AMERICAN INTERNATIONAL UNIVERSITY-BANGLADESH

Final Project

Assignment Title:	Text Analysis and Topic Modeling			
Assignment No:			Date of Submission:	30 May 2025
Course Title:	Introduction to Data Science			
Course Code:			Section:	G
Semester:	Spring	2024-25	Course Teacher:	DR. ASHRAF UDDIN

Declaration and Statement of Authorship:

1. I/we hold a copy of this Assignment/Case-Study, which can be produced if the original is lost/damaged.
2. This Assignment/Case-Study is my/our original work and no part of it has been copied from any other student's work or from any other source except where due acknowledgement is made.
3. No part of this Assignment/Case-Study has been written for me/us by any other person except where such collaboration has been authorized by the concerned teacher and is clearly acknowledged in the assignment.
4. I/we have not previously submitted or currently submitting this work for any other course/unit.
5. This work may be reproduced, communicated, compared and archived for the purpose of detecting plagiarism.
6. I/we give permission for a copy of my/our marked work to be retained by the faculty for review and comparison, including review by external examiners.
7. I/we understand that Plagiarism is the presentation of the work, idea or creation of another person as though it is your own. It is a formofcheatingandisaveryseriousacademicoffencethatmayleadtoexpulsionfromtheUniversity. Plagiarized material can be drawn from, and presented in, written, graphic and visual form, including electronic data, and oral presentations. Plagiarism occurs when the origin of their arterial used is not appropriately cited.
8. I/we also understand that enabling plagiarism is the act of assisting or allowing another person to plagiarize or to copy my/our work.

** Student(s) must complete all details except the faculty use part.*
** Please submit all assignments to your course teacher or the office of the teacher concerned.

Group Name/No.: 07

No	Name	ID	Program	Contribution
1	ARSHAD ABEDIN ABIR	22-47188-1	BSc [CSE]	Text Preprocessing
2	KAMRUZZAMAN SONY	22-46797-1	BSc [CSE]	Topic Modeling
3	MRIDUL KANTI KUNDU	22-47182-1	BSc [CSE]	Exploratory Text Analysis
4	MAYSHA MARIUM	22-47197-1	BSc [CSE]	Web Scraping
5			Choose an item.	
6			Choose an item.	
7			Choose an item.	
8			Choose an item.	
9			Choose an item.	
10			Choose an item.	

Faculty use only		
FACULTY COMMENTS	**Marks Obtained**	
	Total Marks	

1.Introduction

The purpose of this project is to analyze real-world news articles through web scraping, text preprocessing, visual analytics, and topic modeling. We have extracted news content from the Bengali news portal **https://www.banglanews24.com/**, which provides a rich source of current news articles in Bengal. It involves extracting news content from a Bengali news portal, cleaning and analyzing the text, creating meaningful visualizations, and uncovering hidden topics within the news dataset. This approach aims to gain valuable insights from the information and apply them to natural language data sourced from online media, thereby enhancing understanding of the news content.

2.Web Scraping

The first step of the project involves selecting a news portal, preferably a Bengali news website, to extract news articles. The web scraping process entails collecting news article texts from multiple pages linked on the homepage of the portal. After successfully scraping the data, the extracted content will be saved into a CSV file. This file will contain at least two essential columns: the URL of the news article and the corresponding article text. This structured data will then serve as the basis for further preprocessing and analysis.

Data Preview:

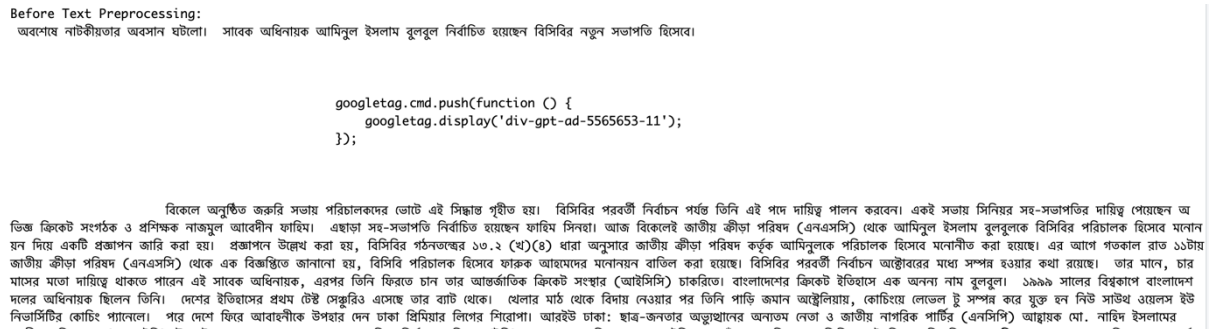
url (character)	article_text (character)
https://www.banglanews24.com/banglanews-special...	রাজধানীর প্রাণকেন্দ্রে নির্মল বাতাসে প্রাণভরে শ্বাস নেওয়ার জা...
https://www.banglanews24.com/politics/news/bd/15...	ঢাকা: প্রধান উপদেষ্টা অধ্যাপক ড. মুহাম্মদ ইউনূসের একটি সা...
https://www.banglanews24.com/climate-nature/new...	ঢাকা: স্থল গভীর নিম্নচাপ ধীরে ধীরে দুর্বল হলেও এর প্রভাবে এ...
https://www.banglanews24.com/national/news/bd/1...	প্রধান উপদেষ্টা অধ্যাপক ড. মুহাম্মদ ইউনূস ও জাপানের প্রধা...
https://www.banglanews24.com/environment-biodiv...	ঢাকা: গভীর নিম্নচাপ স্থলভাগে উঠে আসার পর দেশে মৌসুমে...
https://www.banglanews24.com/saradesh/news/bd/...	নোয়াখালী: বঙ্গোপসাগরে সৃষ্ট গভীর নিম্নচাপের প্রভাবে নোয়াখা...
https://www.banglanews24.com/cricket/news/bd/15...	জাতীয় ক্রীড়া পরিষদ (এনএসসি) কর্তৃক ফারুক আহমেদের প...
https://www.banglanews24.com/politics/news/bd/15...	ঢাকা: সমগ্র জাতি একটি সুষ্ঠু, সুন্দর, নিরপেক্ষ নির্বাচনের জ...
https://www.banglanews24.com/opinion/news/bd/15...	শহীদ প্রেসিডেন্ট জিয়াউর রহমান আমাদের জাতীয় ইতিহাসের ...
https://www.banglanews24.com/saradesh/news/bd/...	ভোলা: নিম্নচাপের প্রভাবে ভোলায় নদ-নদীর পানি অস্বাভাবিক...
https://www.banglanews24.com/saradesh/news/bd/...	দিনাজপুর: দিনাজপুরের পার্বতীপুরে ট্রাক আটকে চাঁদাবাজির স...
https://www.banglanews24.com/daily-chittagong/ne...	চট্টগ্রাম: জোড়া লাগানো যমজ শিশু রিয়াশাদ ও রেনিশকে জ...
https://www.banglanews24.com/politics/news/bd/15...	ঢাকা: মহান স্বাধীনতার ঘোষক ও বিএনপির প্রতিষ্ঠিতা শহীদ প্রে...

Previewing first 50 entries.

Figure 01: CSV Preview of Scraped News Articles

3.Text Preprocessing

Text preprocessing is a fundamental step to preparing raw text data for meaningful analysis. It involves cleaning and transforming the text to reduce noise, standardizing the format, and converting it into a suitable form for modeling. The main steps include cleaning the text, removing stop words, and tokenization.


Before Text Preprocessing:
অবশেষে নাটকীয়তার অবসান ঘটলো। সাবেক অধিনায়ক আমিনুল ইসলাম বুলবুল নির্বাচিত হয়েছেন বিসিবির নতুন সভাপতি হিসেবে।

```
googletag.cmd.push(function () {
    googletag.display('div-gpt-ad-5565653-11');
});
```

বিকেলে অনুষ্ঠিত জরুরি সভায় পরিচালকদের ভোটে এই সিদ্ধান্ত গৃহীত হয়। বিসিবির পরবর্তী নির্বাচন পর্যন্ত তিনি এই পদে দায়িত্ব পালন করবেন। একই সভায় সিনিয়র সহ-সভাপতির দায়িত্ব পেয়েছেন অ ভিজ্ঞ ক্রিকেট সংগঠক ও প্রশিক্ষক নাজমুল আবেদীন ফাহিম। এছাড়া সহ-সভাপতি নির্বাচিত হয়েছেন ফাহিম সিনহা। আজ বিকেলেই জাতীয় ক্রীড়া পরিষদ (এনএসসি) থেকে আমিনুল ইসলাম বুলবুলকে বিসিবির পরিচালক হিসেবে মনোন য়ন দিয়ে একটি প্রজ্ঞাপন জারি করা হয়। প্রজ্ঞাপনে উল্লেখ করা হয়, বিসিবির গঠনতন্ত্রের ১৩.২ (খ)(৪) ধারা অনুসারে জাতীয় ক্রীড়া পরিষদ কর্তৃক আমিনুলকে পরিচালক হিসেবে মনোনীত করা হয়েছে। এর আগে গতকাল রাত ১১টায় জাতীয় ক্রীড়া পরিষদ (এনএসসি) থেকে এক বিজ্ঞপ্তিতে জানানো হয়, বিসিবি পরিচালক হিসেবে ফারুক আহমেদের মনোনয়ন বাতিল করা হয়েছে। বিসিবির পরবর্তী নির্বাচন অক্টোবরের মধ্যে সম্পন্ন হওয়ার কথা রয়েছে। তার মানে, চার মাসের মতো দায়িত্ব থাকতে পারেন এই সাবেক অধিনায়ক, এরপর তিনি ফিরতে চান তার আন্তর্জাতিক ক্রিকেট সংস্থার (আইসিসি) চাকরিতে। বাংলাদেশের ক্রিকেট ইতিহাসে এক অনন্য নাম বুলবুল। ১৯৯৯ সালের বিশ্বকাপে বাংলাদেশ দলের অধিনায়ক ছিলেন তিনি। দেশের ইতিহাসের প্রথম টেস্ট সেঞ্চুরিও এসেছে তার ব্যাট থেকে। খেলার মাঠ থেকে বিদায় নেওয়ার পর তিনি পাড়ি জমান অস্ট্রেলিয়ায়, কোচিংয়ে লেভেল টু সম্পন্ন করে যুক্ত হন নিউ সাউথ ওয়েলস ইউ নিভার্সিটির কোচিং প্যানেলে। পরে দেশে ফিরে আবাহনীকে উপহার দেন ঢাকা প্রিমিয়ার লিগের সিরোপা। আরইউ ঢাকা: ছাত্র-জনতার অভ্যুত্থানের অন্যতম নেতা ও জাতীয় নাগরিক পার্টির (এনসিপি) আহ্বায়ক মো. নাহিদ ইসলামের জাতীয় পরিচয়পত্র (এনআইডি) উপদেষ্টা পদে থাকাবস্থায় লক করে রেখেছিল নির্বাচন কমিশন (ইসি)। সূত্রগুলো জানিয়েছে, এনআইডি তথ্য ফাঁসের অভিযোগের ভিত্তিতে ওই সিদ্ধান্ত নিয়েছিল সংস্থাটি। ২০২৪ সালে ৩৬ দিনের অভূতপূর্ব আন্দোলনের মুখে প্রধানমন্ত্রীর পদ থেকে শেখ হাসিনা পদত্যাগ করে দেশ ছেড়ে ৫ আগস্ট ভারতে পালিয়ে গেলে যে অন্তর্বর্তী সরকার গঠন হয়, সেখানে তথ্য ও সম্প্রচার মন্ত্রণালয় এবং ডাক ও টেলিযোগাযোগ মন্ত্রণালয়ের উপদেষ্টা পদে আসীন হন নাহিদ। আর সেই পদে থাকাকালীনই তার বিরুদ্ধে একটি অনলাইন প্লাটফর্মে এনআইডি তথ্য ফাঁস করার অভিযোগ আসে। জানা গেছে, ন্যাশনাল টেলিকমিউনিকেশন মনিটরিং সেন্টারের (এনটিএমসি) তৎকালীন মহাপরিচাল ক মৌখিকভাবে মো. নাহিদ ইসলামের বিরুদ্ধে 'ভন্ড বাবা' নামে এক হোয়াটসঅ্যাপ গ্রুপের মাধ্যমে নাগরিকদের তথ্য সংগ্রহ করে বাইরে পাচারের অভিযোগ আনেন। সেই অভিযোগ আমলে নিয়ে তদন্ত করার সিদ্ধান্ত নেয় ইসি। তবে সহ কারী প্রোগ্রামার আমিনুল ইসলামের নেতৃত্বে গঠিত দুই সদস্যের তদন্ত কমিটি প্রতিবেদন দেওয়ার আগেই নাহিদের এনআইডি লক করার সুপারিশ করেন সিস্টেম অ্যানালিষ্ট মোহাম্মদ আরিফুল ইসলাম। সেই সুপারিশের ভিত্তিতে গত বছরের ১৭ সেপ্টেম্বর এনআইডি লক করার নির্দেশ দেন তৎকালীন জাতীয় পরিচয় নিবন্ধন (এনআইডি) অনুবিভাগের মহাপরিচালক মো. মাহবুব আলম তালুকদার। এদিকে তদন্ত কমিটি দুদিন পর অর্থাৎ ২০২৪ সালের ১৯ সেপ্টেম্বর প্রতিবেদন দাখিল করে। এতে উল্লেখ করা হয়, এনটিএমসি-এর মহাপরিচালক কর্তৃক মৌখিকভাবে অভিযোগ পাওয়া যায় যে, মো. নাহিদ ইসলাম 'ভন্ড বাবা' নামক হোয়াটসঅ্যাপ গ্রুপের এডমিন হিসেবে কাজ করে নাগরিকদের তথ্যাদি বাহিরে সরবরাহ ক রেন। তদন্ত চলাকালীন সময়ে অনলাইনের বিভিন্ন সোশ্যাল মিডিয়ার গ্রুপ যাচাই বাছাই করে দেখা যায় যে, 'ভন্ড বাবা' গ্রুপটি হোয়াটসঅ্যাপ-এর কোনো গ্রুপ নয়, বরং এটি টেলিগ্রামের একটি গ্রুপ। ওই গ্রুপের অ্যাডমিন তিনি নন। তিনি এই গ্রুপ হতে ডাটা সংগ্রহ করে অন্য কোথাও সরবরাহ করেন এমন অভিযোগটি সত্য নয় এবং ডাটা সরবরাহ করার বিষয়ে তার কোনো সম্পৃক্ততা পাওয়া যায়নি। প্রতিবেদনে বলা হয়, যেহেতু এই ভোটার কর্তৃক ডাটা সরবরাহ করার বিষয়ে কোনো সম্পৃক্ততা পাওয়া যায়নি এবং অভিযোগটি মিথ্যা প্রমাণিত হয়েছে, সেহেতু মো. নাহিদ ইসলামের এনআইডি আনলক করার জন্য সুপারিশ করে কমিটি। তাদের প্রতিবেদনের ভিত্তিতে নথি উত্থাপন করা হলে ২২ সেপ্টেম্বর এন আইডি মহাপরিচালক তখন এনআইডিটি আনলক করার সিদ্ধান্ত দেন। এভাবেই পাঁচ দিনের জন্য লক থাকে সাবেক উপদেষ্টা নাহিদ ইসলামের এনআইডি। ইসি কর্মকর্তারা জানিয়েছেন, এনটিএমসি থেকে এনআইডি নম্বর পাঠিয়ে সংশ্লিষ্ট ভোটারের বিরুদ্ধে তথ্য পাচারের অভিযোগ তোলা হয়। আর সেই এনআইডি যে উপদেষ্টা নাহিদ ইসলামের সংশ্লিষ্টরা সেটা বুঝতেই পারেননি। বরং লক করার পর বোধোদয় হয়েছে যে এটা কার এনআইডি। পরবর্তীতে দ্রুতই আবার সেটা আনলক করা হয়। এই কাজের সঙ্গে যুক্ত দায়িত্বশীল একজন কর্মকর্তা বলেন, বিষয়টি সঠিক যে একজন ছাত্র উপদেষ্টার এনআইডি তখন লক করা হয়েছিল। তবে খুব দ্রুত সময়ের ব্যবধানে আবার আনলক করা হয়। ৫ আগষ্ট আও য়ামী লীগ সরকার পতনের চার মাস আগে ১ এপ্রিল বেসামরিক বিমান চলাচল কর্তৃপক্ষের সদস্য অতিরিক্ত সচিব মাহবুব আলম তালুকদারকে এনআইডি মহাপরিচালক হিসেবে পদায়ন করা হয়। রাজনৈতিক পট পরিবর্তনের পর ৬ নভেম্বর বিশেষ ভারপ্রাপ্ত কর্মকর্তা হিসেবে তাকে সরিয়ে নেওয়া হলেও বর্তমানে তিনি চট্টগ্রাম পোর্ট অথরিটির সদস্য হিসেবে দায়িত্ব পালন করছেন। নাহিদ ইসলামের এনআইডি লক করার বিষয়ে তিনি বাংলানিউজকে বলেন, যখন কোনো অভিযোগ আসে তখন তদন্তকালীন এনআইডি লক করা হয়। এটা একটা প্রসিডিউর। যদি তদন্তে অভিযোগ প্রমাণিত না হয়, তখন আবার আনলক করা হয়। এটা প্রক্রিয়া। তো সেই ছাত্রনেতার ক্ষেত্রেও এমনটি হয়েছিল। হয়তো কোনো এ জেন্সি থেকে অভিযোগ এসেছিল। তবে সেটা প্রমাণিত হয়নি। তাই কয়েকদিনের জন্য এনআইডি লক ছিল। ইসি কর্মকর্তারা জানিয়েছেন, সাধারণত কারও এনআইডি লক করা হলে কমিশনের সিদ্ধান্ত নেওয়া হয়। সরকার পরিবর্তনের ঠিক এক মাসের মাথায় গত বছরের ৫ সেপ্টেম্বর কাজী হাবিবুল আউয়ালের নেতৃত্বাধীন কমিশন পদত্যাগ করেন। এক্ষেত্রে কমিশন শূন্য পরিস্থিতিতে এনআইডি মহাপরিচালক ওই সিদ্ধান্ত নেন। বর্তমান এনআইডি মহাপরিচালক এসএম হুমায়ুন কবীর বলেন, সে সময় আমি ছিলাম না। তাই সেটা আমার বিবেচনার বিষয় নয়। আর পুরোনো বিষয় যেটার সঙ্গে আমার কোনো সম্পৃক্ততা নেই, সেটার মতামতও দিতে চাই না। এনআইডি লক করা হলে সংশ্লিষ্ট ব্যক্তির আর নাগরিক সেবা মেলে না। সম্প্রতি শেখ হাসিনা ও পরিবারের ১০ সদস্য এবং সাবেক এনআইডি মহাপরিচালক সুলতানুজ্জামান মো. সালেহ উদ্দীনের এনআইডিও লক করেছে ইসি। অতীতেও অনেকের এনআইডি নানা প্রেক্ষাপটে লক ক রে সংস্থাটি। ফ্যাসিবাদবিরোধী আন্দোলনের অগ্রভাগে থাকা সংগঠকদের নিয়ে নতুন দল গঠনের জন্য গত ২৫ ফেব্রুয়ারি সরকারের উপদেষ্টা পরিষদ থেকে পদত্যাগ করেন নাহিদ ইসলাম। এরপর ২৮ ফেব্রুয়ারি এনসিপির আত্মপ্রকাশ অ নুষ্ঠানে তাকে আহ্বায়ক ঘোষণা করা হয়। ইইউডি/এইচএ/ ঢাকা: দেশের ছয়টি বিভাগের ওপর ছড়িয়ে পড়েছে দক্ষিণ-পশ্চিম মৌসুমি বায়ু, অর্থাৎ বর্ষা। এর প্রভাবে এসব এলাকায় হতে পারে অতিভারী বৃষ্টি। শুক্রবার (৩০ মে) এমন পূর্বাভাস দিয়েছে আবহাওয়া অধিদফতর। আবহাওয়াবিদ ড. মুহাম্মদ আবুল কালাম মল্লিক জানান, দক্ষিণ-পশ্চিম মৌসুমি বায়ু বরিশাল, চট্টগ্রাম, সিলেট, ময়মনসিংহ, ঢাকা ও রংপুর বিভাগ পর্যন্ত অগ্রসর হয়েছে। শনিবার (৩১ মে) স ন্ধ্যা পর্যন্ত ময়মনসিংহ, ঢাকা, খুলনা, বরিশাল, চট্টগ্রাম ও সিলেট বিভাগের অধিকাংশ জায়গায় এবং রংপুর ও রাজশাহী বিভাগের অনেক জায়গায় অস্থায়ীভাবে দমকা হাওয়াসহ বিদ্যুৎ চমকানো হালকা থেকে মাঝারি ধরনের বৃষ্টি/বজ্রসহ বৃষ্টি হতে পারে। একইসঙ্গে ময়মনসিংহ, ঢাকা, খুলনা, বরিশাল, চট্টগ্রাম ও সিলেট বিভাগের কোথাও কোথাও ভারী থেকে অতিভারী বর্ষণ হতে পারে। সারাদেশে দিন ও রাতের তাপমাত্রা প্রায় অপরিবর্তিত থাকতে পারে। রোববার (১ জুন) সন্ধ্যা থেকে পরবর্তী ২৪ ঘণ্টার পূর্বাভাসে বলা হয়েছে, বরিশাল, চট্টগ্রাম ও সিলেট বিভাগের অধিকাংশ জায়গায়, ময়মনসিংহ ও ঢাকা বিভাগের অনেক জায়গায় এবং রংপুর, রাজশাহী ও খুলনা বিভাগের কিছু কিছু জায়গায় অ স্থায়ীভাবে দমকা হাওয়াসহ বিদ্যুৎ চমকানো হালকা থেকে মাঝারি ধরনের বৃষ্টি/বজ্রসহ বৃষ্টি হতে পারে। সেইসঙ্গে বরিশাল, চট্টগ্রাম ও সিলেট বিভাগে মাঝারি থেকে ভারী বর্ষণ হতে পারে। সারাদেশে দিনের তাপমাত্রা ১-৩ ডিগ্রি সেলসিয়াস

Figure 02: Raw Bengali Text Before Preprocessing

3.1 Clean the Text

The cleaning step removes unnecessary characters and standardizes the text. The following actions are performed:

- **Before Clean Text:**

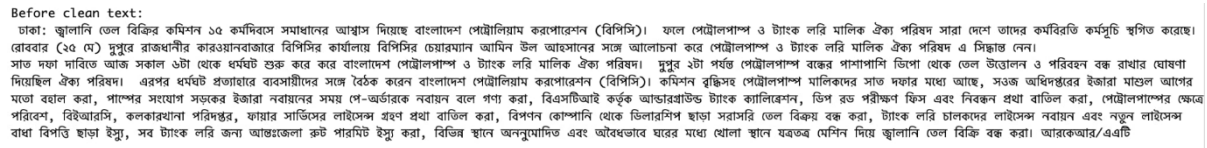

Before clean text:
ঢাকা: জ্বালানি তেল বিক্রির কমিশন ১৫ কর্মদিবসে সমাধানের আশ্বাস দিয়েছে বাংলাদেশ পেট্রোলিয়াম করপোরেশন (বিপিসি)। ফলে পেট্রোলপাম্প ও ট্যাংক লরি মালিক ঐক্য পরিষদ সারা দেশে তাদের কর্মবিরতি কর্মসূচি স্থগিত করেছে। রোববার (২৫ মে) দুপুরে রাজধানীর কারওয়ানবাজারে বিপিসির কার্যালয়ে বিপিসির চেয়ারম্যান আমিন উল আহসানের সঙ্গে আলোচনা করে পেট্রোলপাম্প ও ট্যাংক লরি মালিক ঐক্য পরিষদ এ সিদ্ধান্ত নেন।
সাত দফা দাবিতে আজ সকাল ৬টা থেকে ধর্মঘট শুরু করে করে বাংলাদেশ পেট্রোলপাম্প ও ট্যাংক লরি মালিক ঐক্য পরিষদ। দুপুর ২টা পর্যন্ত পেট্রোলপাম্প বন্ধের পাশাপাশি ডিপো থেকে তেল উত্তোলন ও পরিবহন বন্ধ রাখার ঘোষণা দিয়েছিল ঐক্য পরিষদ। এরপর ধর্মঘট প্রত্যাহারে বাবসায়ীদের সঙ্গে বৈঠক করেন বাংলাদেশ পেট্রোলিয়াম করপোরেশন (বিপিসি)। কমিশন বৃদ্ধিসহ পেট্রোলপাম্প মালিকদের সাত দফার মধ্যে আছে, সউজ অধিদপ্তরের ইজারা মাশুল আগের মতো বহাল করা, পাম্পের সংযোগ সড়কের ইজারা নবায়নের সময় পে-অর্ডারকে নবায়ন বলে গণ্য করা, বিএসটিআই কর্তৃক আন্ডারগ্রাউন্ড ট্যাংক ক্যালিব্রেশন, ডিপ রড পরীক্ষণ ফিস এবং নিবন্ধন প্রথা বাতিল করা, পেট্রোলপাম্পের ক্ষেত্রে পরিবেশ, বিইআরসি, কলকারখানা পরিদপ্তর, ফায়ার সার্ভিসের লাইসেন্স গ্রহণ প্রথা বাতিল করা, বিপণন কোম্পানি থেকে ডিলারশিপ ছাড়া সরাসরি তেল বিক্রয় বন্ধ করা, ট্যাংক লরি চালকদের লাইসেন্স নবায়ন এবং নতুন লাইসেন্স বাধা বিপত্তি ছাড়া ইস্যু, সব ট্যাংক লরি জন্য আন্তঃজেলা রুট পারমিট ইস্যু করা, বিভিন্ন স্থানে অননুমোদিত এবং অবৈধভাবে ঘরের মধ্যে খোলা স্থানে যত্রতত্র মেশিন দিয়ে জ্বালানি তেল বিক্রি বন্ধ করা। আরকেআর/এএটি

Figure 03: Bengali Text Before Cleaning Text

- **Remove non-Bengali characters:** Any character that is not a Bengali letter or whitespace is removed.

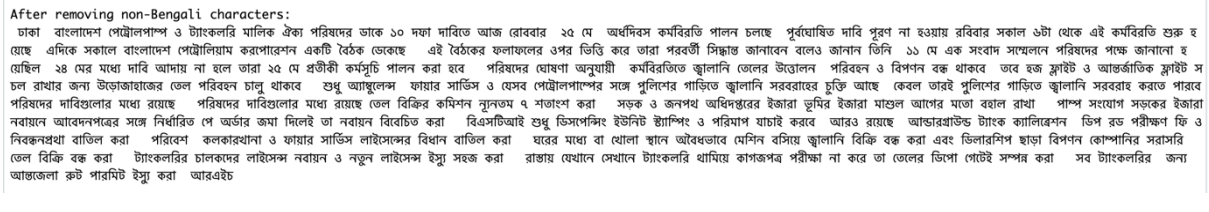

After removing non-Bengali characters:
ঢাকা বাংলাদেশ পেট্রোলপাম্প ও ট্যাংকলরি মালিক ঐক্য পরিষদের ডাকে ১০ দফা দাবিতে আজ রোববার ২৫ মে অর্ধদিবস কর্মবিরতি পালন চলছে পূর্বঘোষিত দাবি পূরণ না হওয়ায় রবিবার সকাল ৬টা থেকে এই কর্মবিরতি শুরু হ য়েছে এদিকে সকালে বাংলাদেশ পেট্রোলিয়াম করপোরেশন একটি বৈঠক ডেকেছে এই বৈঠকের ফলাফলের ওপর ভিত্তি করে তারা পরবর্তী সিদ্ধান্ত জানাবেন বলেও জানান তিনি ১১ মে এক সংবাদ সম্মেলনে পরিষদের পক্ষ জানানো হ য়েছিল ২৪ মের মধ্যে দাবি আদায় না হলে তারা ২৫ মে প্রতীকী কর্মসূচি পালন করা হবে পরিষদের ঘোষণা অনুযায়ী কর্মবিরতিতে জ্বালানি তেলের উত্তোলন পরিবহন ও বিপণন বন্ধ থাকবে তবে হজ ফ্লাইট ও আন্তর্জাতিক ফ্লাইট স চল রাখার জন্য উড়োজাহাজের তেল পরিবহন চালু থাকবে শুধু অ্যাম্বুলেন্স ফায়ার সার্ভিস ও যেসব পেট্রোলপাম্পের সঙ্গে পুলিশের গাড়িতে জ্বালানি সরবরাহের চুক্তি আছে কেবল তারই পুলিশের গাড়িতে জ্বালানি সরবরাহ করতে পারবে পরিষদের দাবিগুলোর মধ্যে রয়েছে পরিষদের দাবিগুলোর মাধ্য রয়েছে তেল বিক্রির কমিশন ন্যূনতম ৭ শতাংশ করা সড়ক ও জনপথ অধিদপ্তরের ইজারা ভূমির ইজারা মাশুল আগের মতো বহাল রাখা পাম্প সংযোগ সড়কের ইজারা নবায়নে আবেদনপত্রের সঙ্গে নির্ধারিত পে অর্ডার জমা দিলেই তা নবায়ন বিবেচিত করা বিএসটিআই শুধু ডিসপেন্সিং ইউনিট স্ট্যাম্পিং ও পরিমাপ যাচাই করবে আরও রয়েছে আন্ডারগ্রাউন্ড ট্যাংক ক্যালিব্রেশন ডিপ রড পরীক্ষণ ফি ও নিবন্ধনপ্রথা বাতিল করা পরিবেশ কলকারখানা ও ফায়ার সার্ভিস লাইসেন্সের বিধান বাতিল করা ঘরের মধ্য বা খোলা স্থানে অবৈধভাবে মেশিন বসিয়ে জ্বালানি বিক্রি বন্ধ করা এবং ডিলারশিপ ছাড়া বিপণন কোম্পানির সরাসরি তেল বিক্রি বন্ধ করা ট্যাংকলরির চালকদের লাইসেন্স নবায়ন ও নতুন লাইসেন্স ইস্যু সহজ করা রাস্তায় যেখানে সেখানে ট্যাংকলরি থামিয়ে কাগজপত্র পরীক্ষা না করে তা তেলের ডিপো গেটেই সম্পন্ন করা সব ট্যাংকলরির জন্য আন্তজেলা রুট পারমিট ইস্যু করা আরএইচ

Figure 04: Bengali Text After Remove Non-Bengali Characters

- **Remove punctuation marks:** All punctuation symbols are eliminated.

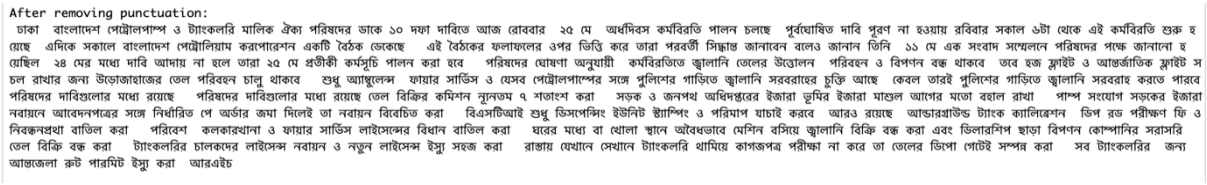


Figure 05: Bengali Text After Remove punctuation marks

- **Remove Bengali digits:** Bengali numerals (০ to ৯) are removed.

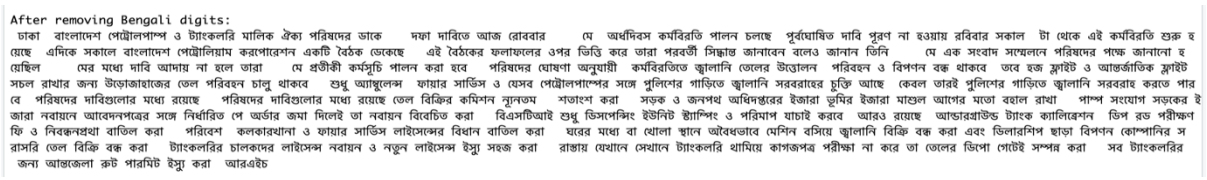


Figure 06: Bengali Text After Remove Bengali digits

- **Remove short words:** Words with lengths between 1 to 3 characters are removed to reduce noise.

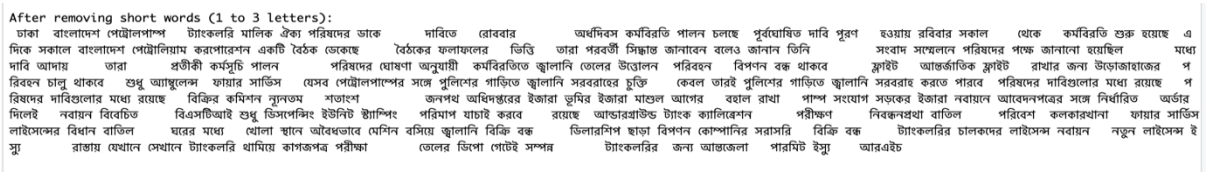
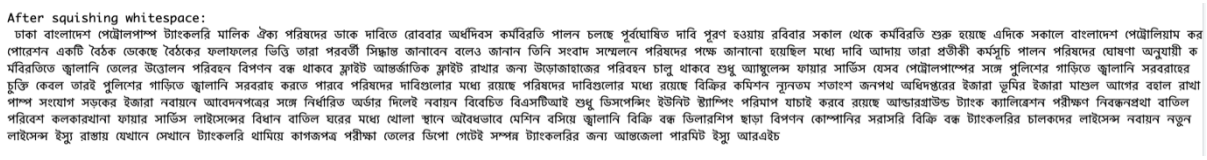
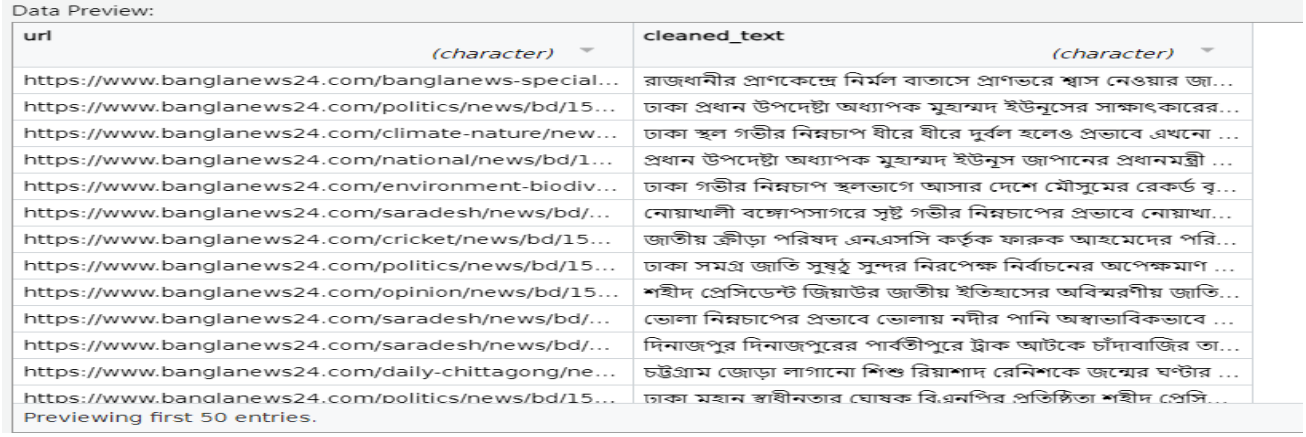


Figure 07: Bengali Text After Remove short words

- **Squishing whitespace:** Multiple spaces are replaced by a single space, and leading/trailing spaces are trimmed.




Data Preview:

url (character)	cleaned_text (character)
https://www.banglanews24.com/banglanews-special...	রাজধানীর প্রাণকেন্দ্রে নির্মল বাতাসে প্রাণভরে শ্বাস নেওয়ার জা...
https://www.banglanews24.com/politics/news/bd/15...	ঢাকা প্রধান উপদেষ্টা অধ্যাপক মুহাম্মদ ইউনূসের সাক্ষাৎকারের...
https://www.banglanews24.com/climate-nature/new...	ঢাকা স্থল গভীর নিম্নচাপ ধীরে ধীরে দুর্বল হলেও প্রভাবে এখনো ...
https://www.banglanews24.com/national/news/bd/1...	প্রধান উপদেষ্টা অধ্যাপক মুহাম্মদ ইউনূস জাপানের প্রধানমন্ত্রী ...
https://www.banglanews24.com/environment-biodiv...	ঢাকা গভীর নিম্নচাপ স্থলভাগে আসার দেশে মৌসুমের রেকর্ড বৃ...
https://www.banglanews24.com/saradesh/news/bd/...	নোয়াখালী বঙ্গোপসাগরে সৃষ্ট গভীর নিম্নচাপের প্রভাবে নোয়াখা...
https://www.banglanews24.com/cricket/news/bd/15...	জাতীয় ক্রীড়া পরিষদ এনএসসি কর্তৃক ফারুক আহমেদের পরি...
https://www.banglanews24.com/politics/news/bd/15...	ঢাকা সমগ্র জাতি সুষ্ঠু সুন্দর নিরপেক্ষ নির্বাচনের অপেক্ষমাণ ...
https://www.banglanews24.com/opinion/news/bd/15...	শহীদ প্রেসিডেন্ট জিয়াউর জাতীয় ইতিহাসের অবিস্মরণীয় জাতি...
https://www.banglanews24.com/saradesh/news/bd/...	ভোলা নিম্নচাপের প্রভাবে ভোলায় নদীর পানি অস্বাভাবিকভাবে ...
https://www.banglanews24.com/saradesh/news/bd/...	দিনাজপুর দিনাজপুরের পার্বতীপুরে ট্রাক আটকে চাঁদাবাজির তা...
https://www.banglanews24.com/daily-chittagong/ne...	চট্টগ্রাম জোড়া লাগানো শিশু রিয়াশাদ রেনিশকে জন্মের ঘণ্টার ...
https://www.banglanews24.com/politics/news/bd/15...	ঢাকা মহান স্বাধীনতার ঘোষক বি.এনপির প্রতিষ্ঠিতা শহীদ প্রেসি...

Previewing first 50 entries.

Figure 09: CSV Preview of Cleaned Bengali News Text

This process helps reduce irrelevant content and prepares the text for deeper analysis.

3.2 Remove Stop Words

A custom list of commonly occurring but non-informative Bengali stop words is defined. These include words such as

```
bangla_stopwords <- c(
  "অবশ্য", "অবশ্যই", "অল্প", "অপর", "অথচ", "অথবা", "অনেক", "অন্য", "অন্যান্য", "অন্যান্যদের", "অনেকে", "অনেকেই",
  "অতএব", "অতি", "অতিরিক্ত", "আগে", "আগেও", "আগে থেকেই", "আগেই", "আগুনে", "আছে", "আজ", "আবার", "আসলে", "ই",
  "ইত্যাদি", "ইতিমধ্যে", "উনি", "উপরে", "উপেক্ষা", "ও", "ওই", "ওদিকে", "ওরা", "কত", "কখনও", "কখনো", "কখনোই",
  "কবলা", "কবে", "কবেই", "কবি", "কম", "কম্পিউটার", "কয়েক", "কয়েকজন", "কয়েকটি", "কয়েকটা", "কয়েকজনই", "কয়েকজনের",
  "কয়েকদিন", "কয়েকটা", "কারণ", "কারণে", "কারো", "কারোই", "কারোকে", "কারোকেও", "কাশ্মীর", "কাজেই", "কাজেই",
  "কাজে", "কাজে লাগানো", "কাজের", "কাজের জন্য", "কাজের সময়", "কাজের কাজে", "কাজেই", "কাজে লাগানো", "কাজের জন্য",
  "কাজে লাগানো", "কাজেই", "কাজের জন্য", "কাজের কাজ", "কাজের কাজেই", "কাজের জন্য", "কাজের জন্যই", "কাজের জন্যে", "কাজের সময়",
  "কারও", "কারো", "কারোই", "কারোকে", "কারোকেও", "কারণ", "কারণে", "কখনও", "কখনো", "কখনোই", "কখনোও", "কখনোওই","অবশেষে",
  "বাংলাদেশের","নিয়ে","পর","রহমান","করতে","মে","কোনো","র","দেশের","আমাদের","যে","পারে","মধ্যে","হিবে","টাকা",
  "এর","এবং","না","থেকে","তিনি","করা","এ","হয়","তার","বলেন","কিন্তু", "তবে", "হয়েছে", "ছিল", "যা", "সে",
  "এই", "হয়", "জন্য", "না", "ওরে", "তুমি", "আমি", "তারা", "আমরা", "আপনি", "তাহলে", "করেন", "হয়েছে", "ড"
)
```

Figure 10: Custom Bengali Stop Words List

among others. Removing these stop words from the cleaned text helps to focus the analysis on meaningful terms that contribute more significantly to the text's content.


After removing stop words:
নাটকীয়তার অবসান ঘটলো সাবেক অধিনায়ক আমিনুল ইসলাম বুলবুল নির্বাচিত হয়েছেন বিসিবির নতুন সভাপতি হিসেবে বিকেলে অনুষ্ঠিত জরুরি সভায় পরিচালকদের ভোটে সিদ্ধান্ত গৃহীত বিসিবির পরবর্তী নির্বাচন পর্যন্ত দায়িত্ব পালন করবেন সভায় সিনিয়র সভাপতির দায়িত্ব পেয়েছেন অভিজ্ঞ ক্রিকেট সংগঠক প্রশিক্ষক নাজমুল আবেদীন ফাহিম এছাড়া সভাপতি নির্বাচিত হয়েছেন ফাহিম সিনহা বিকেলেই জাতীয় ক্রীড়া পরিষদ এনএসসি আমিনুল ইসলাম বুলবুলকে বিসিবির পরিচালক হিসেবে মনোনয়ন দিয়ে প্রজ্ঞাপন জারি প্রজ্ঞাপনে উল্লেখ বিসিবির গঠনতন্ত্রের ধারা অনুসারে জাতীয় ক্রীড়া পরিষদ কর্তৃক আমিনুলকে পরিচালক হিসেবে মনোনীত গতকাল জাতীয় ক্রীড়া পরিষদ এনএসসি বিজ্ঞপ্তিতে জানানো বিসিবি পরিচালক হিসেবে ফারুক আহমেদের মনোনয়ন বাতিল বিসিবির পরবর্তী নির্বাচন আক্টোবরের সম্পন্ন হওয়ার রয়েছে মানে মাসের দায়িত্বে থাকতে পারেন সাবেক অধিনায়ক এরপর ফিরতে আন্তর্জাতিক ক্রিকেট সংস্থার আইসিসি চাকরিতে ক্রিকেট ইতিহাসে অনন্য বুলবুল সালের বিশ্বকাপে দলের অধিনায়ক ছিলেন ইতিহাসের প্রথম টেস্ট সেঞ্চুরিও এসেছে ব্যাট খেলার বিদায় নেওয়ার পাড়ি জমান আন্টেলিয়ায় কোচিংয়ে লেভেল সম্পন্ন যুক্ত সাউথ ওয়েলস ইউনিভার্সিটির কোচিং প্যানেলে দেশে ফিরে আবাহনীকে উপহার ঢাকা প্রিমিয়ার লিগের শিরোপা আরইউ চাকা ছাত্র জনতার অভ্যুত্থানের অন্যতম নেতা জাতীয় নাগরিক পার্টির এনসিপি আহ্বায়ক নাহিদ ইসলামের জাতীয় পরিচয়পত্র এনআইডি উপদেষ্টা থাকাবস্থায় রেখেছিল নির্বাচন কমিশন সূত্রগুলো জানিয়েছে এনআইডি তথ্য ফাঁসের অভিযোগের ভিত্তিতে সিদ্ধান্ত নিয়েছিল সংস্থাটি সালে দিনের অভূতপূর্ব আন্দোলনের মুখে প্রধানমন্ত্রীর হাসিনা পদত্যাগ ছেড়ে আগষ্ট ভারতে পালিয়ে গেলে অ ন্তর্বর্তী সরকার সেখানে তথ্য সম্প্রচার মন্ত্রণালয় টেলিযোগাযোগ মন্ত্রণালয়ের উপদেষ্টা আসীন নাহিদ থাকাকালীনই বিরুদ্ধে অনলাইন প্লাটফর্মে এনআইডি তথ্য ফাঁস করার অভিযোগ জানা গেছে ন্যাশনাল টেলিকমিউনিকেশন মনিটরিং সেন্টারের এনটিএমসি তৎকালীন মহাপরিচালক মৌখিকভাবে নাহিদ ইসলামের বিরুদ্ধে ভন্ড বাবা নামে হোয়াটসঅ্যাপ গ্রুপের মাধ্যমে নাগরিকদের তথ্য সংগ্রহ বাইরে পাচারের অভিযোগ আনেন অভিযোগ আমলে তদন্ত করার সিদ্ধান্ত সহকারী প্রোগ্রামার আমিনুল ইসলামের নেতৃত্বে গঠিত সদস্যের তদন্ত কমিটি প্রতিবেদন দেওয়ার নাহিদের এনআইডি করার সুপারিশ সিস্টেম অ্যানালিস্ট মোহাম্মদ আরিফুল ইসলাম। সুপারিশের ভিত্তিতে বছরের সেপ্টেম্বর এনআইডি করার নির্দেশ তৎ কালীন জাতীয় পরিচয় নিবন্ধন এনআইডি অনুবিভাগের মহাপরিচালক মাহবুব তালুকদার এদিকে তদন্ত কমিটি দুদিন অর্থাৎ সালের সেপ্টেম্বর প্রতিবেদন দাখিল উল্লেখ এনটিএমসি মহাপরিচালক কর্তৃক মৌখিকভাবে অভিযোগ পাওয়া যায় নাহিদ ইসলাম ভন্ড বাবা নামক হোয়াটসঅ্যাপ গ্রুপের এডমিন হিসেবে নাগরিকদের তথ্যাদি বাহিরে সরবরাহ তদন্ত চলাকালীন সময়ে অনলাইনের বিভিন্ন সোশ্যাল মিডিয়ার গ্রুপ যাচাই বাছাই দেখা যায় ভন্ড বাবা গ্রুপটি হোয়াটসঅ্যাপ গ্রুপ টেলিগ্রামের গ্রুপ গ্রুপের অ্যাডমিন গ্রুপ ডাটা সংগ্রহ কোথাও সরবরাহ অভিযোগটি সত্য ডাটা সরবরাহ করার বিষয়ে সম্পৃক্ততা পাওয়া যায়নি প্রতিবেদনে যেহেতু ভোটার কর্তৃক ডাটা সরবরাহ করার বিষয়ে সম্পৃক্ততা পাওয়া যায়নি অভিযোগটি মিথ্যা প্রমাণিত সেহেতু নাহিদ ইসলামের এনআইডি আনলক করার সুপারিশ কমিটি তাদের প্রতিবেদনের ভিত্তিতে উপ্তাপন সেপ্টেম্বর এনআইডি মহাপরিচালক এনআইডিটি আনলক করার সিদ্ধান্ত এভাবেই পাঁচ দিনের থাকে সাবেক উপদেষ্টা নাহিদ ইসলামের এনআইডি কর্মকর্তারা জানিয়েছেন এনটিএমসি এনআইডি নম্বর পাঠিয়ে সংশ্লিষ্ট ভোটারের বিরুদ্ধে তথ্য পাচারের অভিযোগ তোলা এনআইডি উপদেষ্টা নাহিদ ইসলামের সংশ্লিষ্টরা সেটা বুঝতেই পারেননি করার বোধোদয় এন আইডি পরবর্তীতে দ্রুতই সেটা আনলক যুক্ত দায়িত্বশীল একজন কর্মকর্তা বিষয়টি সঠিক একজন ছাত্র উপদেষ্টার এনআইডি হয়েছিল দ্রুত সময়ের ব্যবধানে আনলক আগষ্ট আওয়ামী সরকার পতনের এপ্রিল বেসামরিক বিমান চলাচল কর্তৃপ ক্ষের সদস্য সচিব মাহবুব তালুকদারকে এনআইডি মহাপরিচালক হিসেবে পদায়ন রাজনৈতিক পরিবর্তনের নভেম্বর বিশেষ ভারপ্রাপ্ত কর্মকর্তা হিসেবে তাকে সরিয়ে নেওয়া হলেও বর্তমানে চট্টগ্রাম পোর্ট অথরিটির সদস্য হিসেবে দায়িত্ব পালন কর ছেন নাহিদ ইসলামের এনআইডি করার বিষয়ে বাংলানিউজকে অভিযোগ তদন্তকালীন এনআইডি একটা প্রসিডিউর তদন্তে অভিযোগ প্রমাণিত আনলক প্রক্রিয়া ছাত্রনেতার ক্ষেত্রেও এমনটি হয়েছিল হয়তো এজেন্সি অভিযোগ এসেছিল সেটা প্র মাণিত হয়নি কয়েকদিনের এনআইডি কর্মকর্তারা জানিয়েছেন সাধারণত এনআইডি কমিশনের সিদ্ধান্ত নেওয়া সরকার পরিবর্তনের মাসের মাথায় বছরের সেপ্টেম্বর কাজী হাবিবুল আউয়ালের নেতৃত্বাধীন কমিশন পদত্যাগ এক্ষেত্রে কমিশন শূন্য প রিস্থিতিতে এনআইডি মহাপরিচালক সিদ্ধান্ত বর্তমান এনআইডি মহাপরিচালক এসএম হুমায়ুন কবীর ছিলাম সেটা আমার বিবেচনার বিষয় পুরোনো বিষয় যেটার আমার সম্পৃক্ততা সেটার মতামতও দিতে এনআইডি সংশ্লিষ্ট ব্যক্তির নাগরিক সেবা মেলে সম্প্রতি হাসিনা পরিবারের সদস্য সাবেক এনআইডি মহাপরিচালক সুলতানুজ্জামান সালেহ উদ্দীনের এনআইডিও করেছে অতীতেও অনেকের এনআইডি নানা প্রেক্ষাপটে সংস্থাটি ফ্যাসিবাদবিরোধী আন্দোলনের অগ্রভাগে থাকা সংগঠ কদের নতুন গঠনের ফেব্রুয়ারি সরকারের উপদেষ্টা পরিষদ পদত্যাগ নাহিদ ইসলাম এরপর ফেব্রুয়ারি এনসিপির আত্মপ্রকাশ অনুষ্ঠানে তাকে আহ্বায়ক ঘোষণা ইইউডি এইচএ ঢাকা ছয়টি বিভাগের ছড়িয়ে পড়েছে দক্ষিণ পশ্চিম মৌসুমি বায়ু অ র্থাৎ বর্ষা প্রভাবে এলাকায় অতিভারী বৃষ্টি শুক্রবার পূর্বাভাস দিয়েছে আবহাওয়া অধিদফতর আবহাওয়াবিদ মুহাম্মদ আবুল কালাম মল্লিক জানান দক্ষিণ পশ্চিম মৌসুমি বায়ু বরিশাল চট্টগ্রাম সিলেট ময়মনসিংহ ঢাকা রংপুর বিভাগ পর্যন্ত অগ্র সর শনিবার সন্ধ্যা পর্যন্ত ময়মনসিংহ ঢাকা খুলনা বরিশাল চট্টগ্রাম সিলেট বিভাগের অধিকাংশ জায়গায় রংপুর রাজশাহী বিভাগের জায়গায় অস্থায়ীভাবে দমকা হাওয়াসহ বিদ্যুৎ চমকানো হালকা মাঝারি ধরনের বৃষ্টি বজ্রসহ বৃষ্টি একইসঙ্গে ময় মনসিংহ ঢাকা খুলনা বরিশাল চট্টগ্রাম সিলেট বিভাগের কোথাও কোথাও ভারী অতিভারী বর্ষণ সারাদেশে রাতের তাপমাত্রা প্রায় অপরিবর্তিত থাকতে রোববার সন্ধ্যা পরবর্তী ঘণ্টার পূর্বাভাসে বরিশাল চট্টগ্রাম সিলেট বিভাগের অধিকাংশ জায় গায় ময়মনসিংহ ঢাকা বিভাগের জায়গায় রংপুর রাজশাহী খুলনা বিভাগের কিছু কিছু জায়গায় অস্থায়ীভাবে দমকা হাওয়াসহ বিদ্যুৎ চমকানো হালকা মাঝারি ধরনের বৃষ্টি বজ্রসহ বৃষ্টি সেইসঙ্গে বরিশাল চট্টগ্রাম সিলেট বিভাগে মাঝারি ভারী ব র্ষণ সারাদেশে দিনের তাপমাত্রা ডিগ্রি সেলসিয়াস বৃদ্ধি পেতে রাতের তাপমাত্রা সামান্য বাড়তে সোমবার সন্ধ্যা পরবর্তী ঘণ্টার পূর্বাভাসে ময়মনসিংহ সিলেট বিভাগের জায়গায় ঢাকা বরিশাল চট্টগ্রাম বিভাগের কিছু কিছু জায়গায় রংপুর রাজ শাহী খুলনা বিভাগের জায়গায় অস্থায়ীভাবে দমকা হাওয়াসহ বিদ্যুৎ চমকানো হালকা মাঝারি ধরনের বৃষ্টি বজ্রসহ বৃষ্টি সারাদেশে দিনের তাপমাত্রা সামান্য হ্রাস পেতে রাতের তাপ মাত্রা প্রায় অপরিবর্তিত থাকতে বর্ধিত পাঁচ দিনের আবহাওয়ায় উল্লেখযোগ্য পরিবর্তনের সম্ভাবনা ইইউডি এসআইএস চলতি পর্যন্ত যুক্তরাষ্ট্র হাজারেরও বেশি ভারতীয় দেশে ফিরেছেন তাদের যুক্তরাষ্ট্র দেশে ফেরত পাঠানো ভারতের পররাষ্ট্র মন্ত্র ণালয় এমনটি জানিয়েছে পররাষ্ট্র মন্ত্রণালয়ের মুখপাত্র রনধীর জয়সোয়াল এদের প্রায় শতাংশ বাণিজ্যিক ফ্লাইটে দেশে ফিরেছেন ফেরত এসেছেন প্রক্রিয়ায় বিষয়ে বিস্তারিত কিছু জাননানি যুক্তরাষ্ট্রে অবৈধ অভিবাসীদের বিরুদ্ধে কড়া অবস্থান নিচ্ছে ট্রাম্প প্রশাসন প্রেসিডেন্ট ডোনাল্ড ট্রাম্প একবার বলেছিলেন ভারত অবৈধ অভিবাসীদের ফিরিয়ে নেওয়ার ক্ষেত্রে সঠিক সিদ্ধান্তই নেবে চলতি বছরের ফেব্রুয়ারিতে যুক্তরাষ্ট্র সামরিক ফ্লাইটে শতাধিক ভারতীয় নাগরিককে ফেরত পাঠায় বিভিন্ন প্রতিবেদনে তাদের কয়েকজনকে বেঁধে দেশে পাঠানো হয়েছিল ভারতের পররাষ্ট্র মন্ত্রণালয়ের মুখপাত্র সাপ্তাহিক ব্রিফিংয়ে অভিবাসন ইস্যুতে ভারত যুক্তরাষ্ট্রের ঘনিষ্ঠ সহযাগিতা রয়েছে নাগরিককে ফেরত নেওয়ার জাতীয়তা যাচাই যুক্তরাষ্ট্র পর্যন্ত প্রায় হাজার ভারতীয় নাগরিকে শনাক্ত করেছে যাদের অবৈধভাবে প্রবেশকারী চলতি মাসের শুরুতে ভারত যুক্তরাষ্ট্র দূতাবাস সতর্ক জানায় যারা ভিসা নিয়ে যুক্তরাষ্ট্রে প্রবেশ করেছেন নির্ধারিত সময়ের বেশি সেখানে অবস্থান করছেন তাদের বহিষ্কার আজীবনের যুক্তরাষ্ট্রে প্রবেশে নিষেধাজ্ঞার মুখে পড়তে প্রসঙ্গে ভারতের পররাষ্ট্র মন্ত্রণালয়ের মুখপাত্র ট্রাম্প প্রশাসন শিক্ষার্থীদের ভিসা নিয়ে যেসব নতুন নীতিমালা এনেছে যুক্তরাষ্ট্রে উচ্চশিক্ষার পরিকল্পনায় থাকা ভারতীয় শিক্ষার্থীর সরাসরি প্রভাব ফেলতে আরএইচ বঙ্গোপসাগরের গভীর নিম্নচাপটি স্থলে ওঠার ধীরে ধীরে অগ্রসর দুর্বল লঘুচাপে পরিণত মেঘালয়ে অবস্থান করছে প্রভাবে একদিন অভ্যন্তর অতিভারী বৃষ্টি শুরুবার পূর্বাভাস দিয়েছে আবহাওয়া অ ফিস আবহাওয়াবিদ মুহাম্মদ আবুল কালাম মল্লিক জানিয়েছেন সাতক্ষীরা তৎসংলগ্ন এলাকায় অবস্থিত স্থল গভীর নিম্নচাপটি উত্তর উত্তরপূর্ব দিকে অগ্রসর ক্রমান্বয়ে দুর্বল হয় প্রথমে স্থল নিম্নচাপ দুর্বল হয়ে সুস্পষ্ট লঘুচাপ হিসেবে শেরপুর ত ৎসংলগ্ন মেঘালয়ে অবস্থান করেছে প্রভাবে শনিবার সন্ধ্যা পর্যন্ত রংপুর খুলনা বরিশাল ঢাকা ময়মনসিংহ চট্টগ্রাম সিলেট বিভাগের জায়গায় ভারী ঘণ্টা অতিভারী ঘণ্টা বর্ষণ অতিভারী বর্ষণের চট্টগ্রাম রাঙামাটি খাগড়াছড়ি বান্দরবন কক্স বাজার জেলাসমূহের পাহাড়ি এলাকার কোথাও কোথাও ভূমিধসের সম্ভাবনা রয়েছে সেইসঙ্গে ভারী বর্ষণজনিত ঢাকা চট্টগ্রাম মহানগরীর কোথাও কোথাও অস্থায়ীভাবে জলাবদ্ধতা তৈরি ইইউডি এসএএইচ বঙ্গোপসাগরে সৃষ্ট গভীর নিম্নচাপের প্রবাল দ্বীপ সেন্টমার্টিনে ভারী বৃষ্টি দমকা হাওয়া যাচ্ছে জোয়ারের পানি উচ্চতায় প্রবাহিত দ্বীপের চারপাশে ভাঙন দেখা দিয়েছে সেন্টমার্টিন ইউপির ভারপ্রাপ্ত চেয়ারম্যান ফয়েজুল ইসলাম বঙ্গোপসাগরে গভীর নিম্নচাপ সৃষ্টি হওয়ায় স্বাভাবিকের চেয়ে পাঁচ জোয়ারের পানি বেড়ে যাওয়ায় দ্বীপের গলাচিপা কোনাপাড়া দক্ষিণপাড়া পানিতে তলিয়ে গেছে অন্তত অর্ধশতাধিক ঘরবাড়ি পানিবন্দি পড়েছে এছাড়া ধরার ট্রলার জোয়ারের আঘাতে বিধ্বস্ত গেছে রাতে বাতাস একটু কমেছে তারপরও হোটেল মোটেলসহ আশ্রয়কেন্দ্র প্রস্তুত রেখেছি এছাড়া বিএন স্কুলকে আশ্রয়কেন্দ্র এদিকে চারদিন টেকনাফ সেন্টমার্টিন নৌযান চলাচল বন্ধ রয়েছে দ্বীপের মানুষের নিত্যপ্রয়োজনীয় পণ্যের সংকট দেখা দিয়েছে দ্বীপের মানুষের খারাপ যাচ্ছে সকাল ঝোড়ো বাতাস শুরু সাগরের পানির উচ্চতা বেড়েছে সেন্টমার্টিন দ্বীপের বাসিন্দা হাবিবুর আব্দুল মালেক জোয়ারের বঙ্গোপসাগরে পানির উচ্চতা বৃদ্ধি পেয়েছে দ্বীপের সাগরের তীরে থাকা কয়েকটি এলাকায় পানি ঢুকেছে সেন্টমার্টিনের চারদিকে ভাঙন ধ রেছে এছাড়া ঘরবাড়ি পানিবন্দি অবস্থায় টেকনাফ উপজেলা নির্বাহী কর্মকর্তা ইউএনও এহসান উদ্দিন বঙ্গোপসাগরে পানির উচ্চতা বৃদ্ধি পাওয়ায় জোয়ারের পানি সেন্টমার্টিন দ্বীপের লোকালয়ে প্রবেশ করেছে ছাড়া ঘাটে নোঙর থাকা ধরার ট্র

Figure 11: Bengali Text After Removing Bengali Stop Words

3.3 Tokenization

Tokenization is the process of splitting each cleaned document into individual words, called tokens. This allows the text to be analyzed at the word level. For example, using R's tokenize_words() function, each article's text is broken down into a list of words. These tokens are the basic units used for tasks such as frequency analysis, creating document-term matrices, and topic modeling.

[1]	"রাজধানীর"	"প্রাণকেন্দ্রে"	"নির্মল"	"বাতাসে"	"প্রাণভরে"
[6]	"শ্বাস"	"নেওয়ার"	"জায়গার"	"অভাব"	"অভাবের"
[11]	"জায়গা"	"কিছুটা"	"হলেও"	"পূরণ"	"করছিল"
[16]	"পানি"	"সবুজে"	"মেশা"	"হাতিরঝিল"	"রাস্তার"
[21]	"দুধারে"	"সবুজ"	"গাছের"	"ছায়ায়"	"দুদণ্ড"
[26]	"শান্তি"	"খুঁজতে"	"অনেকেই"	"ছুটে"	"আসেন"
[31]	"হাতির"	"ঝিলে"	"ছুটির"	"দিনে"	"ঘুরতে"
[36]	"আসেন"	"পরিবারসহ"	"প্রতিদিন"	"সন্ধ্যার"	"হাতিরঝিলের"
[41]	"কৃত্রিম"	"আলোর"	"ঝলকানিও"	"নগরের"	"মানুষের"
[46]	"কাছে"	"অন্যতম"	"আকর্ষণের"	"কেন্দ্র"	"এখানে"
[51]	"দর্শনার্থী"	"পথচারী"	"কিংবা"	"আশপাশের"	"বাসিন্দা"
[56]	"সবারই"	"চলাচল"	"চেপে"	"আয়তনের"	"হাতিরঝিল"
[61]	"এফডিসি"	"গুলশান"	"পর্যন্ত"	"সংযুক্ত"	"জলাশয়ের"
[66]	"পানি"	"দুর্গন্ধমুক্ত"	"রাখতে"	"কর্তৃপক্ষের"	"উদ্যোগ"
[71]	"থাকলেও"	"দায়িত্ব"	"পাওয়া"	"ঠিকাদারি"	"প্রতিষ্ঠানের"
[76]	"কারণে"	"সেটি"	"দৃশ্যমান"	"ঝিলের"	"পানি"
[81]	"পরিষ্কার"	"রাখতে"	"বছরের"	"ঠিকাদারি"	"প্রতিষ্ঠান"
[86]	"মেহজাবিন"	"এন্টারপ্রাইজকে"	"কোটি"	"বরাদ্দ"	"দেওয়া"
[91]	"অগ্রগতির"	"তেমন"	"প্রমাণ"	"মধুব্যাগ"	"এলাকার"
[96]	"স্থায়ী"	"বাসিন্দা"	"শাওন"	"আহমেদ"	"চলতি"
[101]	"হাতিরঝিলে"	"প্রাণভরে"	"শ্বাস"	"নেওয়ার"	"উপায়"
[106]	"সকালে"	"বাচ্চাদের"	"দিয়ে"	"হাঁটতে"	"হতাম"
[111]	"এমনই"	"অবস্থা"	"নিজের"	"ঠিকানা"	"এলাকা"
[116]	"ছেড়ে"	"যেতাম"	"প্রকট"	"দুর্গন্ধে"	"হাতিরঝিল"
[121]	"এলাকার"	"মানুষের"	"স্বাস্থ্যঝুঁকি"	"বাড়ছে"	"এখানে"
[126]	"বিশুদ্ধ"	"বাতাস"	"নির্মল"	"পরিবেশ"	"স্থানীয়"
[131]	"রফিকুল"	"নামে"	"আরেকজনের"	"হাতিরঝিল"	"প্রকল্পের"
[136]	"শুরুর"	"দিকে"	"ওয়াটার"	"ট্যাক্সিতে"	"গন্তব্যে"
[141]	"ফিরতাম"	"অন্যরকম"	"একটা"	"সতেজতা"	"অনুভব"
[146]	"চলাই"	"অস্বস্তিকর"	"উঠেছে"	"বাচ্চারা"	"সারা"
[151]	"লেকের"	"পাশে"	"বইসা"	"খেলত"	"গন্ধ"
[156]	"পানি"	"পোলাপান"	"আসতেই"	"হাতিরঝিল"	"এলাকার"
[161]	"বাসিন্দা"	"গৃহিণী"	"হাসনা"	"ঝিলের"	"ধারেই"
[166]	"ঝালমুড়ি"	"বিক্রি"	"হাসান"	"মিয়া"	"আগের"
[171]	"চেয়ে"	"মানুষ"	"হাতিরঝিলে"	"দূষণ"	"দুর্গন্ধে"
[176]	"এলাকাবাসীর"	"পাশাপাশি"	"আমরাও"	"অস্বস্তিতে"	"সালে"
[181]	"জনসাধারণের"	"উন্মুক্ত"	"হাতিরঝিল"	"প্রকল্প"	"পানি"
[186]	"পরিষ্কার"	"পরিশোধনের"	"পর্যাপ্ত"	"ব্যবস্থা"	"থাকায়"
[191]	"পানির"	"গুণগত"	"খারাপ"	"প্রকট"	"আকার"
[196]	"ধারণ"	"করেছে"	"পয়োবর্জ্য"	"ময়লা"	"আবর্জনা"
[201]	"ড্রেনের"	"পানি"	"ঢুকে"	"বিষাক্ত"	"উঠেছে"
[206]	"ঝিলের"	"পানি"	"বাতাসে"	"ভাসছে"	"উৎকট"
[211]	"গন্ধ"	"আশপাশ"	"ঘুরে"	"এমনটি"	"দেখেছেন"
[216]	"প্রতিবেদক"	"হাতিরঝিল"	"ঘুরে"	"দেখা"	"গেছে"
[221]	"ঝিলের"	"প্রায়"	"থেকেই"	"পানির"	"দুর্গন্ধ"
[226]	"ভেসে"	"আসছে"	"পানিতে"	"নানা"	"বর্জ্য"
[231]	"পলিথিন"	"প্লাস্টিকের"	"পাইপও"	"ভাসতে"	"দেখা"
[236]	"গেছে"	"কারওয়ান"	"বাজারের"	"প্যান"	"প্যাসিফিক"
[241]	"সোনারগাঁও"	"হোটেলের"	"পেছনে"	"বর্জ্যের"	"মাত্রা"
[246]	"বেশি"	"কাজী"	"নজরুল"	"ইসলাম"	"অ্যাভিনিউ"
[251]	"থেকেও"	"দুর্গন্ধ"	"পাওয়া"	"ঝিলে"	"মাছও"
[256]	"ভাসতে"	"দেখা"	"পানির"	"ওপরে"	"বর্জ্যের"
[261]	"পুরু"	"স্তর"	"তৈরি"	"ওয়াটার"	"ট্যাক্সিতে"
[266]	"চলার"	"সময়"	"স্পষ্টভাবে"	"দেখা"	"যাচ্ছিল"
[271]	"হাতির"	"ঝিলের"	"পানি"	"জীববৈচিত্র্য"	"পরিবেশের"
[276]	"হুমকি"	"বলছেন"	"পরিবেশবিদরা"	"রাজউকের"	"কর্মকর্তারা"
[281]	"বলছেন"	"প্রায়"	"নর্দমার"	"পানি"	"হাতিরঝিলের"
[286]	"পানিতে"	"মিশে"	"নিষেধাজ্ঞা"	"থাকলেও"	"মানুষ"
[291]	"আবর্জনা"	"পলিথিন"	"এমনকি"	"শিল্প"	"বর্জ্য"
[296]	"ফেলছে"	"ঝিলের"	"পানিতে"	"পানি"	"দূষিত"
[301]	"হচ্ছে"	"ময়লা"	"আবর্জনা"	"মিথেন"	"গ্যাস"
[306]	"সৃষ্টি"	"হচ্ছে"	"কারণেই"	"দুর্গন্ধ"	"ছড়িয়ে"
[311]	"পড়ছে"	"রাজউকের"	"তত্ত্বাবধায়ক"	"প্রকৌশলী"	"যান্ত্রিক"
[316]	"সাব্বির"	"তাহের"	"বাংলানিউজকে"	"হাতিরঝিলের"	"চারপাশে"
[321]	"প্রায়"	"নর্দমা"	"গর্ত"	"রয়েছে"	"প্রতি"
[326]	"মাসে"	"একবার"	"সেগুলো"	"পরিষ্কার"	"কঠিন"
[331]	"বর্জ্য"	"গর্তের"	"পেছনে"	"তারপরে"	"লেকের"
[336]	"পানিতে"	"ভেসে"	"রামপুরা"	"কাঁঠালবাগানে"	"দুটি"
[341]	"সলুইস"	"রয়েছে"	"বর্ষাকালে"	"এগুলো"	"খুলতে"
[346]	"কঠিন"	"বর্জ্য"	"প্রবেশ"	"প্রতি"	"রক্ষণাবেক্ষণের"
[351]	"বরাদ্দের"	"অংশই"	"বিদ্যুৎ"	"খাতে"	"যায়"
[356]	"পানি"	"পরিশোধনের"	"বরাদ্দ"	"রাসায়নিক"	"প্রায়"
[361]	"হয়ে"	"সালের"	"আগের"	"অবশিষ্ট"	"রাসায়নিক"
[366]	"অল্প"	"পরিমাণে"	"ব্যবহার"	"হচ্ছে"	"বছরে"
[371]	"মাত্র"	"একবার"	"সালে"	"হাতিরঝিল"	"সংগৃহীত"
[376]	"নমুনায়"	"ক্ষতিকর"	"রাসায়নিক"	"পিএফওএ"	"পারফ্লুরোঅকটানোয়িক"
[381]	"অ্যাসিড"	"পিএফওএস"	"পারফ্লুরোঅকটেনসালফোনিক"	"অ্যাসিড"	"উভয়ই"
[386]	"দীর্ঘমেয়াদি"	"বিষাক্ততার"	"দায়ী"	"পিএফওএসের"	"স্তর"
[391]	"পরামর্শমূলক"	"স্তরের"	"চেয়ে"	"বেশি"	"মাত্রায়"
[396]	"পাওয়া"	"গেছে"	"বিভিন্ন"	"গবেষণায়"	"হাতিরঝিলে"
[401]	"পানি"	"দূষণের"	"বিষয়টি"	"এসেছে"	"এরপরও"
[406]	"পানি"	"দূষণ"	"যাচ্ছে"	"বাংলানিউজের"	"বলেছে"

Figure 12: Tokenized Bengali Words

4.Exploratory Text Analysis

A Term Document Matrix (TDM) was created to analyze term frequencies across documents. Word frequencies were calculated and the most common words identified. Visualizations such as word clouds and bar plots of the top 20 words were generated to aid intuitive exploration of the data.

Data Preview:

word (character)	freq (double)
জিয়াউর	189
হিসেবে	159
ঢাকা	153
ঘোষণা	141
তাদের	134
জাতীয়	131
জিয়া	123
সালের	122
দিয়ে	121
সরকার	117
শহীদ	116
শুরু	115
বিভিন্ন	111
সালে	111
করেছেন	111
প্রধান	111
করার	109
ছিলেন	109
রহমানের	109
হচ্ছে	107
পানি	101
করেছে	96
মাধ্যমে	93
রয়েছে	92
কিছু	91
সরকারের	91
নির্বাচন	90
চট্টগ্রাম	90
স্বাধীনতার	89
রাজনৈতিক	88
কারণে	87
পর্যন্ত	86
বিএনপির	85
বৃহস্পতিবার	81
প্রথম	80
দলের	79
বিএনপি	79
মেজর	79
বৃষ্টি	78
গেছে	77
জানান	76
দেখা	76
নতুন	76
মার্চ	73
সাবেক	71
তাকে	70
কাছে	69
দেওয়া	69
দাবি	69
অনেক	66

Figure 13: CSV Preview of Bengali Words by Frequency

4.1 **Word Cloud with Most Frequent Word**

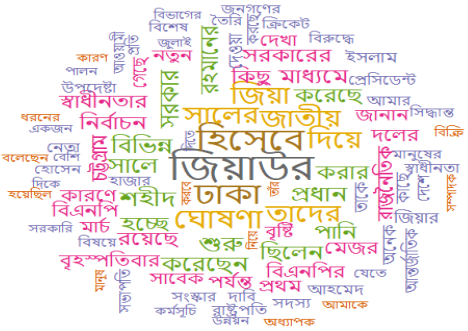


Figure 14: Word Cloud of Most Frequent Bengali Words

This word cloud provides a visual summary of word frequencies found in a collection of Bengali text documents, most likely related to news or political topics.

1. **Most Frequent Words:**
   - জিয়াউর (Ziaur): This is the largest word, showing it is the most commonly mentioned term in the text. It probably refers to "Ziaur Rahman" or the Zia political faction, indicating a political focus.
   - সরকার (Government): Indicates discussions involving government policies, activities, or official statements.
   - ঢাকা (Dhaka): The capital of Bangladesh, frequently mentioned in political news and national events.
   - ঘোষণা (Announcement): Points to frequent mentions of official declarations or statements.
2. **Political Context:**
   - Words such as নির্বাচন (Election), বিএনপি (BNP - Bangladesh Nationalist Party), সদস্য (Member), দলের (Party's) suggest that the text mainly covers political events, party actions, and elections.
   - Terms like স্বাধীনতা (Independence), শ্রদ্ধা (Respect/Honor), শহীদ (Martyr) reflect themes of nationalism and historical significance within political narratives.
3. **Geographical and Social References:**
   - The prominence of ঢাকা (Dhaka) highlights the central role of the capital in the news or discussion topics.
   - Other terms like পানি (Water) and বৃষ্টি (Rain) hint at social or environmental issues alongside political reporting.
4. **Other Frequent Terms:**
   - Names and titles like প্রধান (Chief/Head), মেয়র (Mayor), আহমেদ (Ahmed) suggest mentions of specific leaders or officials.
   - Words such as কারণ (Reason), বিরুদ্ধে (Against), প্রথম (First) indicate the presence of narrative or argumentative elements typical in news stories.

This word cloud reflects a text dataset rich in political and governmental themes, focusing on prominent figures likely connected to the Zia family or faction, the government, and national matters centered mainly on Dhaka. It also includes related topics such as elections, official announcements, social/environmental concerns, and expressions of national pride and historical remembrance.

4.2 **Bar Chart of Top 20 Words**

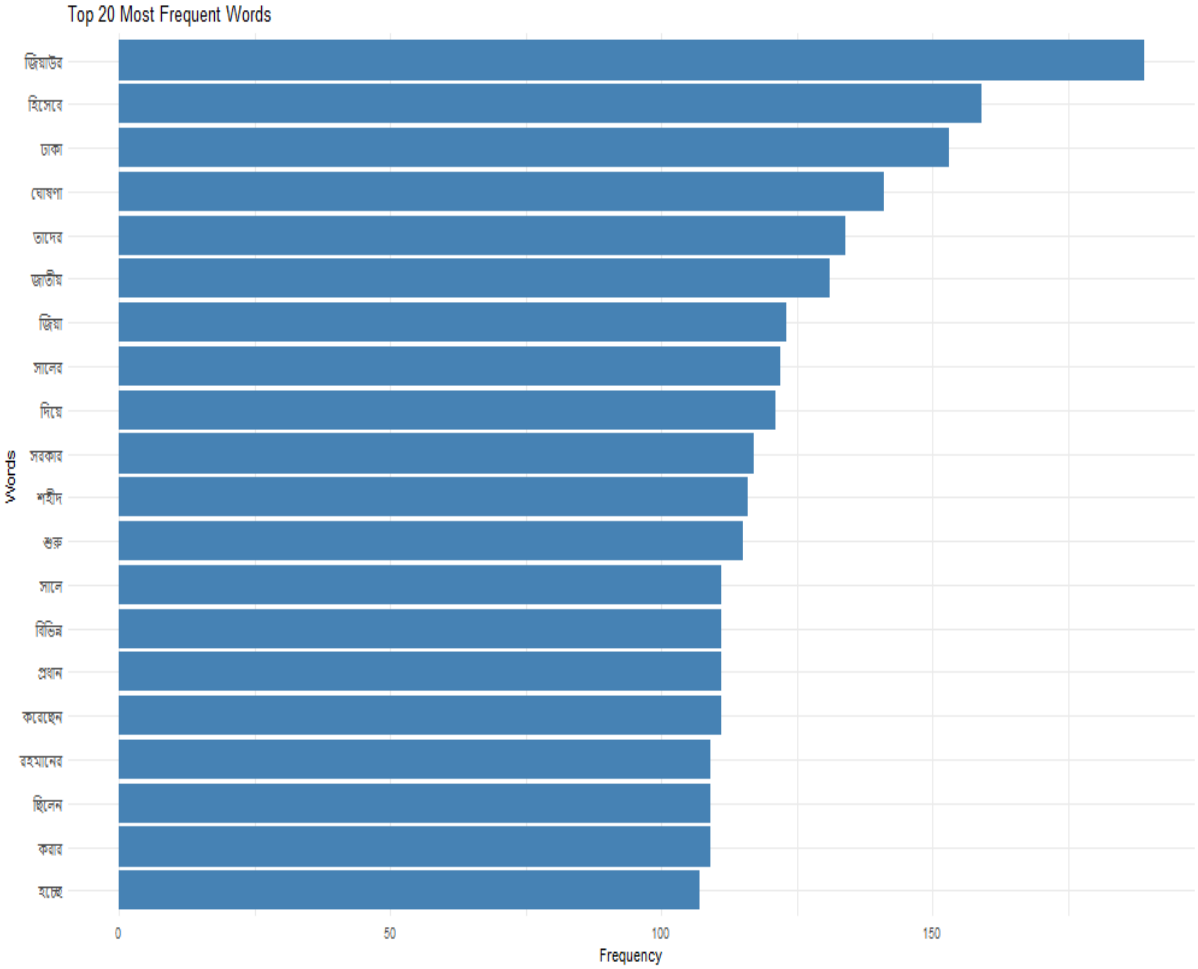


Figure 15: Bar Chart of Top 20 Most Frequent Bengali Words

This bar chart presents the top 20 most frequent Bengali words from a text corpus, likely centered on political and social topics. The word "জিয়াউর" (referring to Ziaur Rahman or his political faction) appears most frequently, followed by terms like "তারা" (they), "যেখান" (where), and "ঢাকা" (the capital city), indicating a strong focus on political figures, places, and government-related discussions. Words such as "সরকার" (government), "নির্বাচন" (election), and "বিএনপি" (BNP) further highlight themes of political events, party activities, and electoral processes. Additionally, nationalistic terms like "স্বাধীনতা" (independence) and "শ্রদ্ধা" (respect) appear alongside social and environmental references such as "পানি" (water) and "বৃষ্টি" (rain), reflecting a broader context of news coverage. The presence of official titles and narrative words like "প্রধান" (chief), "মেয়র" (mayor), "কারণ" (reason), and "বিরুদ্ধে" (against) suggests detailed reporting and argumentation within the texts. Overall, the chart offers a clear quantitative visualization that complements the word cloud, emphasizing the political and social nature of the dataset with a strong geographic focus on Dhaka.

5.Topic Modeling

Topic modeling is a technique used to automatically discover the main themes or topics present in a large collection of text documents. It helps to organize and summarize the text by grouping words that frequently appear together into meaningful topics.

5.1 Document-Term Matrix (DTM)

This step involves creating a structured matrix where rows represent documents and columns represent terms (words). Each cell in the matrix contains the frequency of a particular term in a specific document. This matrix serves as the foundational input for topic modeling algorithms, allowing them to analyze the distribution of words across documents systematically.

5.2 Latent Dirichlet Allocation (LDA)

Latent Dirichlet Allocation (LDA) is a probabilistic model used to uncover hidden thematic structures in a large collection of text documents. By applying LDA to the Document-Term Matrix, the algorithm identifies 3 to 5 main topics. Each topic is represented as a distribution over words that frequently occur together, revealing underlying themes or subjects within the dataset.

5.3 Top Words per Topic

After the topics are identified, bar plots are generated to display the top significant words for each topic. These visualizations make it easier to interpret the content of each topic by highlighting the most influential words, aiding in understanding the essence of the themes discovered.

Figure 16: Bar Charts of Top Words per Topic

Topic 1 (Red Bar) Word Analysis:

- জিয়াউর, জিয়া: These words refer to prominent political figures in Bangladesh, specifically related to Ziaur Rahman and his political legacy, indicating a political theme.
- দিয়ে: A verb indicating an action, likely used in narratives describing events or processes.
- পানি: Refers to water, suggesting environmental or natural resource topics.
- তাদের, সালে, শুরু: Words related to time and groups of people, indicating the start or timeline of certain events.
- সালেন, জাতীয়, শহীদ: Words that relate to national events or commemorations, such as martyrdom or national days.

Topic 2 (Olive Bar) Word Analysis:

- ঘোষণা: Means "announcement," indicating official statements or declarations.
- জিয়াউর, জিয়া, রহমানের: Names related to political leaders, linking this topic to political history or current politics.
- স্বাধীনতার: Refers to "independence," indicating discussions about national freedom or historical struggles.
- মেজর, মার্চ, সালের: Military rank and time indicators, likely referencing specific dates or periods.
- হিসেবে: Means "as" or "in the capacity of," showing roles or positions.

Topic 3 (Green Bar) Word Analysis:

- শনি, বৃষ্টি: Related to weather, specifically rain and days of the week (Saturday).
- ঢাকা: The capital city of Bangladesh, a central location for political, social, and environmental news.
- বৃহস্পতিবার, গেছেকারণে: Time references and causal words, indicating when and why something happened.
- বিভিন্ন, করেছে, প্রধান: Describing various events or actions with importance.

Topic 4 (Blue Bar) Word Analysis:

- বিএনপি: Refers to the Bangladesh Nationalist Party, a major political party.
- হিসেবে: Denotes roles or perspectives.
- জিয়াউর, রহমানের: Political leader names, indicating political discussions.
- শহীদ: Refers to martyrs, highlighting historical or commemorative themes.
- করেছেন, হচ্ছে: Action verbs indicating ongoing or completed activities.
- সারেক, ঢাকা: Names of places or organizations, indicating local context.

Topic 5 (Purple Bar) Word Analysis:

- সরকার, সরকারের: Refers to government and related administrative matters.
- ঢাকা: Political and administrative hub.
- ক্রিকেট: Refers to the sport, indicating social or entertainment-related topics.
- নির্বাচন: Refers to elections, pointing to democratic processes and political events.
- কিছু, ফারুক, সংস্কার, করার, জুলাই: Various terms relating to actions, individuals, reforms, and time (July).

Each topic's key words reflect important aspects of Bangladesh's political landscape, social environment, and daily life. Political figures and parties such as জিয়াউর and বিএনপি, as well as terms like সরকার and নির্বাচন, highlight the political discourse. Environmental and social themes emerge from words like পানি, বৃষ্টি, and ঢাকা, while entertainment and cultural life appear through words like ক্রিকেট. Time-related words such as মার্চ, শনি, and জুলাই emphasize temporal contexts within the data. Overall, the dataset provides a rich representation of the multifaceted realities of Bangladesh's society, politics, and environment.

5.4 Document-Topic Distribution

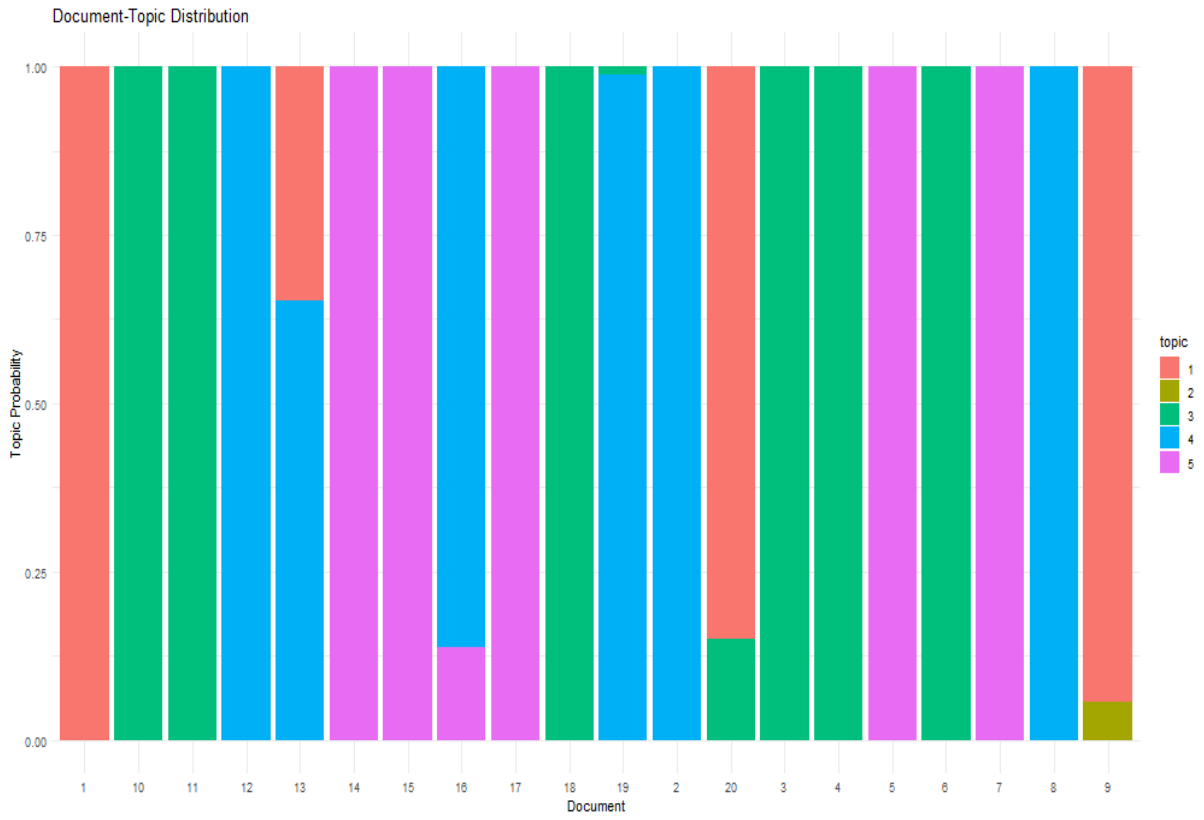


Figure 17: Document–Topic Distribution

To understand how topics are spread across individual documents, charts like stacked bar plots or pie charts are created. These visualizations show the proportion or probability of each topic within each document, illustrating the thematic composition of the documents and highlighting dominant topics.