectively.**

***Index Terms*— Bengali Document Summarization; Text Extraction; Information Retrieval; Word Tokenization; Word Stemming; Sentence Scoring; Sentence Ranking**

I. INTRODUCTION

Bengali is one of the most spoken languages in the world. It is the national language of Bangladesh and second-most spoken language in India. And the seventh most spoken language in the world. Day by day Bengali documents are increasing on the WWW and it is becoming a problem for the increasingly large number of web users in order to review and reduce the information overload problem. Moreover it will also take lots of effort and time-consuming task to evaluate these documents. Therefore, that a automated document summarization system can be an indispensable tool. The summarized document can help readers by giving an overview of an entire document.

As the discussion in the introduction part of the study shows the primordial goal to achieve the method as well as the algorithm by which we can automatically summarize the input document. Many studies have been conducted on automated single or multi-document summarization[1][2] on the Bengali language. The proposed method does not need a pre-processed document for performing summarization. It can be applied to uncleaned or raw document directly. The proposed method can be automatically processed uncleaned document without any human intervention

Bengali is an inflected language, generating the summarization from Bengali document is a difficult task in the field of natural language processing. If we achieve the most effective method to extractive based Bengali document summarization, it will be better for the readers to extract an overview of the document reduces the reading time and also reduces the information overload problems.

The rest of the paper is organized as follows: In section II describes related works. Section III illustrates proposed methodology. In section IV discussed about Experiment, Results and Comparison with other methodology. Finally, the conclusion is turned in section V.

II. RELATED WORK

In this section of the research study, an overview of single document summarization for Bengali language is presented. Although new research on the domain of Bengali text summarization has been started many years ago. In previous, most of the works on the domain of text summarization has been done based on the sentence extraction.

The article[3] proposed a method based on the survey of different text summarization techniques. They accomplished a survey on different text techniques and implemented an extraction based Bengali summarizer. They presented five features in the proposed methodology:

(i)Location method (ii)Cue method, (iii)Title, (iv)Term frequency and (v)Implementation.

K. Sarkar[4] proposed a method, which is providing an overview of the theme of a document without reading top to down of a document. The proposed method has three major steps (i)Preprocessing, (i)Sentence Ranking, (iii)Summary Generation. It is also based on Term Frequency(TF) ,Inverse Document Frequency(IDF), Positional Value(PV).

M. M. Haque et al.[5] presented a method, they proposed an extraction based summarization technique for summarized Bengali document. Their method has four major steps: (i)Preprocessing, (ii)Sentence Ranking, (ii)Sentence Clustering, (iv)Summary Generation. The noticeable features of this proposed method are the incorporation of the sentence frequency and sentence clustering on the basis of similarity ratio among sentences.

M. I. A. Efat et al.[6] proposed an extraction based summarization method which works on the Bengali documents. It is able to summarize single document at the same time. Their proposed method has two major steps: (i)Pre-processing, (ii)Sentence Ranking & Summarization.

A. Das and S. Bandyopadhyay[7] mentioned a method, which identifies the sentiment from the documents, aggregates them and finally represent as a summary document. They utilized a topic sentiment model for sentiment retrieve and aggregation. The aggregation is based on the theme clustering (K-means)and document level theme relational graph algorithm representation and finally generate summary selecting by standard page rank algorithms for Information retrieval.

III. PROPOSED METHODOLOGY

The Bengali document summarizer is a Natural Language Processing(NLP) application which is proposed to extract the most important information of the document. In automatic summarization, there are two distinct techniques either text extraction or text abstraction.

Extraction is a summary consisting of a number of sentences selected from the input document. An abstraction based summary is generated where some text units are not present in the input document. With extraction based summary technique, some features are added based on Information Retrieval.

The total system is segmented into four parts:

- Pre-processing
- Sentence Ranking and Summarization
- Combining Parameters for Sentence Ranking
- Summary Generator

The fig. 1 demonstrates the block diagram of the proposed methodology. The input to summarization process can be made from one or more text documents. When only one document is the input, it is called single document text summarization but in multi- document summarization the input is a group of related text documents.

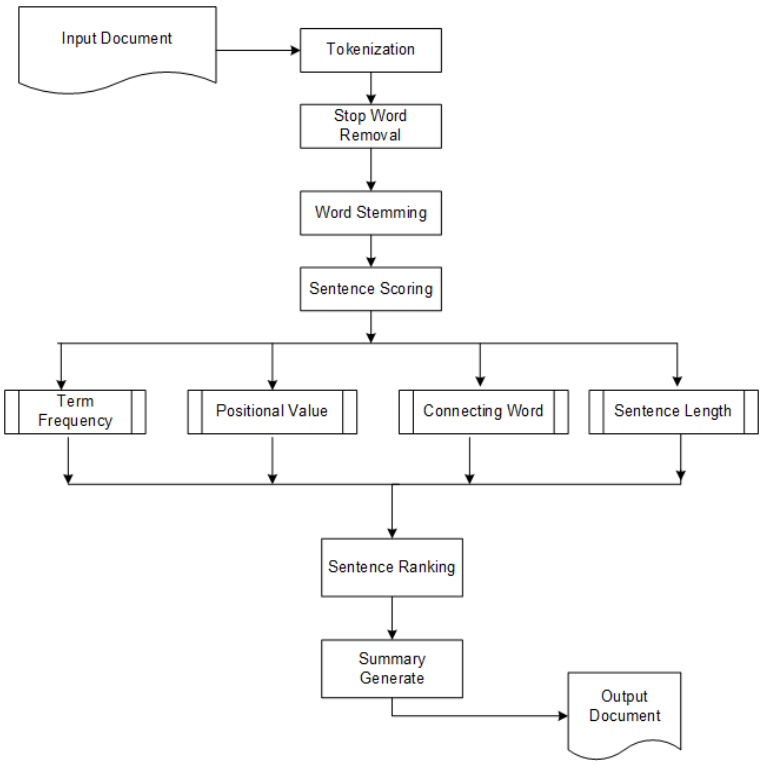

Fig. 1. Block diagram of the Proposed methodology

1) Pre-processing: In Bengali document summarization process, some preprocessing is needed before executing the sentence scoring algorithm. By the pre-processing, the document is prepared for ranking and summary generation. The preprocessing is done on the documents are as follows:

Document: Document is a large collection of texts. In the field of natural language processing document is also called corpus. Number of documents has been collected for testing the summarizer tool from various popular bengali daily news portal such as The Prothom Alo, Anandabazar Patrika, Bangladesh Protidin etc.

Tokenization: Tokenization is the process of chopping it up into segments, called tokens. And also a document is the collection of sentences and a sentence is the collection of words. Here every word is measured as a token. The input document is broken into a collection of sentences.

Stop words removal: In Bengali words like অথবা(Or), কিন্তু(But), etc. are used frequently in sentences which have little importance in a document. These words can simply be removed for the classification process.

Word Stemming: In NLP, word stemming is the process of finding the root word. The main objective of word stemming is to reduce the inflections forms of a word. There are many research has been conducted on word stemming[8][9][10][11]. The lightweight Bengali rule-based word[12][13] stemming algorithm has been used for word stem. For instance in Bengali word like গাছগুলো(Trees) should be গাছ(Tree), গরুগুলো(Cows) should be গরু(Cow).

2) Sentence Ranking and Summarization: After an input document is tokenized the sentences are ranked based on four important features:

- Term Frequency(TF)
- Positional Value(PV)
- Connecting Words(CW)
- Sentence length of the document(SLD)

TF: Frequency is the number of times a word appears in the document. If a word's frequency is high then it can be said that the word has a significant effect on the document.The total frequency of a sentence is calculated by sum up the frequency value of every word in particular sentence. The equation (1) used to estimate the total frequency value of a sentence k is n

$$STF_k = \sum_{i=1}^{n} WF_i \tag{1}$$

where WF is the total frequency of a word in the input document. n is the number of words in the sentence and STF means sentence total frequency. K stands for a sentence.

PV: The position of a sentence in a document has a considerable influence on the content of the document. The positional value of a sentence is calculated by assigning the highest value of the first sentence and the lowest value to the last sentence of the document. The position value PV is calculated using the formula (2)

$$PV_k = \frac{1}{\sqrt{k}} \tag{2}$$

Where, k is the actual positional value of a sentence in the document.

CW: Connecting words are connective expressions(such as therefore, hence, lastly, finally, meanwhile or on the other hand) that link spans of communication and signals semantic relations in a text. This is one of the summarization techniques that helps to select the important sentences in a document. The examples of "connecting Words" in Bengali are এককথায়, মোটকথা, অবশেষে, ইতিমধ্যে, ইত:পূর্বে etc.

SLD: The length of a sentence is considered as a feature because if a sentence is too short, but it occurs in the beginning part of a document it is sometimes selected due to its positional advantage. On the other hand, if a sentence is too long, it is sometimes selected due to the reason that it contains many words. The sentences which are too short are removed at the time of candidate summary sentence.

3) Combining Parameters for Sentence Ranking: The final score of a sentence is computed by using a linear combination of the values of three score STF(k), PV(k), CWS(k). The final score of a sentence k is (3)

$$SCORE_k = \alpha STF_k + \beta PV_k + CWS_k, 0 <= \alpha, \beta <= 1 \tag{3}$$

Where STF means total term frequency of a sentence, PV means the positional value of a sentence and CWS means Connecting word score of a sentence.

4) Summary Generator: A summary is produced after ranking the candidate summary sentences based on their scores and selecting K top-ranked sentence where k is computed by the equation (4)

$$TRS = Ceil(\sqrt{N}) \tag{4}$$

In the equation no. 4, TRS is stand for top-ranked sentence, N is the number of total candidate sentences.

For testing the method number of Bengali documents has been collected from the Bengali daily newspapers, such as Daily Prothom Alo, Anandabazar Patrika, Kaler Kantho, Bangladesh Protidin. The documents are scraped and saved in the text files using the UTF-8 format that is successfully done in order to test our proposed methodology.

IV. EXPERIMENT, RESULT and COMPARISON

A. Experimental Setup:

For experimental purpose there are required some materials to finish the proposed method. Those are Bengali Document, Python, Flask(A Python Microframework), PyCharm Edu and a Workstation.

Python: Python is the well-known programming language in the field of research and decelopment works and we have utilized it's advantages. It is one of the most used programming languages in technical computing for it's batteries. It has many rich and matured libraries for natural language processing such as NLTK, Spacy, TexBlob and so on. In this experiment Python 3 has been used.

Flask: Flask is python's microframework for building web applications. In order to experiment, there used the advantages of Flask for building a web application to show the experimental work over the internet.

PyCharm Edu: PyCharm Edu is an Integrated Development Environment(IDE) developed by Jet Brains. It is free for academia only and used it is used in this experiment for rapid development and testing.

Workstation: The local workstation compatible with the mentioned version of Python, Flask, PyCharm Edu. In the time of experirment the workstation configurations was:

- Operating System: Ubuntu 16.04
- CPU: Corei7, 7th Generation
- RAM: DDR4 8GB
- Secondary Storage: SSD 128GB with 1TB HDD
- Graphics: NVIDIA Geoforce GTX 4GB

B. Datasets:

In order to test the proposed method Bengali document has been collected from various online Bengali news portal such as Prothom Alo, Kaler Kantho, Bangladesh Protidin, Anandabazar Patrika.

C. Case Study:

To achieve the objectives of proposed methodology few stages such as Pre-processing the test document, Word scoring, Sentence scoring finally summarization based on sentence ranking has been done. The proposed method has a novel thing that summarizes the given Bengali document properly.

To summarize Bengali document text extraction method has been used. Extraction is a summary consisting of a number of sentences selected from the input document. The total system is segmented into four parts for getting the better outcome for summarizing the document.

In Preprocessing portion, input a raw document and cleaned punctuation after then tokenized,stop words remove and finally stem the tokens. When document preprocessing once done then Scoring The Words, here we count the stemmed word in a document, count the positional value, Cue Words and Sentence Length. After then scoring and ranking the sentences. And finally generating the summary of the given document as input.

D. Results:

A system generated summary compared to a reference summary and the unigram based Recall score, Precision and Final score are computed. In tab. 1 illustrates the comparison between the proposed method and the two summarization approaches presented in the research articles

This example has taken from The Daily Prothom-alo Dated ২৭ ডিসেম্বর ২০১৭, is shown below: Original text:

কক্সবাজারের মহেশখালীতে বুধবার সন্ধ্যায় বাংলাদেশ বিমানবাহিনীর দুটি প্রশিক্ষণ বিমান বিধ্বস্ত হয়েছে। বিমানে থাকা চার বৈমানিককেই জীবিত উদ্ধার করা সম্ভব হয়েছে। তাঁরা আশঙ্কামুক্ত। চার বৈমানিকই হাসপাতালে চিকিৎসাধীন আছেন।

এ ঘটনায় ফায়ার সার্ভিসের কর্মীসহ আরও ছয়জন আহত হয়েছে বলে জানা গেছে। আন্তঃবাহিনী জনসংযোগ পরিদপ্তরের (আইএসপিআর) পরিচালক লে. কর্নেল রাশিদুল হাসান প্রথম আলোকে বলেন, বিমানবাহিনীর দুটি প্রশিক্ষণ বিমান ইয়াক-১৩০ বিএএফ বেস জহুর থেকে ছেড়ে যায়। সন্ধ্যা সাড়ে ছয়টার দিকে বিমান দুটি বিধ্বস্ত হয়। বিধ্বস্ত হওয়ার আগে বিমান দুটির সঙ্গে রাডারের যোগাযোগ বিচ্ছিন্ন হয়ে যায়। দুর্ঘটনার কারণ এখন পর্যন্ত জানা যায়নি। তবে বিমানে থাকা চারজন বৈমানিককে জীবিত উদ্ধার করা গেছে। তাঁরা সবাই আশঙ্কামুক্ত।

মহেশখালী উপজেলা নির্বাহী কর্মকর্তা (ইউএনও) আবুল কালাম প্রথম আলোকে বলেন, সন্ধ্যা সাড়ে ছয়টার দিকে তিনি বিকট একটা শব্দ শুনতে পান। এর মিনিট দুয়েকের মাথায় বিমান বিধ্বস্ত হয়। সংঘর্ষের কারণে বিমান দুটি বিধ্বস্ত হতে পারে। একটি বিমান মহেশখালীর পুটিবিলায় এবং অন্যটি ছোট মহেশখালী মাইজপাড়া পাহাড়ি এলাকার পানের বরজে বিধ্বস্ত হয়।

মহেশখালীতে ফায়ার সার্ভিসের দায়িত্বপ্রাপ্ত কর্মকর্তা ধীমান বড়ুয়া বলেন, পুটিবিলার বিমানের আগুন নেভাতে তাঁদের প্রায় দুই ঘণ্টা সময় লাগে। তাঁরা বিমানটি বিমানবাহিনীর কাছে বুঝিয়ে দেবেন। মাইজপাড়ায় বিধ্বস্ত বিমানটি পাহাড়ে ঢুকে পড়ে। প্রত্যক্ষদর্শী ব্যক্তিরা শুধু বিমানটির ডানা দুটি দেখতে পেয়েছেন।

মহেশখালীর পুটিবিলার বাসিন্দারা বলেন, সন্ধ্যার দিকে তাঁরা হঠাৎ আকাশে কিছু একটা জ্বলতে দেখেন। এর একটু পরেই বিমানটি আবদুল কাদিরের দুই কক্ষের বাড়িতে আছড়ে পড়ে। ওই কক্ষে থাকা দুই ভাইবোন সুরভী (১৪) ও ফয়সাল (১০) আহত হয়। আহত হন জিয়াউর রহমান (৩৯), মোহাম্মদ হাসান (১৮) নুরুন্নাহার (৩০) ও তিন পথচারী। উদ্ধার অভিযান চলার সময় আহত হন ফায়ার সার্ভিসের কর্মী আরিফ উল্লাহ (৩০)।

উপজেলা স্বাস্থ্য কর্মকর্তা মো. সাজ্জাদ হোসেন চৌধুরী প্রথম আলোকে বলেন, রাত সাড়ে আটটা পর্যন্ত ছয়জন চিকিৎসা নিয়েছেন। মাথা ও হাতে আঘাত নিয়ে চিকিৎসাধীন আছে দুই ভাইবোন। ইয়াক-১৩০ মডেলের যুদ্ধবিমান ১৯৯৬ সালে প্রথম আকাশে উড্ডয়ন করে। এরপর ২০০২ সালে একে রুশ সামরিক পাইলটদের প্রশিক্ষণের জন্য প্রধান আকাশযান হিসেবে নির্বাচিত করা হয়। রাশিয়ার সঙ্গে বাংলাদেশের যে এক বিলিয়ন মার্কিন ডলারের ঋণ হয়েছে, তার আওতায় ২০১৫ সালে বিমানগুলো কেনা হয়। রাশিয়ার সমরাস্ত্র প্রস্তুতকারী প্রতিষ্ঠান ইরকুত করপোরেশন এই অত্যাধুনিক যুদ্ধবিমান নির্মাণ করে থাকে। এই যুদ্ধবিমান মৌলিক ও যুদ্ধ প্রশিক্ষণের জন্য উপযোগী। পাশাপাশি একে আক্রমণের কাজেও ব্যবহার করা যায়।

E. Human generated summary:

কক্সবাজারের মহেশখালীতে বুধবার সন্ধ্যায় বাংলাদেশ বিমানবাহিনীর দুটি প্রশিক্ষণ বিমান বিধ্বস্ত হয়েছে। বিমানে থাকা চার বৈমানিককেই জীবিত উদ্ধার করা সম্ভব হয়েছে।আন্তঃবাহিনী জনসংযোগ পরিদপ্তরের (আইএসপিআর) পরিচালক লে. কর্নেল রাশিদুল হাসান প্রথম আলোকে বলেন, বিমানবাহিনীর দুটি প্রশিক্ষণ বিমান ইয়াক-১৩০ বিএএফ বেস জহুর থেকে ছেড়ে যায়।সন্ধ্যা সাড়ে ছয়টার দিকে বিমান দুটি বিধ্বস্ত হয়। এ ঘটনায় ফায়ার সার্ভিসের কর্মীসহ আরও ছয়জন আহত হয়েছে বলে জানা গেছে। দুর্ঘটনার কারণ এখন পর্যন্ত জানা যায়নি

F. Summary Generated By Proposed Proposed Method:

কক্সবাজারের মহেশখালীতে বুধবার সন্ধ্যায় বাংলাদেশ বিমানবাহিনীর দুটি প্রশিক্ষণ বিমান বিধ্বস্ত হয়েছে। বিমানে থাকা চার বৈমানিককেই জীবিত উদ্ধার করা সম্ভব হয়েছে। আন্তঃবাহিনী জনসংযোগ পরিদপ্তরের (আইএসপিআর) পরিচালক লে. কর্নেল রাশিদুল হাসান প্রথম আলোকে বলেন, বিমানবাহিনীর দুটি প্রশিক্ষণ বিমান ইয়াক-১৩০ বিএএফ বেস জহুর থেকে ছেড়ে যায়। চার বৈমানিকই হাসপাতালে চিকিৎসাধীন আছেন। এ ঘটনায় ফায়ার সার্ভিসের কর্মীসহ আরও ছয়জন আহত হয়েছে বলে জানা গেছে

G. Summary Generated By M. I. A. Efat[14]

কক্সবাজারের মহেশখালীতে বুধবার সন্ধ্যায় বাংলাদেশ বিমানবাহিনীর দুটি প্রশিক্ষণ বিমান বিধ্বস্ত হয়েছে। বিমানে থাকা চার বৈমানিককেই জীবিত উদ্ধার করা সম্ভব হয়েছে। তাঁরা আশঙ্কামুক্ত। আন্তঃবাহিনী জনসংযোগ পরিদপ্তরের (আইএসপিআর) পরিচালক লে. কর্নেল রাশিদুল হাসান প্রথম আলোকে বলেন, বিমানবাহিনীর দুটি প্রশিক্ষণ বিমান ইয়াক-১৩০ বিএএফ বেস জহুর থেকে ছেড়ে যায়। চার বৈমানিকই হাসপাতালে চিকিৎসাধীন আছেন। এ ঘটনায় ফায়ার সার্ভিসের কর্মীসহ আরও ছয়জন আহত হয়েছে বলে জানা গেছে

H. Summary Generated By K. Sarkar[8]

কক্সবাজারের মহেশখালীতে বুধবার সন্ধ্যায় বাংলাদেশ বিমানবাহিনীর দুটি প্রশিক্ষণ বিমান বিধ্বস্ত হয়েছে। বিমানে থাকা চার বৈমানিককেই জীবিত উদ্ধার করা সম্ভব হয়েছে। আন্তঃবাহিনী জনসংযোগ পরিদপ্তরের (আইএসপিআর) পরিচালক লে. কর্নেল রাশিদুল হাসান প্রথম আলোকে বলেন, বিমানবাহিনীর দুটি প্রশিক্ষণ বিমান ইয়াক-১৩০ বিএএফ বেস জহুর থেকে ছেড়ে যায়।সন্ধ্যা সাড়ে ছয়টার দিকে বিমান দুটি বিধ্বস্ত হয়। বিধ্বস্ত হওয়ার আগে বিমান দুটির সঙ্গে রাডারের যোগাযোগ বিচ্ছিন্ন হয়ে যায়।চার বৈমানিকই হাসপাতালে চিকিৎসাধীন আছেন। এ ঘটনায় ফায়ার সার্ভিসের কর্মীসহ আরও ছয়জন আহত হয়েছে বলে জানা গেছে

TABLE I
Comparison between the proposed system to other system

Methodology	Recall Score	Precision Score	Final Score
Proposed System	0.67	0.80	0.72
K. Sarkar[4]	0.67	0.57	0.61
M. I. A. Efat[6]	0.50	0.60	0.50

I. Comparison of Recall Score:

$$RecallScore = \frac{HGS \cap SGS}{HGS} \tag{5}$$

Where,
HGS = Human Generated Summary
SGS = System Generated Summary

The fig. 2 describes the comparison of recall score between proposed system with other summarization approaches. And the formulae(5) for calculating Recall Score. In the data visualization, there recall score of the proposed methodology is 0.67 that means 67% and also Kamal Sarkar's proposed methods achieved 0.67 means 67%. But there are one the method achieved 0.50 means 50% which is proposed by Iftekharul.

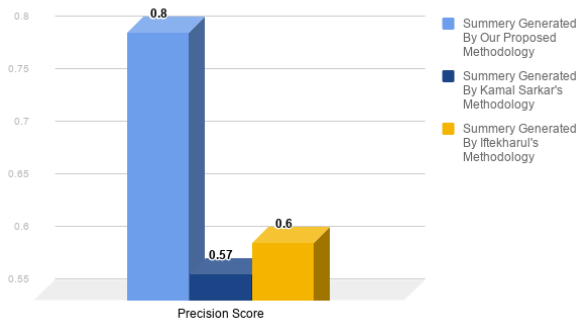

Fig. 2. Result of Recall Score comparison between proposed system with other summarization approaches.

J. Comparison of Precision Score:

$$PrecisionScore = \frac{SGS \cap HGS}{SGS} \tag{6}$$

Where,
HGS = Human Generated Summary
SGS = System Generated Summary

The fig. 3 illustrates the difference of precision score between proposed system with other bengali document summarization approaches. The equation(6) used to calculated the Precision Score of the proposed methodology and compare with others. In the following bar plot we can see that there 0.80 precision score achieved by this proposed method, 0.57 precision score achieved by proposed method of Kamal Sarkar and 0.60 precision score achieved by proposed method of Iftekharul.

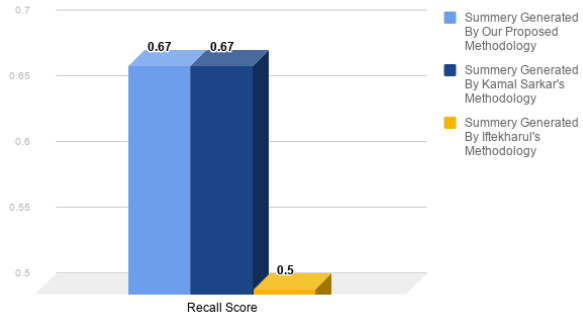

Fig. 3. Result of Precision Score comparison between proposed system with other summarization approaches.

K. Comparison of Final Score:

$$FinalScore = \frac{2 \star RecallScore \star PrecisionScore}{RecallScore + PrecisionScore} \tag{7}$$

Where,
HGS = Human Generated Summary
SGS = System Generated Summary

The equation(7) has been used for calculating the Final Score of the proposed methodology and compare with others. The fig. 4 shows the comparison of final score between proposed system with other summarization approaches. In the following fig. 4 there 0.72 final score achieved by the proposed method, 0.61 final score achieved by proposed method of Kamal Sarkar and 0.50 final score achieved by proposed method of Iftekharul.

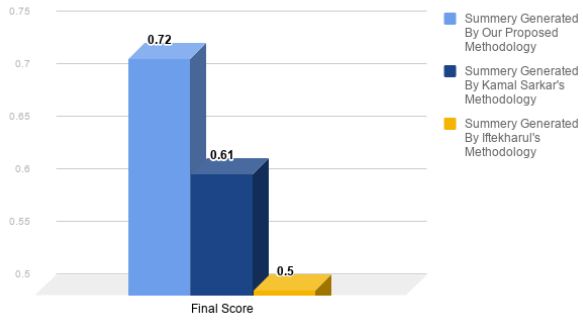

Fig. 4. Result of Final Score comparison between proposed system with other summarization approaches.

V. CONCLUSIONS

In this research, a study has been done for the extraction of Bengali single document summarization technique. The performance of the proposed methodology is 80% comparing with human-generated summaries.

Automatically Bengali document summarization performs the significant activities on the Bengali document in the modality of the NLP field. Finaly the proposed methodology will be able to prove that the approach will be influenced in the linguistic department. Although having the limitations, the methodology proved it's expediency to be included in the modality. And hopefully throughout the future work it will initiate revolutionary changes of the services on the NLP field.

The next target to in introduce more features for word stemming, sentence ranking and scoring. It is also expected to extend abstraction based summarization. The perfomance of the proposed system may further be improved by implementing machine learning model.

References

[1] M. N. Uddin and S. A. Khan, "A study on text summarization techniques and implement few of them for Bangla language," in Computer and information technology, 2007. iccit 2007. 10th international conference on, 2007, pp. 1–4.

[2] A. Das and S. Bandyopadhyay, "Opinion Summarization in Bengali: A Theme Network Model.," in SocialCom/PASSAT, 2010, pp. 675–682.

[3] M. N. Uddin and S. A. Khan, "A study on text summarization techniques and implement few of them for Bangla language," in Computer and information technology, 2007. iccit 2007. 10th international conference on, 2007, pp. 1–4.

[4] K. Sarkar, "An approach to summarizing Bengali news documents," in proceedings of the International Conference on Advances in Computing, Communications and Informatics, 2012, pp. 857–862.

[5] M. M. Haque, S. Pervin, and Z. Begum, "Automatic Bengali news documents summarization by introducing sentence frequency and clustering," in Computer and Information Technology (ICCIT), 2015 18th International Conference on, 2015, pp. 156–160.

[6] M. I. A. Efat, M. Ibrahim, and H. Kayesh, "Automated Bangla text summarization by sentence scoring and ranking," in Informatics, Electronics & Vision (ICIEV), 2013 International Conference on, 2013, pp. 1–5.

[7] A. Das and S. Bandyopadhyay, "Topic-based Bengali opinion summarization," in Proceedings of the 23rd International Conference on Computational Linguistics: Posters, 2010, pp. 232–240.

[8] M. Islam, M. Uddin, and M. Khan, "A light weight stemmer for Bengali and its Use in spelling Checker," 2007.

[9] T. T. Urmi, J. J. Jammy, and S. Ismail, "A corpus based unsupervised Bangla word stemming using N-gram language model," in Informatics, Electronics and Vision (ICIEV), 2016 5th International Conference on, 2016, pp. 824–828.

[10] M. H. Seddiqui, A. A. M. Maruf, and A. N. Chy, "Recursive Suffix Stripping to Augment Bangla Stemmer."

[11] A. K. Saha, M. F. Mridha, and J. K. Das, "Analysis of Bangla Root Word for Universal Networking Language (UNL)," International Journal of Computer Applications, vol. 89, 2014.

[12] S. Sarkar and S. Bandyopadhyay, "Design of a rule-based stemmer for natural language text in bengali," in Proceedings of the IJCNLP-08 Workshop on NLP for Less Privileged Languages, 2008.

[13] M. R. Mahmud, M. Afrin, M. A. Razzaque, E. Miller, and J. Iwashige, "A rule based bengali stemmer," in Advances in Computing, Communications and Informatics (ICACCI, 2014 International Conference on, 2014, pp. 2750–2756.