

Bengali Text Summarization using TextRank, Fuzzy C-Means and Aggregate Scoring methods

Alvee Rahman

Department of Computer Science and
Engineering
BRAC University
Dhaka, Bangladesh
alvee9@gmail.com

Fahim Md Rafiq

Department of Computer Science and
Engineering
BRAC University
Dhaka, Bangladesh
fahim6331006@gmail.com

Ramkrishna Saha

Department of Computer Science and
Engineering
BRAC University
Dhaka, Bangladesh
mukutsaha123@gmail.com

Ruhit Rafian

Department of Computer Science and
Engineering
BRAC University
Dhaka, Bangladesh
ruhitrafian661@gmail.com

Hossain Arif

Department of Computer Science and
Engineering
BRAC University
Dhaka, Bangladesh
hossain.arif@bracu.ac.bd

Abstract—In this world, it is very difficult and time consuming for humans to summarize large documents, reports, news and research articles. Multiple text summarization techniques play vital roles in picking the important points and sentences thus reducing the time and effort required to read a whole article. Numerous summarization techniques have been applied to the English language but works on Bangla text summarization is still limited. Furthermore, in our country, Bangladesh, all summarization is mainly done by humans. Keeping that in mind we aim to find a simple way of summarizing Bengali texts with the technology at hand. Text summarization can be of two types, abstractive and extractive. In this paper, we will use extractive text summarization to summarize Bengali passages, using Fuzzy C-Means, TextRank and Aggregate Sentence Scoring methodologies. We have also done a comparative study, among the three methodologies to find out that Fuzzy C-Means out performs the other two methods to generate a more concise and accurate summary.

Keywords— *Text Summarization, Bengali, FCM, TextRank, ROUGE, Extractive Text Summarization*

I. INTRODUCTION

In an era where everyone needs to be updated every second, but no one has the adequate time to read and stay informed. So the world needs real-time automatic text summarization to help them stay informed with the least time consumed. Text summarization can also be used to skim through large Bengali documents and then deciding which one to read if it seems interesting enough. With the advancement in technologies in Bangladesh, the language Bengali is increasingly being used in almost all online platforms, hence the need of Bengali Text Summarization. Here in this proposed system, the objective is to take in articles written in Bengali and convert them into a shorter version, preserving the true meaning of the article. Text Summarization is primarily divided into two major sections: Extractive and Abstractive. In the Extractive approach, the system simply omits the sentences that possess the least weight in the true meaning of the given text, and generate a shorter and more precise version of the passage [1, 2]. Whereas, abstractive summarization as described in [26] has to use linguistic methods to “understand” the text and then generate a summarized text that has fewer words than the original. Due to these more computationally complex steps, abstractive summarization is much more difficult. The summary built will be much like a one written by a human [3]. The main key points of a text are identified and then understandable sentences are constructed in a concise manner. This paper

namely discusses the extractive approach which has been widely used over the years for summarization purpose.

News articles have been manually fed into the system which were collected from the national daily “The Daily Prothom Alo”, the system then processes the data before it can be summarized; in preprocessing, the system tokenizes the extract, and removes the stopwords from the extract so that they have no influence in the summary generation. After the removal of stop words, the system stems the words to their root forms, so that all the words generated from a common root is considered as a single unit.

The system primarily focuses on Fuzzy C-Means Clustering Algorithm [4] to generate an optimal summary. Along with FCM, TextRank [5] and Aggregate Sentence Scoring [2, 6, 7, 8] has also been implemented to provide a comparative study at the end. For a uniform and accurate evaluation for the comparative study, the system uses the ROUGE [9] scoring method and later calculates F-Measure to provide an understandable illustrative comparative study.

The rest of the paper is organized as follow: section II describes the background study done for the project, section III illustrates the proposed model along with a block diagram, section IV describes the comparison and result analysis, lastly section V concludes the paper.

II. BACKGROUND STUDY

First, Automatic text summarization was first introduced by Luhn [10] in 1958 where he proposed the idea of calculating the word frequencies in sentences and later using those to score sentences, ultimately selecting the highest ranked sentences for the summary. In recent years there have been numerous approaches toward automatic text summarization. Some of the approaches included abstract summarization techniques [3] and extractive summarization techniques [1, 2].

Despite the fact that Bengali text summarization is not as widely popular as English Text summarization, it did get an ample importance for being an emerging field of research in recent decades. First work in the Bengali text summarization field was done by Islam et al. in [11] in the year 2004. They proposed a keyword search-based technique for multiple documents where their corpus-based search engine searches the keyword in multiple documents and then makes a summary of the relevant documents. Later on, it was followed by Uddin and Khan [12] who implemented a summarizer where they used location method, cue method, title, and term

frequency to rank the sentences. Efat et al. in [13] did a similar research using word frequency, cue words, sentence positional value and skeleton of the document for sentence scoring purposes.

Furthermore, Das and Bandyopadhyay in [14] where they made a topic based opinion summarization system. Their system does two tasks: 1) Finding the theme of the document. 2) Finding the summary of the document. It does the first part by finding out the sentiment information in a document by following a topic-sentiment model which uses clustering model such as K-means and uses a theme relational graph technique for finding out the document level summary. Another work was done by Sarkar in [15] who only used the TF*IDF model along with positional value and sentence length to generate a summary of a single document. He only used a single reference summary generated from LEAD baseline for evaluation which undermines the accuracy of the summarizer. Akter et al. in their paper [16] used a different approach in selecting sentences for generating a summary. They used K-means clustering after sentence ranking to choose the best and worst n-sentences for generating the summary. This shone a different light on selecting the sentences for generating the summary as worst scored sentences were not used in generating summaries before for Bangla text summarization.

Out of the handful of brilliant groundbreaking researches that have been done on Bangla Text Summarization none have used the concepts of Fuzzy C-Means. However, implementation of the FCM algorithm, have been in talks for English language processing for quite some time now. In [17] Patil et. al. have proposed a text miner which is based on the Fuzzy C-means algorithm. After the sentence clustering, it has been seen that clustering done using the Fuzzy C-Means algorithm outperforms the traditional K-means algorithm. The FCM algorithm was developed by Dunn [4] in 1973 and was later modified by Bezdek [18] in 1981.

Another process of generating text summarization is the highly accepted TextRank Algorithm. TextRank [5] is a graph based unsupervised algorithm derived from the PageRank algorithm [19]. Before any summarization techniques can be applied on any sort of text, a proper preprocessing is required, similar preprocessing techniques were discussed in [6, 16], where they conducted tokenization, stop word removal and stemming; making the text ready for mathematical analysis. Akter et. al. in [16] used the concept of TF*IDF for word scoring and incorporated cue/skeleton words concepts into their sentence scoring mechanism which further improve their system's accuracy. Having problems in manipulating data for having an unfavorable number of dimensions is not a new problem in this field of research. Nonetheless, it was quite remarkably handled by Tian et al, in their work [20] where they discussed the method of how Principal Component Analysis (PCA) can be used to reduce the dimension of the data, to make complex data more susceptible for manipulation and visualization.

In this paper, a comparative study and a thorough analysis is performed, between the techniques TextRank, Aggregate Scoring method and FCM integrated with PCA to summarize Bengali Text and news articles by the extractive method into concise and meaningful texts.

III. PROPOSED MODEL

The system proposed in this paper uses three popular text summarization methodologies to summarize Bengali text documents and provides a comparative study on the outputs generated.

At the very primary stage Bengali text article is fed into the system; once the file is read it undergoes the immersive preprocessing procedures to prepare the text document for scoring. During preprocessing, the system removes the stopwords present, splits the text into paragraphs, sentences and later into words (tokenization). The system also stems the words to its root version so that in circumstance a word is not misinterpreted as different words in case of occurrences at multiple instances in different forms of the root version.

After the preprocessing is done the system moves on to the feature extraction part which is the scoring mechanism of the sentences in order to generate the Extractive summary of the input text. For the scoring mechanism, the system is equipped with 6 different scoring techniques. The scores used for the system are TF-IDF, Numerical Value, Sentence length, Cue/Skeleton word, Topic Sentence and Sentence Position Scoring.

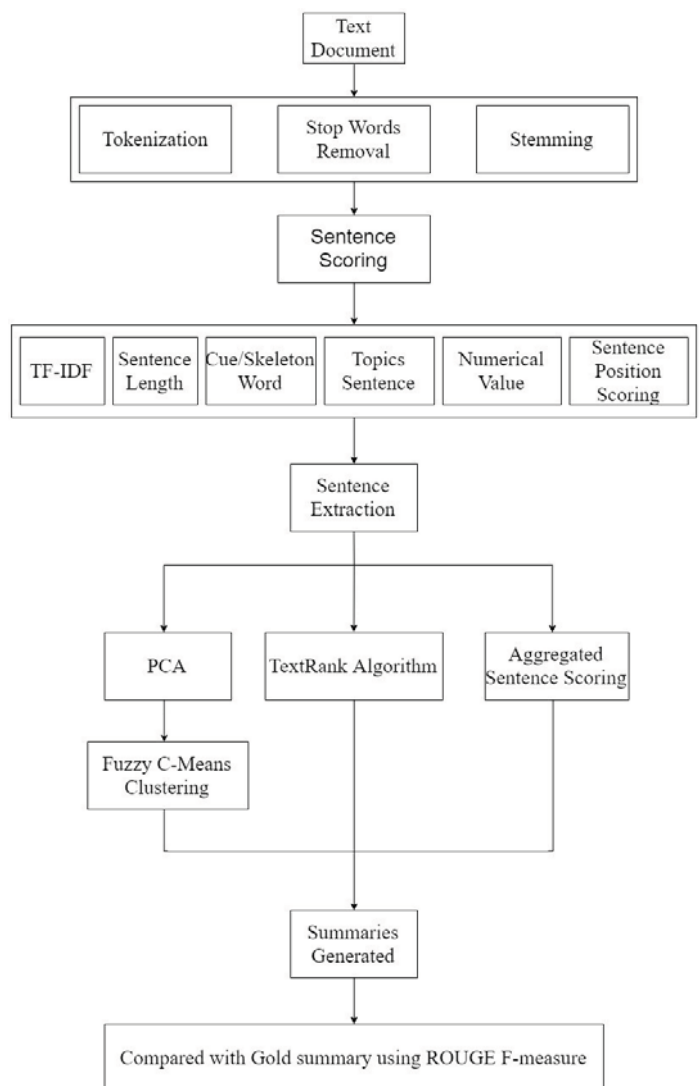


Fig. 1. Workflow of the system

Upon the successful extraction of the features, the system produces a 6-dimensional array upon which Principal Component Analysis (PCA) is performed to reduce the 6-dimensional data to a 2-dimensional data. It was done so that visualization of the data and clustering was better. As well as that, the algorithm used for this FCM study required 2D data but since this system's scoring technique had a 6D output, it was necessary to reduce the dimensions. The 2-dimensional data is then subjected to Fuzzy C-Means (FCM), to classify the sentences into 2 clusters, later the cluster having a greater F-measure value is printed as the output summary.

Apart from FCM, the system also uses TextRank and Aggregate Scoring techniques to generate 2 more summaries for each article. TextRank is a form of summary generation that is derived from PageRank where sentence similarity is used to find the most important sentences [19]. Next, the system finds the Aggregate Scores of the sentences and create a third summary using the most important sentences from the set. The F-Measure is calculated for each of the summaries, comparing it with the Gold Summary (Human Generated Summary) that is manually fed into the system, and a comparative study is conducted exhibiting the classifying methodology with the maximum accuracy. This proposed system has been tested with numerous news articles from different Bengali national dailies. For illustration of this comparative study two news articles have been used and the algorithms were compared i.e Test Article 1 [24] and Test Article 2 [25].

A. Preprocessing

1) *Stemming*: Simply converting the text words into their root forms so that multiple words originating from the same root word is counted to be the same word. A rule-based generic Bengali stemmer as implemented in [21] has been used which converts a Bengali word into its stemmed form. If the following words are taken as an example, 'কাজের', 'কাজটি' etc will all be converted to 'কাজ'.

TABLE I. STEMMING OUTPUT

Unstemmed Word	Stemmed Word
তেলের	তেল
যুক্তরাষ্ট্রে	যুক্তরাষ্ট্র
সালের	সাল
অক্টোবরের	অক্টোবর
কারণে	কারণ

2) *Stop-word Removal*: Removes words from text that itself does not hold much significance but are vital part of the sentence. Words such as 'অবশ্য', 'এই', 'কয়েক' are merely few of words from the enormous list of stopwords that has been installed in the system.

TABLE II. REPRESENTATION OF STOPWORD REMOVAL

Sentences with stopword	Sentences after stopword is removed
অন্য গত অক্টোবরেই তেলের দাম ব্যারেলপ্রতি ৭৬ ডলারে উঠেছিল।	গত অক্টোবরেই তেলের দাম ব্যারেলপ্রতি ৭৬ ডলারে উঠেছিল।

3) *Paragraph Splitting*: Paragraphs are split to help scoring mechanisms, like Topic Sentence, and Sentence Position Scoring to work optimally.

4) *Sentence Splitting*: Sentences are split to create instances of the sentences so that scores can be recorded in each instance of the sentence.

5) *Word Splitting (Tokenization)*: Word Splitting is done to help with the scorings that require the checking of word occurrences.

B. Feature Extraction

1) *TF-IDF*: TF-IDF stands for Term Frequency-Inverse Document Frequency; this score represents the importance or significance of a specific word in the entire document [10].

2) *Numeric Value Based Sentence Scoring*: Sentences containing any sort of numerical value is given a higher priority compared to the others.

3) *Sentence Length Based*: Sentences are compared to the average length of sentences in the text extract and scored accordingly.

4) *Cue/ Skeleton Word Scoring*: Sentences containing cue/skeleton word is given a score increment.

5) *Topic Sentence Scoring*: Sentences containing the words that have occurred in the topic sentence of the extract or the topic sentence of the given paragraph is given higher score.

6) *Sentence Position Based Scoring*: The sentences in the first and last 10% of paragraph is given a higher sentence position score.

C. Algorithms

1) *Fuzzy C-means (FCM) algorithm*: It is a soft clustering algorithm based on fuzzy logic. In soft clustering, a data point may belong to multiple clusters in contrary to hard clustering where a datapoint may belong to only one cluster. Each data point is associated with a membership function, which expresses the degree of its membership to a specific cluster. In this proposed system, two clusters have been used- sentences with high importance and sentences with low importance. The datapoints i.e the sentences, which belong to the high importance cluster are selected and extracted to create a summary. This system uses the FCM algorithm with a fuzziness parameter of value 2 in [22] for implementation.

2) *TextRank Algorithm*: Based on the popular PageRank Algorithm; this algorithm creates a cosine similarity matrix of each sentence and chooses the top 40% of sentences as the Extractive Summary. The system uses an algorithm which was implemented using PHP [23].

3) *Aggregate Sentence Scoring*: All the six score of each sentence is added all together and order in a descending order and the top 40% of the sentences are printed as the Extractive summary. Table III. shows a sample (top five sentences) of the aggregate scores calculated for Test Article 1.

TABLE III. SENTENCES WITH THEIR AGGREGATE SCORES FOR TEST ARTICLE 1

Sentence	Aggregate Score
এফএক্সটিএমের বিশ্লেষক লুকমান ওতুনাগা বলেন, ‘তেলের সরবরাহ একদিকে বাড়ছে, অন্যদিকে চাহিদা কমছে—এই দুই কারণে তেলের বাজারে বিপর্যয় নেমে আসছে	4.627
২০১৭ সালের অক্টোবরের পর এই প্রথম তেলের দাম ব্যারেলপ্রতি ৫০ দশমিক ৪২ ডলারে নেমে এল	3.858
কিন্তু অতি সরবরাহ নিয়ে শঙ্কা, চাহিদা পড়ে যাওয়া—এসব কারণে এক মাসের মধ্যে তেলের দাম এতটা কমে গেছে	3.515
সোসিয়েট জেনারেলের পণ্য গবেষণা বিভাগের প্রধান মাইকেল হেইগ বলেন, ছয় সপ্তাহ ধরে দাম যে হারে কমছে, তাতে বিনিয়োগকারীদের নান্দ্রিয়ার উঠে যাওয়ার জোগাড়	3.44

IV. COMPARISON AND RESULT ANALYSIS

ROUGE [9] is a metric system to compare machine generated summaries or translation against a reference summaries a.k.a Gold Summary. ROUGE tends to generate a metric value that determines the accuracy of the generated summary by generating a ratio of overlapping sentences.

For the evaluation of the system’s summary generated from the 3 different methods, the ROUGE measure was used. What it does is, it compares the summary generated by the system with the reference summary (Human-produced). It has two criteria for evaluation: 1) Recall and 2) Precision. Lastly, the F1 measure which is a measure of a test’s accuracy is calculated using both recall and precision values. The values are calculated with equations (1), (2) and (3). A score of 0 means the test yielded the worst result while 1 stands for the best.

“অপরিশোধিত তেলের দাম আরও এক দফা কমল। ২০১৭ সালের অক্টোবরের পর এই প্রথম তেলের দাম ব্যারেলপ্রতি ৫০ দশমিক ৪২ ডলারে নেমে এল। অথচ গত অক্টোবরেই তেলের দাম ব্যারেলপ্রতি ৭৬ ডলারে উঠেছিল। কিন্তু অতি সরবরাহ নিয়ে শঙ্কা, চাহিদা পড়ে যাওয়া—এসব কারণে এক মাসের মধ্যে তেলের দাম এতটা কমে গেছে। এক মাস আগেই পর্যবেক্ষকেরা দিন গুনছিলেন, তেলের দাম কবে ব্যারেলপ্রতি ১০০ ডলারে উঠবে। তেলের দামের বৈশ্বিক মানদণ্ড হচ্ছে অপরিশোধিত ব্রেন্ট তেলের দাম। শুক্রবার এই তেলের দাম কমেছে ৫ দশমিক ৫ শতাংশ। শুক্রবার ২০১৮ সালের মধ্যে ব্রেন্টের দাম সর্বনিম্ন ৫৯ ডলারে নেমে আসে। তেল কোম্পানির শেয়ারের দাম পড়ে যাওয়ায় শুক্রবার ডাও সূচকের মান ১৭৮ পয়েন্ট কমে যায়। ইরানের ওপর নিষেধাজ্ঞা আসছে এই আশঙ্কায় সৌদি আরবসহ ওপেকভুক্ত দেশগুলো তেলের উৎপাদন বাড়িয়ে দেয়। কিন্তু যুক্তরাষ্ট্র এরপর ভারত, চীনসহ বেশ কয়েকটি দেশকে ইরান থেকে তেল কেনার বেলায় ছাড় দিলে বাজারে তেলের সরবরাহ অনেকটা বেড়ে যায়। এতে বাজারে তেলের দাম ক্রমেই কমে কমে এ জায়গায় এসে দাঁড়িয়েছে। অন্যদিকে বৈশ্বিক আর্থিক বাজারে প্রবৃদ্ধি নিয়ে আবারও আশঙ্কা তৈরি হয়েছে। তবে তেলের পড়তি দাম ভোগ্যপণ্য ক্রেতাদের জন্য আশীর্বাদ হয়ে এসেছে। এফএক্সটিএমের বিশ্লেষক লুকমান ওতুনাগা বলেন, ‘তেলের সরবরাহ একদিকে বাড়ছে, অন্যদিকে চাহিদা কমছে—এই দুই কারণে তেলের বাজারে বিপর্যয় নেমে আসছে। সোসিয়েট জেনারেলের হিসাব মতে, চলতি প্রান্তিকে বড় বড় তহবিলের ক্ষতির পরিমাণ ৭৭০ কোটি ডলার ছাড়িয়ে গেছে। যুক্তরাষ্ট্রে শুক্রবার এক গ্যালন স্থানীয় দাম ছিল ২ দশমিক ৫৮ ডলার, যা এক মাস আগেও ছিল ২.৮৪ ডলার। এই পরিস্থিতিতে তেল উৎপাদনকারী দেশগুলো আগামী মাসে ভিয়েনায় ওপেক ও সহযোগী দেশগুলোর বৈঠকের দিকে তাকিয়ে আছে।”

Fig. 2. Gold Summary for Test Article 1

TABLE IV. NUMBER OF COMMON SENTENCES IN THE SUMMARIES GENERATED

	TextRank	Aggregate Scoring	FCM
Article 1	10	10	12
Article 2	7	9	10

TABLE V. COMPARISON BETWEEN F-NUMBER, PRECISION AND RECALL FOR TEST ARTICLE 2

	TextRank	Aggregate Scoring	FCM
F1 measure	0.350	0.500	0.606
Precision	0.304	0.473	0.625
Recall	0.411	0.529	0.588

TABLE VI. COMPARISON BETWEEN F-NUMBER, PRECISION AND RECALL FOR TEST ARTICLE 1

	TextRank	Aggregate Scoring	FCM
F1 measure	0.625	0.588	0.685
Precision	0.714	0.625	0.705
Recall	0.555	0.555	0.667

$$Recall = \frac{\text{Number of overlapping sentences}}{\text{Total number of Sentences in reference summary}} \quad (1)$$

$$Precision = \frac{\text{Number of overlapping sentences}}{\text{Total number of Sentences in system summary}} \quad (2)$$

$$F1 = 2 * \frac{Precision * Recall}{Precision + Recall} \quad (3)$$

As represented by Table IV and visualized by Fig. 3, on both test cases (Test Article 1 and 2), FCM yields a higher number of common sentences. This means that FCM algorithm returns sentences that have a higher probability of carrying more importance from the input article.

When judging the accuracy of summaries, as illustrated in Fig. 4 and Fig. 5 we can look at two factors, the F1 Measure and Common Sentences. For the first article, we notice a higher F1 measure for FCM summary than both Aggregate scoring and TextRank. This is backed up by the fact that FCM summary generates more common sentences than both TextRank and Aggregate scoring. FCM generated 2 more relevant sentences than TextRank and Aggregate Scoring, which results in a higher F1 measure. The summaries of Test Article 1 generated by FCM, TextRank and Aggregate Scoring method are shown in Fig. 6, Fig. 7 and Fig.8 respectively.

Number of Common Sentences

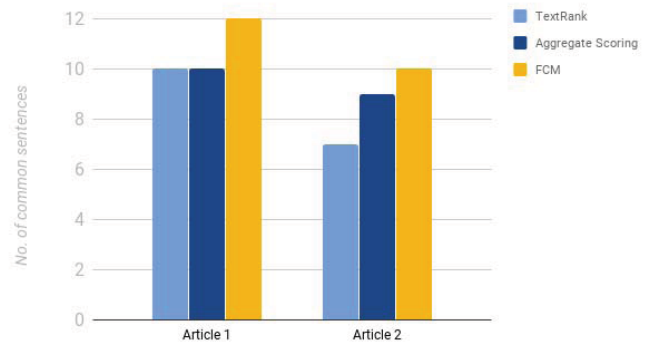
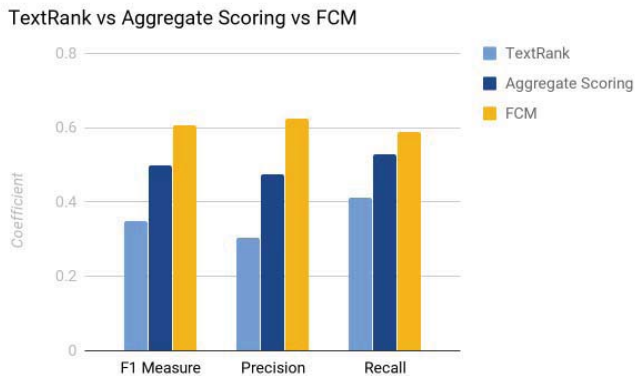
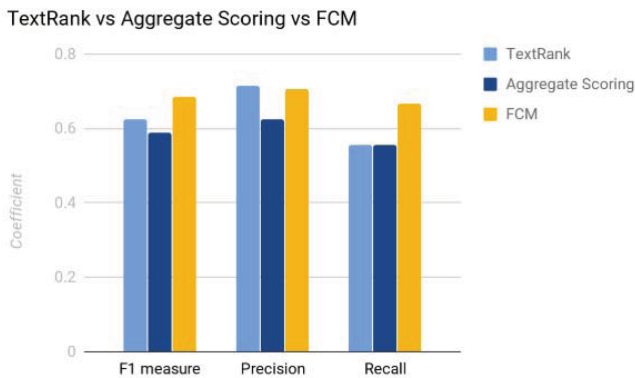


Fig. 3. Bar chart comparing the number of common sentences in the summaries



“২০১৭ সালের অক্টোবরের পর এই প্রথম তেলের দাম ব্যারেলপ্রতি ৫০ দশমিক ৪২ ডলারে নেমে এল। কিন্তু অতি সরবরাহ নিয়ে শঙ্কা, চাহিদা পড়ে যাওয়া—এসব কারণে এক মাসের মধ্যে তেলের দাম এতটা কমে গেছে। এক মাস আগেই পর্যবেক্ষকেরা দিন গুনছিলেন, তেলের দাম কবে ব্যারেলপ্রতি ১০০ ডলারে উঠবে। সোসিয়েট জেনারেলের পণ্য গবেষণা বিভাগের প্রধান মাইকেল হেইগ বলেন, ছয় সপ্তাহ ধরে দাম যে হারে কমছে, তাতে বিনিয়োগকারীদের নান্ধিশ্বাস উঠে যাওয়ার জোগাড়। তেল খাতসংশ্লিষ্ট ব্যক্তিরা আশা করছেন, আগামী মাসে ভিয়েনায় ওপেক ও সহযোগী দেশগুলোর বৈঠকে সৌদি আরববহ অনুরা তেলের সরবরাহ যথেষ্ট হারে কমাবে এবং তাতে বাজার কিছুটা সামগ্রী হবে। তবে দাম কমে যাওয়া সত্ত্বেও মার্কিন প্রেসিডেন্ট ডোনাল্ড ট্রাম্প ওপেকসহ সৌদি আরবকে উৎপাদন না কমাতে চাপ দিচ্ছেন। আর সম্প্রতি তিনি যেভাবে সৌদি আরবের প্রশংসা করলেন, তাতে বিনিয়োগকারীদের মনে শঙ্কা, সৌদি আরব সম্ভবত উৎপাদন তেমন একটা কমাবে না। শুরুবার ২০১৮ সালের মধ্যে ব্রেন্টের দাম সর্বনিম্ন ৫৯ ডলারে নেমে আসে। তেল কোম্পানির শেয়ারের দাম পড়ে যাওয়ায় শুরুবার ডাও সূচকের মান ১৭৮ পয়েন্ট কমে যায়। ইরানের ওপর নিষেধাজ্ঞা আসছে এই আশঙ্কায় সৌদি আরবসহ ওপেকভুক্ত দেশগুলো তেলের উৎপাদন বাড়িয়ে দেয়। কিন্তু যুক্তরাষ্ট্র এরপর ভারত, চীনসহ বেশ কয়েকটি দেশকে ইরান থেকে তেল কেনার বেলায় ছাড় দিলে বাজারে তেলের সরবরাহ অনেকটা বেড়ে যায়। অন্যদিকে বৈশ্বিক আর্থিক বাজারে প্রবৃদ্ধি নিয়ে আবারও আশঙ্কা তৈরি হয়েছে। এসব কারণে বৈশ্বিক অর্থনীতির চালিকা শক্তি জ্বালানি তেলের বাজার রমরমা হওয়ার সম্ভাবনা নেই। এক্সপ্রটিএমের বিশ্লেষক লুকমান ওতুন্গুগা বলেন, ‘তেলের সরবরাহ একদিকে বাড়ছে, অন্যদিকে চাহিদা কমছে—এই দুই কারণে তেলের বাজারে বিপর্যয় নেমে আসছে। সোসিয়েট জেনারেলের হিসাব মতে, চলতি প্রান্তিকে বড় বড় তহবিলের ক্ষতির পরিমাণ ৭৭০ কোটি ডলার ছাড়িয়ে গেছে। যুক্তরাষ্ট্রে শুরুবার এক গ্যালন জ্বালানির দাম ছিল ২ দশমিক ৫৮ ডলার, যা এক মাস আগেও ছিল ২.৮৪ ডলার। এই পরিস্থিতিতে তেল উৎপাদনকারী দেশগুলো আগামী মাসে ভিয়েনায় ওপেক ও সহযোগী দেশগুলোর বৈঠকের দিকে তাকিয়ে আছে।”

“অপরিশোধিত তেলের দাম আরও এক দফা কমল। গতকাল যুক্তরাষ্ট্রে তেলের দাম ৭ শতাংশ কমেছে। ২০১৭ সালের অক্টোবরের পর এই প্রথম তেলের দাম ব্যারেলপ্রতি ৫০ দশমিক ৪২ ডলারে নেমে এল। অথচ গত অক্টোবরেই তেলের দাম ব্যারেলপ্রতি ৭৬ ডলারে উঠেছিল। কিন্তু অতি সরবরাহ নিয়ে শঙ্কা, চাহিদা পড়ে যাওয়া—এসব কারণে এক মাসের মধ্যে তেলের দাম এতটা কমে গেছে। এক মাস আগেই পর্যবেক্ষকেরা দিন গুনছিলেন, তেলের দাম কবে ব্যারেলপ্রতি ১০০ ডলারে উঠবে। এখন তেলের এই পড়তি দাম দেখে তাঁদের কপালে চিল্লার ভাঁজ পড়েছে। তেলের দামের বৈশ্বিক মানদণ্ড হচ্ছে অপরিশোধিত ব্রেন্ট তেলের দাম। শুক্রবার এই তেলের দাম কমেছে ৫ দশমিক ৫ শতাংশ। শেভরন ও কলোকাফিলিপসের শেয়ারের দাম ৩ শতাংশ পড়ে যায়। কিন্তু যুক্তরাষ্ট্রে এরপর ভারত, চীনসহ বেশ কয়েকটি দেশকে ইরান থেকে তেল কেনার বেলায় ছাড় দিলে বাজারে তেলের সরবরাহ অনেকটা বেড়ে যায়। এতে বাজারে তেলের দাম ক্রমেই কমতে কমতে এ জায়গায় এসে দাঁড়িয়েছে। এফএক্সটিএমের বিশ্লেষক লুকমান ওতুন্সুগা বলেন, ‘তেলের সরবরাহ একদিকে বাড়ছে, অন্যদিকে চাহিদা কমেছে—এই দুই কারণে তেলের বাজারে বিপর্যয় নেমে আসছে.’ তেলের দাম এভাবে কমার কারণে অনেকই হতবুদ্ধ হয়ে গেছেন। তবে তেলের পড়তি দাম ভোগ্যপণ্য ক্রেতাদের জন্য আশীর্বাদ হয়ে এসেছে।”

“২০১৭ সালের অক্টোবরের পর এই প্রথম তেলের দাম ব্যারেলপ্রতি ৫০ দশমিক ৪২ ডলারে নেমে এল। কিন্তু অতি সরবরাহ নিয়ে শঙ্কা, চাহিদা পড়ে যাওয়া—এসব কারণে এক মাসের মধ্যে তেলের দাম এতটা কমে গেছে। সোসিয়েটে জেনারেলের পণ্য গমষণা বিভাগের প্রধান মাইকেল হেইগ বলেন, ছয় সপ্তাহ ধরে দাম যে হারে কমছে, তাতে বিনিয়োগকারীদের নাভিস্বাস উঠে যাওয়ার জোগাড়। তেল খাতসংশ্লিষ্ট ব্যক্তির আশা করছেন, আগামী মাসে ভিয়েনায় ওপেক ও সহযোগী দেশগুলোর বৈঠকে সৌদি আরবসহ অন্যরা তেলের সরবরাহ যথেষ্ট হারে কমাবে এবং তাতে বাজার কিছুটা শান্তশী হবে। তবে দাম কমে যাওয়া সত্ত্বেও মার্কিন প্রেসিডেন্ট ডোনাল্ড ট্রাম্প ওপেকসহ সৌদি আরবকে উৎপাদন না কমাতে চাপ দিচ্ছেন। আর সম্প্রতি তিনি যেভাবে সৌদি আরবের প্রশংসা করলেন, তাতে বিনিয়োগকারীদের মনে শঙ্কা, সৌদি আরব সম্ভবত উৎপাদন তেমন একটা কমাবে না। শুরুবার ২০১৮ সালের মধ্যে ব্রেন্টের দাম সর্বনিম্ন ৫২ ডলারে নেমে আসে। তেল কোম্পানির শেয়ারের দাম পড়ে যাওয়ায় শুরুবার ডাও সূচকের মান ১৭৮ পয়েন্ট কমে যায়। ইরানের ওপর নিষেধাজ্ঞা আসছে এই আশঙ্কায় সৌদি আরবসহ ওপেকভুক্ত দেশগুলো তেলের উৎপাদন বাড়িয়ে দেয়। কিন্তু যুক্তরাষ্ট্রে এরপর ভারত, চীনসহ বেশ কয়েকটি দেশকে ইরান থেকে তেল কেনার বেলায় ছাড় দিলে বাজারে তেলের সরবরাহ অনেকটা বেড়ে যায়। অন্যদিকে বৈশ্বিক আর্থিক বাজারে প্রবৃদ্ধি নিয়ে আবারও আশঙ্কা তৈরি হয়েছে। এসব কারণে বৈশ্বিক অর্থনীতির চালিকা শক্তি জ্বালানি তেলের বাজার রমরমা হওয়ার সম্ভাবনা নেই। এফএন্ডটিএমের বিশ্লেষক লুকমান ওতুন্গা বলেন, ‘তেলের সরবরাহ একদিকে বাড়ছে, অন্যদিকে চাহিদা কমছে—এই দুই কারণে তেলের বাজারে বিপর্যয় নেমে আসছে।’ তেলের দাম এভাবে কমার কারণে অনেকেই হতবুদ্ধ হয়ে গেছেন। যুক্তরাষ্ট্রে শুরুবার এক গ্যালন জ্বালানির দাম ছিল ২ দশমিক ৫৮ ডলার, যা এক মাস আগেও ছিল ২.৮৪ ডলার। এই পরিস্থিতিতে তেল উৎপাদনকারী দেশগুলো আগামী মাসে ভিয়েনায় ওপেক ও সহযোগী দেশগুলোর বৈঠকের দিকে তাকিয়ে আছে।”

As the world progresses in this Information Technology era, research in the Bengali language becomes more and more important. A text summarization system holds significance because of the importance of saving time, effort and also data. Text summarization can have two schools of thoughts: extractive summarization, and abstractive summarization. While the output from the abstractive method of summarization is more natural and coherent, it needs more processing and the complexity of the program is too high. As such, the extractive method to summarization provides a greater trade-off value due to the lower computational requirements. In this paper, a completely new approach to Bengali Text Summarization has been proposed, which uses the FCM algorithm. The FCM based algorithm uses the 6

sentence scoring methods to find the most important sentences. For the purpose of a comparative study, TextRank algorithm was used to generate a summary along with an Aggregate scoring algorithm. TextRank algorithm is a popular name in the field of NLP, but this is the first time it has been used for Bengali Text Summarization. TextRank algorithm uses a similarity measure to find the most useful sentences in an article. The Aggregate Scoring algorithm also uses the 6 scoring methods and the sum of all 6 scores from these methods are added to calculate aggregate scores for each sentence. These scores are then finally sorted in descending order, and then the top scoring sentences are then printed in original order to form an extracted summary. An FCM based algorithm tends to return higher F-number as well as a higher number of relevant sentences (sentences that are also found in the Gold summary).

REFERENCES

- [1] Andhale, N., & Bewoor, L. A. (2016, August). An overview of text summarization techniques. In *Computing Communication Control and automation (ICCUBE), 2016 International Conference on* (pp. 1-7). IEEE.
- [2] Moratanch, N., & Chitrakala, S. (2017, January). A survey on extractive text summarization. In *Computer, Communication and Signal Processing (ICCCSP), 2017 International Conference on* (pp. 1-6). IEEE.
- [3] Moratanch, N., & Chitrakala, S. (2016, March). A survey on abstractive text summarization. In *Circuit, Power and Computing Technologies (ICCPCT), 2016 International Conference on* (pp. 1-7). IEEE.
- [4] Dunn, J. C. (1973). A fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters.
- [5] Mihalcea, R., & Tarau, P. (2004). TextRank: Bringing order into text. In *Proceedings of the 2004 conference on empirical methods in natural language processing*.
- [6] Abujar, S., Hasan, M., Shahin, M. S. I., & Hossain, S. A. (2017, July). A heuristic approach of text summarization for Bengali documentation. In *Computing, Communication and Networking Technologies (ICCCNT), 2017 8th International Conference on* (pp. 1-8). IEEE.
- [7] Krishnaveni, P., & Balasundaram, S. R. (2017, July). Automatic text summarization by local scoring and ranking for improving coherence. In *Computing Methodologies and Communication (ICCMC), 2017 International Conference on* (pp. 59-64). IEEE.
- [8] Vijay, S., Rai, V., Gupta, S., Vijayvargia, A., & Sharma, D. M. (2017, December). Extractive text summarisation in hindi. In *Asian Language Processing (IALP), 2017 International Conference on* (pp. 318-321). IEEE.
- [9] Lin, C. Y. (2004). Rouge: A package for automatic evaluation of summaries. Text Summarization Branches Out.
- [10] Luhn, H. P. (1958). The automatic creation of literature abstracts. *IBM Journal of research and development*, 2(2), 159-165.
- [11] Islam, M. T., & Al Masum, S. M. (2004, December). Bhasa: A corpus-based information retrieval and summariser for bengali text. In *Proceedings of the 7th International Conference on Computer and Information Technology*.
- [12] Uddin, M. N., & Khan, S. A. (2007, December). 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* (pp. 1-4). IEEE.
- [13] Efat, M. I. A., Ibrahim, M., & Kayesh, H. (2013, May). Automated Bangla text summarization by sentence scoring and ranking. In *Informatics, Electronics & Vision (ICIEV), 2013 International Conference on* (pp. 1-5). IEEE.
- [14] Das, A., & Bandyopadhyay, S. (2010, August). Topic-based Bengali opinion summarization. In *Proceedings of the 23rd International Conference on Computational Linguistics: Posters* (pp. 232-240). Association for Computational Linguistics.
- [15] Sarkar, K. (2012). Bengali text summarization by sentence extraction. *arXiv preprint arXiv:1201.2240*.
- [16] Akter, S., Asa, A. S., Uddin, M. P., Hossain, M. D., Roy, S. K., & Afjal, M. I. (2017, February). An extractive text summarization technique for Bengali document (s) using K-means clustering algorithm. In *Imaging, Vision & Pattern Recognition (icIVPR), 2017 IEEE International Conference on* (pp. 1-6). IEEE.
- [17] Patil, D. B., & Dongre, Y. V. (2015). A fuzzy approach for text mining. *IJ Mathematical Sciences and Computing*, 4, 34-43.
- [18] Bezdek, J. C. (1981). Objective function clustering. In *Pattern recognition with fuzzy objective function algorithms* (pp. 43-93). Springer, Boston, MA.
- [19] Langville, A. N., & Meyer, C. D. (2011). *Google's PageRank and beyond: The science of search engine rankings*. Princeton University Press.
- [20] Tian, S. (2017). A hybrid debris flow hazard degree analysis model based on PCA and SFLA-FCM. *Revista de la Facultad de Ingeniería*, 31(9).
- [21] Kamal, R. (2014). rafi-kamal/Bangla-Stemmer. [online] GitHub. Available at: <https://github.com/rafi-kamal/Bangla-Stemmer>.
- [22] Josh Warner, Jason Sexauer, scikit-fuzzy, twmeggs, Alexandre M. S., Aishwarya Unnikrishnan, ... Himanshu Mishra. (2017, October 6). JDWarner/scikit-fuzzy: Scikit-Fuzzy 0.3.1 (Version 0.3.1). Zenodo. doi:10.5281/zenodo.1002946
- [23] DavidBelicza. (2018, October 08). DavidBelicza/PHP-Science-TextRank. Retrieved from <https://github.com/DavidBelicza/PHP-Science-TextRank>
- [24] ৫০ ডলারে নেমে এল তেলের দাম. (2018, November 26). The Daily Prothom Alo. Retrieved from <https://www.prothomalo.com/economy/article/1566899/৫০-ডলারে-নেমে-এল-তেলের-দাম>
- [25] ইসলাম, ত. (2018, November 27). শিক্ষাবিলম্বে মরিয়া চীন. The Daily Prothom Alo. Retrieved from <https://www.prothomalo.com/international/article/1566945/শিক্ষা-বিলম্বে-মরিয়া-চীন>
- [26] Dalal, V., & Malik, L. (2013, December). A survey of extractive and abstractive text summarization techniques. In *2013 6th International Conference on Emerging Trends in Engineering and Technology* (pp. 109-110). IEEE.