

HYBRID TEXT SUMMARIZER FOR BANGLA DOCUMENT

Asadullahil Galib	15.02.04.022
Mahimul Islam	15.02.04.047
Fariha Nuzhat Majumdar	15.02.04.054
Najmul Huda Auvy	13.02.04.081



**DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING
AHSANULLAH UNIVERSITY OF SCIENCE AND TECHNOLOGY**

Dhaka, Bangladesh

January 2020

HYBRID TEXT SUMMARIZER FOR BANGLA DOCUMENT

A thesis

Submitted in partial fulfillment of the requirements for the Degree of
Bachelor of Science in Computer Science and Engineering

Submitted by

Asadullahil Galib	15.02.04.022
Mahimul Islam	15.02.04.047
Fariha Nuzhat Majumdar	15.02.04.054
Najmul Huda Auvy	13.02.04.081

Supervised by

Mohammad Moinul Hoque
Associate Professor



Department of Computer Science and Engineering
Ahsanullah University of Science and Technology

Dhaka, Bangladesh

January 2020

CANDIDATES' DECLARATION

We, hereby, declare that the thesis presented in this report is the outcome of the investigation performed by us under the supervision of Mohammad Moinul Hoque, Associate Professor, Department of Computer Science and Engineering, Ahsanullah University of Science and Technology, Dhaka, Bangladesh. The work was spread over two final year courses, CSE4100: Project and Thesis-I and CSE4250: Project and Thesis-II, in accordance with the course curriculum of the Department for the Bachelor of Science in Computer Science and Engineering program.

It is also declared that neither this thesis nor any part thereof has been submitted anywhere else for the award of any degree, diploma or other qualifications.

Asadullahhil Galib

15.02.04.022

Mahimul Islam

15.02.04.047

Fariha Nuzhat Majumdar

15.02.04.054

Najmul Huda Auvy

13.02.04.081

CERTIFICATION

This thesis titled, “**HYBRID TEXT SUMMARIZER FOR BANGLA DOCUMENT**”, submitted by the group as mentioned below has been accepted as satisfactory in partial fulfillment of the requirements for the degree B.Sc. in Computer Science and Engineering in January 2020.

Group Members:

Asadullahhil Galib	15.02.04.022
Mahimul Islam	15.02.04.047
Fariha Nuzhat Majumdar	15.02.04.054
Najmul Huda Auvy	13.02.04.081

Mohammad Moinul Hoque
Associate Professor & Supervisor
Department of Computer Science and Engineering
Ahsanullah University of Science and Technology

Dr. Kazi A. Kalpoma
Professor & Head
Department of Computer Science and Engineering
Ahsanullah University of Science and Technology

ACKNOWLEDGEMENT

Foremost, we would like to express our sincere gratitude to our supervisor, Mohammad Moinul Hoque, Associate Professor, Department of CSE, AUST, for his continuous support on our thesis research and study and for his patience, motivation, enthusiasm, and immense knowledge. His guidance helped us in all the time of research and writing of this thesis. We could not have asked for a better supervisor and mentor for our thesis study. Besides our advisor, we would like to thank the faculties of the department for their support and help. We would also like to thank our mentor Shah Siam Ansary Dhrubo for his wonderful co-operation and support. We would like to thank our families for supporting us spiritually throughout our lives. Last but not the least, we are greatly thankful to the Almighty who granted us patience and motivation to complete our work to our hearts content.

Dhaka
January 2020

Asadullahhil Galib

Mahimul Islam

Fariha Nuzhat Majumdar

Najmul Huda Auvy

ABSTRACT

Automatic text summarization is needed to concisely extract a small subset of text portion from a large text where the isolated text may have sentences which are more significant compared to other sentences in the text. Although there have been a lot of approaches on English text summarization, very few works have been done on automatic Bengali text summarization. For the evaluation purpose, a data set was formulated from the scratch with Bengali news documents from a reputed newspaper. The evaluation data set was classified in four different classes with benchmark standard summary text generated by random people for each of the documents. The current work presents a hybrid approach for dealing with summarization process of Bengali text documents. The hybrid model is introduced with a goal to improve the overall accuracy of the summary text generation. The proposed model is expected to generate a summary text based on sentence scoring, sentiment analysis and interconnection of sentences. Empirical verification with other similar systems shows that, the proposed model can be used as an alternative system to address the Text Summarization problem of Bengali documents.

Contents

<i>CANDIDATES' DECLARATION</i>	i
<i>CERTIFICATION</i>	ii
<i>ACKNOWLEDGEMENT</i>	iii
<i>ABSTRACT</i>	iv
List of Figures	viii
List of Tables	ix
1 Introduction	1
1.1 Objective	1
1.2 Definition of Text Summarization	1
1.3 Types Of Text Summarization	2
1.3.1 Extractive Summarization	2
1.3.2 Abstractive Summarization	3
1.4 Methods of Automatic Text Summarization	3
1.4.1 Generic Summarization	3
1.4.2 Topic Centric Summarization	4
1.5 Steps in text summarization	4
1.6 Summary of the Chapter	5
2 State Of The Art	6
2.1 Literature Review	6
2.2 Research Problems	9
2.3 Research Opportunities	10
2.4 Summary of the Chapter	11
3 Proposed Model	12
3.1 Introduction	12
3.2 Proposed Statement	12
3.2.1 Sentiment Scoring	13

3.2.2	Keyword Ranking	13
3.2.3	Text Ranking	13
3.3	Necessity of a Hybrid Method	15
3.4	Detailed Analysis of our proposed method	16
3.4.1	Input Document	16
3.4.2	Pre-processing	17
3.4.3	Document Type Separation	17
3.4.4	Sentence Ranking	17
3.4.5	Generating Output Summary	17
3.5	Steps and algorithms related to Hybrid Summarizer	17
3.5.1	Preparing Datasets for Training and Testing	17
3.5.1.1	Collection of Data sets manually	17
3.5.1.2	Collection of data sets from BNLPC	18
3.5.2	Splitting Sentences	18
3.5.3	Keyword Extraction	19
3.5.4	Classification of Documents	19
3.5.5	Score Calculation (IDF) of words	21
3.5.6	Individual Scoring for Each Sentence	21
3.5.7	Sentiment Based Ranking	22
3.5.8	Text Ranking	23
3.5.8.1	Generating Vector Model from training data sets	23
3.5.8.2	Scoring sentences based on similarities	24
3.5.9	Hybrid Scoring 1 (Keyword Ranking and Sentiment Scoring)	24
3.5.10	Hybrid Scoring 2 (Keyword Ranking, Sentiment Scoring and Text Ranking)	25
3.5.11	Hybrid Scoring 3 (Keyword Ranking, Sentiment Scoring and Text Ranking)	25
3.6	Sample Generated Output	26
3.7	Summary of the Chapter	27
4	Evaluation Measures of Summary	28
4.1	Introduction	28
4.2	Evaluation Measures	28
4.3	Intrinsic Measure	30
4.3.1	Text Quality Evaluation	30
4.3.2	Content Evaluation	30
4.4	Summary of the Chapter	31
5	Experimental Verification	32
5.1	Introduction	32

5.2	Experimental Setup and Verification	32
5.2.1	Experimental setup on Manual Data set	33
5.2.2	Experimental Verification with Evaluation on Manual Data set . . .	33
5.2.3	Experimental Setup on BNLPC Data Set	42
5.2.4	Experimental Verification with Evaluation on BNLPC Data set . . .	43
5.2.5	Comparison with Existing Models	47
5.3	Time Analysis for Summary Generation	48
5.4	Limitations	50
5.5	Summary of the Chapter	51
6	Conclusion and Future Works	52
6.1	Conclusion	52
6.1.1	Brief Survey	52
6.1.2	Difficulties Encountered	52
6.2	Future Work	53
6.3	Summary of the Chapter	54
	References	55
A	Examples of Summary with Evaluation Scores	57
B	Resources and Data sets	73

List of Figures

3.1	Inter connectivity between sentences.	14
3.2	Proposed Design for Hybrid Summarizer for Bengali Document	15
3.3	A flow chart of the proposed model for Hybrid summarizer for Bengali Document	16
4.1	Classification of summary evaluation measures	29
5.1	Average Rouge-1 Scores of the system for different methods (1st Setup)	39
5.2	Average Rouge-1 Scores of the system for different methods (2nd Setup)	40
5.3	Average Rouge-2 Scores of the system for different methods (1st Setup)	41
5.4	Average Rouge-2 Scores of the system for different methods (2nd Setup)	42

List of Tables

2.1	Scores related to Summarization in several researches	9
3.1	Categories of news and number of documents collected from “Daily Prothom Alo”	18
3.2	Classification result in 1st setup	21
3.3	Classification result in 2nd setup	21
5.1	Categories and number of documents for training and testing	33
5.2	Average Rouge Score of the system for Category Accidents (1st Setup).	34
5.3	Average Rouge Score of the system for Category Economics (1st Setup).	34
5.4	Average Rouge Score of the system for Category Entertainment (1st Setup).	35
5.5	Average Rouge Score of the system for Category Politics (1st Setup).	36
5.6	Average Rouge Score of the system for Category Accident (2nd Setup).	36
5.7	Average Rouge Score of the system for Category Accident(2nd Setup).	37
5.8	Average Rouge Score of the system for Category Entertainment(2nd Setup).	37
5.9	Average Rouge Score of the system for Category Politics (2nd Setup).	38
5.10	Classification and ROUGE Scores of summaries for different categories (for BNLPC data set 1 and 1st model summary).	43
5.11	Classification and ROUGE Scores of summaries for different categories (for BNLPC data set 2 and 1st model summary).	43
5.12	Average ROUGE Scores of summaries for BNLPC data sets.	44
5.13	Classification and ROUGE Scores of summaries for different categories (for BNLPC data set 1 and 2nd model summary).	44
5.14	Classification and ROUGE Scores of summaries for different categories (for BNLPC data set 2 and 2nd model summary).	45
5.15	Average ROUGE Scores of summaries for BNLPC data sets.	45
5.16	Classification and ROUGE Scores of summaries for different categories (for BNLPC data set 1 and 3rd model summary).	45
5.17	Classification and ROUGE Scores of summaries for different categories (for BNLPC data set 2 and 3rd model summary).	46
5.18	Average ROUGE Scores of summaries for BNLPC data sets.	46
5.19	Average ROUGE Scores of the proposed system summaries for BNLPC data sets.	46

5.20 Average ROUGE Scores of our proposed system and other existing systems for BNLPC data sets.	47
5.21 Average ROUGE Scores of summaries our proposed system and other existing web based system for our data set.	47
5.22 Time needed to generate summary of 10 documents for Category Accident . . .	48
5.23 Time needed to generate summary of 10 documents for Category Economics . .	49
5.24 Time needed to generate summary of 10 documents for Category Entertainment	49
5.25 Time needed to generate summary of 10 documents for Category Politics	50
A.1 ROUGE 1 and ROUGE 2 Score of the key word generated summary for the news.	58
A.2 ROUGE 1 and ROUGE 2 Score of the sentiment score generated summary for the news.	58
A.3 ROUGE 1 and ROUGE 2 Score of the Text Rank generated summary for the news.	59
A.4 ROUGE 1 and ROUGE 2 Score of the Hybrid summary (combining KR, SS, TR) for the news.	59
A.5 ROUGE 1 and ROUGE 2 Score of other web based system generated summary for the news.	60
A.6 ROUGE 1 and ROUGE 2 Score of the key word generated summary for the news.	62
A.7 ROUGE 1 and ROUGE 2 Score of the sentiment score generated summary for the news.	63
A.8 ROUGE 1 and ROUGE 2 Score of the Text Rank generated summary for the news.	64
A.9 ROUGE 1 and ROUGE 2 Score of the Hybrid summary (combining KR, SS, TR) for the news.	64
A.10 ROUGE 1 and ROUGE 2 Score of other web based system generated summary for the news.	65
A.11 ROUGE 1 and ROUGE 2 Score of the key word generated summary for the news.	66
A.12 ROUGE 1 and ROUGE 2 Score of the sentiment score generated summary for the news.	67
A.13 ROUGE 1 and ROUGE 2 Score of the Text Rank generated summary for the news.	67
A.14 ROUGE 1 and ROUGE 2 Score of the Hybrid summary (combining KR, SS, TR) for the news.	68
A.15 ROUGE 1 and ROUGE 2 Score of other web based system generated summary for the news.	68
A.16 ROUGE 1 and ROUGE 2 Score of the key word generated summary for the news.	70
A.17 ROUGE 1 and ROUGE 2 Score of the sentiment score generated summary for the news.	70
A.18 ROUGE 1 and ROUGE 2 Score of the Text Rank generated summary for the news.	71
A.19 ROUGE 1 and ROUGE 2 Score of the Hybrid summary (combining KR, SS, TR) for the news.	71

A.20 ROUGE 1 and ROUGE 2 Score of other web based system generated summary for the news.	72
---	----

Chapter 1

Introduction

In this chapter we will broadly discuss about text summarization and its different types. Also we will get to know about automatic text summarizer, how it works and the steps involved in text summarization.

1.1 Objective

Text summarization requires a short, accurate, and fluent summary of a longer text document. From summary, important information can be gained which makes the overall procedure more comfortable and less resources are needed in the long run.

Automatic text summarization ideas are needed to address huge amount of text data available online to discover relevant information faster. Although there have been a lot of approaches on English summarization, very few works have been done on automatic Bengali text summarization.

We have explored a new method for the generation of summary from Bengali document. We had to consider all the major topics present in the document. If the summary contains sentences from the major topics in the document, that has a better chance of giving better perspective of the document. Our summary generation approach is extractive, i.e., the summaries contain exact sentences present in the document.

1.2 Definition of Text Summarization

Text summarization is the process of shortening a text document for creating a summary from the original document by extracting key points. The main objective of summarization is to find a subset of data that includes the crucial information of the entire set. Document summarization

is the process that generates a summary or abstract of the entire document, by finding the most informative sentences.

Automatic summarization is the process of shortening a text document for creating a summary from the original document by extracting key points with software. Automatic Text Summarization is one of the major applications of Natural Language Processing (NLP). The whole summarization process involves huge time consuming efforts that include reading the entire document and extracting the important ideas from the raw text. Automatic summarizing software summarizes texts of hundreds of words in a couple of seconds. Thus, users are able to read less data but still receive the most important information and make impactful conclusions.

Instead of reading the whole news articles that are mostly full of irrelevant information, the summaries of such web pages can be precise and accurate, but still 20 Percent to 40 Percent the size of the original article.

Machine generated summaries are free from bias. While humans can sometimes miss out an important sentence, it is impossible for the computers to miss it and thus important sentences will always be mentioned.

With digitization of media along with publishing, for those who have no time to go through entire article, document or book, this technique saves their reading time as they do not have to read huge amount of useless and redundant data.

1.3 Types Of Text Summarization

There are generally two categories of text summarization. They are:

- Extractive Summarization
- Abstractive Summarization

1.3.1 Extractive Summarization

In extraction based summarization, sentences with the most important points that are relevant to the topics are extracted from a piece of text and combined to make a summary. In extractive summarization process, the automatic system pulls out objects from the entire data set, without modifying the objects. This include key phrase extraction, where the goal is to select individual words or phrases to tag a document, and in document summarization, where the objective is to select the sentences without modifying them to create a short summary. Different types of algorithms and methods can be used to calculate the weights of the sentences and then rank them according to their relevance and similarity with one another-and further joining them to

generate a summary. In extraction based summarization, scores are provided to sentences using some methods and the sentences that achieve highest scores are included in the summaries. This kind of summary extracts the sentences that contain essential information but it is not always smooth or fluent. Sometimes in the generated summary, there may be lacking of connectivity between adjacent sentences. Our summarizer only generates extractive summaries.

1.3.2 Abstractive Summarization

In abstraction based summarization, advanced deep learning techniques are applied to paraphrase and shorten the original document, just like humans do. Here in the abstraction based summary, there can be sentences or words and phrases that are not present in the original document. Abstractive summarization systems generate new phrases, possibly rephrasing or using words that are not in the original text. Naturally abstractive approaches are harder. In order to generate a perfect abstractive summary, it is important to understand the document first and then express that understanding in short using new sentences and words. It has complex capabilities like generalization, paraphrasing and incorporating real-world knowledge. Abstraction involves paraphrasing sections of the source document. Abstractive summary is important as the textual structure present in the summary can vary significantly from the original. While some work has been done in abstractive summarization, the majority of summarization systems are extractive. Structure based approach encodes most vital data from the document(s) through psychological feature schemes like templates, extraction rules, alternative structures such as tree, ontology, lead and body rule, graph based structure etc. Semantic based approach emphasizes on the linguistic illustration of document(s) to feed into Natural Language Generation system. This technique specializes in identifying noun phrases and verb phrases by processing linguistic data.

1.4 Methods of Automatic Text Summarization

There are generally two methods of automatic text summarization:

- Generic Summarization
- Topic Centric Summarization

1.4.1 Generic Summarization

Generic summary provides the overall sense of document. It typically contains the core information of a document. A document normally consists of sentences that give information to users

about a single broad or narrow matter but understanding the matter requires information from other domains as well.

1.4.2 Topic Centric Summarization

Topic centric summaries are generally restricted to one topic. Topic is typically based on the key words that are supplied by human. Summary can be created from single or multiple documents. This kind of summary are going to be impactful in future search engines.

1.5 Steps in text summarization

According to Kamal Sarker [1] text summarization involves preprocessing, stemming, sentence ranking, summary generation. The preprocessing step includes stop-word removal, stemming and breaking the input document in to a collection of sentences.

Using stemming, a word is split into its stem and affix. The design of a stemmer is language specific, and requires some significant linguistic expertise in the language. A typical simple stemmer algorithm involves removing suffixes using a list of frequent suffixes, while a more complex one would use morphological knowledge to derive a stem from the words. Since Bengali is a highly inflectional language, stemming is necessary while computing frequency of a term.

After an input document is formatted and stemmed, the document is broken into a collection of sentences and the sentences are ranked based on two important features: thematic term and position.

The thematic terms are the terms which are related to the main theme of a document. We define the thematic terms are the terms whose TFIDF values are greater than a predefined threshold. The TFIDF value of a term is measured by the product of TF and IDF, where TF (term frequency) is the number of times a word occurs in a document and IDF is Inverse Document Frequency. The positional score of a sentence is computed in such a way that the first sentence of a document gets the highest score and the last sentence gets the lowest score.

Md. Nizam Uddin and Shakil Akter Khan [2] described extraction based methods for summarizing Bengali documents.

Different methods are used that has been found in the survey to rank the sentences. The methods are as following:

Location method: sentences under headings have higher score. Sentences near beginning or end of document and/or paragraphs have higher score.

Cue method: The Cue method is based on the hypothesis that the probable relevance of a sentence is affected by the presence of pragmatic words.

Title: Words in title and in following sentences gives high score.

Term frequency: Words or terms which are frequent in the text are more important than the less frequent. Open class terms are words that change over time.

Numerical data: Sentences containing numerical data is scored higher than the ones without numerical values.

Implementation: Based on the methods mentioned above the Bengali summarizer is implemented. For the implementation the summary size is 40 percent of the actual content.

1.6 Summary of the Chapter

We have gathered a thorough idea about text summarization after discussing different terms related to summarization from this chapter. Text summarization extracts important information from a document and ignores irrelevant information. So there is significant importance of summarization and we have discussed its importance in this chapter. Text summarization is of different types: extractive and abstractive. Different methods can be applied to summarize a document: generic and topic centric. Text or document does through several stages and then a proper summary is generated. In the next chapter we will discuss about the state of the art regarding automatic text summarization.

Chapter 2

State Of The Art

In the previous chapter we discussed about text summarization and different terms related to it. In this chapter we will get an overview of several researches, their limitations and research gaps relevant to automatic text summarization.

2.1 Literature Review

In today's fast emerging world text summarization is one of the most required tools for understanding significant information of a document. Sometimes due to large amount of data and various amount of sources it becomes really tough to understand the gist of documents. Selecting information from very large data is difficult for human beings. To manually summarize information available on the internet is really challenging, complicated and difficult task. There have been some researches on automatic Bengali text summarization. We have studied some papers regarding this topic.

Kamal Sarkar [1] [3] has discussed text summarization for single document of Bengali language signifying the impact of thematic term feature and position feature of sentences. In linguistics, thematic feature means to relate to the theme of a writing. In the experiment, LEAD baseline has been used for comparison of produced summary. LEAD has been defined in DUC 2001 and DUC 2002 conferences. The discussed work mainly is of three phases. The phases are preprocessing, sentence ranking and summary generation. Stop words removal, stemming and tokenization have been done as preprocessing. Sentence ranking has been done with thematic terms and sentence position. Thematic terms have been defined as the terms with higher TFIDF value than predefined values. As for position of sentences, first to last sentences have been scored as higher to lower. Too short and too long sentences have not been preferred as summary sentences. After combination of scores, sentences have been ranked and order of sentences has been maintained as input document. It has been ensured in the work that the length of produced

summary and reference summary remains same. Also in terms of recall, precision and f-score, the proposed system has given better results. It has been argued that the proposed model has outperformed the compared systems. The average unigram based Recall score: 0.4122. The score for LEAD baseline: 0.3991.

Md. Nizam Uddin and Shakil Akter Khan [2] experimented with Bengali text summarization taking into consideration several criteria. They have put significance on sentence location, cue phrase presence, title word presence, term frequency and numerical data. They have argued that sentences which appear in the first or last of a passage, are of more importance. They have said that presence of cue phrases, words from title, words with high frequency and numerical data also put importance on a text. They have tested their approach using documents from the newspaper the “Daily Prothom Alo”. They have considered two facts for evaluating a summary and that are the information in a summary and its size. They have concluded that the summary should be of forty percent of the main document. They have achieved average accuracy of 71.3 percent.

Amitava Das and Sivaji Bandyopadhyay [4] have summarized Bengali documents using sentiment information. They have tried to identify the sentiment information in a document and they aggregate that for generating summary. They have used a classifier based on support vector machine. Three kinds of feature have been considered which are lexico-syntactic, syntactic and of discourse level. Parts of speech, SentiWordNet, frequency, stemming, chunk label, dependency parsing depth, title of document, first paragraph, term distribution, collocation have been used as features in the work. After theme detection, theme clustering has been done and with theme relational graph using PageRank algorithm, summary sentences have been identified. For corpus, documents from Bengali newspaper “AnandabazarPatrika” have been used. Annotation of sentence level subjectivity and discourse level theme words has been done manually from portion of the corpus. It has been said that the summarization system has achieved precision: 72.15 percent, recall: 67.32 percent, f-score: 69.65 percent.

Md. Iftekharul Alam Efat, Mohammad Ibrahim, Humayun Kayesh [5] have discussed Bengali summarization taking into consideration of several attributes. They have calculated sentences’ scores based on aspects such as frequency, sentence position, cue phrases’ presence and skeleton of document. Before scoring of sentences, pre-processing such as tokenization, stop words removal, stemming have been done. After calculating scores based on various aspects, final sentences’ scores have been calculated as weighted summation of the scores of individual features. For experiments, dataset has been made using articles from various newspapers like “The Daily Prothom Alo”, “The Daily Ittefaq”, “The Daily Jugantor” etc. They have argued that 83.57 percent of summary sentences match to human summaries.

Haque, Pervin and Begum [6] have done text summarization with Bengali documents using sentence ranking and clustering. The summarization has been done through several phases. In the preprocessing phase, stop words removal and stemming have been done. Then sentences have been ranked with term frequency calculation for each sentence and sentence frequency. Here,

sentence frequency has been introduced for removing redundant sentences. If overlap ratio of two sentences have been shown over or equal to sixty percent, then the smaller sentence falls out of consideration and importance of larger sentence increases. Sentences have been clustered using cosine similarity to group similar sentences. Then summary has been generated by selecting sentences from clusters based on volumes of clusters. To build the corpus for experiment, Bengali newspapers “The Daily Prothom Alo” and “The Daily Jugantor” have been used. After evaluation, precision, recall and f-score values have been calculated as 0.608, 0.664 and 0.632 respectively.

Haque, Pervin and Begum [7] have discussed Bangla summarization using key phrases. They have argued that key phrases with three or four terms are better for summarization tasks. They have discussed that from first to last, sentences’ scores should decrease from more to less and sentences with numerical figures should be given importance. After combining the scores, sentences have been ranked. Data set has been made with four hundred newspaper documents that are of wide varieties. Using ROUGE-1 and ROUGE-2, they said that summaries’ quality has improved.

Akter et al. [8] proposed an extractive summarization using K-means clustering algorithm. They have used clustering to tackle both single and multiple document summarization issues. At first some pre-processing such as noise words and stop words removal, tokenization and stemming are done. Later, using TFIDF word scores have been calculated. Afterword, sentences have been scored with word scores, position and cue or skeleton words presence. Later, the ranked sentences have been saved in a file. For multi document summarization, all sentences from different inputs have been saved in a file and after merging, sort has been done in descending order. Then K-means clustering has been done taking the score of greatest and lowest score of sentences. After that summary sentences have been picked as the top thirty percent of the input document(s).

Paul et al. [9] have discussed about summarization process with sentence ranking based on clustering. For this, at first noise words, stop words have been removed and stemming has been done. Then, clustering of sentences has been done. Term frequency matrix has been used for summarization. Sentences have been given scores based on four different aspects. Based on term frequency, term frequency and clustering, TFIDF, TFIDF and clustering the four methods have been implemented. For experiments, corpus has been made from newspaper articles. Using ROUGE-N, it has been argued that better results has been gained with term frequencies. Table 2.1 shows the scores of summarization of different systems.

Table 2.1: Scores related to Summarization in several researches

Author	Document Resource	Evaluation
Kamal Sarkar	“Ananda Bazar Patrika”	Unigram based Recall score 0.4122 The score for LEAD baseline 0.3991
Md. Nizam Uddin and Shakil Akter Khan	“Daily Prothom Alo”	Average accuracy 71.3 Percent
Amitava Das and Sivaji Bandopadhyay	“Ananda Bazar Patrika”	Precision: 72.15 Percent, Recall: 67.32 Percent, F-score: 69.65 Percent
Md. Iftekharul Alam Efat, Mohammad Ibrahim, Humayun Kayesh	“The Daily Prothom Alo”, “The Daily Ittefaq”, “The Daily Jugantor”	83.57 percent of summary sentences match
Md. Majharul Haque, Suraiya Pervin and Zerina Begum	“The Daily Prothom Alo” and “The Daily Jugantor”	Precision: 0.608 Recall: 0.664 F-score: 0.632

From table 2.1 we can see that different authors have used different data sets and different evaluation measures.

2.2 Research Problems

According to Kamal Sarkar [1] [3] the performance of the proposed system depends on preprocessing, stemming, sentence ranking. It may be improved by improving stemming process, exploring more features and applying learning algorithm for effective feature combination. From the table shown in the previous section, we clearly see that there is an opportunity to improve the unigram based recall score and LEAD baseline score in the summarizer they have proposed. According to Md. Nizam Uddin and Shakil Akter Khan [2] the main limitation of the Bangla summarizer is it just extracts some sentences from the given text which is much more different from the human generated summary. Another limitation is sometime the sentences which come early in the text have higher possibility to appear in the summary.

Amitava Das and Sivaji Bandyopadhyay [4] stated that, the evaluation result of their present summarization system for Bengali documents is reasonably good but still not outstanding. The main limitation of their system occurs for subjectivity identifier. The recall value of the classifier is higher than it which order to present them so that the whole text makes sense for the user. They prefer the original order of sentences as they occurred in original document.

Our aim is to generate summaries from given Bengali document(s) that are good and close to human created ones. To gain better scores in evaluation matrices we would explore in which way we should proceed to make better summaries.

The Problems with extractive text summarization [10] [11] [12] are:

- Sentences selected for summary generally longer, so unnecessary parts of the sentences for summary also get included they consume space.
- If summary size is not long enough, the important information scattered in various statements cannot captured using extractive summarization.
- Information which is clashing may not be presented accurately.
- Sentences frequently contain pronouns. They lose their referents when used out of context. If irrelevant sentences are clubbed together, may lead to confusing understanding of anaphors which will result in erroneous representation of original information.
- The same problem is with multi document summarization, because extraction of text is performed on different sources. Post processing can be used to deal with these troubles, for example, replacing pronouns with their background, replacing relative temporal expression with actual dates etc.

Our goal is to work on the above problems. Besides, **feature selection and combination** is also an important factor. All of the features cannot be used in a good summary as some features may worsen the quality of the summary. Again combining two or more features according to hierarchy of significance results in good summarization. So, we focused on this factor.

We have proposed a model to generate summaries so that the features are selected in a way that the scores such as precision, recall, f-measure are improved and the generated summaries match with the gold summaries which are human created.

2.3 Research Opportunities

Research opportunity refers to research problems which have not been answered appropriately or at all in a given field of study and there are scopes of improvement on those specific fields. Research opportunities make others' research publishable. Because it shows that existing research is not being duplicated and it shows that research has been conducted in order to fulfill that gap in the literature.

We have discussed about various types of research problems in the previous section. In our perspective, combining different methods can reduce the problems. So, we have applied different methods of scoring in our summarizer so that it can overcome the problems and generate better summaries.

2.4 Summary of the Chapter

We have discussed about the state of the art related to text summarization for Bengali document. In this chapter we came to know different literature reviews and the problems and issues that have been faced regarding text summarization. More focus is given to extractive approaches of text summarization. In the next chapter we will have a broad discussion about our proposed model and the methods we used in our summarizer for Bengali document.

Chapter 3

Proposed Model

In the previous chapter we talked about several literature reviews and problems regarding automatic text summarizer. In this chapter we will discuss about the model we proposed for our summarizer for Bengali document and different methods we used.

3.1 Introduction

We have devised a summarizer that generates extractive summary of Bengali documents. We give score to the sentences based on some criteria to select the best ones that represent the gist of a given document.

3.2 Proposed Statement

We have combined three approaches to determine the ranking of the sentences of the given document:

- Sentiment Ranking
- Keyword Ranking
- Text Ranking

We have added weight to each of them and the final ranking is based on the combination of the three.

Say,

$$SS*W1 + KR*W2 + TR*W3 = \text{Final Score of a sentence}$$

Where,

SS = Sentiment based score.

KR = Keyword ranking score.

TR = Text ranking based score.

W1 = A percentage of the total sentiment based score.

W2 = A percentage of the total keyword based score.

W3 = A percentage of the total text rank based score.

3.2.1 Sentiment Scoring

“PolyGlott” is a natural language pipeline that supports massive multilingual applications. It is a python’s library that supports Sentiment Analysis for 136 Languages. Unique Words Polarity can be calculated, and then sentences can have a scoring based on the neutral words (the words with polarity score 0) present in each sentence. So in a sentence, more neutral words mean more sentiment score.

3.2.2 Keyword Ranking

Keywords are extracted based on each category. Then according to their frequency, score can be given. More appearances of words in each category means higher score. After analyzing the data sets it can be determined that key phrases don’t put extra value on top of keywords. As there are minimal number of key phrases in our data sets.

3.2.3 Text Ranking

Text Rank, is a similarity-based ranking model for text processing which can be used in order to find the most relevant sentences in text and also to find keyword. Sentences are like nodes connected with each other based on their matching words. For example in cosine similarity, two sentences are plotted in a graph . The angle between them depends on their cosine similarity. The more cosine similarity will mean the lesser angle between them.

We have used the method “word2vec” from python’s library “gensim.models” as it was really important in finding correlated sentences. It is an effective way to represent words as vectors and create a vector model with them. The whole text is encapsulated in lower dimension space and in this space all vectors have certain orientation. So it is possible to define their relationship with each other. The representation of words as vectors helps to find the words in many applications in NLP. The angle that determines the correlation perhaps gets closer than any other method that is used in the case, e.g., cosine similarity which measures the similarity between two non-zero vectors. It involves measuring the cosine of the angle between them. It is most

used in word2vec. Two vectors with the same orientation have a cosine similarity of 1, two vectors at 90 degree orientation have a cosine similarity of 0, and two vectors in opposite direction have a similarity of -1 (independent of their magnitude). Figure 3.1 shows the inter connectivity between sentences.

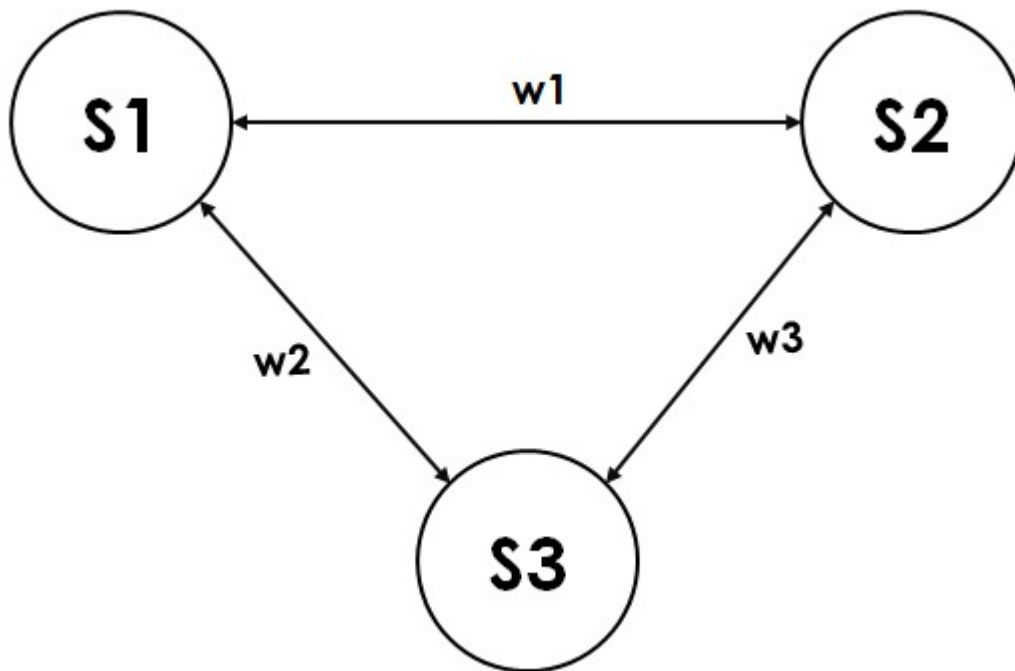


Figure 3.1: Inter connectivity between sentences.

Where,

$w1 = \text{Similarity}(S1, S2)$

$w2 = \text{Similarity}(S1, S3)$

$w3 = \text{Similarity}(S2, S3)$

In figure 3.1, S1, S2 and S3 refer to three sentences of a document. If we consider these three sentences as nodes and if we draw edges between them, then the inter connectivity between the three sentences can be shown as a graph and $w1$, $w2$ and $w3$ can be described as the similarities between (S1, S2) sentence pair, (S1, S3) sentence pair and (S2, S3) sentence pair respectively.

3.3 Necessity of a Hybrid Method

Our aim is to combine the mentioned three approaches to generate output summary. Therefore, our proposed design is a combination of Sentiment Scoring, Keyword Ranking and Text Ranking. So, we can consider our summarizer for Bengali document a hybrid one. Combining several methods performs well than the individual methods most of the times and the result is quite satisfactory and close to human generated summaries. Figure 3.2 shows the proposed design of our system.

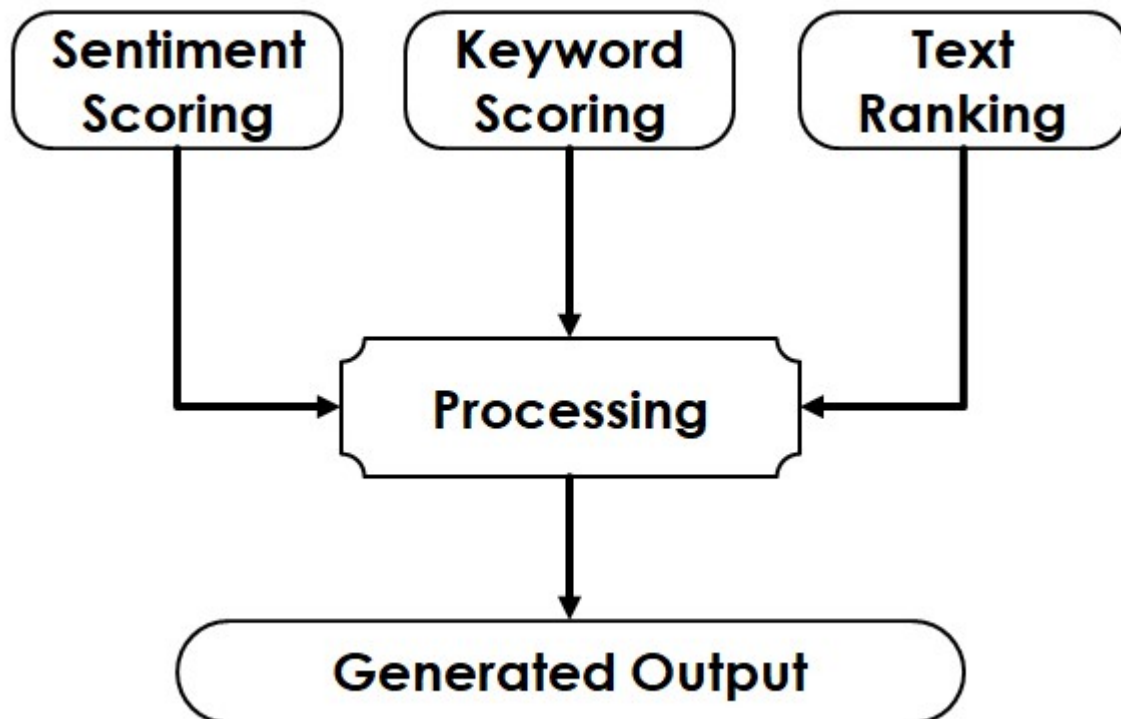


Figure 3.2: Proposed Design for Hybrid Summarizer for Bengali Document

3.4 Detailed Analysis of our proposed method

We will have a thorough idea about the steps we have followed in our devised summarizer. Figure 3.3 shows the workflow of our hybrid summarizer.

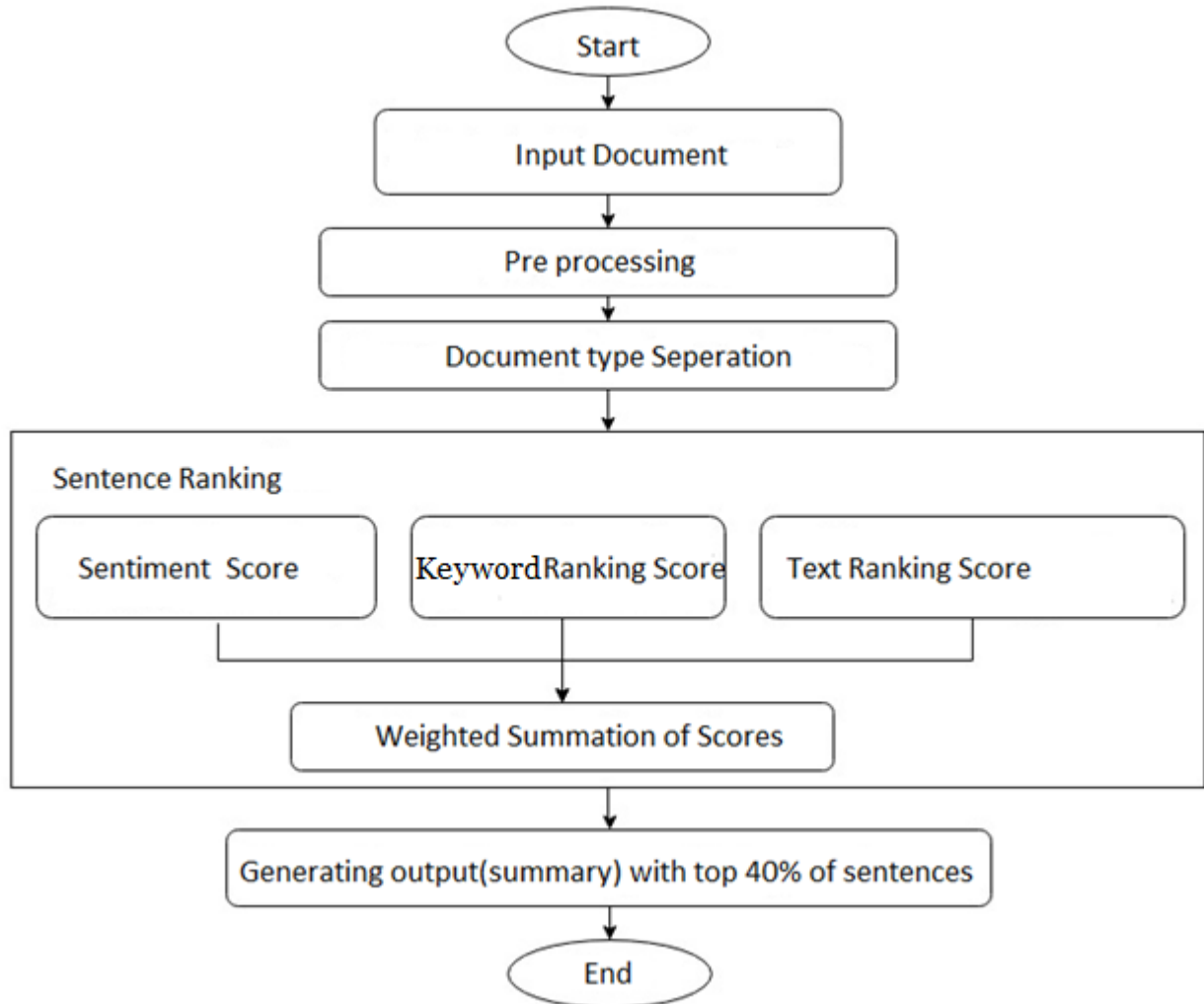


Figure 3.3: A flow chart of the proposed model for Hybrid summarizer for Bengali Document

3.4.1 Input Document

We can use any Bengali news document as input of the summarizer. The Document can be from a specific category but it's not absolutely necessary. For starters, it can be any news from our data sets, that are reserved for testing or training purpose.

3.4.2 Pre-processing

Pre-processing involves any type of processing performed on raw data to prepare it for another processing procedure. We performed pre-processing by splitting the input documents into words. We have used stopwords removal to remove irrelevant sentences that won't be added to the generated summaries.

3.4.3 Document Type Separation

Document Type separation involves separating documents in different categories. Before processing, the input document has been categorized as a specific type. We have worked on summarizing four categories of news: Politics, Economics, Entertainment and Accidents. So, the input document are categorized in any of the stated document types.

3.4.4 Sentence Ranking

Sentence ranking involves ranking the sentences of a document based on different scoring techniques. We have stated that we have worked on three methods: sentiment scoring, phrase ranking and text ranking. To get the highest ranked sentences, we have combined these three approaches. We have put weights to the different approaches and by adding the weighted scores we got a weighted summation of scores.

3.4.5 Generating Output Summary

After sentences are ranked, they are sorted in descending order. Highest ranked sentences get the highest probability of being selected. After basic filtering the top 40 Percent of the highest ranked sentences are selected to generate summary.

3.5 Steps and algorithms related to Hybrid Summarizer

3.5.1 Preparing Datasets for Training and Testing

3.5.1.1 Collection of Data sets manually

There is no open accessible data set for our topic for a proper verification of the summarization methods. We researched on different domains and selected the most important categories

to collect news from. We found out that people are more interested in reading accident, entertainment, economics and politics related news. So we chose top four categories: Accident, Entertainment, Economics and Politics to conduct our experiment. We prepared our own data sets by directly extracting news from “Daily Prothom Alo” newspaper without any modification. We have collected about 400 news documents from the mentioned categories. The data sets and news documents are available on our GitHub link [13].

From each category we collected 100 news, e.g., "দাউদকান্দিতে সড়ক দুর্ঘটনায় নিহত ৪ "কুমিল্লার দাউদকান্দি উপজেলায় যাত্রীবাহী বাস দুর্ঘটনায় চারজন নিহত এবং কমপক্ষে ১৫ জন আহত হয়েছে। উপজেলার সিংগুলাদীঘির পাড় এলাকায় ঢাকা-চট্টগ্রাম মহাসড়কে গতকাল রোববার দিবাগত রাত দেড়টার দিকে এ দুর্ঘটন ঘটে।. . . .

Extractive summaries have been generated by random people and are considered as gold summaries. Each summary is around 40 percent of the original document. The summaries generated by the summarizer are then compared with the gold summaries for generating rouge scores. Table 3.1 shows the categories and number of documents we collected from each category.

Table 3.1: Categories of news and number of documents collected from “Daily Prothom Alo”

News Category	Number of Documents
Politics	100
Economics	100
Entertainment	100
Accident	100

So, in total 400 documents from four different categories were used.

3.5.1.2 Collection of data sets from BNLPC

Apart from manual data set creation, we collected the data sets from "Bangla Natural Language Processing Community" (BNLPC) [14]. There are two data sets for the evaluation of Bangla text summarization system. Each set contains 100 Bangla news documents and three human generated model summaries for each document. Two groups of scholars (each group consists of three human judges) of Bangla language generated these summaries. From the data set we separated the documents into four categories.

3.5.2 Splitting Sentences

We need to separate the sentences. We separated all the sentences from all documents from each class. Then we prepared some text documents for each class by using sentence splitter of Polyglot where all the appropriate sentences from the news of each class are present. The

Polyglot splitter automatically splits each sentence finding ('|' or '.') from the whole document. We used a Stop-word list to remove the unnecessary and irrelevant sentences. We used the Stop-word list of Bengali-Sentiment-Analysis [15]. We extracted sentences from 50 documents and added them in a text file for each category e.g., accidentsentences.txt.

3.5.3 Keyword Extraction

We have found out that there are very few key phrases in Bengali documents. So, we worked on keywords instead of key phrases. There are four categories of documents in our system. We have extracted the keywords from all the documents in each category and calculated their frequencies. Tokenization in polyglot relies on the Unicode Text Segmentation algorithm. We created a list for storing all the unique words and a list for storing their frequencies. We stored the keywords with their respective scores in a document.e.g., 'সড়ক', 'গত', 'দুর্ঘটনায়', 'দুই', 'নিহত' and their respective frequencies are 163, 105, 105, 101, 98,

The algorithm is given below:

Algorithm 1 Unique words detection and frequency calculation

Result: Detect Unique Words and Count their frequencies

```

1 initialization
2 for each word in the list do
3   if Word Not in Stopwords then
4     if The word was existing then
5       Counter=Counter+1
6     else
7       Add word in the list.
8       Counter=1
9     end
10  else
11    do nothing
12  end
13 Sort words in descending order according to frequency.
```

After calculating frequency score, the words are sorted in descending order (most frequently used words to least frequently used words).

3.5.4 Classification of Documents

Classification involves determining class for unknown documents. For classifying an unknown document, we have crosschecked each and every word of the document with the four documents (one document from each category) prepared earlier that stored the keywords with scores. If a

match is found, the corresponding score in that category (same word can have different score in different category) is added. Thus, we have chosen the class with the highest score among those four scores. We can understand the concept by the following algorithm:

Algorithm 2 Classification of unknown documents

Result: Classify Unknown Document

```

1 for each word in unknown document do
2   if Word exists in keyword with score doc for accident then
3     | accident score = accident score + keyword score
4   else
5     | flag1=0
6   end
7   if Word exists in keyword with score doc for economics then
8     | economics score = economics score + keyword score
9   else
10    | flag2=0
11  end
12  if Word exists in keyword with score doc for entertainment then
13    | entertainment score = entertainment score + keyword score
14  else
15    | flag3=0
16  end
17  if Word exists in keyword with score doc for politics then
18    | politics score=politics score+ keyword score
19  else
20    | flag4=0
21  end
22 end
23 Classify doc to the category having the highest score
  
```

Table 3.2 and 3.3 show the classification results we got using our classifier for both setups.

Table 3.2: Classification result in 1st setup

Actual Class/Predicted Class	Accident	Economics	Entertainment	Politics
Accident	49	1	0	0
Economics	0	46	0	4
Entertainment	3	2	41	4
Politics	0	13	0	37

From the table 3.2 we can say that for the first setup the overall accuracy of our classifier is $(49 + 46 + 41 + 37) / (400 * 0.5) = 0.865$ or 86.5 Percent.

Table 3.3: Classification result in 2nd setup

Actual Class/Predicted Class	Accident	Economics	Entertainment	Politics
Accident	53	17	0	0
Economics	0	70	0	0
Entertainment	2	51	15	2
Politics	0	20	0	50

From the table 3.3 we can say that for the second setup the overall accuracy of our classifier is $(53 + 70 + 15 + 50) / (400 * 0.7) = 0.671$ or 67.1 Percent.

3.5.5 Score Calculation (IDF) of words

We calculated the score of each word by using the following formula:

Word score = Word Frequency / No of words present in all documents in each category. e.g., 'সড়ক', '0.0044033822', 'গত', '0.0028365346', 'দুর্ঘটনায়', '0.0028365346', 'দুই', '0.0027284761', 'নিহত', '0.0026474323',

3.5.6 Individual Scoring for Each Sentence

Using text function of polyglot, we stored the previously scored words in a variable (textual). For each document from each category, every sentence is read and stored in a variable (totaltext). The sentences are split by polyglot splitter. Every word in each sentence is matched with the words previously stored with frequency score and the score is taken. For each word in a sentence, the scores are added and the sentence score is stored in a new list.

E.g., উপজেলার সিংগুলাদীঘির পাড় এলাকায় ঢাকা-চট্টগ্রাম মহাসড়কে গতকাল রোববার দিবাগত রাত দেড়টার দিকে এ দুর্ঘটনা ঘটে।

Score: 0.0115352405.

The algorithm is given below:

Algorithm 3 Individual Sentence Scoring based on Keyword and Decision of Summary

Result: Scoring of each sentence based on Keyword,summary decision

```

1 initialization
2 for each sentence in all sentences do
3     sentence count=sentence count+1
4     sentence score = 0
5     for each word in each sentence do
6         if Word is in scored word text then
7             Extract keyword score
8             sentence score=sentence score + keyword score
9         else
10            do nothing
11        end
12    end
13 Sort the sentences in the doc in descending order.
14 for Top 40 percent of sorted sentence do
15     if length of sentence is >= 10 then
16         Take the sentence in the summary document
17     else
18         do nothing
19     end
20 end

```

After individual scoring of sentences, they are sorted in descending order (higher scored sentences to lower scored sentences). Then top 40 percent sentences are selected for summarization.

3.5.7 Sentiment Based Ranking

For sentiment scoring, we have used a predefined library Polyglot's function ("Word.Polarity()"). By using this polarity of each unique word has been calculated. Words have been scored based on their polarity. Using a function of polyglot we calculated the polarity of the words (+1 for positive words, -1 for negative words and 0 for neutral words). E.g., "", 0, 'বারবার', 0, 'কেন', 0, 'ডুবছে', 0, 'নৌযান', 0, '!', 0, "", 0,

For sentiment wise ranking, we used the following algorithm:

Algorithm 4 Ranking Based on Sentiment and Decision of Summary

Result: Sentiment Scoring of Each Sentence, Summary Decision

```

1 initialization
2 for each sentence in all sentences do
3   sentence count=sentence count+1
   sentiment score = 0
   for each word in polyglot sentiment score do
4     if Word With Sentiment Score = 0 then
5       | Sentiment Score= Sentiment Score + 0.0001
6     else
7       | do nothing
8     end
9   end
10 end
11 Sort the sentences in the doc in descending order.
   for Top 40 percent of sorted sentence do
12   if length of sentence is >= 10 then
13     | Take the sentence in the summary document
14   else
15     | do nothing
16   end
17 end

```

After the words are ranked sentiment wise, for each sentence in a document the sentiment scores of the words are added to score the sentences sentiment wise. Then sentences are sorted in descending order (sentences with higher sentiment score to sentences with lower sentiment score). After that top 40 percent sentences are selected for summarization.

3.5.8 Text Ranking

Text Ranking is another method we used in our hybrid summarizer. Summaries were generated using text ranking and were compared with the summaries generated by the hybrid summarizer.

3.5.8.1 Generating Vector Model from training data sets

This method includes tokenization of each sentence from training data sets (both setups for each category) and converting them into sentence vectors. In order to do so we used “word2vec” imported from “gensim.models”. Words which are not in the stop word list were tokenized using “nltk.wordtokenize”. Then a training vector model 'trainedsentence.model' was created. The algorithm is given below:

Algorithm 5 Genrating the vector model**Result:** Vector Model

```

1 initialization
2 for Each Sentence in all documents from training do
3   | TokenizeCorpus = nltk.wordtokenize(Sentence)
4 end
5 Model=Word2Vec(TokenizeCorpus,sg=1,size=100>window=5,mincount=5,workers=4,iter=100)
   Model.save("TrainedModel.model")

```

3.5.8.2 Scoring sentences based on similarities

In order to score sentences using text rank method, we need to find similarities between sentences. We had earlier created and saved a model using (Word2Vec) and the vector model was loaded. After that each sentence was converted to vector using our user defined function “getvector” from the class “sentence2vec”. Then the sentences were given score based on their similarities. Similarities between sentences were determined using a user defined function “similarity”. After that the size of each vector was checked and the score was calculated for the nonzero pairs. The similarity score was calculated dividing the dot product (.) of two vector sentences by the multiplication of the normal of those two vectors and the final score was divided by ten. We divided the score by 10 for easier calculation, but later we normalized the score. The algorithm is given below:

Algorithm 6 Individual Sentence Scoring Based on sentence inter connectivity/similarity

```

1: procedure Similarity(sent1,sent2)
2:   vector1=self.getvector(sent1);
3:   vector2=self.getvector(sent2);
4:   score=0 ;
5:   if vector1.size and vector2.size>0 and norm(vector1)!=0 and norm(vector2)!=0 then
6:     score=dot(vector1,vector2)/norm(vector1)*norm(vector2)
7:   return score/10
8: end procedure

```

The above scoring method was applied to the testing data sets (in both setups for each category) and the sentences were sorted in descending order (highest to lowest) and thus summary based on text ranking was generated.

3.5.9 Hybrid Scoring 1 (Keyword Ranking and Sentiment Scoring)

In this model we have combined the two methods: keyword ranking and sentiment score ranking and calculated a combined score. We have added a weight to keyword ranking of a sentence and the weight is 40 percent of the sentence's total key word ranking. Similarly, a weight of 60

percent of the total sentiment score is added to the sentence's sentiment score and both weighted scores are summed up to calculate the hybrid score.

Let,

Keyword Ranking * W1 + Sentiment scoring * W2 = Combined Score of a sentence.

Where,

W1 = 40 percent of the total keyword ranking score.

W2 = 60 percent of the total sentiment score.

3.5.10 Hybrid Scoring 2 (Keyword Ranking, Sentiment Scoring and Text Ranking)

In this model we generated summaries combining Keyword Ranking, Sentiment Scoring and Text Ranking. We have added a weight to keyword ranking of a sentence and the weight is 30 percent of the sentence's total key word ranking. Similarly, a weight of 20 percent of the total sentiment score is added to the sentence's sentiment score and a weight of 50 percent of the total text rank score is added to the sentence's text rank score. The weighted scores are summed up to calculate the hybrid score.

Let,

Keyword Ranking * W1 + Sentiment scoring * W2 + Text Ranking * W3 = Combined Score of a sentence.

Where,

W1 = 30 percent of the total keyword ranking score.

W2 = 20 percent of the total sentiment score.

W3 = 50 percent of the total Text Rank score.

3.5.11 Hybrid Scoring 3 (Keyword Ranking, Sentiment Scoring and Text Ranking)

We conducted another experiment where we have added a weight to keyword ranking of a sentence and the weight is 50 percent of the sentence's total key word ranking and a weight of 30 percent of the total text rank score is added to the sentence's text rank score. We did not change the weight of sentiment score. The weighted scores are summed up to calculate the hybrid score.

Let,

Keyword Ranking * W1 + Sentiment scoring * W2 + Text Ranking * W3 = Combined Score of

a sentence.

Where,

W1 = 50 percent of the total keyword ranking score.

W2 = 20 percent of the total sentiment score.

W3 = 30 percent of the total Text Rank score.

3.6 Sample Generated Output

We will now explain the scoring methods of different systems with an example.

To generate summary based on Keyword Ranking, let us calculate the score of the following sentence.

নোয়াখালীতে সড়ক দুর্ঘটনায় নিহত দুই

Here individual score of each word of the sentence

'নোয়াখালীতে', '0.0002701462'

'সড়ক', '0.0044033822'

'দুর্ঘটনায়', '0.0028365346'

'নিহত', '0.0026474323',

'দুই', '0.0027284761'

The total score of this sentence (based on keyword ranking) after adding the individual score of each word is 0.0102385391.

Similarly, the score of the following sentence based on keyword ranking is,

নোয়াখালীর সুবর্ণচর উপজেলার পরিষ্কার বাজার রাস্তার মাথা নামক স্থানে সোনাপুর চরজব্বার সড়কে যাত্রীবাহী সিএনজিচালিত অটোরিকশা ও ইজিবাইকের মধ্যে সংঘর্ষে দুজন নিহত হয়েছে।

0.0062133614

To generate summary based on Sentiment Scoring, let us calculate the score of the following sentence.

নোয়াখালীর সুবর্ণচর উপজেলার পরিষ্কার বাজার রাস্তার মাথা নামক স্থানে সোনাপুর চরজব্বার সড়কে যাত্রীবাহী সিএনজিচালিত অটোরিকশা ও ইজিবাইকের মধ্যে সংঘর্ষে দুজন নিহত হয়েছে।

We select the neutral words for summary and for each neutral word the score is 0.0001. Finally the score of each neutral word is added and we get the score of the sentence.

Here in this sentence there are 21 neutral words.

নোয়াখালীর সুবর্ণচর উপজেলার বাজার রাস্তার মাথা নামক স্থানে সোনাপুর চরজব্বার সড়কে যাত্রীবাহী সিএনজিচালিত অটোরিকশা ও ইজিবাইকের মধ্যে সংঘর্ষে দুজন নিহত হয়েছে।

So, total sentiment score of this sentence is $(21 * 0.0001)$ 0.0021000000000000003.

To generate summary based on Text Ranking, we calculate the score of a sentence by comparing the sentence with other sentences of the document. gensim.model calculates the score of each sentence using model.similarity. Let us consider the following sentence.

নোয়াখালীর সুবর্ণচর উপজেলার পরিষ্কার বাজার রাস্তার মাথা নামক স্থানে সোনাপুর চরজব্বার সড়কে যাত্রীবাহী সিএনজিচালিত অটোরিকশা ও ইজিবাইকের মধ্যে সংঘর্ষে দুজন নিহত হয়েছে।

The total score of the sentence based on text ranking is 0.5952268240310518

To generate summary based on Hybrid Ranking (KR+SS+TR), first we normalize the score we got from keyword ranking, sentiment scoring and text ranking and then combine the three scores each multiplied by a percentage of the total weight. For the document mentioned above, we got total keyword score (summing up the keyword scores of each sentence in the document) 0.04854526270000003, total sentiment score (summing up the sentiment scores of each sentence in the document) 0.012099999999999982 and total text rank score (summing up the text rank scores of each sentence in the document) 4.780813066878625 for the document. We divide the individual keyword, sentiment and text rank score of the sentence by the total keyword score, sentiment score and text rank score respectively. The result is then multiplied with 0.3 (keyword), 0.2 (sentiment) and 0.5 (text rank) and we get the final score by adding these three scores. So the final score of the following sentence is:

নোয়াখালীর সুবর্ণচর উপজেলার পরিষ্কার বাজার রাস্তার মাথা নামক স্থানে সোনাপুর চরজব্বার সড়কে যাত্রীবাহী সিএনজিচালিত অটোরিকশা ও ইজিবাইকের মধ্যে সংঘর্ষে দুজন নিহত হয়েছে।

$$0.3*(0.0062133614/0.04854526270000003) + 0.2*(0.0021000000000000003/0.012099999999999982) + 0.5*(0.5952268240310518/4.780813066878625) = 0.13049065159304923$$

These are just the samples. We have uploaded all the resources i.e., original news documents, summaries and their scores along with individual sentence and their ROUGE scores on GitHub [13].

3.7 Summary of the Chapter

In this chapter, we have discussed about the proposed methods we have used in our summarizer. We have also given solid reasons why we named it a Hybrid summarizer. We have mentioned the steps we are following to generate summaries. In the next chapter we will gather knowledge about different types of evaluation measures.

Chapter 4

Evaluation Measures of Summary

In the previous chapter we have talked about the proposed model of our hybrid summarizer for Bengali document and the steps along with the algorithms we used in our system. Here in this chapter our topic of discussion will be different types of summary evaluation techniques.

4.1 Introduction

Automatic summarization takes an information source, extract content from it, and users can get the gist content of a document. As automatic summarization is out of bias and not human created, it may lack the importance and expression of human mind and sometimes it cannot express the main idea of the document properly. So, it is important to understand what makes summary a good summary. To identify good summaries, it is important to evaluate summaries using several measures. So, evaluation of summary is an important task.

4.2 Evaluation Measures

There have been many researches on summary evaluation and many measures have been invented. Josef Steinberger, Karel Jeřek [16] stated that, intrinsic content evaluation is the main approach for determining summary quality that is often done by comparison with an ideal or reference or gold summary. For sentence extractions, it is measured by co-selection. It detects the number of ideal sentences in the automatic summary. Content-based measures compare the words that are present in a sentence. Their advantage is that they can compare both human and automatic summaries and can extract with human abstracts that contain newly written sentences. Another major group are extrinsic methods. They are also known as task-based methods. They measure the performance of using the summaries for a certain task. So basically, there are two

types of measures:

- Intrinsic Measures
- Extrinsic Measures

Saiyed Saziabegum, Dr. Priti S. Sajja [10] have defined intrinsic and extrinsic measures. Intrinsic measures check the summarization system itself. It determines the quality of summary by comparing the automatically generated summary and the human made summary. Summary is evaluated regarding two aspects: quality or informativeness. The informativeness of a summary is evaluated by comparing it with a human-made summary.

Extrinsic measures check the summarization on how it can influence the completion of some other task such as text classification, information retrieval, answering of question etc. It evaluates the quality of summarization using different tasks like reading comprehension, relevance assessment etc.

Our focus is on mainly Intrinsic Measure.

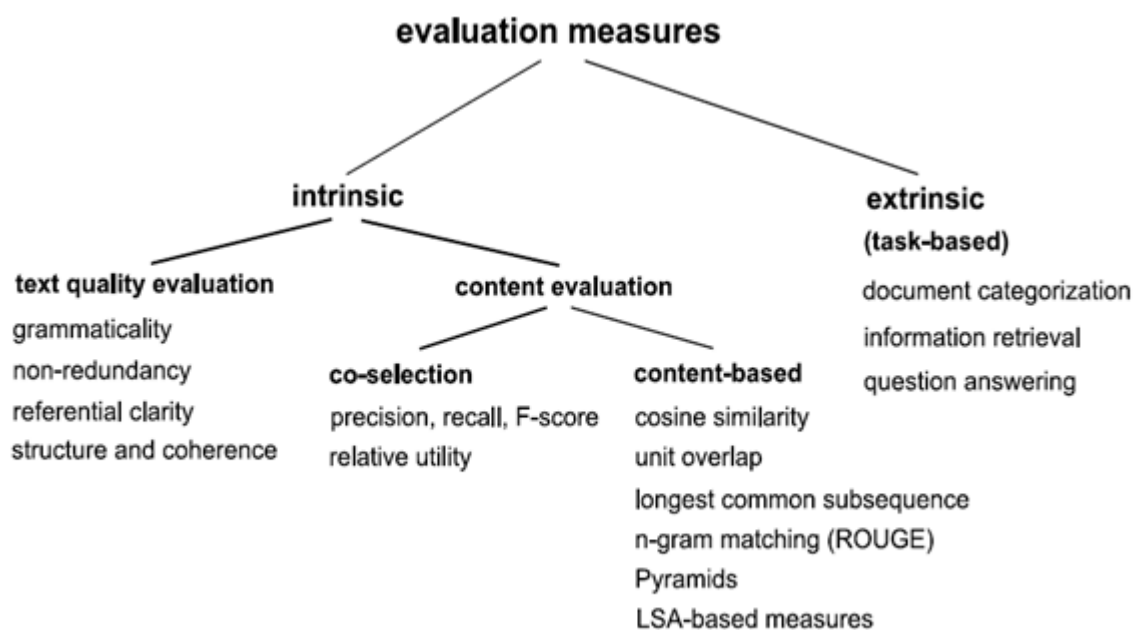


Figure 4.1: Classification of summary evaluation measures

4.3 Intrinsic Measure

From the figure 4.1, we see that intrinsic measures can be classified into two another classes:

- Text Quality Evaluation
- Content Evaluation

We will have a detailed idea about different types of intrinsic measures from this section.

4.3.1 Text Quality Evaluation

Under this class we see text quality can be evaluated in different ways. Such as,

Grammaticality: The text contains textual items and punctuation errors or incorrect words should not be contained in the text.

Non-redundancy: The text should not contain unnecessary information.

Reference clarity: The nouns and pronouns of sentences should be clearly referred to in the summary e.g., the pronoun 'he/ she' has to mean somebody in the context of the summary.

Coherence and structure: The summary should have good structure and the sentences need to be coherent. This cannot be done automatically. Marks are assigned to each summary.

4.3.2 Content Evaluation

Under this class we see that contents can be evaluated by co-selection and based on the contents.

Co-selection involves precision, recall, f-score and relative utility.

- **Precision** : Precision is defined as the number of words present in both system generated and gold summaries divided by the total number of words in the system generated summary.
- **Recall** : Recall is defined as the number of overlapping words occurring in both system generated and gold summaries divided by the total number of words in the gold summary.
- **F-score** : F-score is a evaluation measure that combines precision and recall. The basic way to compute the F-score is to calculate a harmonic mean of precision and recall. The formula is given below: $F = 2 * P * R / (P+R)$
- Content based evaluation involves cosine similarity, ROUGE etc. measures.

- **Cosine Similarity:** Cosine similarity is a commonly used approach to match similar documents. It refers to counting the maximum number of common words between the documents. It is basically a metric that is used to measure how similar the documents are in respect to their size. In terms of mathematics, this method measures the cosine of the angle between two vectors that are projected in a multi-dimensional axis.
- **ROUGE :** It is the abbreviation of **Recall-Oriented Understudy for Gisting Evaluation**. It refers to a set of metrics and a software package. It is mainly used for evaluating summarization. The metrics compare an automatically produced summary or translation against a reference or a set of references (human-produced) summary or translation. It measures the quality of summary by counting overlapping units such as the n-gram, sequences of words i.e., which word is appearing after which and word pairs between the generated and reference summary.
- **ROUGE-N:** It is defined as the overlapping of N-grams between the system's generated and reference summaries.
- **ROUGE-1:** It determines the overlap of 1-gram (each word) between the system generated and reference summaries. In this measure, summary is evaluated comparing the precision, recall and f-measure scores for 1-gram.
- **ROUGE-2:** It determines the overlap of bi-grams between the system generated and reference summaries. In this measure, summary is evaluated comparing the precision, recall and f-measure scores for bi-grams.

4.4 Summary of the Chapter

We discussed about the different measures to evaluate whether a summary is good or bad. We broadly explained the main classifications evaluation measure: intrinsic and extrinsic and their different types. In the next chapter we will get to know how we measured the performance and evaluation scores of our hybrid system.

Chapter 5

Experimental Verification

In the previous chapter we discussed about several types of evaluation measure techniques of summary. In this chapter we will explain how we evaluated our summarizer and the output of ROUGE scores.

5.1 Introduction

We are discussing about our work to inform our supervisor and associates about the progress we've made on our summarizer in this chapter. We will provide our recipients with a brief look at the findings and work of the thesis.

5.2 Experimental Setup and Verification

We have stated earlier that we have generated summaries based on keyword, sentiment scoring and text ranking and also combining these three approaches. Previously we generated another hybrid that combines only keyword and sentiment ranking. We have evaluated our generated summaries based on ROUGE score. It measures summary quality by counting overlapping units such as the n-gram, word sequences and word pairs between the candidate summary and reference summary.

We have used two different module for data sets. In the first module we have used manually created data sets, and in the second module we have used data sets collected from “Bangla Natural Language Processing Community” [14].

5.2.1 Experimental setup on Manual Data set

We have evaluated our generated summaries by taking 50 percent, 30 percent of the total documents respectively for training purpose and 50 percent, 70 percent of the total documents for testing in keyword extraction and hybrid method.

That means, we have done evaluation in two setups. In 1st setup, the system is trained with 50 percent documents and has been tested for the rest 50 percent documents.

In 2nd setup, the system is trained with 30 percent documents and has been tested for the rest 70 percent documents. Table 5.1 shows the number of documents in each category and also the number of documents selected for training and testing for both setups.

Table 5.1: Categories and number of documents for training and testing

Document Category	Number of Documents	Doc no. for testing(1st setup)	Doc no. for Training(1st setup)	Doc no. for testing(2nd setup)	Doc no. for Training(2nd setup)
Accidents	100	50	50	70	30
Economics	100	50	50	70	30
Entertainment	100	50	50	70	30
Politics	100	50	50	70	30

For summary based on sentiment we have used 100 percent of the documents because we don't have to train the data set as we have used python's existing library polyglot's polarity scores for calculating positive, negative and neutral sentences.

We calculated Rouge-1 and Rouge-2 precision, recall and f-measures for both experiments. We have used the following formulas to calculate ROUGE precision, recall and f-measure.

ROUGE recall = No. of overlapping words / No. of words in the gold summary

ROUGE precision = No. of overlapping words / No. of words in the generated summary

F-measure = $2 * \text{Precision} * \text{Recall} / (\text{Precision} + \text{Recall})$.

We can get a clear overview if we see the tables and charts regarding ROUGE evaluation.

5.2.2 Experimental Verification with Evaluation on Manual Data set

We have evaluated our generated summaries using Rouge-1 and Rouge-2 measures. We have already stated that, there are two types of setups we have experimented (training 50 percent , 30 percent and testing 50 percent , 70 percent respectively). In this section we will see the ROUGE scores in both setups along with their bar charts.

ROUGE Scores of different categories for **1st Setup** are shown in the following tables.

For Category Accidents: Rouge-1 and Rouge-2 precision, recall and f-measure scores are shown in table 5.2.

Table 5.2: Average Rouge Score of the system for Category Accidents (1st Setup).

Scoring Criteria	Keyword Rank-ing	Sentiment Scoring	Text Rank-ing	Hybrid Ranking 1(.4KR+.6SS)	Hybrid Ranking 2(.3KR+.2SS+.5TR)	Hybrid Ranking 3(.5KR+.2SS+.3TR)
Precision	0.6615	0.5308	0.5712	0.6636	0.6737	0.6679
	0.5726	0.4488	0.4745	0.5882	0.5984	0.5878
Recall	0.5949	0.6970	0.6768	0.6434	0.6589	0.6376
	0.5241	0.6073	0.5990	0.5783	0.5934	0.5701
F-measure	0.6188	0.5841	0.6127	0.6456	0.6606	0.6454
	0.5403	0.4961	0.5224	0.5762	0.5902	0.5722

In table 5.2, for precision, recall and f-measure, upper part of the row refers to Rouge-1 scores and lower part of the row refers to Rouge-2 scores. So we can see that Rouge scores of the hybrid systems beat individual ranking methods. And we can clearly see that hybrid combining three methods (KR, SS, TR) gives better result than hybrid combining two methods (KR, SS). We also calculated the ROUGE scores swapping the weights of keyword and text rank and definitely we get better F-score if the text rank is given 50 percent of the total weight.

For Category Economics: Rouge-1 and Rouge-2 precision, recall and f-measure scores are shown in table 5.3.

Table 5.3: Average Rouge Score of the system for Category Economics (1st Setup).

Scoring Criteria	Keyword Rank-ing	Sentiment Scoring	Text Rank-ing	Hybrid Ranking 1(.4KR+.6SS)	Hybrid Ranking 2(.3KR+.2SS+.5TR)	Hybrid Ranking 3(.5KR+.2SS+.3TR)
Precision	0.5159	0.4659	0.4584	0.5080	0.5001	0.5103
	0.4330	0.3806	0.3720	0.4273	0.4206	0.4311
Recall	0.6333	0.6922	0.7028	0.6511	0.6599	0.6572
	0.5219	0.5568	0.5841	0.5361	0.5456	0.5441
F-measure	0.5602	0.5506	0.5489	0.5619	0.5613	0.5656
	0.4680	0.4473	0.4509	0.4702	0.4705	0.4756

In table 5.3, for precision, recall and f-measure, upper part of the row refers to Rouge-1 scores and lower part of the row refers to Rouge-2 scores. So we can see that Rouge scores of the hybrid systems beat individual ranking methods. Here we see that hybrid combining three methods (KR, SS, TR) where keyword ranking has a weight of 50 percent gives better result than hybrid combining two methods (KR, SS) and the hybrid combining three methods where text ranking

has a weight of 50 percent.

For Category Entertainment: Rouge-1 and Rouge-2 precision, recall and f-measure scores are shown in table 5.4.

Table 5.4: Average Rouge Score of the system for Category Entertainment (1st Setup).

Scoring Criteria	Keyword Rank-ing	Sentiment Scoring	Text Rank-ing	Hybrid Ranking 1(.4KR+.6SS)	Hybrid Ranking 2(.3KR+.2SS+.5TR)	Hybrid Ranking 3(.5KR+.2SS+.3TR)
Precision	0.3725	0.4255	0.4128	0.3882	0.3940	0.3792
	0.2794	0.3468	0.3229	0.3016	0.3092	0.2893
Recall	0.5110	0.6347	0.6588	0.5706	0.5818	0.5467
	0.3944	0.5173	0.5471	0.4558	0.4665	0.4277
F-measure	0.4099	0.4893	0.4841	0.4403	0.4482	0.4262
	0.3077	0.3981	0.3839	0.3417	0.3503	0.3244

In table 5.4, for precision, recall and f-measure, upper part of the row refers to Rouge-1 scores and lower part of the row refers to Rouge-2 scores. So we can see that Rouge-1 and Rouge-2 F-score for sentiment scoring and text ranking is better than the hybrid system 1 and hybrid system 2. Here we see that hybrid combining three methods (KR, SS, TR) where keyword ranking has a weight of 50 percent gives better F-score for Rouge-1 than hybrid combining two methods (KR, SS) and the hybrid combining three methods where text ranking has a weight of 50 percent. From the data set we have observed the documents from entertainment class were of very poor qualities, and they need a lot of pre-processing before feeding to the summarizer, which was not done specifically. So performance slightly falls for hybrid systems in case of entertainment category.

For Category Politics: Rouge-1 and Rouge-2 precision, recall and f-measure scores are shown in table 5.5.

Table 5.5: Average Rouge Score of the system for Category Politics (1st Setup).

Scoring Criteria	Keyword Rank-ing	Sentiment Scoring	Text Rank-ing	Hybrid Ranking 1(.4KR+.6SS)	Hybrid Ranking 2(.3KR+.2SS+.5TR)	Hybrid Ranking 3(.5KR+.2SS+.3TR)
Precision	0.6205	0.6124	0.5921	0.6233	0.6210	0.6221
	0.5051	0.5088	0.4741	0.5098	0.5102	0.5088
Recall	0.6885	0.7503	0.7321	0.7264	0.7057	0.7137
	0.5836	0.6435	0.6336	0.6035	0.6247	0.6144
F-measure	0.6500	0.6708	0.6507	0.6594	0.6667	0.6622
	0.5385	0.5645	0.5382	0.5498	0.5585	0.5524

In table 5.5, for precision, recall and f-measure, upper part of the row refers to Rouge-1 scores and lower part of the row refers to Rouge-2 scores. So we can see that ROUGE scores of the hybrid systems beat individual ranking methods Keyword ranking, text ranking but cannot beat sentiment scoring. The difference is negligible. And we can clearly see that hybrid combining three methods (KR, SS, TR) gives better result than hybrid combining two methods (KR, SS). We also calculated the ROUGE scores swapping the weights of keyword and text rank and definitely we get better F-score for politics category if the text rank is given 50 percent of the total weight.

For 2nd setup, we haven't calculated ROUGE scores for sentiment-based ranking in our second setup because sentiment scores were calculated using polyglot polarity scores for 100 percent of the total document in both setups. So the result is same as the 1st setup which we have shown in previous tables. ROUGE scores of Different categories for **2nd Setup** are shown in the following tables.

For Category Accidents: Rouge-1 and Rouge-2 precision, recall and f-measure scores are shown in table 5.6.

Table 5.6: Average Rouge Score of the system for Category Accident (2nd Setup).

Scoring Criteria	Keyword Ranking	Text Rank-ing	Hybrid Ranking 1(.4KR+.6SS)	Hybrid Ranking 2(.3KR+.2SS+.5TR)	Hybrid Ranking 3(.5KR+.2SS+.3TR)
Precision	0.5927	0.5238	0.6000	0.6103	0.6056
	0.4975	0.4225	0.5125	0.5276	0.5164
Recall	0.6195	0.6791	0.6551	0.6801	0.6567
	0.5382	0.5878	0.5804	0.6060	0.5790
F-measure	0.5875	0.5793	0.6095	0.6280	0.6132
	0.4986	0.4787	0.5267	0.5470	0.5280

In table 5.6, for precision, recall and f-measure, upper part of the row refers to Rouge-1 scores and lower part of the row refers to Rouge-2 scores. So we can see that ROUGE scores of the hybrid systems beat individual ranking methods. We also calculated the ROUGE scores swapping the weights of keyword and text rank and definitely we get better F-score in this setup for Accident category category if the text rank is given 50 percent of the total weight.

For Category Economics: Rouge-1 and Rouge-2 precision, recall and f-measure scores are shown in table 5.7.

Table 5.7: Average Rouge Score of the system for Category Accident(2nd Setup).

Scoring Criteria	Keyword Ranking	Text Ranking	Hybrid Ranking 1(.4KR+.6SS)	Hybrid Ranking 2(.3KR+.2SS+.5TR)	Hybrid Ranking 3(.5KR+.2SS+.3TR)
Precision	0.5336	0.4816	0.5327	0.5222	0.5276
	0.4479	0.3938	0.4492	0.4401	0.4454
Recall	0.6497	0.7332	0.6662	0.6831	0.6760
	0.5402	0.6212	0.5543	0.5691	0.5643
F-measure	0.5788	0.5761	0.5845	0.5853	0.5854
	0.4850	0.4788	0.4914	0.4921	0.4934

In table 5.7, for precision, recall and f-measure, upper part of the row refers to Rouge-1 scores and lower part of the row refers to Rouge-2 scores. So we can see that Rouge scores of the hybrid systems beat individual ranking methods. We also calculated the Rouge scores swapping the weights of keyword and text rank and we get slightly better F-score in this setup for Economics category category if the text rank is given 30 percent of the total weight than the hybrid system where text rank has 50 percent of the total weight. But the difference is negligible.

For Category Entertainment: Rouge-1 and Rouge-2 precision, recall and f-measure scores are shown in table 5.8.

Table 5.8: Average Rouge Score of the system for Category Entertainment(2nd Setup).

Scoring Criteria	Keyword Ranking	Text Ranking	Hybrid Ranking 1(.4KR+.6SS)	Hybrid Ranking 2(.3KR+.2SS+.5TR)	Hybrid Ranking 3(.5KR+.2SS+.3TR)
Precision	0.4050	0.4076	0.4023	0.4026	0.3998
	0.3136	0.3146	0.3142	0.3160	0.3122
Recall	0.5302	0.6175	0.5635	0.5705	0.5524
	0.4149	0.5055	0.4452	0.4537	0.4350
F-measure	0.4387	0.4702	0.4487	0.4509	0.4432
	0.3401	0.3695	0.3501	0.3538	0.3457

In table 5.8, for precision, recall and f-measure, upper part of the row refers to Rouge-1 scores and lower part of the row refers to Rouge-2 scores. Here we can see that Rouge scores of the

text ranking system give better F-score than the hybrid systems. We also calculated the Rouge scores swapping the weights of keyword and text rank and we get better F-score in this setup for Entertainment category if the text rank is given 50 percent of the total weight than the hybrid system where text rank has 30 percent of the total weight. But the F-score of each hybrid system beats the F-score of other individual systems. In general there were a downgrade of f-measure for all the methods and we have already stated the reason.

For Category Politics: Rouge-1 and Rouge-2 precision, recall and f-measure scores are shown in table 5.9.

Table 5.9: Average Rouge Score of the system for Category Politics (2nd Setup).

Scoring Criteria	Keyword Ranking	Text Ranking	Hybrid Ranking 1(.4KR+.6SS)	Hybrid Ranking 2(.3KR+.2SS+.5TR)	Hybrid Ranking 3(.5KR+.2SS+.3TR)
Precision	0.6145	0.5898	0.6122	0.6135	0.6121
	0.5007	0.4727	0.4995	0.5044	0.5000
Recall	0.6881	0.7343	0.6996	0.7224	0.7044
	0.5838	0.6373	0.5950	0.6195	0.6004
F-measure	0.6461	0.6495	0.6530	0.6601	0.6519
	0.5357	0.5382	0.5397	0.5525	0.5422

In table 5.9, for precision, recall and f-measure, upper part of the row refers to Rouge-1 scores and lower part of the row refers to Rouge-2 scores. So we can see that Rouge scores of the hybrid systems beat individual ranking methods. And we can clearly see that hybrid combining three methods (KR, SS, TR) gives better result than hybrid combining two methods (KR, SS). We also calculated the Rouge scores swapping the weights of keyword and text rank and definitely we get better F-score if the text rank is given 50 percent of the total weight.

We have calculated the average Rouge-1 score of all four categories. We will see tables and bar diagrams of average Rouge-1 Scores (both setups) for different methods. Bar charts along with the table showing average Rouge-1 scores based on Keyword, Sentiment, Text rank and Hybrid rank in 1st Setup is shown in figure 5.1.

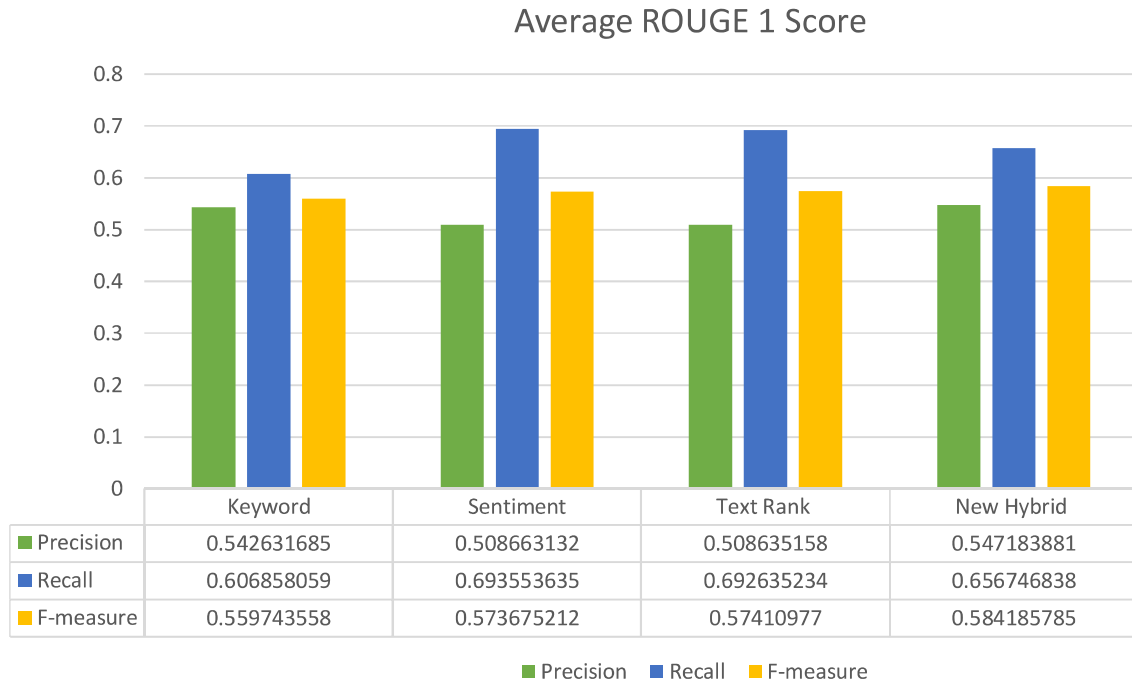


Figure 5.1: Average Rouge-1 Scores of the system for different methods (1st Setup)

From the chart we can say in terms of Precision, Hybrid ranking gets slightly better scores than Keyword ranking. In terms of Recall, Sentiment scoring and text rank are the highest, but in terms of F-Measure Hybrid ranking again beats them all.

Bar charts along with the table showing average Rouge-1 scores based on Keyword, Sentiment, Text rank and Hybrid rank in 2nd Setup is shown in figure 5.2.

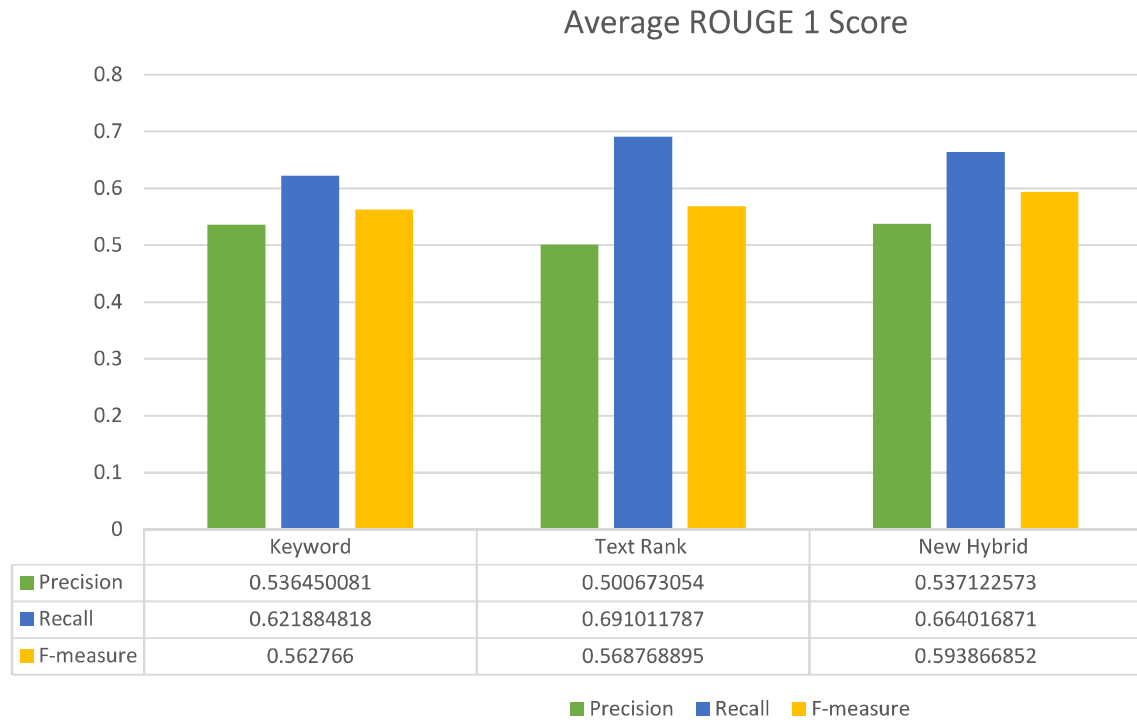


Figure 5.2: Average Rouge-1 Scores of the system for different methods (2nd Setup)

From the chart we can say in terms of Precision, Hybrid ranking gets slightly better scores than Keyword ranking and Text ranking. In terms of Recall, Text rank scoring is the highest, but in terms of F-Measure Hybrid scoring beats them all.

Similarly, we have calculated the average Rouge-2 score of all four categories. We will see tables and bar diagrams of average Rouge-2 Scores (both setups) for different methods. Bar charts along with the table showing average Rouge-2 scores based on Keyword, Sentiment, Text rank and Hybrid rank in 1st Setup is shown in figure 5.3.

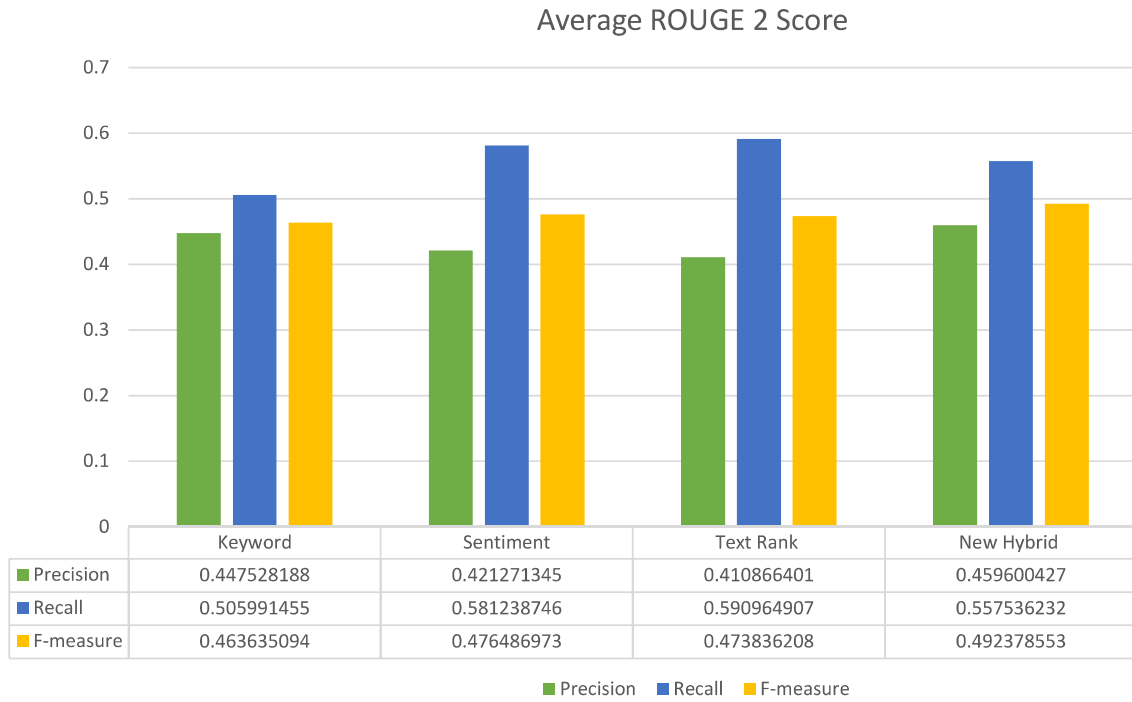


Figure 5.3: Average Rouge-2 Scores of the system for different methods (1st Setup)

From the chart we can say in terms of Precision, Hybrid ranking gets slightly better scores than Keyword ranking and Text ranking. In terms of Recall, text rank is the highest, but in terms of F-Measure Hybrid ranking again beats them all. Keyword based ranking seems not too good.

Bar chart along with the table for average Rouge-2 scores based on Keyword, Sentiment, Text rank and Hybrid rank in 2nd Setup is shown in figure 5.4.

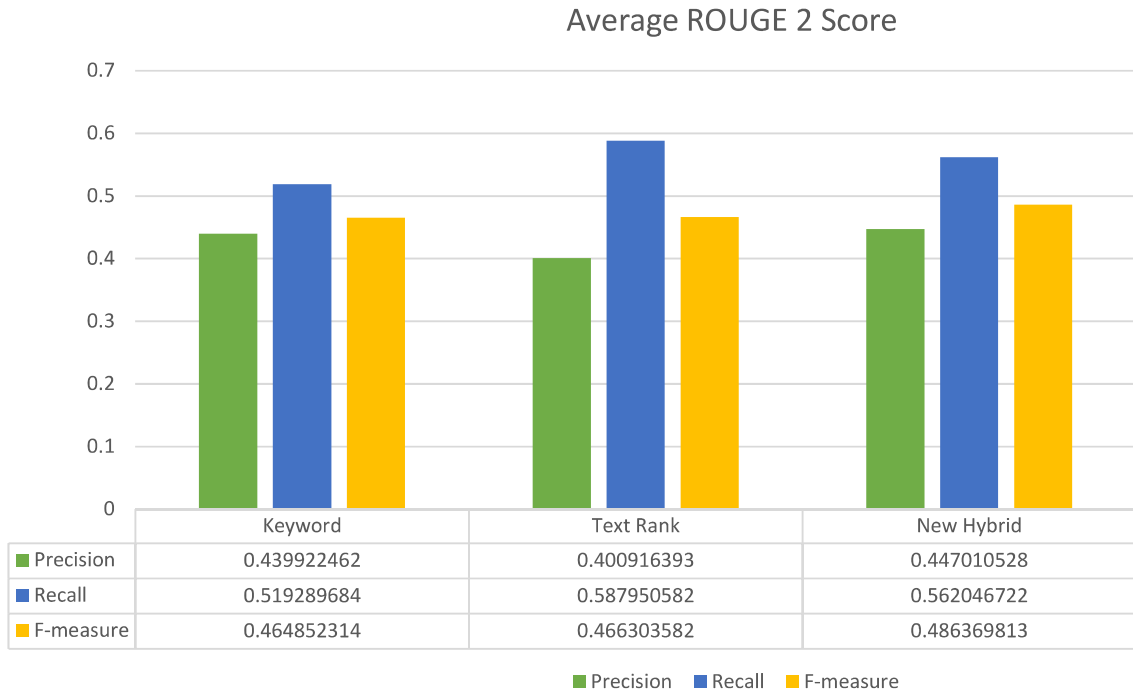


Figure 5.4: Average Rouge-2 Scores of the system for different methods (2nd Setup)

From the chart we can say in terms of Precision, Hybrid ranking gets slightly better scores than Keyword ranking. In terms of Recall, text ranking is the highest, also in terms of F-Measure Hybrid scoring again beats them all.

5.2.3 Experimental Setup on BNLPC Data Set

We have already stated in section 3.5.1.2 that, for our 2nd module we have used data sets from “Bangla Natural Language Processing Community” [14], where two datasets for evaluating Bangla text summarization system have been used and each set contains 100 Bangla news documents including three human generated model summaries for each document. Two groups of scholars (each group consists of three human judges) of Bangla language generated these summaries.

5.2.4 Experimental Verification with Evaluation on BNLPC Data set

We generated summaries for both BNLPC data sets using our best hybrid method where text ranking weighs 50 percent of the total document, keyword ranking weighs 30 percent and sentiment scoring weighs 20 percent and evaluated them (with provided gold summaries). We have already trained the model using our own data set. We have not used any documents from this data set to train our model. Firstly, the documents from both data sets were classified into four categories: Accident, Economics, Entertainment and Politics. Then for each category, the generated summaries were evaluated using Rouge-1 and Rouge-2 measures.

Classification of documents (for BNLPC data set 1) along with ROUGE scores that have been calculated comparing with 1st human generated model summary for different categories are shown in table 5.10.

Table 5.10: Classification and ROUGE Scores of summaries for different categories (for BNLPC data set 1 and 1st model summary).

Category	No. of Documents	Rouge-1 Precision	Rouge-1 Recall	Rouge-1 F-measure	Rouge-2 Precision	Rouge-2 Recall	Rouge-2 F-measure
Accident	29	0.6258	0.8083	0.7017	0.5561	0.7508	0.6348
Economics	35	0.6415	0.7783	0.6994	0.5693	0.7181	0.6319
Entertainment	12	0.5387	0.6897	0.6024	0.4600	0.6125	0.5238
Politics	24	0.5749	0.7703	0.6492	0.5088	0.7030	0.5814

Classification of documents (for BNLPC data set 2) along with ROUGE scores that have been calculated comparing with 1st human generated model summary for different categories are shown in table 5.11.

Table 5.11: Classification and ROUGE Scores of summaries for different categories (for BNLPC data set 2 and 1st model summary).

Category	No. of Documents	Rouge-1 Precision	Rouge-1 Recall	Rouge-1 F-measure	Rouge-2 Precision	Rouge-2 Recall	Rouge-2 F-measure
Accident	18	0.5691	0.7759	0.6511	0.4928	0.6986	0.5727
Economics	19	0.5819	0.7396	0.6458	0.5140	0.6747	0.5777
Entertainment	6	0.5338	0.7113	0.6084	0.4753	0.6459	0.5460
Politics	57	0.5702	0.7745	0.6517	0.5015	0.7121	0.5842

From table 5.10 and 5.11 we see that the Rouge-1 and Rouge-2 f-measure scores for each category are quite satisfactory for both data sets and 1st model summary.

We have also calculated the average Rouge-1 and Rouge-2 scores of both setups individually

and also the average ROUGE scores of the two setups. For individual ROUGE score calculation, the summations of ROUGE scores multiplied with the number of documents for each category are divided by total number of documents in the data set. For example, the average Rouge-1 precision of data set 1 is calculated in the following way.

$((29 * 0.6258) + (35 * 0.6415) + (12 * 0.5387) + (24 * 0.5749)) / 100$. The average ROUGE scores for BNLPC data sets using 1st model summary are shown in table 5.12.

Table 5.12: Average ROUGE Scores of summaries for BNLPC data sets.

Data set	Rouge-1 Precision	Rouge-1 Recall	Rouge-1 F-measure	Rouge-2 Precision	Rouge-2 Recall	Rouge-2 F-measure
Data set 1	0.6086	0.7744	0.6763	0.5378	0.7113	0.6076
Data set 2	0.5700	0.7644	0.6479	0.5008	0.6986	0.5786
Average	0.5893	0.7694	0.6621	0.5193	0.7050	0.5931

From table 5.12 we see that the average ROUGE scores of our proposed system are good enough i.e., the system generated summaries quite similar to the human generated ones.

Classification of documents (for BNLPC data set 1) along with ROUGE scores that have been calculated comparing with 2nd model summary for different categories are shown in table 5.13.

Table 5.13: Classification and ROUGE Scores of summaries for different categories (for BNLPC data set 1 and 2nd model summary).

Category	No. of Documents	Rouge-1 Precision	Rouge-1 Recall	Rouge-1 F-measure	Rouge-2 Precision	Rouge-2 Recall	Rouge-2 F-measure
Accident	29	0.5737	0.7743	0.6543	0.5007	0.7107	0.5828
Economics	35	0.5599	0.7138	0.6231	0.4762	0.6333	0.5400
Entertainment	12	0.4843	0.6544	0.5531	0.4059	0.5758	0.4726
Politics	24	0.5311	0.7341	0.6110	0.4685	0.6567	0.5420

Classification of documents (for BNLPC data set 2) along with ROUGE scores that have been calculated comparing with 2nd model summary for different categories are shown in table 5.14.

Table 5.14: Classification and ROUGE Scores of summaries for different categories (for BNLPC data set 2 and 2nd model summary).

Category	No. of Documents	Rouge-1 Precision	Rouge-1 Recall	Rouge-1 F-measure	Rouge-2 Precision	Rouge-2 Recall	Rouge-2 F-measure
Accident	18	0.5448	0.7600	0.6314	0.4665	0.6770	0.5491
Economics	19	0.5293	0.7396	0.6125	0.4588	0.6667	0.5393
Entertainment	6	0.5657	0.7747	0.6526	0.4999	0.7067	0.5836
Politics	57	0.5333	0.7438	0.6143	0.4629	0.6752	0.5430

From table 5.13 and 5.14 we see that the Rouge-1 and Rouge-2 f-measure scores for each category are quite satisfactory for both data sets and 2nd model summary.

We have also calculated the average Rouge-1 and Rouge-2 scores of both setups individually and also the average ROUGE scores of the two setups. The average ROUGE scores for BNLPC data sets using 2nd model summary are shown in table 5.15.

Table 5.15: Average ROUGE Scores of summaries for BNLPC data sets.

Data set	Rouge-1 Precision	Rouge-1 Recall	Rouge-1 F-measure	Rouge-2 Precision	Rouge-2 Recall	Rouge-2 F-measure
Data set 1	0.5479	0.7291	0.6208	0.4730	0.6544	0.5448
Data set 2	0.5365	0.7477	0.6193	0.4649	0.6758	0.5458
Average	0.5422	0.7384	0.6201	0.4690	0.6651	0.5453

From table 5.15 we see that the average ROUGE scores of our proposed system are good enough i.e., the system generated summaries are quite similar to the human generated ones.

Classification of documents (for BNLPC data set 1) along with ROUGE scores that have been calculated comparing with 3rd model summary for different categories are shown in table 5.16.

Table 5.16: Classification and ROUGE Scores of summaries for different categories (for BNLPC data set 1 and 3rd model summary).

Category	No. of Documents	Rouge-1 Precision	Rouge-1 Recall	Rouge-1 F-measure	Rouge-2 Precision	Rouge-2 Recall	Rouge-2 F-measure
Accident	29	0.5918	0.7885	0.6712	0.5207	0.7255	0.6016
Economics	35	0.5685	0.7183	0.6300	0.4945	0.6534	0.5589
Entertainment	12	0.5306	0.7089	0.6054	0.4566	0.6265	0.5263
Politics	24	0.5558	0.7472	0.6291	0.4875	0.6745	0.5579

Classification of documents (for BNLPC data set 2) along with ROUGE scores that have been

calculated comparing with 3rd model summary for different categories are shown in table 5.17.

Table 5.17: Classification and ROUGE Scores of summaries for different categories (for BNLPC data set 2 and 3rd model summary).

Category	No. of Documents	Rouge-1 Precision	Rouge-1 Recall	Rouge-1 F-measure	Rouge-2 Precision	Rouge-2 Recall	Rouge-2 F-measure
Accident	18	0.5355	0.7577	0.6242	0.4663	0.6819	0.5498
Economics	19	0.6066	0.7770	0.6737	0.5363	0.7128	0.6048
Entertainment	6	0.4681	0.6879	0.5552	0.4045	0.62207	0.4886
Politics	57	0.5126	0.7367	0.5967	0.4379	0.6570	0.5175

From table 5.16 and 5.17 we see that the Rouge-1 and Rouge-2 f-measure scores for each category are quite satisfactory for both data sets and 3rd model summary.

We have also calculated the average Rouge-1 and Rouge-2 scores of both setups individually and also the average ROUGE scores of the two setups. The average ROUGE scores for BNLPC data sets using 3rd model summary are shown in table 5.18.

Table 5.18: Average ROUGE Scores of summaries for BNLPC data sets.

Data set	Rouge-1 Precision	Rouge-1 Recall	Rouge-1 F-measure	Rouge-2 Precision	Rouge-2 Recall	Rouge-2 F-measure
Data set 1	0.5677	0.7445	0.6388	0.4959	0.6761	0.5671
Data set 2	0.5319	0.7452	0.6138	0.4597	0.6699	0.5381
Average	0.5498	0.7449	0.6263	0.4778	0.6730	0.5526

From table 5.18 we see that the average ROUGE scores of our proposed system are good enough i.e., the system generated summaries are quite similar to the human generated ones.

Finally we have evaluated our system combining all the scores found from different models and the results can be found in table 5.19.

Table 5.19: Average ROUGE Scores of the proposed system summaries for BNLPC data sets.

Model Summary	Rouge-1 Precision	Rouge-1 Recall	Rouge-1 F-measure	Rouge-2 Precision	Rouge-2 Recall	Rouge-2 F-measure
Model Summary 1	0.5893	0.7694	0.6621	0.5139	0.7050	0.5931
Model Summary 2	0.5422	0.7384	0.6201	0.4690	0.6651	0.5453
Model Summary 3	0.5498	0.7449	0.6263	0.4778	0.6730	0.5526
Average	0.5658	0.7539	0.6362	0.4869	0.6810	0.5636

5.2.5 Comparison with Existing Models

We have considered two systems for comparison. One where we have used the BNLPC data sets. These data sets have been used previously by Kamal Sarkar [3] and also Md. Majharul Haque, Suraiya Pervin, Zerina Begum [6] in their summarizer.

Table 5.20 shows the average ROUGE scores of our proposed system and other published systems [3], [6].

Table 5.20: Average ROUGE Scores of our proposed system and other existing systems for BNLPC data sets.

Methods	Precision	Recall	F-measure
Proposed System	0.5658	0.7539	0.6362
Majharul Haque et al., 2015 [6]	0.6080	0.6640	0.6320
Kamal Sarkar et al., 2012 [3]	0.5380	0.5560	0.5460

From table 5.20 we see that the ROUGE f-measure score of our proposed hybrid system outperforms the previously existed model [3] substantially and also it is slightly better than the existing system [6].

The generated summaries along with their ROUGE scores and the relevant resources i.e., original news documents, summaries collected from BNLPC etc. can be found on our GitHub source [13]. Another one is a web based solution where we have used our data set and the generated summaries were evaluated based on our gold summaries.

Table 5.21 shows the average ROUGE scores of our proposed system and other existing web based system [17].

Table 5.21: Average ROUGE Scores of summaries our proposed system and other existing web based system for our data set.

Methods	Rouge-1 Precision	Rouge-1 Recall	Rouge-1 F-measure	Rouge-2 Precision	Rouge-2 Recall	Rouge-2 F-measure
Proposed System	0.5472	0.6567	0.5842	0.4596	0.5575	0.4923
Existing Web Based System [17]	0.6220	0.4019	0.4652	0.5249	0.3253	0.3769

Again from table 5.21 we see that the ROUGE f-measure score of our proposed system is way higher than the f-measure score of a similar web based system [17]. Thus we can say that our hybrid system combining three methods certainly generates better extractive summaries that are close to human generated ones.

5.3 Time Analysis for Summary Generation

We selected ten random test documents from each category for both setups and observed the time needed to generate summary for our hybrid summarizer and other individual methods. We have used the configuration of Intel Core i5 Processor, 8 Gb Ram and 256 GB SSD to calculate the time required to generate summaries.

For Accident category: Time needed to generate summary of 10 documents for both setups are shown in table 5.22.

Table 5.22: Time needed to generate summary of 10 documents for Category Accident

Setup	Methods	Time needed to summarize 10 documents (in seconds)	Per Unit Time (in seconds)
1st Setup	Keyword Ranking	40.3269	4.0327
	Sentiment Scoring	0.1462	0.0146
	Text Ranking	1.8191	0.1819
	Hybrid System (.3KR+.2SS+.5TR)	85.6817	8.568
2nd Setup	Keyword Ranking	30.0367	3.0037
	Sentiment Scoring	0.1462	0.0146
	Text Ranking	2.6004	0.2600
	Hybrid System (.3KR+.2SS+.5TR)	74.2117	7.421

From tables 5.22 we see that, summary based on sentiment scoring takes lowest time in both setups as we trained this system with 100 percent data. Text ranking performs well in both setups than keyword ranking. But hybrid system takes a good amount of time because it combines three methods. So naturally the process is a bit slow.

For Economics category: Time needed to generate summary of 10 documents for both setups are shown in table 5.23.

Table 5.23: Time needed to generate summary of 10 documents for Category Economics

Setup	Methods	Time needed to summarize 10 documents (in seconds)	Per Unit Time (in seconds)
1st Setup	Keyword Ranking	195.7745	19.5775
	Sentiment Scoring	8.3297	0.8329
	Text Ranking	42.6732	0.4267
	Hybrid System (.3KR+.2SS+.5TR)	548.6031	54.8603
2nd Setup	Keyword Ranking	409.2141	40.9214
	Sentiment Scoring	8.3297	0.8329
	Text Ranking	53.6738	5.3674
	Hybrid System (.3KR+.2SS+.5TR)	985.4700	98.5470

From tables 5.23 we see that, summary based on sentiment scoring takes lowest time in both setups as we trained this system with 100 percent data. Text ranking performs well in both setups than keyword ranking and hybrid system takes a good amount of time because it combines three methods. So naturally the process is a bit slow. But we see that for economics category time needed to generate summary is greater than time needed to generate summary for accident category. One reason behind it may be the size of document i.e., news we selected for economics category is much bigger in size than the news from accident category.

For Entertainment category: Time needed to generate summary of 10 documents for both setups are shown in table 5.24.

Table 5.24: Time needed to generate summary of 10 documents for Category Entertainment

Setup	Methods	Time needed to summarize 10 documents (in seconds)	Per Unit Time (in seconds)
1st Setup	Keyword Ranking	66.2184	0.6622
	Sentiment Scoring	0.5999	0.0599
	Text Ranking	6.3414	0.6341
	Hybrid System (.3KR+.2SS+.5TR)	154.5379	15.4538
2nd Setup	Keyword Ranking	54.8148	5.4815
	Sentiment Scoring	0.5999	0.0599
	Text Ranking	6.0225	0.6022
	Hybrid System (.3KR+.2SS+.5TR)	113.0268	11.3027

From tables 5.24 we see that, summary based on sentiment scoring takes lowest time in both setups. Text ranking performs well in both setups than keyword ranking and hybrid system takes

a good amount of time. In this category time needed to generate summary is greater than time needed to generate summary for accident category. But compared to economics category the time is much less.

For Politics category: Time needed to generate summary of 10 documents for both setups are shown in table 5.25.

Table 5.25: Time needed to generate summary of 10 documents for Category Politics

Setup	Methods	Time needed to summarize 10 documents (in seconds)	Per Unit Time (in seconds)
1st Setup	Keyword Ranking	278.1268	27.8127
	Sentiment Scoring	12.0796	1.2079
	Text Ranking	56.3719	5.6372
	Hybrid System (.3KR+.2SS+.5TR)	705.4806	70.5481
2nd Setup	Keyword Ranking	189.5014	18.9501
	Sentiment Scoring	12.0796	1.2079
	Text Ranking	56.2164	5.6216
	Hybrid System (.3KR+.2SS+.5TR)	580.9077	58.0908

From table 5.25 we see that, summary based on sentiment scoring takes lowest time in both setups. Text ranking performs well in both setups than keyword ranking and hybrid system takes a good amount of time to generate summary. In this category time needed to generate summary is greater than time needed to generate summary for accident category and entertainment category. But the time is much less compared to economics category.

5.4 Limitations

There are some limitations in our system. We have found out some problems in the second method i.e., sentiment scoring.

- We have stated before that, in sentiment scoring method, we have given the neutral words in sentences higher scores in order to select the sentence for summary generation i.e., more neutral words mean more score. We have given higher priority to the sentences that have neutral words. If a sentence has only neutral words that should be selected for the summary. But a sentence that has only neutral words gets less score than a sentence that has positive or negative words along with more neutral words than the previous sentence. So, the sentence with only neutral words doesn't get selected which should not be the case.

- We have used polyglot in our system to determine the polarity of the words. But the sentiment of polyglot cannot always determine the polarity correctly. So, it is a limitation of our system.
- Due to poorly organized news in the entertainment category, our summaries aren't up to the mark, and also the classification was not good either.
- Even though our system outperforms some of the similar kind of systems, it is still not ideal and cannot achieve perfect score.

5.5 Summary of the Chapter

In this chapter we gained knowledge and a clear overview of the system which we are working on for generating automatic summary. We have given a brief idea about our experimental results and our limitations. We have attached our sample generated outputs as well from which a clear idea about how the outputs will be generated can be seen. From this chapter, the recipients will get a chance to evaluate our work and to suggest if any changes need to be done or not. In the next chapter we will discuss about the difficulties we faced along with the future work.

Chapter 6

Conclusion and Future Works

In the previous chapter we discussed about the performance measures of our system and compared the ROUGE scores of the system with the ROUGE scores of different systems. In this chapter we will discuss about the difficulties we have faced and also how we plan to improve the limitations our summarizer has.

6.1 Conclusion

We have discussed about the overview of our proposed system. We also have discussed about the evaluation measures, experimental verification and how accurate result our system provides. We have encountered some difficulties in our works throughout the process.

6.1.1 Brief Survey

We have completed our work on generating automatic extractive summaries of Bengali documents. We have proposed a system that combines three methods of generating automatic summary. We have prepared our data sets and have been able to generate summaries based on key phrase ranking, sentiment scoring, text ranking and combining all of the methods.

6.1.2 Difficulties Encountered

We have encountered some difficulties in our work. Some of them are mentioned below:

- We have mentioned that we have created our own summaries and considered them as gold summaries. But we cannot assure that the summaries are 100 percent correct.

- Bengali is a sophisticated language. But very few works have been done for text summarization for Bengali documents. As a result, there are very few resources as well as predefined libraries are available for Bengali documents.
- We have used polyglot library for determining the polarity of words. Polyglot is a pre-defined, open source library that covers 136 languages including Bengali. So naturally it is less trained library than the library which works for only one language. Polyglot sometimes lacks crucial words that must be included in the summary.
- Another vital point is that, polyglot most of the times splits sentences with full stops (.). Polyglot will split the sentence into two sentences: one before the full stop has occurred and other one after the full stop to ”|”.

It has a huge impact on the generated summary. As a result, we get less ROUGE score. This is a great issue.

6.2 Future Work

We have proposed a hybrid model and our work is complete. Still in future we will try to perform the following tasks:

- FastText in comparison with polyglot - We have used **polyglot** for now but in future we may use another library **FastText** for sentiment analysis depending on the result. **FastText** is a Library that uses Word to Vector method for 157 languages. Pre-trained word vectors are distributed over 157 languages using FastText. We will compare the result and see the outcome.
- We will train our summarizer for many other different categories like literature, international, editorial etc.
- We will also provide more importance on sentences that are based on headlines and numerical values.
- We need to introduce “Named Entity Recognition”, in our system, which will be helpful in finding important sentences.
- We will search for better approaches which can upgrade our rouge scores.

6.3 Summary of the Chapter

In this chapter, we have discussed about the whole idea of our work shortly. We have given a clear overview of what difficulties we have faced and we will be doing in future. We hope our summarizer generate better summaries for Bengali document.

References

- [1] K. Sarkar, “Bengali text summarization by sentence extraction,” *arXiv preprint arXiv:1201.2240*, 2012.
- [2] M. N. Uddin and S. A. Khan, “A study on text summarization techniques and implement few of them for bangla language,” in *2007 10th international conference on computer and information technology*, pp. 1–4, IEEE, 2007.
- [3] K. Sarkar, “An approach to summarizing bengali news documents,” in *proceedings of the International Conference on Advances in Computing, Communications and Informatics*, pp. 857–862, ACM, 2012.
- [4] A. Das and S. Bandyopadhyay, “Topic-based bengali opinion summarization,” in *Proceedings of the 23rd International Conference on Computational Linguistics: Posters*, pp. 232–240, Association for Computational Linguistics, 2010.
- [5] M. I. A. Efat, M. Ibrahim, and H. Kayesh, “Automated bangla text summarization by sentence scoring and ranking,” in *2013 International Conference on Informatics, Electronics and Vision (ICIEV)*, pp. 1–5, IEEE, 2013.
- [6] M. M. Haque, S. Pervin, and Z. Begum, “Automatic bengali news documents summarization by introducing sentence frequency and clustering,” in *2015 18th International Conference on Computer and Information Technology (ICCIT)*, pp. 156–160, IEEE, 2015.
- [7] M. M. Haque, S. Pervin, and Z. Begum, “Enhancement of keyphrase-based approach of automatic bangla text summarization,” in *2016 IEEE Region 10 Conference (TENCON)*, pp. 42–46, IEEE, 2016.
- [8] S. Akter, A. S. Asa, M. P. Uddin, M. D. Hossain, S. K. Roy, and M. I. Afjal, “An extractive text summarization technique for bengali document (s) using k-means clustering algorithm,” in *2017 IEEE International Conference on Imaging, Vision & Pattern Recognition (icIVPR)*, pp. 1–6, IEEE, 2017.

- [9] A. Paul, M. T. Imtiaz, A. H. Latif, M. Ahmed, F. A. Adnan, R. Khan, I. Kadery, and R. M. Rahman, “Bangla news summarization,” in *International Conference on Computational Collective Intelligence*, pp. 479–488, Springer, 2017.
- [10] S. Saziyabegum and P. S. Sajja, “Literature review on extractive text summarization approaches,” *International Journal of Computer Applications*, vol. 156, no. 12, 2016.
- [11] J. Lin, “Summarization,” *Encyclopedia of database systems*, pp. 2884–2889, 2009.
- [12] J. C. Cheung, “Comparing abstractive and extractive summarization of evaluative text: controversiality and content selection,” *B. Sc.(Hons.) Thesis in the Department of Computer Science of the Faculty of Science, University of British Columbia*, 2008.
- [13] M. Islam, “Hybrid text summarizer for bangla document.” <https://github.com/mahimulislam/ResourcesHybridTextSummarization>, January 2020.
- [14] BNLPC, “Dataset for evaluating bangla text summarization system.” <http://www.bnlpc.org/research.php>, January 2015.
- [15] Imran, “Bengali-sentiment-analysis.” <https://github.com/Imran-cse/Bengali-Sentiment-Analysis>, June 2019.
- [16] J. Steinberger and K. Ježek, “Evaluation measures for text summarization,” *Computing and Informatics*, vol. 28, no. 2, pp. 251–275, 2012.
- [17] P. Chondro, “Web based bengali document summarizer.” <https://bengali-document-summarizer.herokuapp.com>, January 2018.

Appendix A

Examples of Summary with Evaluation Scores

Accident Category:

Actual News:

"নোয়াখালীতে সড়ক দুর্ঘটনায় নিহত ২"

নোয়াখালীর সুবর্ণচর উপজেলার পরিষ্কার বাজার রাস্তার মাথা নামক স্থানে সোনাপুর চরজব্বার সড়কে যাত্রীবাহী সিএনজিচালিত অটোরিকশা ও ইজিবাইকের মধ্যে সংঘর্ষে দুজন নিহত হয়েছে। আহত তিনজন। গতকাল মঙ্গলবার রাত ১১টার দিকে এ দুর্ঘটনা ঘটে।

নিহত ব্যক্তির হলে সুবর্ণচরের চরভাটার আবুল কালাম (৪৫) ও লক্ষ্মীপুরের রামগতির জসিম উদ্দিন (৩৫)। সুবর্ণচরের চরজব্বার থানার পরিদর্শক (তদন্ত) ইকবাল হোসেন জানান, রাত ১১টার দিকে যাত্রীবাহী সিএনজিচালিত অটোরিকশা সুবর্ণচরের দিকে যাওয়ার পথে বিপরীত দিক থেকে আসা ব্যাটারিচালিত ইজিবাইকের সঙ্গে সংঘর্ষ হয়। এতে ঘটনাস্থলে সিএনজিচালিত অটোরিকশার পাঁচ যাত্রী আহত হন। তাঁদের মধ্যে আবুল কালাম ও জসিম উদ্দিনকে নোয়াখালী জেনারেল হাসপাতালে ভর্তির পর রাত সাড়ে ১২টার দিকে তাঁরা মারা যান। আহত ব্যক্তির স্থানীয় ক্লিনিকে ভর্তি রয়েছেন। এ ঘটনায় কোনো মামলা হয়নি।

Provided Gold summary generated by random human:

নোয়াখালীর সুবর্ণচর উপজেলার পরিষ্কার বাজার রাস্তার মাথা নামক স্থানে সোনাপুর চরজব্বার সড়কে যাত্রীবাহী সিএনজিচালিত অটোরিকশা ও ইজিবাইকের মধ্যে সংঘর্ষে দুজন নিহত হয়েছে।

গতকাল মঙ্গলবার রাত ১১টার দিকে এ দুর্ঘটনা ঘটে।

সুবর্ণচরের চরজব্বার থানার পরিদর্শক (তদন্ত) ইকবাল হোসেন জানান, রাত ১১টার দিকে যাত্রীবাহী সিএনজিচালিত অটোরিকশা সুবর্ণচরের দিকে যাওয়ার পথে বিপরীত দিক থেকে আসা ব্যাটারিচালিত ইজিবাইকের সঙ্গে সংঘর্ষ হয়। এতে ঘটনাস্থলে সিএনজিচালিত অটোরিকশার পাঁচ যাত্রী আহত হন। তাঁদের মধ্যে আবুল কালাম ও জসিম উদ্দিনকে নোয়াখালী জেনারেল হাসপাতালে ভর্তির পর রাত সাড়ে ১২টার দিকে তাঁরা মারা যান। আহত ব্যক্তির

স্থানীয় ক্লিনিকে ভর্তি রয়েছেন। এ ঘটনায় কোনো মামলা হয়নি।

Summary based on Keyword Ranking:

"নোয়াখালীতে সড়ক দুর্ঘটনায় নিহত ২"

গতকাল মঙ্গলবার রাত ১১টার দিকে এ দুর্ঘটনা ঘটে। সুবর্ণচরের চরজব্বার থানার পরিদর্শক (তদন্ত) ইকবাল হোসেন জানান, রাত ১১টার দিকে যাত্রীবাহী সিএনজিচালিত অটোরিকশা সুবর্ণচরের দিকে যাওয়ার পথে বিপরীত দিক থেকে আসা ব্যাটারিচালিত ইজিবাইকের সঙ্গে সংঘর্ষ হয়। নোয়াখালীর সুবর্ণচর উপজেলার পরিষ্কার বাজার রাস্তার মাথা নামক স্থানে সোনাপুর চরজব্বার সড়কে যাত্রীবাহী সিএনজিচালিত অটোরিকশা ও ইজিবাইকের মধ্যে সংঘর্ষে দুজন নিহত হয়েছে।

Table A.1: ROUGE 1 and ROUGE 2 Score of the key word generated summary for the news.

ROUGE 1 Precision	ROUGE 1 Recall	ROUGE 1 F-measure	ROUGE 2 Precision	ROUGE 2 Recall	ROUGE 2 F-measure
0.9020	0.5897	0.7132	0.8771	0.5556	0.6803

Summary based on Sentiment Scoring:

সুবর্ণচরের চরজব্বার থানার পরিদর্শক (তদন্ত) ইকবাল হোসেন জানান, রাত ১১টার দিকে যাত্রীবাহী সিএনজিচালিত অটোরিকশা সুবর্ণচরের দিকে যাওয়ার পথে বিপরীত দিক থেকে আসা ব্যাটারিচালিত ইজিবাইকের সঙ্গে সংঘর্ষ হয়। নোয়াখালীর সুবর্ণচর উপজেলার পরিষ্কার বাজার রাস্তার মাথা নামক স্থানে সোনাপুর চরজব্বার সড়কে যাত্রীবাহী সিএনজিচালিত অটোরিকশা ও ইজিবাইকের মধ্যে সংঘর্ষে দুজন নিহত হয়েছে। তাঁদের মধ্যে আবুল কালাম ও জসিম উদ্দিনকে নোয়াখালী জেনারেল হাসপাতালে ভর্তির পর রাত সাড়ে ১২টার দিকে তাঁরা মারা যান। নিহত ব্যক্তির হলে সুবর্ণচরের চরভাটার আবুল কালাম (৪৫) ও লক্ষ্মীপুরের রামগতির জসিম উদ্দিন (৩৫)।

Table A.2: ROUGE 1 and ROUGE 2 Score of the sentiment score generated summary for the news.

ROUGE 1 Precision	ROUGE 1 Recall	ROUGE 1 F-measure	ROUGE 2 Precision	ROUGE 2 Recall	ROUGE 2 F-measure
0.8769	0.7308	0.7972	0.7949	0.6889	0.7381

Summary based on Text Ranking:

নোয়াখালীর সুবর্ণচর উপজেলার পরিষ্কার বাজার রাস্তার মাথা নামক স্থানে সোনাপুর চরজব্বার সড়কে যাত্রীবাহী সিএনজিচালিত অটোরিকশা ও ইজিবাইকের মধ্যে সংঘর্ষে দুজন নিহত হয়েছে "নোয়াখালীতে সড়ক দুর্ঘটনায় নিহত ২" গতকাল মঙ্গলবার রাত ১১টার দিকে এ দুর্ঘটনা ঘটে আহত তিনজন নিহত ব্যক্তির হালাত সুবর্ণচরের চরভাটার আবুল কালাম (৪৫) ও লক্ষ্মীপুরের রামগতির জসিম উদ্দিন (৩৫) সুবর্ণচরের চরজব্বার থানার পরিদর্শক (তদন্ত) ইকবাল হোসেন জানান, রাত ১১টার দিকে যাত্রীবাহী সিএনজিচালিত অটোরিকশা সুবর্ণচরের দিকে যাওয়ার পথে বিপরীত দিক থেকে আসা ব্যাটারিচালিত ইজিবাইকের সঙ্গে সংঘর্ষ হয়

Table A.3: ROUGE 1 and ROUGE 2 Score of the Text Rank generated summary for the news.

ROUGE 1 Precision	ROUGE 1 Recall	ROUGE 1 F-measure	ROUGE 2 Precision	ROUGE 2 Recall	ROUGE 2 F-measure
0.7500	0.6154	0.6761	0.6439	0.5222	0.5767

Summary based on Hybrid Ranking (KR+SS+TR):

সুবর্ণচরের চরজব্বার থানার পরিদর্শক (তদন্ত) ইকবাল হোসেন জানান, রাত ১১টার দিকে যাত্রীবাহী সিএনজিচালিত অটোরিকশা সুবর্ণচরের দিকে যাওয়ার পথে বিপরীত দিক থেকে আসা ব্যাটারিচালিত ইজিবাইকের সঙ্গে সংঘর্ষ হয়। নোয়াখালীর সুবর্ণচর উপজেলার পরিষ্কার বাজার রাস্তার মাথা নামক স্থানে সোনাপুর চরজব্বার সড়কে যাত্রীবাহী সিএনজিচালিত অটোরিকশা ও ইজিবাইকের মধ্যে সংঘর্ষে দুজন নিহত হয়েছে। তাঁদের মধ্যে আবুল কালাম ও জসিম উদ্দিনকে নোয়াখালী জেনারেল হাসপাতালে ভর্তির পর রাত সাড়ে ১২টার দিকে তাঁরা মারা যান। "নোয়াখালীতে সড়ক দুর্ঘটনায় নিহত ২"

Table A.4: ROUGE 1 and ROUGE 2 Score of the Hybrid summary (combining KR, SS, TR) for the news.

ROUGE 1 Precision	ROUGE 1 Recall	ROUGE 1 F-measure	ROUGE 2 Precision	ROUGE 2 Recall	ROUGE 2 F-measure
0.9180	0.7180	0.8058	0.8857	0.6889	0.7750

Summary generated by other web based system :

"নোয়াখালীতে সড়ক দুর্ঘটনায় নিহত ২"

নোয়াখালীর সুবর্ণচর উপজেলার পরিষ্কার বাজার রাস্তার মাথা নামক স্থানে সোনাপুর চরজব্বার সড়কে যাত্রীবাহী সিএনজিচালিত অটোরিকশা ও ইজিবাইকের মধ্যে সংঘর্ষে দুজন নিহত হয়েছে। সুবর্ণচরের চরজব্বার থানার পরিদর্শক (তদন্ত) ইকবাল হোসেন জানান, রাত ১১টার দিকে যাত্রীবাহী সিএনজিচালিত অটোরিকশা সুবর্ণচরের দিকে যাওয়ার পথে বিপরীত দিক থেকে আসা ব্যাটারিচালিত ইজিবাইকের সঙ্গে সংঘর্ষ হয়। নিহত ব্যক্তির হলে সুবর্ণচরের চরভাটার আবুল কালাম (৪৫) ও লক্ষ্মীপুরের রামগতির জসিম উদ্দিন (৩৫)

Table A.5: ROUGE 1 and ROUGE 2 Score of other web based system generated summary for the news.

ROUGE 1 Precision	ROUGE 1 Recall	ROUGE 1 F-measure	ROUGE 2 Precision	ROUGE 2 Recall	ROUGE 2 F-measure
0.7895	0.5769	0.6667	0.6923	0.5000	0.5806

Economics Category:

Actual News:

‘ভালো’ ব্যাংকেও খেলাপি বাড়ছে

সরকারি-বেসরকারি ২১ ব্যাংকের খেলাপি ঋণ ৭ শতাংশের বেশি। এর মধ্যে খেলাপিতে শীর্ষে থাকা পুরোনো ব্যাংকগুলো সঙ্গে নতুন করে যুক্ত হয়েছে বেসরকারি খাতের ভালো বেশ কয়েকটি ব্যাংকও। সর্বশেষ গত সেপ্টেম্বর শেষে খেলাপি ঋণের হালনাগাদ তথ্য বিশ্লেষণ করে এমন তথ্য পাওয়া গেছে।

বেসরকারি খাতের ভালো ব্যাংকের মধ্যে ৭ শতাংশের বেশি খেলাপি ঋণ রয়েছে আইএফআইসি, প্রাইম, স্ট্যান্ডার্ড, ইউসিবিএল, সোস্যাল ইসলামী ব্যাংকসহ আরও কয়েকটির। গত সেপ্টেম্বর শেষে দেশের ব্যাংক খাতে খেলাপি ঋণের হার দাঁড়িয়েছে ১১ দশমিক ৪৫ শতাংশে। টাকার অঙ্কে যা ৯৯ হাজার ৩৭০ কোটি টাকা। বাংলাদেশের ব্যাংক খাতের দুরবস্থার চিত্রের কিছুটা খেলাপি ঋণের তথ্যের মাধ্যমে ক্রমেই প্রকাশ পাচ্ছে। যদিও খেলাপি ঋণের তথ্য নিয়ে ব্যাংকগুলোতে নানা রকমের ছলচাতুরী করা হয়। কেউ পুনঃ তফসিল করে খেলাপি ঋণ কমায়, আবার কেউ কেউ ঋণ পরিশোধ না হলেও তা খেলাপি হিসাবে দেখায় না।

কেন্দ্রীয় ব্যাংকের সাবেক ডেপুটি গভর্নর খোন্দকার ইব্রাহিম খালেদ এ নিয়ে বলেন, সরকারি ব্যাংকগুলোর যে সমস্যা রয়েছে, তা বেসরকারি ব্যাংকেও ছড়িয়ে পড়েছে। বেসরকারি ব্যাংকের পরিচালকেরা এখন জমিদারের মতো আচরণ করছেন। এ জন্য কেন্দ্রীয় ব্যাংককে শক্তিশালী ভূমিকা নিতে হবে।

কেন্দ্রীয় ব্যাংকের হিসাবে, সেপ্টেম্বর শেষে অগ্রণী ব্যাংকের খেলাপি ঋণ ৫ হাজার ৯৬২ কোটি টাকা বা ১৯ শতাংশ, রূপালী ব্যাংকের ৪ হাজার ৯৭০ কোটি টাকা বা ২২ শতাংশ। জনতা ব্যাংকের খেলাপি ঋণ ১৪ হাজার ৩০৭ কোটি টাকা বা ৩১ শতাংশ ও সোনালী ব্যাংকের ১২ হাজার ৭৭১ কোটি টাকা বা ৩৪ শতাংশ। এ ছাড়া বেসিক ব্যাংকের খেলাপি ঋণের হার ৬২ শতাংশ বা ৯ হাজার ১৪৩ কোটি টাকা, বিডিবিএলের ৮৪৮ কোটি টাকা বা ৫৫ শতাংশ। কৃষি ব্যাংকের খেলাপি ঋণ ২১ শতাংশ ও রাজশাহী কৃষি উন্নয়ন ব্যাংকের ২৫ শতাংশ।

বেসরকারি ব্যাংকগুলোর মধ্যে এবি ব্যাংকের খেলাপি ঋণের হার ৭ শতাংশ বা ১ হাজার ৬৩০ কোটি টাকা, আইএফআইসি ব্যাংকের ৮ শতাংশ বা ১ হাজার ৪৮০ কোটি টাকা। ন্যাশনাল ব্যাংকের খেলাপি ঋণ ৮ শতাংশ বা ২ হাজার ২৯১ কোটি টাকা, প্রাইম ব্যাংকের সাড়ে ৮ শতাংশ বা ১ হাজার ৩৯৭ কোটি টাকা। স্ট্যান্ডার্ড ব্যাংকের খেলাপি ঋণের হার ৯ শতাংশ বা ১ হাজার ১৬৮ কোটি টাকা, ইউনাইটেড কমার্শিয়াল ব্যাংকের ৯ শতাংশ বা ২ হাজার ১৭৮ কোটি টাকা, উত্তরা ব্যাংকের ৮ শতাংশ বা ৮৩২ কোটি টাকা। এদিকে মালিকানা পরিবর্তন হওয়া সোস্যাল ইসলামী ব্যাংক ক্রমেই খারাপ হতে শুরু করেছে। গত সেপ্টেম্বর শেষে ব্যাংকটির খেলাপি ঋণের হার দাঁড়িয়েছে ৭ শতাংশ বা ১ হাজার ৫৬২ কোটি টাকা।

সংকটে থাকা বাংলাদেশ কমার্স ব্যাংকের খেলাপি ঋণের হার ২৬ শতাংশ, টাকার অঙ্কে যা ৫৫৭ কোটি টাকা, আইসিবি ইসলামিক ব্যাংকের খেলাপি ঋণ ৭০৮ কোটি টাকা বা ৮৪ শতাংশ।

এদিকে ফারমার্স ব্যাংকের খেলাপি ঋণের হার ৫৮ শতাংশ, টাকার অঙ্কে যা ৩ হাজার ৭০ কোটি টাকা। মেঘনা ব্যাংকের খেলাপি ঋণ ২৩৫ কোটি টাকা বা ৮ শতাংশ, এনআরবি কমার্শিয়াল ব্যাংকের ৩৫০ কোটি টাকা বা ৮ শতাংশ। বিদেশি ব্যাংকগুলোর মধ্যে হাবিব ব্যাংকের খেলাপি ঋণের হার ১০ শতাংশ বা ৪২ কোটি টাকা, ন্যাশনাল ব্যাংক অব পাকিস্তানের ৯৪ শতাংশ বা ১ হাজার ৩৬২ কোটি টাকা ও স্টেট ব্যাংক অব ইন্ডিয়া ৭ শতাংশ বা ৪২ কোটি টাকা।

অ্যাসোসিয়েশন অব ব্যাংকার্স বাংলাদেশের (এবিবি) চেয়ারম্যান সৈয়দ মাহবুবুর রহমান বলেন, এখন নতুন ঋণ সেভাবে বাড়ছে না। কিন্তু খেলাপি বাড়ছে। রিট ভ্যাকেট হওয়ায় অনেক ঋণ খেলাপি হিসাবে যুক্ত হচ্ছে। এ কারণে সার্বিকভাবে খেলাপি ঋণের হার বাড়ছে। সার্বিকভাবে ব্যাংকের সুশাসন নিশ্চিত করতে না পারলে পরিস্থিতি আরও খারাপ হবে।

Provided Gold summary generated by random human:

সরকারি-বেসরকারি ২১ ব্যাংকের খেলাপি ঋণ ৭ শতাংশের বেশি এর মধ্যে খেলাপিতে শীর্ষে থাকা পুরোনো ব্যাংকগুলো সঙ্গে নতুন করে যুক্ত হয়েছে বেসরকারি খাতের ভালো বেশ কয়েকটি ব্যাংকও বেসরকারি খাতের ভালো ব্যাংকের মধ্যে ৭ শতাংশের বেশি খেলাপি ঋণ রয়েছে আইএফআইসি, প্রাইম, স্ট্যান্ডার্ড, ইউসিবিএল, সোস্যাল ইসলামী ব্যাংকসহ আরও কয়েকটির গত সেপ্টেম্বর শেষে দেশের ব্যাংক খাতে খেলাপি ঋণের হার দাঁড়িয়েছে ১১ দশমিক ৪৫ শতাংশে কেন্দ্রীয় ব্যাংকের হিসাবে, সেপ্টেম্বর শেষে অগ্রণী ব্যাংকের খেলাপি ঋণ ৫ হাজার ৯৬২ কোটি টাকা বা ১৯ শতাংশ, রূপালী ব্যাংকের ৪ হাজার ৯৭০ কোটি টাকা বা ২২ শতাংশ জনতা ব্যাংকের খেলাপি ঋণ ১৪ হাজার ৩০৭ কোটি টাকা বা ৩১ শতাংশ ও সোনালী ব্যাংকের ১২ হাজার ৭৭১ কোটি টাকা বা ৩৪ শতাংশ বেসরকারি ব্যাংকগুলোর মধ্যে এবি ব্যাংকের খেলাপি ঋণের হার ৭ শতাংশ বা ১ হাজার ৬৩০ কোটি টাকা, আইএফআইসি ব্যাংকের ৮ শতাংশ বা ১ হাজার ৪৮০ কোটি টাকা ন্যাশনাল ব্যাংকের খেলাপি ঋণ ৮ শতাংশ বা ২ হাজার ২৯১ কোটি টাকা, প্রাইম ব্যাংকের সাড়ে ৮ শতাংশ বা ১ হাজার ৩৯৭ কোটি টাকা স্ট্যান্ডার্ড ব্যাংকের খেলাপি ঋণের হার ৯ শতাংশ বা ১ হাজার ১৬৮ কোটি টাকা, ইউনাইটেড কমার্শিয়াল ব্যাংকের ৯ শতাংশ বা ২ হাজার ১৭৮ কোটি টাকা, উত্তরা ব্যাংকের ৮ শতাংশ বা ৮৩২ কোটি টাকা সংকটে থাকা বাংলাদেশ কমার্স ব্যাংকের খেলাপি ঋণের হার ২৬ শতাংশ, টাকার অঙ্কে যা ৫৫৭ কোটি টাকা, আইসিবি ইসলামিক ব্যাংকের খেলাপি ঋণ ৭০৮ কোটি টাকা বা ৮৪ শতাংশ এদিকে ফারমার্স ব্যাংকের খেলাপি ঋণের হার ৫৮ শতাংশ, টাকার অঙ্কে যা ৩ হাজার ৭০ কোটি টাকা মেঘনা ব্যাংকের খেলাপি ঋণ ২৩৫ কোটি টাকা বা ৮ শতাংশ, এনআরবি কমার্শিয়াল ব্যাংকের ৩৫০ কোটি টাকা বা ৮ শতাংশ বিদেশি ব্যাংকগুলোর মধ্যে হাবিব ব্যাংকের খেলাপি ঋণের হার ১০ শতাংশ বা ৪২ কোটি টাকা, ন্যাশনাল ব্যাংক অব পাকিস্তানের ৯৪ শতাংশ বা ১ হাজার ৩৬২ কোটি টাকা

ও স্টেট ব্যাংক অব ইন্ডিয়ান ৭ শতাংশ বা ৪২ কোটি টাকা

Summary based on Keyword Ranking:

স্ট্যান্ডার্ড ব্যাংকের খেলাপি ঋণের হার ৯ শতাংশ বা ১ হাজার ১৬৮ কোটি টাকা, ইউনাইটেড কমার্শিয়াল ব্যাংকের ৯ শতাংশ বা ২ হাজার ১৭৮ কোটি টাকা, উত্তরা ব্যাংকের ৮ শতাংশ বা ৮৩২ কোটি টাকা। বিদেশি ব্যাংকগুলোর মধ্যে হাবিব ব্যাংকের খেলাপি ঋণের হার ১০ শতাংশ বা ৪২ কোটি টাকা, ন্যাশনাল ব্যাংক অব পাকিস্তানের ৯৪ শতাংশ বা ১ হাজার ৩৬২ কোটি টাকা ও স্টেট ব্যাংক অব ইন্ডিয়ান ৭ শতাংশ বা ৪২ কোটি টাকা। কেন্দ্রীয় ব্যাংকের হিসাবে, সেপ্টেম্বর শেষে অগ্রণী ব্যাংকের খেলাপি ঋণ ৫ হাজার ৯৬২ কোটি টাকা বা ১৯ শতাংশ, রূপালী ব্যাংকের ৪ হাজার ৯৭০ কোটি টাকা বা ২২ শতাংশ। ন্যাশনাল ব্যাংকের খেলাপি ঋণ ৮ শতাংশ বা ২ হাজার ২৯১ কোটি টাকা, প্রাইম ব্যাংকের সাড়ে ৮ শতাংশ বা ১ হাজার ৩৯৭ কোটি টাকা। বেসরকারি ব্যাংকগুলোর মধ্যে এবি ব্যাংকের খেলাপি ঋণের হার ৭ শতাংশ বা ১ হাজার ৬৩০ কোটি টাকা, আইএফআইসি ব্যাংকের ৮ শতাংশ বা ১ হাজার ৪৮০ কোটি টাকা। জনতা ব্যাংকের খেলাপি ঋণ ১৪ হাজার ৩০৭ কোটি টাকা বা ৩১ শতাংশ ও সোনালী ব্যাংকের ১২ হাজার ৭৭১ কোটি টাকা বা ৩৪ শতাংশ। সংকটে থাকা বাংলাদেশ কমার্স ব্যাংকের খেলাপি ঋণের হার ২৬ শতাংশ, টাকার অঙ্কে যা ৫৫৭ কোটি টাকা, আইসিবি ইসলামিক ব্যাংকের খেলাপি ঋণ ৭০৮ কোটি টাকা বা ৮৪ শতাংশ। মেঘনা ব্যাংকের খেলাপি ঋণ ২৩৫ কোটি টাকা বা ৮ শতাংশ, এনআরবি কমার্শিয়াল ব্যাংকের ৩৫০ কোটি টাকা বা ৮ শতাংশ। এ ছাড়া বেসিক ব্যাংকের খেলাপি ঋণের হার ৬২ শতাংশ বা ৯ হাজার ১৪৩ কোটি টাকা, বিডিবিএলের ৮৪৮ কোটি টাকা বা ৫৫ শতাংশ। গত সেপ্টেম্বর শেষে ব্যাংকটির খেলাপি ঋণের হার দাঁড়িয়েছে ৭ শতাংশ বা ১ হাজার ৫৬২ কোটি টাকা। এদিকে ফারমার্স ব্যাংকের খেলাপি ঋণের হার ৫৮ শতাংশ, টাকার অঙ্কে যা ৩ হাজার ৭০ কোটি টাকা। কৃষি ব্যাংকের খেলাপি ঋণ ২১ শতাংশ ও রাজশাহী কৃষি উন্নয়ন ব্যাংকের ২৫ শতাংশ।

Table A.6: ROUGE 1 and ROUGE 2 Score of the key word generated summary for the news.

ROUGE 1 Precision	ROUGE 1 Recall	ROUGE 1 F-measure	ROUGE 2 Precision	ROUGE 2 Recall	ROUGE 2 F-measure
0.8532	0.7266	0.7848	0.7622	0.7014	0.0.7306

Summary based on Sentiment Scoring:

বিদেশি ব্যাংকগুলোর মধ্যে হাবিব ব্যাংকের খেলাপি ঋণের হার ১০ শতাংশ বা ৪২ কোটি টাকা, ন্যাশনাল ব্যাংক অব পাকিস্তানের ৯৪ শতাংশ বা ১ হাজার ৩৬২ কোটি টাকা ও স্টেট ব্যাংক অব ইন্ডিয়ান ৭ শতাংশ বা ৪২ কোটি টাকা। স্ট্যান্ডার্ড ব্যাংকের খেলাপি ঋণের হার ৯ শতাংশ বা ১ হাজার ১৬৮ কোটি টাকা, ইউনাইটেড কমার্শিয়াল ব্যাংকের ৯ শতাংশ বা ২ হাজার ১৭৮ কোটি টাকা, উত্তরা ব্যাংকের ৮ শতাংশ বা ৮৩২ কোটি টাকা। কেন্দ্রীয় ব্যাংকের হিসাবে, সেপ্টেম্বর শেষে অগ্রণী ব্যাংকের খেলাপি ঋণ ৫ হাজার ৯৬২ কোটি টাকা বা ১৯ শতাংশ, রূপালী ব্যাংকের ৪ হাজার ৯৭০ কোটি টাকা বা ২২ শতাংশ। বেসরকারি ব্যাংকগুলোর মধ্যে এবি ব্যাংকের খেলাপি ঋণের হার ৭ শতাংশ বা ১ হাজার ৬৩০ কোটি টাকা, আইএফআইসি ব্যাংকের ৮ শতাংশ বা ১ হাজার ৪৮০ কোটি টাকা। সংকটে থাকা বাংলাদেশ কমার্স ব্যাংকের খেলাপি ঋণের হার ২৬ শতাংশ, টাকার অঙ্কে যা ৫৫৭ কোটি

টাকা, আইসিবি ইসলামিক ব্যাংকের খেলাপি ঋণ ৭০৮ কোটি টাকা বা ৮৪ শতাংশ। জনতা ব্যাংকের খেলাপি ঋণ ১৪ হাজার ৩০৭ কোটি টাকা বা ৩১ শতাংশ ও সোনালী ব্যাংকের ১২ হাজার ৭৭১ কোটি টাকা বা ৩৪ শতাংশ। এ ছাড়া বেসিক ব্যাংকের খেলাপি ঋণের হার ৬২ শতাংশ বা ৯ হাজার ১৪৩ কোটি টাকা, বিডিবিএলের ৮৪৮ কোটি টাকা বা ৫৫ শতাংশ। ন্যাশনাল ব্যাংকের খেলাপি ঋণ ৮ শতাংশ বা ২ হাজার ২৯১ কোটি টাকা, প্রাইম ব্যাংকের সাড়ে ৮ শতাংশ বা ১ হাজার ৩৯৭ কোটি টাকা। কেন্দ্রীয় ব্যাংকের সাবেক ডেপুটি গভর্নর খোন্দকার ইব্রাহিম খালেদ এ নিয়ে বলেন, সরকারি ব্যাংকগুলোর যে সমস্যা রয়েছে, তা বেসরকারি ব্যাংকেও ছড়িয়ে পড়েছে। বেসরকারি খাতের ভালো ব্যাংকের মধ্যে ৭ শতাংশের বেশি খেলাপি ঋণ রয়েছে আইএফআইসি, প্রাইম, স্ট্যান্ডার্ড, ইউসিবিএল, সোস্যাল ইসলামী ব্যাংকসহ আরও কয়েকটির। মেঘনা ব্যাংকের খেলাপি ঋণ ২৩৫ কোটি টাকা বা ৮ শতাংশ, এনআরবি কমার্শিয়াল ব্যাংকের ৩৫০ কোটি টাকা বা ৮ শতাংশ। কেউ পুনঃ তফসিল করে খেলাপি ঋণ কমায়, আবার কেউ কেউ ঋণ পরিশোধ না হলেও তা খেলাপি হিসাবে দেখায় না।

Table A.7: ROUGE 1 and ROUGE 2 Score of the sentiment score generated summary for the news.

ROUGE 1 Precision	ROUGE 1 Recall	ROUGE 1 F-measure	ROUGE 2 Precision	ROUGE 2 Recall	ROUGE 2 F-measure
0.7226	0.7734	0.7472	0.6916	0.7363	0.7133

Summary based on Text Ranking:

সরকারি-বেসরকারি ২১ ব্যাংকের খেলাপি ঋণ ৭ শতাংশের বেশি গত সেপ্টেম্বর শেষে দেশের ব্যাংক খাতে খেলাপি ঋণের হার দাঁড়িয়েছে ১১ দশমিক ৪৫ শতাংশে বেসরকারি খাতের ভালো ব্যাংকের মধ্যে ৭ শতাংশের বেশি খেলাপি ঋণ রয়েছে আইএফআইসি, প্রাইম, স্ট্যান্ডার্ড, ইউসিবিএল, সোস্যাল ইসলামী ব্যাংকসহ আরও কয়েকটির এর মধ্যে খেলাপিতে শীর্ষে থাকা পুরোনো ব্যাংকগুলো সঙ্গে নতুন করে যুক্ত হয়েছে বেসরকারি খাতের ভালো বেশ কয়েকটি ব্যাংকও সর্বশেষ গত সেপ্টেম্বর শেষে খেলাপি ঋণের হালনাগাদ তথ্য বিশ্লেষণ করে এমন তথ্য পাওয়া গেছে ‘ভালো’ ব্যাংকেও খেলাপি বাড়ছে

বাংলাদেশের ব্যাংক খাতের দুরবস্থার চিত্রের কিছুটা খেলাপি ঋণের তথ্যের মাধ্যমে ক্রমেই প্রকাশ পাচ্ছে টাকার অঙ্কে যা ৯৯ হাজার ৩৭০ কোটি টাকা যদিও খেলাপি ঋণের তথ্য নিয়ে ব্যাংকগুলোতে নানা রকমের ছলচাতুরী করা হয় কেন্দ্রীয় ব্যাংকের হিসাবে, সেপ্টেম্বর শেষে অগ্রণী ব্যাংকের খেলাপি ঋণ ৫ হাজার ৯৬২ কোটি টাকা বা ১৯ শতাংশ, রূপালী ব্যাংকের ৪ হাজার ৯৭০ কোটি টাকা বা ২২ শতাংশ কেউ পুনঃ তফসিল করে খেলাপি ঋণ কমায়, আবার কেউ কেউ ঋণ পরিশোধ না হলেও তা খেলাপি হিসাবে দেখায় না জনতা ব্যাংকের খেলাপি ঋণ ১৪ হাজার ৩০৭ কোটি টাকা বা ৩১ শতাংশ ও সোনালী ব্যাংকের ১২ হাজার ৭৭১ কোটি টাকা বা ৩৪ শতাংশ কেন্দ্রীয় ব্যাংকের সাবেক ডেপুটি গভর্নর খোন্দকার ইব্রাহিম খালেদ এ নিয়ে বলেন, সরকারি ব্যাংকগুলোর যে সমস্যা রয়েছে, তা বেসরকারি ব্যাংকেও ছড়িয়ে পড়েছে এ ছাড়া বেসিক ব্যাংকের খেলাপি ঋণের হার ৬২ শতাংশ বা ৯ হাজার ১৪৩ কোটি টাকা, বিডিবিএলের ৮৪৮ কোটি টাকা বা ৫৫ শতাংশ বেসরকারি ব্যাংকের পরিচালকেরা এখন জমিদারের মতো আচরণ করছেন বেসরকারি ব্যাংকগুলোর মধ্যে এবি ব্যাংকের খেলাপি ঋণের হার ৭ শতাংশ বা ১ হাজার ৬৩০ কোটি টাকা, আইএফআইসি ব্যাংকের ৮ শতাংশ বা ১ হাজার ৪৮০ কোটি টাকা কৃষি ব্যাংকের খেলাপি ঋণ ২১ শতাংশ ও রাজশাহী কৃষি উন্নয়ন ব্যাংকের ২৫ শতাংশ ন্যাশনাল ব্যাংকের খেলাপি ঋণ ৮ শতাংশ বা ২ হাজার ২৯১ কোটি টাকা, প্রাইম ব্যাংকের সাড়ে ৮ শতাংশ বা ১ হাজার ৩৯৭ কোটি টাকা স্ট্যান্ডার্ড ব্যাংকের

খেলাপি ঋণের হার ৯ শতাংশ বা ১ হাজার ১৬৮ কোটি টাকা, ইউনাইটেড কমার্শিয়াল ব্যাংকের ৯ শতাংশ বা ২ হাজার ১৭৮ কোটি টাকা, উত্তরা ব্যাংকের ৮ শতাংশ বা ৮৩২ কোটি টাকা এ জন্য কেন্দ্রীয় ব্যাংককে শক্তিশালী ভূমিকা নিতে হবে

Table A.8: ROUGE 1 and ROUGE 2 Score of the Text Rank generated summary for the news.

ROUGE 1 Precision	ROUGE 1 Recall	ROUGE 1 F-measure	ROUGE 2 Precision	ROUGE 2 Recall	ROUGE 2 F-measure
0.5625	0.7734	0.6513	0.5189	0.6816	0.5892

Summary based on Hybrid Ranking (KR+SS+TR):

বিদেশি ব্যাংকগুলোর মধ্যে হাবিব ব্যাংকের খেলাপি ঋণের হার ১০ শতাংশ বা ৪২ কোটি টাকা, ন্যাশনাল ব্যাংক অব পাকিস্তানের ৯৪ শতাংশ বা ১ হাজার ৩৬২ কোটি টাকা ও স্টেট ব্যাংক অব ইন্ডিয়া ৭ শতাংশ বা ৪২ কোটি টাকা। স্ট্যান্ডার্ড ব্যাংকের খেলাপি ঋণের হার ৯ শতাংশ বা ১ হাজার ১৬৮ কোটি টাকা, ইউনাইটেড কমার্শিয়াল ব্যাংকের ৯ শতাংশ বা ২ হাজার ১৭৮ কোটি টাকা, উত্তরা ব্যাংকের ৮ শতাংশ বা ৮৩২ কোটি টাকা। কেন্দ্রীয় ব্যাংকের হিসাবে, সেপ্টেম্বর শেষে অগ্রণী ব্যাংকের খেলাপি ঋণ ৫ হাজার ৯৬২ কোটি টাকা বা ১৯ শতাংশ, রূপালী ব্যাংকের ৪ হাজার ৯৭০ কোটি টাকা বা ২২ শতাংশ। বেসরকারি ব্যাংকগুলোর মধ্যে এবি ব্যাংকের খেলাপি ঋণের হার ৭ শতাংশ বা ১ হাজার ৬৩০ কোটি টাকা, আইএফআইসি ব্যাংকের ৮ শতাংশ বা ১ হাজার ৪৮০ কোটি টাকা। ন্যাশনাল ব্যাংকের খেলাপি ঋণ ৮ শতাংশ বা ২ হাজার ২৯১ কোটি টাকা, প্রাইম ব্যাংকের সাড়ে ৮ শতাংশ বা ১ হাজার ৩৯৭ কোটি টাকা। সংকটে থাকা বাংলাদেশ কমার্স ব্যাংকের খেলাপি ঋণের হার ২৬ শতাংশ, টাকার অঙ্কে যা ৫৫৭ কোটি টাকা, আইসিবি ইসলামিক ব্যাংকের খেলাপি ঋণ ৭০৮ কোটি টাকা বা ৮৪ শতাংশ। জনতা ব্যাংকের খেলাপি ঋণ ১৪ হাজার ৩০৭ কোটি টাকা বা ৩১ শতাংশ ও সোনালী ব্যাংকের ১২ হাজার ৭৭১ কোটি টাকা বা ৩৪ শতাংশ। এ ছাড়া বেসিক ব্যাংকের খেলাপি ঋণের হার ৬২ শতাংশ বা ৯ হাজার ১৪৩ কোটি টাকা, বিডিবিএলের ৮৪৮ কোটি টাকা বা ৫৫ শতাংশ। মেঘনা ব্যাংকের খেলাপি ঋণ ২৩৫ কোটি টাকা বা ৮ শতাংশ, এনআরবি কমার্শিয়াল ব্যাংকের ৩৫০ কোটি টাকা বা ৮ শতাংশ। গত সেপ্টেম্বর শেষে ব্যাংকটির খেলাপি ঋণের হার দাঁড়িয়েছে ৭ শতাংশ বা ১ হাজার ৫৬২ কোটি টাকা। এদিকে ফারমার্স ব্যাংকের খেলাপি ঋণের হার ৫৮ শতাংশ, টাকার অঙ্কে যা ৩ হাজার ৭০ কোটি টাকা। বেসরকারি খাতের ভালো ব্যাংকের মধ্যে ৭ শতাংশের বেশি খেলাপি ঋণ রয়েছে আইএফআইসি, প্রাইম, স্ট্যান্ডার্ড, ইউসিবিএল, সোস্যাল ইসলামী ব্যাংকসহ আরও কয়েকটির।

Table A.9: ROUGE 1 and ROUGE 2 Score of the Hybrid summary (combining KR, SS, TR) for the news.

ROUGE 1 Precision	ROUGE 1 Recall	ROUGE 1 F-measure	ROUGE 2 Precision	ROUGE 2 Recall	ROUGE 2 F-measure
0.8898	0.8203	0.8537	0.8144	0.7861	0.7999

Summary generated by other web based system :

‘ভালো’ ব্যাংকেও খেলাপি বাড়ছে

সরকারি-বেসরকারি ২১ ব্যাংকের খেলাপি ঋণ ৭ শতাংশের বেশি। বেসরকারি খাতের ভালো ব্যাংকের মধ্যে ৭ শতাংশের বেশি খেলাপি ঋণ রয়েছে আইএফআইসি, প্রাইম, স্ট্যান্ডার্ড, ইউসিবিএল, সোস্যাল ইসলামী ব্যাংকসহ আরও কয়েকটির। সর্বশেষ গত সেপ্টেম্বর শেষে খেলাপি ঋণের হালনাগাদ তথ্য বিশ্লেষণ করে এমন তথ্য পাওয়া গেছে। গত সেপ্টেম্বর শেষে দেশের ব্যাংক খাতে খেলাপি ঋণের হার দাঁড়িয়েছে ১১ দশমিক ৪৫ শতাংশে। এর মধ্যে খেলাপিতে শীর্ষে থাকা পুরোনো ব্যাংকগুলো সঙ্গে নতুন করে যুক্ত হয়েছে বেসরকারি খাতের ভালো বেশ কয়েকটি ব্যাংকও। টাকার অঙ্কে যা ৯৯ হাজার ৩৭০ কোটি টাকা

Table A.10: ROUGE 1 and ROUGE 2 Score of other web based system generated summary for the news.

ROUGE 1 Precision	ROUGE 1 Recall	ROUGE 1 F-measure	ROUGE 2 Precision	ROUGE 2 Recall	ROUGE 2 F-measure
0.7647	0.4063	0.5306	0.6538	0.2537	0.3656

Entertainment Category:

Actual News:

"আজ বিকেলে ‘সময়ের মুখ’"

জাতীয় অধ্যাপক শিক্ষাবিদ মুস্তাফা নূরউল ইসলামের জীবন অবলম্বনে চলচ্চিত্রকার বেলায়াত হোসেন মামুন তৈরি করেছেন প্রামাণ্য চলচ্চিত্র ‘সময়ের মুখ’। আজ বুধবার বিকেল চারটায় বাংলাদেশ শিল্পকলা একাডেমির জাতীয় চিত্রশালা মিলনায়তনে বাংলাদেশ স্বল্পদৈর্ঘ্য কাহিনি ও প্রামাণ্য চলচ্চিত্র উৎসবে ছবিটির উদ্বোধনী প্রদর্শনী হবে। চলচ্চিত্রটি প্রযোজনা করেছে বাংলাদেশ শিল্পকলা একাডেমি।

‘সময়ের মুখ’ ছবি নিয়ে বেলায়াত হোসেন মামুন জানান, শিক্ষাবিদ, গবেষক ও অধ্যাপক মুস্তাফা নূরউল ইসলামের জীবনের উল্লেখযোগ্য ঘটনার ব্যাখ্যা ও বিশ্লেষণ রয়েছে এই চলচ্চিত্রে। এখানে মুস্তাফা নূরউল ইসলাম নিজেই নিজের জীবনের কথকরূপে হাজির হয়েছেন। নিজের জীবনের নানা ঘটনা বয়ানের মধ্য দিয়ে তিনি বলে গেছেন ব্রিটিশ শাসনাধীন অখণ্ড ভারতের স্বাধীনতাকামী মানুষের সংগ্রামের ইতিহাস আর পূর্ব বাংলার পূর্ব পাকিস্তান হয়ে ওঠার গল্প। মুস্তাফা নূরউল ইসলাম আরও বলেছেন, কেমন করে একজন সাধারণ শিক্ষার্থী থেকে তিনি হয়ে উঠেছেন ছাত্র আন্দোলনের নেতা, বঙ্গবন্ধু শেখ মুজিবুর রহমানের সঙ্গে তাঁর পরিচয়, সখ্য এবং একসঙ্গে ভাষা আন্দোলন গড়ে তোলার ইতিহাস। মুস্তাফা নূরউল ইসলাম দেশভাগের ইতিহাস থেকে শুরু করে স্বাধীন বাংলাদেশ অর্জনের ইতিহাস বলেছেন নিজস্ব দৃষ্টি ও মননশীল ব্যাখ্যায়।

বেলায়াত হোসেন মামুন বলেন, ৪২ মিনিটের এই প্রামাণ্য চলচ্চিত্রে জাতীয় অধ্যাপক মুস্তাফা নূরউল ইসলামের জীবনের ইতিহাস তুলে ধরতে গিয়ে বাংলাদেশের ‘বাংলাদেশ’ হয়ে ওঠার সময়কে ধারণ করা হয়েছে। সেই সময়কে যিনি বলছেন, তিনি জাতীয় অধ্যাপক মুস্তাফা নূরউল ইসলাম।

শিক্ষাবিদ মুস্তাফা নূরউল ইসলাম প্রয়াত হয়েছে ২০১৮ সালের ৯ মে।

Provided Gold summary generated by random human:

জাতীয় অধ্যাপক শিক্ষাবিদ মুস্তাফা নূরউল ইসলামের জীবন অবলম্বনে চলচ্চিত্রকার বেলায়াত হোসেন মামুন তৈরি করেছেন প্রামাণ্য চলচ্চিত্র ‘সময়ের মুখ’। আজ বুধবার বিকেল চারটায় বাংলাদেশ শিল্পকলা একাডেমির জাতীয় চিত্রশালা মিলনায়তনে বাংলাদেশ স্বল্পদৈর্ঘ্য কাহিনি ও প্রামাণ্য চলচ্চিত্র উৎসবে ছবিটির উদ্বোধনী প্রদর্শনী হবে। চলচ্চিত্রটি প্রযোজনা করেছে বাংলাদেশ শিল্পকলা একাডেমি। মুস্তাফা নূরউল ইসলাম দেশভাগের ইতিহাস থেকে শুরু করে স্বাধীন বাংলাদেশ অর্জনের ইতিহাস বলেছেন নিজস্ব দৃষ্টি ও মননশীল ব্যাখ্যায়। বেলায়াত হোসেন মামুন বলেন, ৪২ মিনিটের এই প্রামাণ্য চলচ্চিত্রে জাতীয় অধ্যাপক মুস্তাফা নূরউল ইসলামের জীবনের ইতিহাস তুলে ধরতে গিয়ে বাংলাদেশের ‘বাংলাদেশ’ হয়ে ওঠার সময়কে ধারণ করা হয়েছে।

Summary based on Keyword Ranking:

আজ বুধবার বিকেল চারটায় বাংলাদেশ শিল্পকলা একাডেমির জাতীয় চিত্রশালা মিলনায়তনে বাংলাদেশ স্বল্পদৈর্ঘ্য কাহিনি ও প্রামাণ্য চলচ্চিত্র উৎসবে ছবিটির উদ্বোধনী প্রদর্শনী হবে। ‘সময়ের মুখ’ ছবি নিয়ে বেলায়াত হোসেন মামুন জানান, শিক্ষাবিদ, গবেষক ও অধ্যাপক মুস্তাফা নূরউল ইসলামের জীবনের উল্লেখযোগ্য ঘটনার ব্যাখ্যা ও বিশ্লেষণ রয়েছে এই চলচ্চিত্রে। জাতীয় অধ্যাপক শিক্ষাবিদ মুস্তাফা নূরউল ইসলামের জীবন অবলম্বনে চলচ্চিত্রকার বেলায়াত হোসেন মামুন তৈরি করেছেন প্রামাণ্য চলচ্চিত্র ‘সময়ের মুখ’। বেলায়াত হোসেন মামুন বলেন, ৪২ মিনিটের এই প্রামাণ্য চলচ্চিত্রে জাতীয় অধ্যাপক মুস্তাফা নূরউল ইসলামের জীবনের ইতিহাস তুলে ধরতে গিয়ে বাংলাদেশের ‘বাংলাদেশ’ হয়ে ওঠার সময়কে ধারণ করা হয়েছে।

Table A.11: ROUGE 1 and ROUGE 2 Score of the key word generated summary for the news.

ROUGE 1 Precision	ROUGE 1 Recall	ROUGE 1 F- measure	ROUGE 2 Precision	ROUGE 2 Recall	ROUGE 2 F- measure
0.8154	0.7571	0.7852	0.7403	0.6951	0.7169

Summary based on Sentiment Scoring:

মুস্তাফা নূরউল ইসলাম আরও বলেছেন, কেমন করে একজন সাধারণ শিক্ষার্থী থেকে তিনি হয়ে উঠেছেন ছাত্র আন্দোলনের নেতা, বঙ্গবন্ধু শেখ মুজিবুর রহমানের সঙ্গে তাঁর পরিচয়, সখ্য এবং একসঙ্গে ভাষা আন্দোলন গড়ে তোলার ইতিহাস। বেলায়াত হোসেন মামুন বলেন, ৪২ মিনিটের এই প্রামাণ্য চলচ্চিত্রে জাতীয় অধ্যাপক মুস্তাফা নূরউল ইসলামের জীবনের ইতিহাস তুলে ধরতে গিয়ে বাংলাদেশের ‘বাংলাদেশ’ হয়ে ওঠার সময়কে ধারণ করা হয়েছে। নিজের জীবনের নানা ঘটনা বয়ানের মধ্য দিয়ে তিনি বলে গেছেন ব্রিটিশ শাসনাধীন অখণ্ড ভারতের স্বাধীনতাকামী মানুষের সংগ্রামের ইতিহাস আর পূর্ব বাংলার পূর্ব পাকিস্তান হয়ে ওঠার গল্প। ‘সময়ের মুখ’ ছবি নিয়ে বেলায়াত হোসেন মামুন জানান, শিক্ষাবিদ, গবেষক ও অধ্যাপক মুস্তাফা নূরউল ইসলামের জীবনের উল্লেখযোগ্য ঘটনার ব্যাখ্যা ও বিশ্লেষণ রয়েছে এই চলচ্চিত্রে।

Table A.12: ROUGE 1 and ROUGE 2 Score of the sentiment score generated summary for the news.

ROUGE 1 Precision	ROUGE 1 Recall	ROUGE 1 F-measure	ROUGE 2 Precision	ROUGE 2 Recall	ROUGE 2 F-measure
0.3556	0.4571	0.4000	0.2700	0.3293	0.2967

Summary based on Text Ranking:

জাতীয় অধ্যাপক শিক্ষাবিদ মুস্তাফা নূরউল ইসলামের জীবন অবলম্বনে চলচ্চিত্রকার বেলায়াত হোসেন মামুন তৈরি করেছেন প্রামাণ্য চলচ্চিত্র ‘সময়ের মুখ’ আজ বুধবার বিকেল চারটায় বাংলাদেশ শিল্পকলা একাডেমির জাতীয় চিত্রশালা মিলনায়তনে বাংলাদেশ স্বল্পদৈর্ঘ্য কাহিনি ও প্রামাণ্য চলচ্চিত্র উৎসবে ছবিটির উদ্বোধনী প্রদর্শনী হবে "আজ বিকেলে ‘সময়ের মুখ’" ‘সময়ের মুখ’ ছবি নিয়ে বেলায়াত হোসেন মামুন জানান, শিক্ষাবিদ, গবেষক ও অধ্যাপক মুস্তাফা নূরউল ইসলামের জীবনের উল্লেখযোগ্য ঘটনার ব্যাখ্যা ও বিশ্লেষণ রয়েছে এই চলচ্চিত্রে চলচ্চিত্রটি প্রযোজনা করেছে বাংলাদেশ শিল্পকলা একাডেমি এখানে মুস্তাফা নূরউল ইসলাম নিজেই নিজের জীবনের কথকরূপে হাজির হয়েছেন নিজের জীবনের নানা ঘটনা বয়ানের মধ্য দিয়ে তিনি বলে গেছেন ব্রিটিশ শাসনাধীন অঞ্চল ভারতের স্বাধীনতাকামী মানুষের সংগ্রামের ইতিহাস আর পূর্ব বাংলার পূর্ব পাকিস্তান হয়ে ওঠার গল্প মুস্তাফা নূরউল ইসলাম আরও বলেছেন, কেমন করে একজন সাধারণ শিক্ষার্থী থেকে তিনি হয়ে উঠেছেন ছাত্র আন্দোলনের নেতা, বঙ্গবন্ধু শেখ মুজিবুর রহমানের সঙ্গে তাঁর পরিচয়, সখ্য এবং একসঙ্গে ভাষা আন্দোলন গড়ে তোলার ইতিহাস

Table A.13: ROUGE 1 and ROUGE 2 Score of the Text Rank generated summary for the news.

ROUGE 1 Precision	ROUGE 1 Recall	ROUGE 1 F-measure	ROUGE 2 Precision	ROUGE 2 Recall	ROUGE 2 F-measure
0.4054	0.6428	0.4972	0.3178	0.5000	0.3886

Summary based on Hybrid Ranking (KR+SS+TR):

বেলায়াত হোসেন মামুন বলেন, ৪২ মিনিটের এই প্রামাণ্য চলচ্চিত্রে জাতীয় অধ্যাপক মুস্তাফা নূরউল ইসলামের জীবনের ইতিহাস তুলে ধরতে গিয়ে বাংলাদেশের ‘বাংলাদেশ’ হয়ে ওঠার সময়কে ধারণ করা হয়েছে। ‘সময়ের মুখ’ ছবি নিয়ে বেলায়াত হোসেন মামুন জানান, শিক্ষাবিদ, গবেষক ও অধ্যাপক মুস্তাফা নূরউল ইসলামের জীবনের উল্লেখযোগ্য ঘটনার ব্যাখ্যা ও বিশ্লেষণ রয়েছে এই চলচ্চিত্রে। আজ বুধবার বিকেল চারটায় বাংলাদেশ শিল্পকলা একাডেমির জাতীয় চিত্রশালা মিলনায়তনে বাংলাদেশ স্বল্পদৈর্ঘ্য কাহিনি ও প্রামাণ্য চলচ্চিত্র উৎসবে ছবিটির উদ্বোধনী প্রদর্শনী হবে। জাতীয় অধ্যাপক শিক্ষাবিদ মুস্তাফা নূরউল ইসলামের জীবন অবলম্বনে চলচ্চিত্রকার বেলায়াত হোসেন মামুন তৈরি করেছেন প্রামাণ্য চলচ্চিত্র ‘সময়ের মুখ’।

Table A.14: ROUGE 1 and ROUGE 2 Score of the Hybrid summary (combining KR, SS, TR) for the news.

ROUGE 1 Precision	ROUGE 1 Recall	ROUGE 1 F-measure	ROUGE 2 Precision	ROUGE 2 Recall	ROUGE 2 F-measure
0.8154	0.7571	0.7852	0.7403	0.6951	0.7169

Summary generated by other web based system :

জাতীয় অধ্যাপক শিক্ষাবিদ মুস্তাফা নূরউল ইসলামের জীবন অবলম্বনে চলচ্চিত্রকার বেলায়াত হোসেন মামুন তৈরি করেছেন প্রামাণ্য চলচ্চিত্র ‘সময়ের মুখ’। আজ বুধবার বিকেল চারটায় বাংলাদেশ শিল্পকলা একাডেমির জাতীয় চিত্রশালা মিলনায়তনে বাংলাদেশ স্বল্পদৈর্ঘ্য কাহিনি ও প্রামাণ্য চলচ্চিত্র উৎসবে ছবিটির উদ্বোধনী প্রদর্শনী হবে। ‘সময়ের মুখ’ ছবি নিয়ে বেলায়াত হোসেন মামুন জানান, শিক্ষাবিদ, গবেষক ও অধ্যাপক মুস্তাফা নূরউল ইসলামের জীবনের উল্লেখযোগ্য ঘটনার ব্যাখ্যা ও বিশ্লেষণ রয়েছে এই চলচ্চিত্রে। চলচ্চিত্রটি প্রযোজনা করেছে বাংলাদেশ শিল্পকলা একাডেমি

Table A.15: ROUGE 1 and ROUGE 2 Score of other web based system generated summary for the news.

ROUGE 1 Precision	ROUGE 1 Recall	ROUGE 1 F-measure	ROUGE 2 Precision	ROUGE 2 Recall	ROUGE 2 F-measure
0.7593	0.5857	0.6613	0.6774	0.5122	0.5833

Politics Category:

Actual News:

অর্থমন্ত্রীর বাসায় ইনাম আহমদ অর্থমন্ত্রী আবুল মাল আবদুল মুহিত এবার নির্বাচন করছেন না। তাঁর আসনে (সিলেট-১) প্রার্থী হয়েছেন তাঁর ছোট ভাই এ কে আবদুল মোমেন। গতকাল অর্থমন্ত্রীর কাছে নির্বাচনে সমান সুযোগের দাবি তুললেন ওই আসনের বিএনপির সম্ভাব্য প্রার্থী ইনাম আহমদ চৌধুরী। দলীয় নেতা-কর্মীদের ধরপাকড় এবং বাড়ি বাড়ি পুলিশি তল্লাশিরও প্রতিবাদ জানান তিনি। গতকাল বৃহস্পতিবার সকালে সিলেট শহরে অর্থমন্ত্রীর বাসায় গিয়ে তাঁর সঙ্গে সাক্ষাৎ করেন ইনাম আহমদ চৌধুরী। এ সময় আবদুল মোমেনও উপস্থিত ছিলেন। ইনাম আহমদ প্রথম আলোকে বলেন, ‘একধরনের প্রতিবাদী মন নিয়ে আমি তাঁর বাসায় গিয়েছিলাম। নির্বাচনের বাইরে আছেন অর্থমন্ত্রী। তাই তাঁকে এসব ব্যাপারে নজর রাখতে বলেছি। মুখে গণতন্ত্র বলব, পুলিশ যথেষ্ট আচরণ করবে—তা হতে পারে না। সমান সুযোগের দাবি নিয়ে মূলত তাঁর বাসায় যাওয়া।’ সিলেট জেলা বিএনপির সাংগঠনিক সম্পাদক এমরান আহমদ চৌধুরী জানান, বুধবার মনোনয়নপত্র দাখিলের পর রাতে বিএনপির মনোনয়নপ্রত্যাশী কয়েকজনের বাসায়ও যান ইনাম আহমদ। গতকাল সকালে তাঁর কাছে খবর আসে, ওই দিন রাতে চার নেতা-কর্মীর বাসায় পুলিশি তল্লাশি হয়েছে। এ খবর শুনে তিনি অর্থমন্ত্রীর বাসায় যান। ইনাম আহমদ চৌধুরী বলেন, ‘অর্থমন্ত্রী আমার বিশ্ববিদ্যালয় জীবনের একজন বন্ধু। তাঁর সঙ্গে শিক্ষাজীবন থেকে শুরু করে কর্মজীবনে একনিষ্ঠ হিসেবে কাজ করার সৌভাগ্য হয়েছে আমার। তাই তাঁর সঙ্গে সৌজন্য সাক্ষাৎ করেছি। তাঁকে

জানিয়েছি, আমি এবার সিলেট-১ আসন থেকে বিএনপির প্রার্থী হয়ে নির্বাচনে অংশ নিচ্ছি। তিনি শুনেই আমাকে স্বাগত জানান। পরক্ষণে পুলিশি ধরপাকড়, বাসায় তল্লাশি প্রভৃতি বিষয়ে বলে এসেছি।’ বিএনপির কয়েকজন নেতা জানিয়েছেন, নির্বাচন ঘিরে সিলেটে কোনোভাবেই যাতে রাজনৈতিক সম্প্রীতি নষ্ট না হয়, সেদিকে খেয়াল রেখে নিজ নিজ অবস্থান থেকে কাজ করার জন্য একে অন্যের প্রতি আহ্বান জানান তাঁরা। সিলেট সদরে ইনাম আহমদ ছাড়াও বিএনপির চেয়ারপারসনের উপদেষ্টা খন্দকার আবদুল মুক্তাদির মনোনয়নপত্র দাখিল করেছেন। এক আসনে দলের একাধিক প্রার্থী দেওয়ার বিষয়ে ইনাম আহমদ বলেন, প্রায় ৮০ শতাংশ আসনে বিএনপি একাধিক প্রার্থী দিয়েছে। এটি বিভিন্ন কারণে নীতিগতভাবে করা হয়েছে। শেষ পর্যন্ত সর্বগ্রহণযোগ্য প্রার্থীই থাকবেন।

Provided Gold summary generated by random human:

অর্থমন্ত্রীর বাসায় ইনাম আহমদ অর্থমন্ত্রী আবুল মাল আবদুল মুহিত এবার নির্বাচন করছেন না। তাঁর আসনে (সিলেট-১) প্রার্থী হয়েছেন তাঁর ছোট ভাই এ কে আবদুল মোমেন। গতকাল অর্থমন্ত্রীর কাছে নির্বাচনে সমান সুযোগের দাবি তুললেন ওই আসনের বিএনপির সম্ভাব্য প্রার্থী ইনাম আহমদ চৌধুরী। গতকাল বৃহস্পতিবার সকালে সিলেট শহরে অর্থমন্ত্রীর বাসায় গিয়ে তাঁর সঙ্গে সাক্ষাৎ করেন ইনাম আহমদ চৌধুরী। এ সময় আবদুল মোমেনও উপস্থিত ছিলেন। সিলেট জেলা বিএনপির সাংগঠনিক সম্পাদক এমরান আহমদ চৌধুরী জানান, বুধবার মনোনয়নপত্র দাখিলের পর রাতে বিএনপির মনোনয়নপ্রত্যাশী কয়েকজনের বাসায়ও যান ইনাম আহমদ। বিএনপির কয়েকজন নেতা জানিয়েছেন, নির্বাচন ঘিরে সিলেটে কোনোভাবেই যাতে রাজনৈতিক সম্প্রীতি নষ্ট না হয়, সেদিকে খেয়াল রেখে নিজ নিজ অবস্থান থেকে কাজ করার জন্য একে অন্যের প্রতি আহ্বান জানান তাঁরা। সিলেট সদরে ইনাম আহমদ ছাড়াও বিএনপির চেয়ারপারসনের উপদেষ্টা খন্দকার আবদুল মুক্তাদির মনোনয়নপত্র দাখিল করেছেন। এক আসনে দলের একাধিক প্রার্থী দেওয়ার বিষয়ে ইনাম আহমদ বলেন, প্রায় ৮০ শতাংশ আসনে বিএনপি একাধিক প্রার্থী দিয়েছে।

Summary based on Keyword Ranking:

এক আসনে দলের একাধিক প্রার্থী দেওয়ার বিষয়ে ইনাম আহমদ বলেন, প্রায় ৮০ শতাংশ আসনে বিএনপি একাধিক প্রার্থী দিয়েছে। গতকাল অর্থমন্ত্রীর কাছে নির্বাচনে সমান সুযোগের দাবি তুললেন ওই আসনের বিএনপির সম্ভাব্য প্রার্থী ইনাম আহমদ চৌধুরী। বিএনপির কয়েকজন নেতা জানিয়েছেন, নির্বাচন ঘিরে সিলেটে কোনোভাবেই যাতে রাজনৈতিক সম্প্রীতি নষ্ট না হয়, সেদিকে খেয়াল রেখে নিজ নিজ অবস্থান থেকে কাজ করার জন্য একে অন্যের প্রতি আহ্বান জানান তাঁরা। তাঁকে জানিয়েছি, আমি এবার সিলেট-১ আসন থেকে বিএনপির প্রার্থী হয়ে নির্বাচনে অংশ নিচ্ছি। সিলেট জেলা বিএনপির সাংগঠনিক সম্পাদক এমরান আহমদ চৌধুরী জানান, বুধবার মনোনয়নপত্র দাখিলের পর রাতে বিএনপির মনোনয়নপ্রত্যাশী কয়েকজনের বাসায়ও যান ইনাম আহমদ। তাঁর আসনে (সিলেট-১) প্রার্থী হয়েছেন তাঁর ছোট ভাই এ কে আবদুল মোমেন। দলীয় নেতা-কর্মীদের ধরপাকড় এবং বাড়ি বাড়ি পুলিশি তল্লাশিরও প্রতিবাদ জানান তিনি। গতকাল সকালে তাঁর কাছে খবর আসে, ওই দিন রাতে চার নেতা-কর্মীর বাসায় পুলিশি তল্লাশি হয়েছে। সিলেট সদরে ইনাম আহমদ ছাড়াও বিএনপির চেয়ারপারসনের উপদেষ্টা খন্দকার আবদুল মুক্তাদির মনোনয়নপত্র দাখিল করেছেন। ইনাম আহমদ প্রথম আলোকে বলেন, ‘একধরনের প্রতিবাদী মন নিয়ে আমি তাঁর বাসায় গিয়েছিলাম।

Table A.16: ROUGE 1 and ROUGE 2 Score of the key word generated summary for the news.

ROUGE 1 Precision	ROUGE 1 Recall	ROUGE 1 F-measure	ROUGE 2 Precision	ROUGE 2 Recall	ROUGE 2 F-measure
0.7500	0.8455	0.7949	0.6456	0.7391	0.6892

Summary based on Sentiment Scoring:

বিএনপির কয়েকজন নেতা জানিয়েছেন, নির্বাচন ঘিরে সিলেটে কোনোভাবেই যাতে রাজনৈতিক সম্প্রীতি নষ্ট না হয়, সেদিকে খেয়াল রেখে নিজ নিজ অবস্থান থেকে কাজ করার জন্য একে অন্যের প্রতি আস্থান জানান তাঁরা। সিলেট জেলা বিএনপির সাংগঠনিক সম্পাদক এমরান আহমদ চৌধুরী জানান, বুধবার মনোনয়নপত্র দাখিলের পর রাতে বিএনপির মনোনয়নপ্রত্যাশী কয়েকজনের বাসায়ও যান ইনাম আহমদ। গতকাল সকালে তাঁর কাছে খবর আসে, ওই দিন রাতে চার নেতা-কর্মীর বাসায় পুলিশি তল্লাশি হয়েছে। এক আসনে দলের একাধিক প্রার্থী দেওয়ার বিষয়ে ইনাম আহমদ বলেন, প্রায় ৮০ শতাংশ আসনে বিএনপি একাধিক প্রার্থী দিয়েছে। তাঁর আসনে (সিলেট-১) প্রার্থী হয়েছেন তাঁর ছোট ভাই এ কে আবদুল মোমেন। গতকাল অর্থমন্ত্রীর কাছে নির্বাচনে সমান সুযোগের দাবি তুললেন ওই আসনের বিএনপির সম্ভাব্য প্রার্থী ইনাম আহমদ চৌধুরী। তাঁকে জানিয়েছি, আমি এবার সিলেট-১ আসন থেকে বিএনপির প্রার্থী হয়ে নির্বাচনে অংশ নিচ্ছি। সিলেট সদরে ইনাম আহমদ ছাড়াও বিএনপির চেয়ারপারসনের উপদেষ্টা খন্দকার আবদুল মুজাদির মনোনয়নপত্র দাখিল করেছেন। গতকাল বৃহস্পতিবার সকালে সিলেট শহরে অর্থমন্ত্রীর বাসায় গিয়ে তাঁর সঙ্গে সাক্ষাৎ করেন ইনাম আহমদ চৌধুরী। ইনাম আহমদ প্রথম আলোকে বলেন, ‘একধরনের প্রতিবাদী মন নিয়ে আমি তাঁর বাসায় গিয়েছিলাম।

Table A.17: ROUGE 1 and ROUGE 2 Score of the sentiment score generated summary for the news.

ROUGE 1 Precision	ROUGE 1 Recall	ROUGE 1 F-measure	ROUGE 2 Precision	ROUGE 2 Recall	ROUGE 2 F-measure
0.8115	0.9000	0.8534	0.7250	0.8406	0.7785

Summary based on Text Ranking:

গতকাল অর্থমন্ত্রীর কাছে নির্বাচনে সমান সুযোগের দাবি তুললেন ওই আসনের বিএনপির সম্ভাব্য প্রার্থী ইনাম আহমদ চৌধুরী গতকাল বৃহস্পতিবার সকালে সিলেট শহরে অর্থমন্ত্রীর বাসায় গিয়ে তাঁর সঙ্গে সাক্ষাৎ করেন ইনাম আহমদ চৌধুরী অর্থমন্ত্রীর বাসায় ইনাম আহমদ অর্থমন্ত্রী আবুল মাল আবদুল মুহিত এবার নির্বাচন করছেন না তাঁর আসনে (সিলেট-১) প্রার্থী হয়েছেন তাঁর ছোট ভাই এ কে আবদুল মোমেন দলীয় নেতা-কর্মীদের ধরপাকড় এবং বাড়ি বাড়ি পুলিশি তল্লাশিরও প্রতিবাদ জানান তিনি ইনাম আহমদ প্রথম আলোকে বলেন, ‘একধরনের প্রতিবাদী মন নিয়ে আমি তাঁর বাসায় গিয়েছিলাম এ সময় আবদুল মোমেনও উপস্থিত ছিলেন সিলেট জেলা বিএনপির সাংগঠনিক সম্পাদক এমরান আহমদ চৌধুরী জানান, বুধবার মনোনয়নপত্র দাখিলের পর রাতে বিএনপির মনোনয়নপ্রত্যাশী কয়েকজনের বাসায়ও যান ইনাম আহমদ সমান সুযোগের দাবি নিয়ে মূলত তাঁর বাসায় যাওয়া নির্বাচনের বাইরে আছেন অর্থমন্ত্রী তাই তাঁকে এসব ব্যাপারে নজর রাখতে বলেছি গতকাল সকালে তাঁর কাছে খবর আসে, ওই

দিন রাতে চার নেতা-কর্মীর বাসায় পুলিশি তল্লাশি হয়েছে মুখে গণতন্ত্র বলব, পুলিশ যথেষ্ট আচরণ করবে—তা হতে পারে না এ খবর শুনে তিনি অর্থমন্ত্রীর বাসায় যান

Table A.18: ROUGE 1 and ROUGE 2 Score of the Text Rank generated summary for the news.

ROUGE 1 Precision	ROUGE 1 Recall	ROUGE 1 F-measure	ROUGE 2 Precision	ROUGE 2 Recall	ROUGE 2 F-measure
0.5636	0.5636	0.5636	0.4694	0.5000	0.4842

Summary based on Hybrid Ranking (KR+SS+TR):

এক আসনে দলের একাধিক প্রার্থী দেওয়ার বিষয়ে ইনাম আহমদ বলেন, প্রায় ৮০ শতাংশ আসনে বিএনপি একাধিক প্রার্থী দিয়েছে। বিএনপির কয়েকজন নেতা জানিয়েছেন, নির্বাচন ঘিরে সিলেটে কোনোভাবেই যাতে রাজনৈতিক সম্প্রীতি নষ্ট না হয়, সেদিকে খেয়াল রেখে নিজ নিজ অবস্থান থেকে কাজ করার জন্য একে অন্যের প্রতি আহ্বান জানান তাঁরা। সিলেট জেলা বিএনপির সাংগঠনিক সম্পাদক এমরান আহমদ চৌধুরী জানান, বুধবার মনোনয়নপত্র দাখিলের পর রাতে বিএনপির মনোনয়নপ্রত্যাশী কয়েকজনের বাসায়ও যান ইনাম আহমদ। গতকাল অর্থমন্ত্রীর কাছে নির্বাচনে সমান সুযোগের দাবি তুললেন ওই আসনের বিএনপির সম্ভাব্য প্রার্থী ইনাম আহমদ চৌধুরী। তাঁকে জানিয়েছি, আমি এবার সিলেট-১ আসন থেকে বিএনপির প্রার্থী হয়ে নির্বাচনে অংশ নিচ্ছি। তাঁর আসনে (সিলেট-১) প্রার্থী হয়েছেন তাঁর ছোট ভাই এ কে আবদুল মোমেন। গতকাল সকালে তাঁর কাছে খবর আসে, ওই দিন রাতে চার নেতা-কর্মীর বাসায় পুলিশি তল্লাশি হয়েছে। সিলেট সদরে ইনাম আহমদ ছাড়াও বিএনপির চেয়ারপারসনের উপদেষ্টা খন্দকার আবদুল মুজাদির মনোনয়নপত্র দাখিল করেছেন। দলীয় নেতা-কর্মীদের ধরপাকড় এবং বাড়ি বাড়ি পুলিশি তল্লাশিরও প্রতিবাদ জানান তিনি। ইনাম আহমদ প্রথম আলোকে বলেন, ‘একধরনের প্রতিবাদী মন নিয়ে আমি তাঁর বাসায় গিয়েছিলাম।

Table A.19: ROUGE 1 and ROUGE 2 Score of the Hybrid summary (combining KR, SS, TR) for the news.

ROUGE 1 Precision	ROUGE 1 Recall	ROUGE 1 F-measure	ROUGE 2 Precision	ROUGE 2 Recall	ROUGE 2 F-measure
0.7500	0.8454	0.7949	0.6582	0.7536	0.7027

Summary generated by other web based system :

অর্থমন্ত্রী আবুল মাল আবদুল মুহিত এবার নির্বাচন করছেন না। গতকাল অর্থমন্ত্রীর কাছে নির্বাচনে সমান সুযোগের দাবি তুললেন ওই আসনের বিএনপির সম্ভাব্য প্রার্থী ইনাম আহমদ চৌধুরী। তাঁর আসনে (সিলেট-১) প্রার্থী হয়েছেন তাঁর ছোট ভাই এ কে আবদুল মোমেন। গতকাল বৃহস্পতিবার সকালে সিলেট শহরে অর্থমন্ত্রীর বাসায় গিয়ে তাঁর সঙ্গে সাক্ষাৎ করেন ইনাম আহমদ চৌধুরী। দলীয় নেতা-কর্মীদের ধরপাকড় এবং বাড়ি বাড়ি পুলিশি তল্লাশিরও প্রতিবাদ জানান তিনি

Table A.20: ROUGE 1 and ROUGE 2 Score of other web based system generated summary for the news.

ROUGE 1 Precision	ROUGE 1 Recall	ROUGE 1 F- measure	ROUGE 2 Precision	ROUGE 2 Recall	ROUGE 2 F- measure
0.8113	0.3909	0.5276	0.7667	0.3333	0.4646

Appendix B

Resources and Data sets

We have uploaded the resources and data sets on GitHub [[13](#)].

Generated using Undegraduate Thesis \LaTeX Template, Version 2.0. Department of
Computer Science and Engineering, Ahsanullah University of Science and Technology,
Dhaka, Bangladesh.

This thesis was generated on Saturday 25th January, 2020 at 11:29am.