# A Transformer Based Approach to Detect the Sentiment of Drivers in Ride Sharing Platforms

by

Sovon Chakraborty
22366023

A thesis submitted to the Department of Computer Science and Engineering
in partial fulfillment of the requirements for the degree of
M.Sc. in Computer Science

Department of Computer Science and Engineering
Brac University
April 2024

# Declaration

The Author of this research declares that

1. The work is conducted with sufficient groundwork. All the data are properly gathered. This is an original work for the completion M.Sc. in CSE thesis at Brac University.

2. The thesis does not contain material which has been accepted, or submitted, for any other degree or diploma at a university or other institution.

3. All the references and aiding persons are addressed properly. .

**Full Name of the Student & Signature:**

Sovon Chakraborty

22366023

# Approval

The thesis/project titled "A Transformer-Based Approach to Detect the Sentiment of Drivers in Ride Sharing Platforms" submitted by

1. Sovon Chakraborty (22366023)

Spring 2024 has been accepted as satisfactory in partial fulfillment of the requirement for the degree of M.Sc. in Computer Science on April 29, 2024.

**Examining Committee:**

Supervisor:
(Member)

Farig Yousuf Sadeque

Associate Professor
Department of Computer Science and Engineering
Brac University

M.Sc. Coordinator:
(Member)

Md Sadek Ferdous

Associate Professor
Department of Computer Science and Engineering
Brac University

Head of Department:
(Chair)

Sadia Hamid kazi

Associate Professor
Department of Computer Science and Engineering
Brac University

# Abstract

Globally, ride-sharing is very popular, especially in developed countries. The scheme has been launched in many developing countries, and Bangladesh is no exception. The ongoing transportation problem and traffic jams make this country vulnerable economically. The impact of COVID-19 has snatched the jobs of many people. The ride-sharing platform allowed them to grab a chance to be self-dependent. On the contrary, the increasing hike in daily vehicle accessories, fuels, and parts makes it difficult for a rider to earn his bread and butter. In this research, the author focuses on the impact of ride-sharing and drivers on the Bangladeshi economy. Along with this, many social and economic statuses are analyzed. At first, a dataset was prepared after discussing it with 2234 drivers. Extensive exploratory data analysis was performed to find insightful information from the dataset. Later, the dataset is preprocessed precisely before feeding into numerous Machine Learning and Deep Learning architectures. A comment from each of the riders is also taken to understand the sentiment of these riders. Three sentiments have been considered, namely Positive, Negative, and Neutral. The researchers have adopted an optimized BERT transformer-based approach to validate the dataset and classify Bengali comments correctly. The model can outperform the state-of-the-art architectures in numerous performance metrics. The optimized model shows a 80.63% F1-score in the training dataset, whereas it shows an 84.53% F1-score in the validation set. Finally, the black box model is interpreted with the aid of Explainable Artificial Intelligence.



**Keywords:** NLP; Machine Learning; Deep Learning; Transformer; XAI; Bengali Sentiment Analysis

# Acknowledgement

This research is conducted with the aid of some volunteers. We are grateful to those who collected the data from numerous sources. The annotation procedure was a difficult job. We extend our thanks to the persons who helped us in annotating data. The major credit goes to the Department of Computer Computer Science and Engineering where we have avail some laboratory facilities to perform some research work. A big credit goes to the honorable supervisor Farig Yousuf Sadeque for guiding in each step for the completion of the task. Throughout the whole process, many respected faculty members have given their opinions in order to enhance the quality of the work. We are also thanking them for their support.

# Table of Contents

# List of Figures

# List of Tables

# Chapter 1

# Introduction

## 1.1 Introduction

Bangladesh is a developing country with over 171.2 million people [1]. Although more than 60% of people still reside in village areas, most reside in urban areas. Many people are shifting from village areas to city life to earn bread and butter. One of the significant challenges in urban areas is the transportation system. The capital of Bangladesh is Dhaka, which is overpopulated with 10.2 million people, and the area is only 306.3 square kilometers [2]. The availability of public transportation is limited in this city during office hours. Ride-sharing has emerged as an innovative solution to tackle this issue. Ride-sharing platforms are prevalent globally because of their convenience and efficiency [3]. Affordability is a significant issue for a country like Bangladesh, where the annual income of every person is less than 3000 USD. Ride-sharing provides an alternative solution to public transportation such as Buses in Bangladesh. A study suggests that around 20% of people have changed their transportation medium to save time, especially those who frequently move a great distance [4].

## 1.2 Motivation

Ride-sharing is vital for solving the transportation problem and providing a proper medium for earning the livelihood of many people [5]. The COVID pandemic hit developing countries badly, and Bangladesh is no exception. Many lost their jobs during this period and adopted ride-sharing as their primary income. Although the ride-sharing concept was introduced in 2015, it has become popular in the COVID period. Pathao was the first ride-sharing platform in Bangladesh, and other companies such as Uber, Sohoz, and Obhai introduced one by one within a shorter period [6]. These platforms gather immense responses as they save time, most notably when traffic jam costs more than 10 billion annually [7]. A BRAC Institute of Government and Development study estimated that Bangladesh lost 11.4 billion dollars annually in 2018 due to traffic delays after accounting for the additional operating costs. During COVID, more than 70% of people became jobless or had to accept a pay cut [8]. During that period, a significant number of the person who could drive skillfully switched jobs and became either full-time or part-time riders. Another approach has been found recently, where it has been observed that a drastic number of captains/ drivers are not using ride-sharing platforms for numerous reasons. Instead, they gather passengers from offline mediums, where offline medium denotes bargaining instantly with a passenger, fixing a fare, and then conducting the ride. Understanding the sentiment of these captains is essential as it reflects a big scenario in the socio-economical sector in Bangladesh. Although many studies have been conducted from passengers' perspective [9]–[12], almost no study has focused on this sector due to data availability. Gathering data requires contacting them personally for a good amount of time. Usually, ride-sharing applications contribute significantly to the national economy. When the ride is conducted in an offline medium, it does not contribute to the economy. Understanding the viewpoint of riders will lead to solving the unstated problems hampering the country's finances. Societal and behavioral science can be understood with the aid of research.

## 1.3 Objective

The author focus on key points affecting the current ride-sharing status in this research. The economic conditions will be adequately understood from the gathered data from the riders. With the advent of Computer Science and Technology, Natural Language Processing (NLP) plays a vital role in extracting necessary data from comments [13]. A subsector of NLP is Sentiment Analysis (SA) is vital for understanding the insights of numerous sectors. The authors focus on the application of SA in this particular ride-sharing sector. At first, a good number of data is gathered from the riders, and an extensive Exploratory Data Analysis (EDA) is performed to understand the factors closely related to Bangladesh's socio-economic conditions. Apart from these, many comments are gathered to make a dataset for future research. From the computing viewpoint, many machine learning(ML) and deep learning (DL) architectures have been adopted to extract sufficient information from these communications. As the comments are gathered in Bengali text, they are difficult to process [14]. Furthermore, necessary steps have been taken to optimize these models, and Explainable Artificial Intelligence (XAI) has been used to understand the pattern of the black box models. XAI also introduces the transparency of black box models in such cases.

## 1.4 Common Ride Sharing Platforms

Although many platforms have started them in Bangladesh, very few have sustained them. There are multiple reasons behind not operating. Among them, financial viability is the cardinal reason. On the other hand, poor traffic conditions and privacy concerns are also major factors. However, some companies have understood the market precisely and still operating in Bangladesh. In this subsection, the authors discuss the top ride-sharing platforms in Bangladesh.

1. **Pathao:** Pathao is the first ride-sharing platform to introduce bike sharing in Bangladesh. At this moment, it offers various transportation in Dhaka and Chittagong. According to the source of Pathao.com, more than 8 million users are using this platform. Apart from ride-sharing, it also offers food and parcel delivery.

2. **Uber:** Uber is an international ride-sharing platform that was introduced in Bangladesh in 2016. Among all platforms, Uber charges more commission from the riders. As a result, the fare shown to the passengers is higher than that of other platforms. According to Uber.com's information, Uber has more than 1,75,000 driver partners and more than 4 million users. Although they have introduced Uber Eats in Bangladesh, they shut down the operation because of running losses. Uber is mainly famous for its car ride-sharing service.

3. **InDrive:** In-drive is another platform that offers motorbike services in urban areas. The company started its journey in 2022 with limited resources. It is the only platform not charging riders any commission for the first six months. They operate only inside Dhaka, providing the facility for negotiating between users and drivers. Less than 2 million users are using this app currently.

4. **Obhai:** This is the first app to start offering CNG ride shares in Bangladesh. With a friendly app interface, Obhai gained popularity among users. They are also currently operating in Dhaka. Obhai has approximately 1.5 million users. Like other companies, they offer a digital payment system.

## 1.5 Business Model of Ride Sharing Applications

Primarily, ride-sharing platforms operate in the business-to-consumer (B2C) model, where they directly serve consumers. With time, these platforms started working on the business (B2B), such as delivering parcels to organizations and collaborating numerous times [15]. In this research, the authors focus on the B2C sector, where passengers receive direct service from the riders. The business steps of the ride-sharing platforms can be chunked into the following steps:

1. **Registration of the Users and Riders:** Using a ride-sharing app requires registration of both riders and users. Usually, these platforms maintain two different apps for registering drivers and passengers. If a rider or captain wants to register for the application, that person must provide certain information, including the license number, vehicle number, and legal proof of the vehicle. Other information, such as Name, Phone Number, and Email, are similar for both parties. Usually, verifications are made using a One-Time Password (OTP).

2. **Sending and Accepting Ride Requests:** The second step is establishing contact between the drivers and passengers. Most ridesharing apps maintain a distance radius of 2.5 km for finding a suitable rider for a passenger. First, the ride taker selects the proper location and requests a ride. The request is forwarded to multiple captains for acceptance. The captains can decline or accept the offer. Later, after acceptance, contact information is exchanged between these parties. Usually, after the contact, the captain reaches the rider's location and initiates the ride. Online platforms usually provide multiple options, including bike and car rides. Car rides are usually chunked down into multiple sub-categories. The platforms usually use matching algorithms to resolve all these issues.

3. **Payment Procedure:** Before making a request, a certain fare is calculated based on the distance, traffic in that particular location, weather conditions, and vehicle availability. This fare can be increased or decreased based on the situation. Usually, the final payment is usually shown after reaching the destination, where the ride-sharing platforms take an average of 12% -15% [16]. Usually, the payment is made using cash, whereas online vendors such as Bkash and Nagad are also integrated into these platforms. A surprising factor is that most of the captains are unwilling to take payment using online vendors.

4. **Safety Features:** All platforms are usually considered safety platforms. Almost every platform integrates safety features such as a Real-time Global Positioning System (GPS), buttons for emergencies, driver verification, and regular inspection to ensure a reliable experience from the user's perspective.

5. **Feedback:** After the ride, the feedback of users and captains is considered using a rating system where both can rate between 1 and 5. The two-way rating system aids in giving proper feedback and helps maintain standards and accountability within the community.

Finally, numerous promotions and offers are made to platform users. In very few cases, facilities are offered to the captains. However, the captains are awarded numerous tags, such as gentle, polite, and extraordinary vehicles shown to users before the ride.

## 1.6 Major Concerning Factors for Drivers

Drivers or captains in Bangladesh face multiple problems. While gathering data, the riders stated the reasons hampering them from stabilizing them. Some of the major concerns are stated below:

1. **Cost of Fuel:** The growing fuel cost, such as Petrol and Octen, is hampering the riders' stability. In 2022, fuel costs increased by more than 50% after the petroleum sector subsidy was stopped. On the contrary, the increasing number of riders has stopped increasing the fare. Even online ride-sharing platforms are not increasing the fares due to the passengers' poor financial stability.

2. **Maintenance Cost of The Vehicle:** Drivers usually maintain their vehicles and bear all the repair costs. However, drivers often work long hours, and the vehicle gets dirty and needs to be cleaned. The increasing cost of bike accessories and parts has become a huge burden for drivers.

3. **High Commission:** With the increase in additional costs, the commission has increased on most platforms. After paying commissions, it has been difficult for drivers to earn some money from a ride. In some cases, the commission gets doubled in heavy traffic, causing problems for riders.

4. **Market Competition and Fair Treatment:** Initially, this sector was very profitable for captains because of fewer drivers. After the COVID-19 pandemic, drivers increased by more than 30% [17]. As a result, the number

of drivers is not frequent now. A rider must wait longer to find a suitable ride to his destination. Additionally, these riders are not getting fair treatment from the platforms where incentive procedures are not clarified precisely. Transparency is the demand of many drivers in such cases.

5. **Traffic Regulatory Issues:** Traffic conditions are the main issue in big cities such as Dhaka and Chittagong. The unavailability of proper rules and regulations worsens the conditions of roads. The application of proper rules and regulations can help people earn their livelihood more easily.

## 1.7 Contributions

The contributions of this research paper are discussed as follows:

1. Analyzing ride-sharing's social and economic impact by exploring the numerous factors closely related to transportation accessibility, employment opportunities, and the pros and cons of this concerned sector. These factors intersect in complex ways, influencing numerous aspects of society.

2. We created a dataset after talking with 2234 drivers. For this research, only bike and car riders are considered. All the dataset attributes are precisely understood, along with a comment related to the sentiment of partner drivers considered. The polarities of the comments are annotated with the aid of experts. Furthermore, this dataset will assist in further research in the concerned sector.

3. A modified Bangla BERT model is proposed after necessary hyperparameter tuning and changing the attention layer. The model is compared with state-of-the-art architecture.

4. XAI models are utilized to interpret the model. Multiple XAI models have been utilized for this purpose.

### 1.7.1 Usability of this Research

Primarily, the revenues generated from ridesharing platforms by the government are enormous. Not only the Government but also the ridesharing platforms especially Pathao and Uber are conducting surveys to get the proper ideas as to why drivers are shifting towards offline mode. Our gathered data will assist them in finding out the proper reason behind the scenario. Apart from this, facilitating the drivers to raise their concerns and helping them to achieve their income through proper policy can be achieved also. The drastic shift in the behavioral science of a population can be determined by their economic status and security issues. This research addresses both of the issues that will help the lawgivers to take steps to balance the whole situation. In modern-day society, the gig economy plays a vital role in allowing people to do their work flexibly. This research will address some of the major concerns regarding the Gig economy.

1. **Security Concerns:** Although one of the most important purposes of ridesharing is flexibility, multiple risk factors are understood through the gathered data from the drivers.

2. **Digital Platforms:** A detailed discussion regarding the ridesharing platforms has been provided so that drivers can choose their desired platform to conduct ridesharing with proper security measures.

3. **Pros and Cons of this sector:** Workers from the Gig economy remain insecure about multiple things, especially income. This research provides a clear idea about the earned money from the concerned sector. Drivers willing to enter this area can get a proper idea of whether it will be a sustainable sector for them or not.

Finally, the research is helpful for drivers to understand their concerns whereas clear ideas from this research will allow Government and Ridesharing platforms to find the problematic areas and solve them accordingly.

# Chapter 2

# Related Work

Research in the ride-sharing sector is limited due to the availability of public data in Bangladesh. Although the transportation sector is wide open globally, major studies regarding this sector were performed [23] in 2020, where two leading platforms of the United States of America were considered. The work focused on users' sentiments towards Uber and OLA. A dataset from those sentiments was built, and some symbolic machine-learning algorithms were considered to benchmark the dataset. From there, Random Forest (RF) performed significantly better than other models. In [30], the authors first analyzed Uber's business model, where they precisely analyzed it from the perspective of human-computer interaction (HCI) but did not consider the usage of ML or DL architectures. Furthermore, in 2019, the business perspective of multiple ride-sharing applications was scrutinized in [25], where 164 data were gathered from social media platforms. The concern is that most respondents were users of these platforms rather than partner drivers. A good amount of statistical values were shown, whereas the dataset could have been more comprehensive for concluding anything conclusive. The first ML approach to ride-sharing was adopted in [29], where they have used a Recurrent Neural Network (RNN) model that exhibits greater accuracy in finding the fake reviews provided by the users. Like Uber, in [28], a detailed analysis of Pathao had been done where the reviews of users had been gathered at first. No ML approach has been integrated on this system. Surveys were conducted based on the users of Pathao.
Apart from this, a detailed business analysis had been performed. Another study [20] make their focuses on how the situation drastically changed after the introduction of ride-sharing platforms. Primarily, they focused on analyzing Bangladesh's demographic situation. However, the focus had been shifted towards the ride-sharing and usage of public transportation of the people of Dhaka city. It had been found that users' satisfaction increased on a greater level after using these platforms. But security remained a crucial factor in such cases. No studies have been performed till now from the perspective of riders. The drastic job loss during COVID-19 forcefully allowed people to switch their jobs. In [27], the study aimed to study the shifts and perceptions of ride-sharing services. The number of bikes had drastically increased, and they took their adopted ride-sharing as a professional job. Transportation preferences shifted, and passengers started taking these platforms seriously. Although the primary purpose of these apps was to provide part-time income, most of the riders took it as a full-time job. Initially, riders were interested in taking online payments, but with the advent of time, they became reluctant to take online payments.

The dissatisfaction towards these platforms rose in 2022 when the fuel price increased significantly [21]. Not only the fuel price hike but also the commissions of these platforms had increased significantly. The driver's perspective has not been considered in Bangladesh, but a study in [22] USA considered the driver's concerns when they took the opinion of 200 drivers. The concerns were stated in a tabular format where pros and cons were discussed. They [26] also considered the perspective of part-time riders. The sustainability of these apps had been inspected in [18] where researchers had precisely the factors that could make the system more robust. Table 2.1 represents the current research gap available from the present literature.

Table 2.1: Research Gaps Addressed from the Recent Literature

| **Reference of the Paper** | **Existing Problems** |
|---|---|
| [16] | All the factors have not been considered. |
| [19] | No DL method has been adopted. |
| [20] | Riders' comments were gathered, but too few factors were stated. |
| [26] | Fuel price has been considered, but the impact is not stated properly. |
| [27] | Sustainability has been studied without taking any survey. |

This research focuses on solving the current issues addressed in the literature. First, necessary data has been adopted for EDA. Later, a transformer-based approach captures semantic information from the necessary comments. Finally, XAI is integrated to interpret the proposed model.

# Chapter 3

# Research Methodology

Figure 3.1 represents the overall workflow of this research. At first, data are gathered from multiple places. Multiple persons annotate gathered data to ensure they are not wrongly assigned sentiments. Later, Exploratory data analysis was performed to find proper patterns from the data. Additionally, necessary preprocessing techniques were adopted before feeding data to the RNN, ML, and Bangal BERT models. Different performance metrics have been observed to observe the models' performance. Performance metrics include Precision, Recall, and F1 score. Later, a comparison of energy consumption was also shown. Finally, the proposed model is interpreted using the multiple XAI: Shapley Additive Explanations and Local Interpretable Model-agnostic Explanations. The complexity of the Bengali language makes the tasks tough for the XAI models.

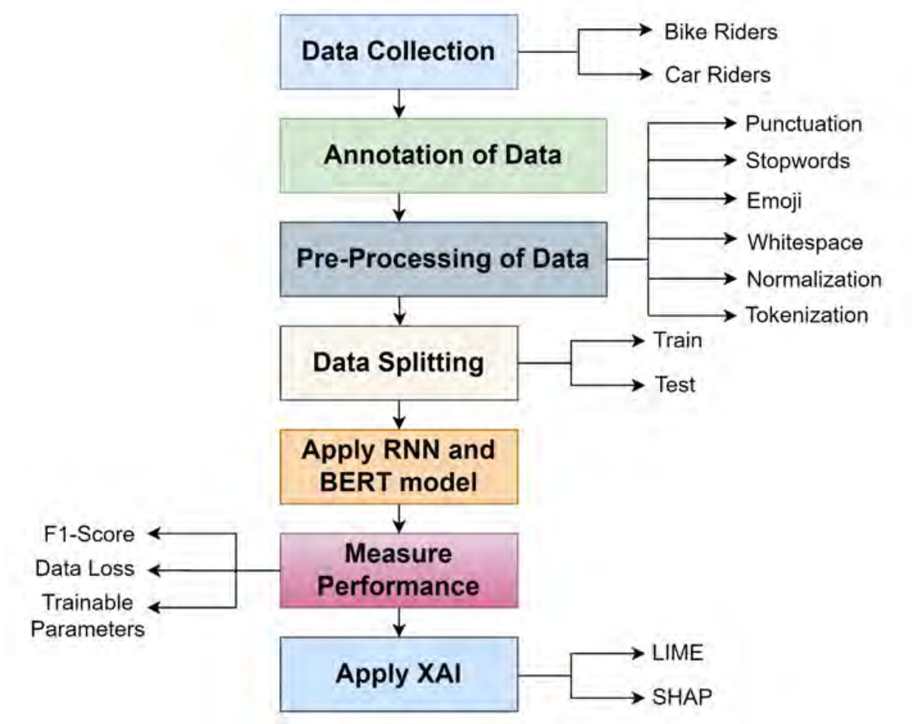



Figure 3.1: Methodology of Proposed Research Work

## 3.1 Data Collection Procedure

Researchers have classified the riders into two sections to gather the necessary data. These sections are:

1. Car Drivers
2. Bike Drivers

The main challenge was finding many drivers who wanted to share their valid words for research purposes. From that perspective, some busy areas of Dhaka city were initially selected. Primarily, two locations were spotted, and the number increased by another two within a shorter span of time. The four locations are namely,

1. Gabtoli Mazar road
2. Gulshan Link road
3. Hatirjheel Bus point (Rampura)
4. Mohammadpur Bus stand

Hence, the versatility allowed us to understand riders' perspectives from numerous places. Although bike riders were available in good numbers, finding car drivers was difficult. They were only found after calling them through an app. Table 3.1 shows the detailed breakdown of data collection. Here, it can be observed that most of the riders were from the offline medium in the case of Bikes, as they preferred it because of the lack of commission and flexibility of moving from one place to another without binding.

Table 3.1: Details break down number of drivers in the dataset

| **Driver Partner Type** | **Pathao** | **Uber** | **In Drive** | **Offline** |
|---|---|---|---|---|
| Bike Riders | 418 | 424 | 475 | 683 |
| Car Riders | 76 | 123 | 9 | 0 |

## 3.2 Major Challenges while Gathering Data

Collecting data from the drivers was not an easy task. The researchers faced numerous challenges while gathering data from the captains. Among them, some the important points are:
The main problem was that drivers were unwilling to provide data as they were unsure about the research. The drivers asked some interesting questions. The table shows some of the concerns and questions raised by the drivers.
As ridesharing is illegal in Bangladesh without using verified app services such as Pathao, Uber, Obhai and In-Drive, riders who do not use these services told us they use apps. The primary reason behind this is that they were doubtful about

Table 3.2: Most Frequent Questions Asked By the Drivers

| **Questions** |
|---|
| What will be the benefit? |
| Will you inform the Government about the result? |
| Are you from an NGO that will compensate us? |
| How will you help us financially and socially? |
| Are you the police who are investigating? |

whether we were from police enforcement or not. Apart from this, some drivers made vague comments that were removed from the dataset while the experts annotated them. The machine would learn misinformation if these things were preserved in the dataset. Some of them have the tendency to speak negatively to police and general people. Here, general people are known as riders.
The data will not immediately benefit the drivers, so some of them ignored it and used bad words sometimes. Some people like to stay within their personal space, so gathering information from them was very difficult.
As the Bengali language changes its nature based on the demographic location, understanding the sentences from the driver's mouth was difficult. Drivers were from numerous places all over the country. Most of them were from the northern part of Bangladesh, where they used to do part-time jobs in their local area. They had shifted to the capital city for better earning opportunities.
These are the major challenges faced by the author while gathering information from different places. One thing was clear from the point, not only poor people but also middle class people are also involved in this profession.

## 3.3 Dataset Description

Later, based on the findings, the dataset was designed, and multiple columns were integrated to understand the social and economic factors. Another column was denoted for gathering a Bengali comment provided by the riders. After that, necessary polarity was assigned with the aid of three volunteers. Every attribute in the dataset is described in detail in Table 3.3.
Here, the Service_type denotes the behavior of the passengers towards them, along with how they feel about that attitude. The attitude is either 0 or 1. 0 denotes a negative attitude, and 1 denotes a positive one. The second attribute is the working type, where we tried to determine whether the riders had taken the job full-time or part-time. Here, 0 denotes part-time job holders, and 1 denotes full-time ones. Then the query was whether they started the job before or During COVID. The notable factor is that we also considered the term post-COVID here. Hence, 0 denotes before COVID-19, and 1 denotes during or post-COVID. Furthermore, a comment is taken that represents the sentiment of the riders. The comment is taken into Bengali, and polarity is finally assigned to that comment. Multiple persons have cast their vote for the polarity. Three polarities, namely Positive, Negative, and Neutral, are considered. Here, 1 denotes positive polarity, 0 denotes negative polarity, and 2 denotes neutral polarity.

Table 3.3: Attribute Description from the Dataset

| **Attribute Name** | **Attribute Description** |
|---|---|
| Service_Type | Do people respect the job or not from riders' perspective? |
| Working_Type | Whether the rider is doing it on a full-time basis or part-time basis. |
| Start_Date | The starting date denotes before COVID or During COVID. |
| Family_Support | Is the earned money sufficient to support a family? |
| Comments | Positive, Negative, and Neutral things about ride-sharing. |
| Polarity | A polarity denotes Positive, Negative, or Neutral. |

Figure 3.2 reflects the original glimpse of the gathered dataset where all the attributes are correctly mentioned.

| Service | Work | Start_Date | Family Support | Comments | Polarity |
|---|---|---|---|---|---|
| 1 | 0 | 1 | 0 | এক্সট্রা সাপোর্ট | 2 |
| 2 | 1 | 0 | 1 | সব মিলিয়ে কস্ট হলেও চালানো লাগে। | 1 |
| 2 | 1 | 1 | 1 | ট্রাফিক পুলিশ সমস্যা। | 0 |
| 0 | 1 | 1 | 0 | অনলাইন কমিশন বেশি। | 0 |
| 2 | 1 | 0 | 1 | স্বাধীন ভাবে কাজ করা যায় | 1 |

Figure 3.2: Outlook of the Dataset

## 3.4 Exploratory Data Analysis

The first priority of performing the EDA is to understand the overall flow of the dataset. This initial phase is where authors gain a comprehensive understanding of the dataset. Extensive EDA can help understand outliers. In the case of this research, EDA provides detailed analysis, such as data quality and integrity. Apart from this, numerous visualization techniques have been adopted to observe the data patterns. Preprocessing steps are also carefully identified with the aid of available programming libraries. The findings of EDA are stated below in subsections.

### 3.4.1 Data Quality Checking using Kappa Score

We have utilized the Kappa score to measure the authors' agreement, providing a quantitative measure of user agreement. The Kappa score continuously observes and compares the agreement between raters. It ensures data consistency and reliability and informs decisions based on the level of agreement. In the socio-economic sector, the Kappa score provides an excellent investigation of inconsistency. In this work, we measure the inter-rater agreement using the formula below.

The formula for Kappa score for three persons is:

$$\kappa = \frac{P_o - P_e}{1 - P_e}$$

Where:

- $P_o$ is the observed agreement between raters.
- $P_e$ is the expected agreement, calculated as the probability of random agreement.

Table 3.4 represents the values associated with the data quality. In our case, the kappa score is 82%, which resides in the category of almost perfect agreement. That reflects that the inter-annotator agreement is strong enough to process the dataset. Furthermore, the comments of the dataset can be fed to the numerous models.

Table 3.4: Interpretation of Cohen's Kappa Score

| **Kappa Score Range** | **Interpretation** |
|---|---|
| 0.01-0.09 | Poor Agreement |
| 0.10 - 0.20 | Slight Agreement |
| 0.21 - 0.40 | Fair Agreement |
| 0.41 - 0.60 | Moderate Agreement |
| 0.61 - 0.80 | Substantial Agreement |
| 0.81 - 1.00 | Almost Perfect Agreement |

### 3.4.2 Statistical Analysis of the Dataset

The statistics of the dataset begin by describing the dataset where authors have observed the mean, standard deviation, min, and max attributes. Table 3.5 describes the dataset using statistical measures. While processing this operation, it has been considered that the Comment attribute was deleted as it bears no significance in these measures. The advantage of understanding the dataset description are many. Primarily, a comprehensive description ensures a clear understanding regarding the data. The structure and context are understandable before feeding into numerous models. omes.

Table 3.5: Dataset Description

| **Attributes** | **Service** | **Work** | **Start (COVID)** | **Family Support** | **Polarity** |
|---|---|---|---|---|---|
| **Count** | 2234 | 2234 | 2234 | 2234 | 2234 |
| **mean** | 1.20 | 0.41 | 0.63 | 0.53 | 1.1 |
| **Std** | 0.72 | 0.48 | 0.49 | 0.48 | 0.67 |
| **min** | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| **25%** | 1.00 | 0.00 | 0.00 | 0.00 | 1.0 |

| | | | | | |
|---|---|---|---|---|---|
| **50%** | 1.00 | 0.00 | 1.00 | 1.00 | 1.00 |
| **75%** | 2.00 | 1.00 | 1.00 | 1.00 | 2.00 |
| **max** | 2.00 | 1.00 | 1.00 | 1.00 | 2.00 |

Later, the covariance and correlation of these attributes are also checked. A strong correlation denotes a strong linear relationship between data points. Feature extraction and selection become easier for DL and ML architectures when the data points inherit strong correlations. The data points available in the dataset also exhibit strong covariance, which means data tends to move forward with each other. Table 3.6 shows the correlation between the data points in the dataset.
Data covariance measures the extent to which two variables vary together. It indicates the direction and strength of the linear relationship between variables. Positive covariance suggests that when one variable increases, the other tends to increase as well, while negative covariance indicates that when one variable increases, the other tends to decrease. However, covariance alone does not provide a standardized measure of the relationship's strength, making it difficult to compare across different datasets. Therefore, covariance is often normalized to a correlation coefficient ranging between -1 and 1 to facilitate meaningful comparisons and interpretations. Table 3.7 shows the covariance among variables.

Table 3.6: Correlation Among Data Points

| Attributes | Service | Work | Start (COVID) | Family Support | Polarity |
|---|---|---|---|---|---|
| **Service** | 1.000 | 0.046 | -0.007 | 0.066 | -0.048 |
| **Work** | 0.046 | 1.000 | -.14 | 0.15061 | -0.014 |
| **Start (COVID)** | -0.007 | -.14 | 1.0 | -.15 | 0.020 |
| **Family Support** | 0.066 | 0.15 | -.15 | 1.0 | -0.08 |
| **Polarity** | -0.048 | -0.014 | 0.020 | -0.08 | 1.0 |

After analyzing the correlation, authors are focusing on the covariance of the dataset. Here, a covariance description allows us to know about the relative movement of datapoints with the increament or decreament of each other.

Table 3.7: Covariance Among Data Points

| Attributes | Service | Work | Start (COVID) | Family Support | Polarity |
|---|---|---|---|---|---|
| **Service** | 0.53 | 0.001 | -0.002 | 0.02 | -0.023 |
| **Work** | 0.001 | 0.0243 | -.14 | -0.034 | 0.037 |
| **Start (COVID)** | -0.002 | -.034 | 0.23 | -.036 | 0.006 |
| **Family Support** | 0.024 | 0.037 | -.036 | 1.0 | .249 |

| | | | | | |
|---|---|---|---|---|---|
| **Polarity** | -0.039 | -0.004 | 0.020 | -0.006 | 0.027 |

### 3.4.3 Data Exploration and Analysis

Later, the authors explore individual data columns and the actual scenario. At first, the working type of the riders is explored. Two labels are considered: Full-time and Part-time. At first, the authors are focused on observing the number of bikers working both ways. Here, it can be viewed from Figure 3.3 that 1376 bikers are working on a full-time basis, whereas only 624 bikers are working on a part-time basis. Although the primary purpose of ride-sharing was to provide part-time income, our dataset shows that most riders work full-time.

Furthermore, the authors are now considering the starting date of these drivers, especially considering the three periods. Broadly, two periods are taken into account: before and after COVID-19. Figure 3.3 demonstrates an exciting study, where it is observed that only 436 people used to provide ride-sharing before the COVID-19 pandemic. The pandemic hit the job sector with many problems. Most companies force their employees to accept a pay cut or remove them from the company. As a result, the number of riders has increased significantly. From our dataset, among 2000 riders, 1534 drivers started their ride-sharing job after COVID-19. That represents the job sector, which has drastically shifted. On the other hand, another vital factor is that a biker earns more or less 30,000 BDT after all monthly costs, whereas average companies provide less than that.

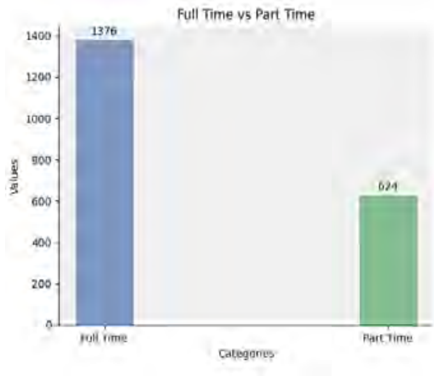
The authors also focus on whether the earned income supports a full family. A family can consist of a variable number of persons, but for this research, we asked them whether their earnings are sufficient to support a family of four. Most respondents replied that their income was not sufficient to live a lavish life but sufficient to live a middle-class family. However, with the ongoing fuel price and necessary price hike, the situation is getting more challenging for them. Figure 3.3 reflects the proportion of riders stating whether their income is sufficient for supporting their family.

Next, the authors focus on the Pairplot, a visualization technique, and another essential EDA for understanding the multivariate relationship in a proper dataset. The plot is vital for identifying patterns and trends in a dataset. Figure 3.3 demonstrates the distribution among the variables in the dataset. The quality assessment is also understood with the help of that figure. From there, we can declare that the dataset can provide semantic meaning to the ML and DL architectures. In some cases, it is also viewed that very few outliers are available in the dataset, which ensures its quality.

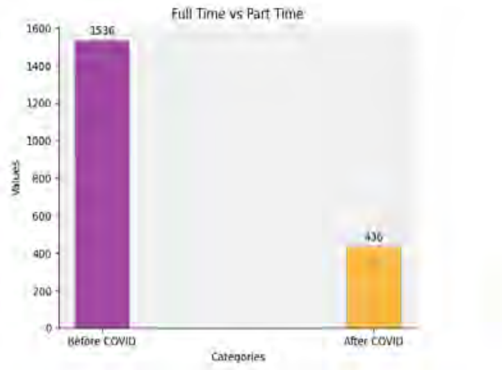
Thus, the above explorations conclude some important facts that are

1. Most bike riders started their profession after COVID-19, whereas very few riders worked before the pandemic. That means the job market has become thinner for most professionals. As a result, many people have taken ride-sharing as their full-time job.

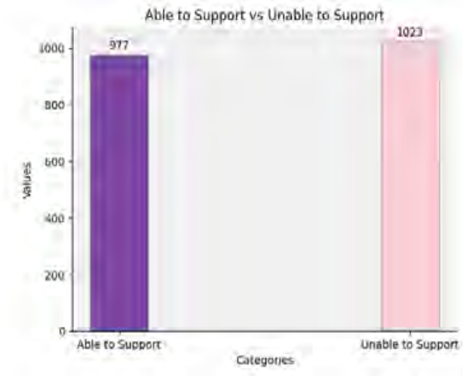
2. Additionally, the income from this sector is adequate on a narrow margin to maintain a family of four persons, whereas a lavish life is not possible for these



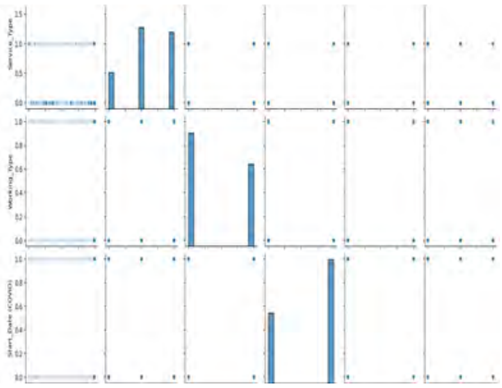

(a) Bar diagram regarding full-time or part-time




(b) Ride-sharing Starting Time




(c) Opinion related to family support of the bike riders



(d) Pair plot distribution diagram for the dataset

Figure 3.3: Some Notable Findings From the Dataset

riders. With the increasing price hike, it is becoming difficult to bear the expense of day-to-day life with this income.

3. The Pairplot, Inter-annotator agreement, Covariance, and Correlation declare the dataset suitable for further investigation and provide it to ML or DL architectures.

In the next phase, researchers have focused on preprocessing the comments, which is difficult because the Bengali language is ambiguous. The main issues with the Bengali language are described below, along with the complexities of why it is difficult for ML and DL models to capture semantics.

1. The goal of employing a certain preprocessing technique in research is not well-explained in the literature. There aren't as many studies available as for other languages. Every area of BNLP research is comparable to the scenario [28].

2. Some preprocessing techniques heavily rely on memory and processing power. Due to resource constraints, identifying the proper techniques for large-scale Bengali text and signal processing was difficult [24].

3. Most studies have gathered unstructured online forum data involving code-switching between Bengali and English. Another difficult task has been real-

izing which preprocessing approaches are appropriate to address this problem [19].

### 3.4.4 Preprocessing of the dataset

Data preprocessing is vital for providing data to numerous architectures to understand semantics precisely. It includes a good number of operations, which depend on the language and pattern of the dataset. Another important part is ensuring the model is not overfitted because of the anomaly in the dataset. For Bengali, numerous preprocessing techniques are available. Hence, the authors have used only the vital operations required to provide the DL and ML architectures. The adopted preprocessing techniques are stated below:

1. **Dropping Null Values:** The very thing is to drop the null values from the dataset. We have utilized the Pyhton nltk library to operate. Other than that, all other columns are eradicated from the dataset except for the comment attribute.

2. **Removing Stopwords and Keywords:** Bengali language consists of many stopwords that do not provide any special meaning to the models. On the contrary, they introduce ambiguity while capturing the inner meaning of these sentences. That is why all the stopwords have been removed. Another vital factor is to eradicate the punctuation marks from the sentences. Punctuation marks also do not provide any special meaning in these sectors.

3. **Special Character Removal:** All the special characters from the comments have been removed. To fulfill this purpose, the RE module of Python is used.

4. **Tokenization:** Tokenization is the process of dividing a written document into smaller pieces so that machine learning methods can be used. We have converted all the big comments into smaller tokens. That allows the models to understand and capture the words that are helpful to understand the semantics. The Bangla BERT tokenizer has been utilized to provide the model to the Bangla BERT model.

5. **Creating Dictionary:** A dictionary has been created to identify unique words. The word definition is also created using this dictionary. As we are mining semantics from the texts, we have created a dictionary to extract the inside information with precision. The word confusion can be resolved with the aid of this dictionary. Figure 3.4 represents the most occurred words encountered approximately 309 times. Apart from this, there are 7404 words available in the dataset, where 507 words are unique. After closely understanding the words, it can be observed that most of the words are related to finance.

6. **Stemming:** After observing closely, the third and fourth words disclose the same meaning but differ in spelling. The primary purpose of stemming is to convert the data into its base form. As a result, it becomes easier for the models to understand the meaning. As a result, the number of variations reduces to a greater number. In this research, stemming also plays a vital part in handling the sparsity of the feature space.

| | |
|---|---|
| Number of Words | 7404 |
| Number of Unique Words | 507 |
| Most Frequent Words | |
| পেশা | 309 |
| হিসেবে | 246 |
| ভাল | 220 |
| ভালো | 171 |
| স্বাধীন | 166 |
| কাজ | 163 |
| এই | 147 |
| করা | 129 |
| আসা | 125 |
| খরচ | 125 |

Figure 3.4: Frequent Words of the Dataset

7. **Word Embedding:** The cardinal purpose of using Word Embedding is to represent data in a dense vector in a vector space. Word embedding is a proper solution for a one-hot encoder that produces high-dimensional vectors. Learning distributed representations of words based on their context in a sizable corpus of text data is the fundamental notion behind word embedding. In this research, authors have utilized neural network word embedding techniques such as Word2Vec and Glove vectorizer. For the Bangla BERT model, no embedding technique has been adopted as this model incorporates word embedding into its architecture.

### 3.4.5 Experimental Models

A good number of ML and DL architectures were utilized to validate the dataset. This section provides a brief summary of these models. While running these architectures, the train-to-test size was 70% % to 30%.

**Machine Learning Models:**

1. **Logistic Regression:** Logistic Regression is a straightforward statistical technique to predict a particular outcome's likelihood. It makes an ideal situation for answering yes or no questions. For instance, it can determine the probability of an email being spam based on features like the words used within it. This method works by inputting data into a formula that employs a special S-shaped curve, ultimately producing a value between 0 and 1. This value represents the probability of the email being spam or not. One of the strengths of logistic regression is its simplicity and ability to provide a binary yes-or-no answer and quantify the confidence in its prediction. However, it does have its limitations. It presupposes a simple linear relationship between the characteristics of the data and the predicted outcome, which does not always come true in complex real-world scenarios. Moreover, it faces difficulties when the data contains a lot of noise or when the instances of yes and no

outcomes are imbalanced.
Regression analysis is popular for predicting continuous data points, whereas logistic regression is widely known for properly detecting supervised data points. Here, two types of variables are considered: independent and dependent variables. Another name for logistic regression is predictive analysis. Here, the dependent variable is mapped with the independent variable. It is especially better for detecting binary and ternary classifications. The data is precisely described with the aid of this algorithm. Binary logistic regression and Multinomial logistic regression usually predict the probability of selecting one of the available classes from the dataset. Usually, the distance between data points is measured at first. Based on the result, the predicted class is selected. Linear regression was first introduced to solve classification problems. However, the primary problem was that the outliers were not properly understood by the linear regression model. Manhattan distance is globally utilised for performing logistic regression on numerous datasets. The dataset must be adequately cleaned before providing data for logistic regression. At first, the train and test data are split properly and in a proper ratio. The logistic regression model learns from the train data and classifies test data based on learning. Numerous hyperparameters must be tuned properly to achieve the best outcome, including learning rate, train-test split, and epoch. After projecting the data points, the probability of the classes is measured. The highest probability is awarded with the mapped class.
The activation function is used in binary classification, and the sigmoid function is used in the activation function. The effect of the input is mapped with the output with the aid of coefficients. The primary assumption is that the observations are independent. Logistic regression never gives a direct yes or no answer. Rather, it provides a probability.

$$P(Y = 1|X) = \frac{1}{1 + e^{-(b_0+b_1X_1+b_2X_2+\ldots+b_nX_n)}} \tag{3.1}$$

2. **Decision Tree Classifier:** A Decision Tree Classifier helps decide what category something belongs to based on its characteristics. It works by repeatedly splitting the data into smaller groups based on certain features, aiming to have each group as pure as possible. It decides these splits by calculating which moves will reduce uncertainty about the data's classification at each step. Benefits It's easy to understand and explain to others. But it complicates if it gets too fixated on the details in the training data, making wrong guesses when it sees new data. To solve the problem, sometimes it's best to limit how many questions it can ask or to use a bunch of decision trees together to make a better guess.

   One of the most readable classification models is the Decision tree classifier. The context of numerous applications, such as sentiment analysis, is also properly worked with the Decision tree classifier. The basic working procedure is to build a decision tree with multiple branches. Each branch represents a decision. Each node receives a firm decision from the descending node. The class levels are represented with the help of each node. The training instances are divided by navigating them from the tree's root to a leaf. Each node represents a probability regarding classes in the dataset. The attributes are also proposed for each branch using a top-down approach. The best-fitting attitude is selected in each step to get the final result. Numerous ways are used to find the most effective way. The most effective way is the Gini impurity indexing. It is an impure way to grow the tree and find the correct and incorrect labels. The main advantages of decision tree classifiers are stated below:

(a) **Interpretation capability:** The cardinal benefit of the decision tree classifier is the easy interpretability of this model. The logic is concise, and easy to make precise decisions. The growing tree shows the logic behind selecting all the branches.

(b) **Understanding versatility:** One of the main issues with Machine learning models is that they can not handle multiple data types. The decision tree classifier is well capable of understanding numerical and categorical data. It is efficient in classification and regression tasks, which allows the decision tree to be flexible in many cases.

(c) **Extraction of Features:** The most important thing is that the decision tree classifier can extract informative features while exploring the data and understand the most important features while training the data relevant to each label. This model precisely captures the non-linear relationships and interactions. Although ML models find datasets with complex relationships difficult to understand, the decision tree classifier is unique in that perspective.

(d) **Handling the outliers:** As a robust model, the decision tree handles missing values properly. Based on the information available, each decision tree node can also handle outliers precisely. The model can work properly with enormous datasets where optimized information is extracted. It is very robust in such cases.

(e) **Non-Parametric Nature:** part from scalability, the decision tree is not necessarily a parametric model that considers the underlying data distribution. The model adopts all the necessary assumptions while classifying data. At each node, it calculates the probability and handles the missing values properly.

(f) **Parallel Execution:** The decision tree is calculated at each branch level, suggesting that the model might take a long time to process. In real-case scenarios, the model executes things parallelly, where trees are constructed individually regardless of the side of the branch. The training time is comparatively the same as that of other ML models

3. **Multinomial Naive Bayes:** The Multinomial Naive Bayes method is widely used for sorting items into various categories by dividing the words or attributes they contain. This technique works independently as an item belonging to each category based on the presence of certain features within the item. It counts how these features are associated with different categories to determine the most fitting category for the item. The fact is that specific categories are expected to feature certain words more frequently. After analyzing how

often words occur in each category, the method predicts the most appropriate category for a new email, for example, by considering its content. One of the calculations involved estimating the chance of a document being part of a category, considering the occurrence rate of each word. Multinomial Naive Bayes stands out for its simplicity and effectiveness, mainly when dealing with vast datasets. Nonetheless, this simplicity comes with a limitation, as it treats all words as equally important and independent of each other, which doesn't always hold in the nuances of language.

$$P(C_k|\mathbf{x}) = \frac{P(C_k) \cdot \prod_{i=1}^{n} P(x_i|C_k)^{x_i}}{\sum_{j=1}^{K} P(C_j) \cdot \prod_{i=1}^{n} P(x_i|C_j)^{x_i}} \tag{3.2}$$

The cardinal purpose of integrating Multinomial Naive Bayes (MNB) in this research is stated below:

(a) **Effectiveness of the model:** The simplicity and straightforward behaviour of the MNB allow the model to be super-efficient. The underlying probabilistic nature effectively classifies data points. The occurrence of each data point is very important here. The simple counting operation is the main reason for optimising the whole procedure. The model is easier to implement than most of the other ML models. This model is easy to integrate with any of the baseline software systems.

(b) **Adaptability with Text Data:** The cardinal reason for integrating MNB in this research is the model's adaptability with text data. The gathered dataset for this research is text data that MNB can understand properly. Three classes are considered in the dataset where MNB can perform well for multiclass classification. It considered the feature counts as words and counted the frequencies.

(c) **Handles Irrelevant Features Properly:** MNB usually works well with feature independence, whereas there are cases where this model performs better where features are not strictly independent. This nature allows the model to be robust in all cases and understand the data pattern properly.

(d) **The capability of understanding spare data:** The words used in this research are not large enough. There are cases where the dictionary gets larger because of the vocabulary size. In all cases, MNB finds the feature vector to be sparse due to the likelihood of each feature. That helps the model to be efficient in handling sparse vectors.

(e) **Easy to Interpret:** The MNB model's detection and prediction are transparent, and the whole procedure is easy. The model understands the classification decisions properly. The model performs some baseline performances for classifying text data. Even if the dataset is complex, the model understands the data pattern very well. The adaptability also allows the model to be flexible.

4. **k-Nearest Classifier:** KNN, denoted as k-Nearest Neighbors, is a way of making predictions on new data by looking at similar examples. KNN predicts without complex models. It stores all training data. It finds the k closest data points based on distance for new data. In classification, the majority class of

those k neighbours becomes the prediction. For regression, the average value of those neighbours is predicted. KNN finds the k closest points on the map. The new point is assigned the most common label from those neighbours in classification. The new point's value is predicted in regression as the average of its neighbours' values. It excels at making predictions based on similar past experiences. It stores all the training data like a reference map, allowing it to quickly find the k closest data points, which means the neighbours to a new arrival. Still, when it avoids complex models, KNN can become cumbersome with massive datasets, as searching the entire map gets time-consuming.

$$\hat{y} = \text{mode}(y_{i_1}, y_{i_2}, \ldots, y_{i_k}) \tag{3.3}$$

5. '**Support Vector Classifier (SVC):** The machine learning took SVC as a Support Vector Classifier, a powerful tool for adapting classification. It finds the best boundary (or hyperplane in higher dimensions) separating different feature space classes. This boundary is chosen to maximize the margin, the distance between the boundary, and the nearest points from each class, known as the support vectors. Essentially, SVC tries to find the broadest possible street, such as the margin between classes, with the support vectors being the buildings as data points closest to the street edges.SVC is powerful and flexible due to its ability to handle linear and non-linear classification using kernels such as linear, polynomial, radial basis functions, etc. However, it might be tricky to choose the right kernel, and tuning the model's parameters, like the regularization parameter and kernel coefficients, can be challenging and requires a good understanding of the data and the model.

$$f(\mathbf{x}) = \mathbf{w}^T\mathbf{x} + b \tag{3.4}$$

The primary reasons for using Support Vector classifier in this research are stated below:

(a) **Effective in High Dimensional Spaces:** Support Vector classifiers can handle dimensions when computing has substantial complexities. The main primary subset of training points in the decision functions are the node points when they affect decision functions. It works very effectively in the case of higher-dimensional data and features, such as geometric data or text data.

(b) **Works Well With a Clear Margin of Separation:** In the case of clearly separable datasets, an SVC really works well, even if it contains many features. Computation complexity can be suppressed to handle dimensions. Hyperplanes are divided into classes by making classifiers and widest margins, which tends to reply to overfitting in this model.

(c) **Avoiding Overfitting:** By choosing kernel functions properly, SVM can lead to avoid overfitting for the resulting output. The regulation parameter of SVC allows the model to optimise in case of generalisations.

(d) **Versatility through the Kernel Trick:** Nonlinear boundaries can be used to trick the kernel and, later, imply mapping inputs into high-dimensional feature spaces. Also, SVC works really well on a variety of data types where features are not linear and class labels are indicated.

(e) **Optimal Solution:** When a model gives an optimal solution, it can be considered a reliable and trustworthy model. An SVC is used to solve complex convex optimization solutions, making sure that the result shows a better global minimum than a local minimum.

(f) **Scalability to Larger Datasets:** Datasets that work as sequential minimal optimisation methods can break the tasks into smaller problems, and then they can be solved analytically. Scalability and efficiency are the main parts of SVC regarding larger datasets.

(g) **Theoretical Guarantees on Performance:** Support Vector Machine usually shows better performance than other algorithms because this specific model can be minimized and bounded under specific conditions, providing a solid foundation for performance.

Apart from this advantages, some of the major disadvantages of SVC are stated here also.

(a) **Poor Performance With Large Datasets:** A noisy dataset, when it comes to overlapping the dataset classes can make the Support Vector Classifier poor in terms of performance. A larger dataset is less practical for SVM in terms of performance.

(b) **Poor Performance With Lots of Noise:** Misleading results can be gained after reading a noisy dataset. Scattered results can give the model results where the support vector points are occupied with other classes. Smaller margins will be allocated to accommodate this dataset where noisy points interrupt.

(c) **No Probabilistic Explanation for Classifications:** Class predictions and probabilities are not provided by standard SVCs; rather, they operate by placing data points near relative to the decision boundary. This denotes that the model can easily generate an output, but the level of prediction is not like that of probabilistic models.

(d) **Sensitive to the Choice of Kernel and Regularization:** Support Vector Machine is highly dependable on the kernel type, selection, and regulation parameters. It is mostly because while choosing it, it can be overfitted or underfitted, making it sensitive towards Kernel choices.

6. **Stochastic Gradient Descent (SGD):** Stochastic Gradient Descent (SGD) represents a robust and efficient optimization technique widely employed in machine learning to refine models by minimizing their errors. This strategy is particularly adept at handling scenarios that involve extensive datasets by optimizing model parameters incrementally, using only a subset of the data or even a single data point at each step. This incremental approach allows for rapid convergence on the optimal model parameters, distinguishing SGD from traditional gradient descent methods that require processing the entire dataset to update parameters once. The benefits of SGD lie in its ability to process and learn from large volumes of data efficiently. However, the method's inherent stochasticity introduces unpredictability; it often results in the model's parameters fluctuating within the vicinity of the optimal values rather than

converging on a precise minimum. This issue can be substantially mitigated through careful adjustment of the learning rate Such strategic manipulation of the learning rate enhances SGD's effectiveness and stabilizes its convergence on the desired outcomes, making it a valuable tool in the machine learning toolkit for optimizing model performance across large and complex datasets.

$$\text{EQN: } \theta_{\text{next step}} = \theta_{\text{current}} - \eta \cdot \nabla_\theta J(\theta; \mathbf{x}^{(i)}, y^{(i)}) \tag{3.5}$$

### 3.4.6 Random Forest:

Usually Random forest is built using the decision tree classifier where Gini impurity is utilized for building the decision trees, It works using bagging procedure. Finally, the results of all the decision trees are integrated. The highest vote count method is utilized for selecting the final result.
Random forest has some advantages over other algorithms.

1. **Higher Accuracy:** The first advantage of the Random Forrest algorithm is that it shows very high accuracy. Mostly in prediction tasks where Random Forrest plays a vital role. Ensemble learning is an example of a random forest, such as a decision tree, which can really give a better output when it comes to multiple learners also, as it is very susceptible to work on single decision trees where it can be exposed to noise, bias, the variance of dataset or averaging the result. Likewise, when it works to build on random data samples, it is also drawn with replacement, and as a bootstrap sample, it shows great results. Higher accuracy is also seen in random forests in case of feature randomness. During the tree construction, when the nodes are split, the split choices are not based totally on the full set of variables, but Random Forrest selects a subset randomly of features for each node. This way of selection maintains the dominant ones to work better. Diversity is also found and increased within the ensemble, which causes the lowering of the model's increased accuracy and variance, which also causes higher accuracy. Furthermore, averaging the given tress causes a reduction in variance. Random Forrest assumes by averaging all the predictions for given trees that if they worked on specific individual trees, it might show some sensitivity towards noise. But this way, it gets cancelled out and gives a huge accuracy. Handling unbalanced data is also a reason for higher accuracy in Random Forrest.

2. **Resistance to Overfitting:** Models like Random Forests, which tend to run constantly across various datasets, are likely reliable in predicting mostly because they don't get distracted by noise. Although they are very generalised, the models tend to get more reliable, making them effective. Without using an excessive set of training datasets, Random Forrest can really work well. Also, Random forest comes with Flexibility and Scalability, where the algorithm has to handle well overfitting without significant loss. Overall, RF shows a better ability to resist overfitting, which results in better results.

3. **Works Well With Large Datasets:** The large number of records or a wide range of features denotes a large dataset, and in particular, Random Forrest shows its effectiveness by using those datasets and giving a better

output. During the training process, it selects the features and makes its subsets for training. Random Forrest can precisely determine which features are important and which will impact the algorithm. Random Forrest can detect irrelevant or erroneous noise and stop the model from producing misleading pieces of information. Moreover, a good balance when dealing with a dataset between bias and variance can cause generalisation is also a vital advantage of Random Forrest. Random Forrest does not require Normalisation or scaling data for any kind of dataset. Using feature values and numerical or categorical data it can work in different ranges. That's why Random Forrest can quickly recognise the pattern of the dataset and show better output.

4. **Handles Missing Values:** Random forest handles the missing values during the training phase and gives better results. By splitting a tree with a missing feature, this algorithm can still make a good decision about the patterns. Random Forrest is also good at estimating missing values based on the data, such as missing values, biases, etc. An imputation method is used for this detection. Although, this algorithm also avoids introducing potential biases that might arise during the running time. Random Forrest also comes with flexibility in real-world applications. It discards random rows where there are missing values and data might lost.

5. **Gives Estimates of Feature Importance:** Random Forrest has a mechanism for detecting which individual feature is most usable. This helps give insights into balancing the dataset because it helps denote higher accuracy. A clear working of Random Forrest while working with dataset features is detecting the underlying dynamics. Observing feature extraction and its importance also enhances model performance and improves visualisation.

Apart from this, there are some problems exist with the Random forest algorithm. These disadvantages disallow the model to be precise in some cases. Although for Bengali language Random forest model is widely utilized, these problems do not allow models to be a proper state-of-the-art algorithm.

1. **Interpretability Issue:** Unlike simple tree algorithm models, Random Forrest models are not easily interpretable since they are an ensemble of many trees, so it is tough for it to explain the model prediction in a simpler way. For complex decision processes, multiple decision trees, such as weighted combined specific individual decisions, suffer from detecting the number of trees and also lack visibility about Random Forest's clusters to make the proper decision on the algorithm. Random Forests also suffer to extract rules for clear paths of nodes and branches. When it averages the majority across many trees, it shows enigmatic decision paths. Regarding model interpretability, it comes with binding legal, and regulatory standards in industries where these models are heavily interpreted for predicting. Black Box refers to the trust section of any model, and Random Forest can create trust issues in that section as well. Moreover, Random Forest has a binding of interpretability barriers, which are also distinguished as lacking.

2. **Issue with the performance:** High computational cost is one reason for performance issues in Random Forest. Running models like these can be computationally expensive, specifically when there are an increasing number and

depth of trees. High memory consumption is also a reason for performance issues. Predicting results from Random Forest is very time-consuming. Moreover, a large number of trees causes complexity. It also shows overfitting in the case of noisy data.

3. **Complexity in Real-Time Prediction:** Random Forst can be slower in real-time predictions, as it requires substantial aggregation. Multiple trees are processing at a time, which is a major reason. Voting for classifications also requires overheading aggregation. Moreover, as it requires high resource utilisation, it is tough to give value from time to time. Scaling values are a major factor in predictions. Latency when working within a certain time frame is a drawback. Load balancing can not work effectively when it comes to optimisation limits.

### 3.4.7 Deep Learning Architectures

For this research, the authors are focused on applying RNN-based architectures, which have excellent semantics-catching capability. The subsection focuses on providing short details about some of the state-of-the-art DL architectures in the sector concerned.

1. **Long Short-Term Memory** LSTM is one of the most significant architectures based on Recurrent Neural Networks (RNNs). The main goal of utilizing this is the ability to solve the vanishing gradient problem. This architecture's memory cells can comprehend the meaning behind a lengthy string of words. To regulate the information flow, three gates are used: the forget gate, the input gate, and the output gate. These gates enable us to update and forget information according to semantics regularly. Two LSTM layers with a hidden layer size '45 are employed for every embedding layer. To produce the probability distribution, the model has an internal Softmax layer. The benefits of LSTM are stated in the below portion:

   (a) **Handling the Dependencies:** The introduction of LSTM ensures the proper address of the vanishing gradient problem, one of the major issues with the Recurrent Neural Network (RNN). The LSTM model properly understands the long-term dependencies. The Cell state available in the LSTM allows for the preservation of information. The long sequence is understood properly with the availability of these cells.

   (b) **Preservation of Memory:** The most important factor is that LSTM can capture long-sequence information. In the proposed research, the dataset has long sentences, and sequence information is the most important factor. The presence of a conveyor belt can update, delete and add important information throughout three gates.

   (c) **Sequence Length:** LSTM can understand sentences of variable sequence length. This allows for understanding real-world data using the mechanisms of the forget gate, input gate, and output gate. The flow of information is understood properly for the most stable flow of information.

(d) **Learning of Features:** Another important advantage of LSTM is that it can understand relevant features from input data. In NLP and Sentiment analysis problems, datasets are relatively complex and consist of high-dimensional data. LSTM has the capability of learning features or extracting features that most ML model fails.

(e) **Parallel Training:** As an RNN architecture, LSTM can be trained in parallel. This lessens the train time in numerous aspects compared to sequential learning. LSTM also observes faster convergence.

(f) **Transfer Learning Capability:** Now, pre-trained LSTM models can easily and precisely understand the data pattern. With proper fine-tuning, LSTM can perform far better than other black box models. This architecture is most used in language translation.

(g) Interpretibility of the Model with Explainable AI: LSTM is a black box model and difficult to comprehend. Hence, Explainable AI such as LIME and SHAP are effective in understanding the working pattern of LSTM. This interpretability allows the users of such models to understand what is happening inside the model.

The parametric details of the LSTM architecture are presented in Table 3.8.

Table 3.8: Hyperparameter details of LSTM

| **Hyperparameter Name** | **Value** |
|---|---|
| Number of epoch | 30 |
| Activation function | Softmax |
| LSTM layers | 16 |
| Embedding vector length | 64 |
| Recurrent Dropout | 0.15 |
| Amount of Train data | 0.70 |

2. **Gated Recurrent Unit:** The Gated Recurrent Unit (GRU), an alternative to the LSTM network, is another popular RNN architecture. The Reset and Update gates are the only two gates in this architecture, but they allow much semantic capturing. While the update gate includes fresh information, the reset gate typically handles previously collected data. The primary benefit of GRU lies in its computational speed. Table 3.9 provides the GRU's parametric details.

   The main properties of GRU are stated below:

   (a) **Generality and Simplicity:** Unlike LSTM, GRU has a simple architecture with fewer trainable parameters than the mentioned architecture. As a result, the required training time is fewer than LSTM. Another important property is that GRU has a tiny reduced computational overhead when the amount of resources is limited.

   (b) **Efficiency in smaller Datasets:** GRU can understand data properly even if the dataset is smaller. It has good sequence-capturing ability and can optimize sequences properly for certain applications.

(c) **Adaptability:** Like LSTM, GRU can also handle sentences of variable length. This architecture can be trained on any dataset and can properly handle sequence information. Sometimes, preprocessing techniques such as padding and truncation are required for flexibility.

(d) **Overfitting Issue:** GRU's blessing is its simpler architecture than most black box models. As a result, this architecture is less prone to overfitting. The fewer trainable parameters prevent the storage of noisy data. GRU also has better generalization performance.

(e) **Scalability and Interpretability:** GRU's performance is good in small and larger datasets. As a result, this dataset is better suited for numerous kinds of datasets. In terms of interpretability, the simpler structure helps to provide clear insight compared to other models.

(f) **Parallel Processing:** GRU can be trained parallelly efficiently. The convergence is also faster, and less hardware requirements are observed in the case of GRU.

(g) **Versatility:** For forecasting, prediction and translations from one language to another, GRU plays a crucial role. This wide range of applications allows this architecture to be integrated properly in numerous cases. Regardless of domains, GRU shows great versatility in performing optimized tasks.

Table 3.9: Hyperparameter details of GRU

| **Hyperparameter Name** | **Value** |
|---|---|
| Number of epoch | 30 |
| Activation function | Softmax |
| GRU layers | 8 |
| Embedding vector length | 64 |
| Recurrent Dropout | 0.18 |
| Percentage of train data | 0.70 |

3. **Bidirectional Long Short-Term Memory:** The Bidirectional Long Short-Term Memory, or BiLSTM, functions similarly to an intelligent assistant by evaluating sounds or words from both the forward and backward directions. This approach is useful for jobs requiring thorough context comprehension, like text generation, language translation, and even sentence-level word prediction. It may be set up to go into data sequences and detect the tiny clues that reveal a more complete tale using tools like TensorFlow or PyTorch. BiLSTM's ability to retain specifics and recent information from far back in the sequence makes it an effective tool for situations where every nuance of context matters. It's similar to talking to someone who hears the final few words and retains the entire conversation. Because of this feature, BiLSTM is the preferred choice for projects where precise forecasting or analysis depends on knowing the details of the situation.

   The BiLSTM architecture provides advantages in many cases. Some of the reasons are stated above:

(a) **Ability to capture Bi-directional Context:** The cardinal advantage of BiLSTM is that it can capture sequence in both forward and backward ways. This mechanism properly understands not only past information but also future information.

(b) **Improved Memory Access:** BiLSTM can access more contextual information from both sides, which improves memory retention. This approach mostly benefits machine translation and sentiment analysis.

(c) **Improved Feature Representation:** BiLSTM can understand and learn features properly considering the future and past information. The sequential patterns help the model to understand features in numerous tasks.

(d) **Less Prone to Overfitting:** By providing a more comprehensive view of the input sequence, BiLSTM can reduce overfitting. It also understands irrelevant features better than LSTM, and noises are properly reduced.

(e) **Flexibility in Numerous Tasks:** Another major advantage of BiLSTM is that it can be mitigated with numerous neural networks. These networks exhibit tremendous f1-score in complex mapping tasks. The time series data are also handled precisely with the aid of BiLSTM architecture.

(f) **Interpretability:** For sequential data, BiLSTM provides proper interpretability with the assist of Explainable AI. Further research can be accomplished after integrating the result of the forward and backward hidden states.

(g) **Robustness:** BiLSTM architecture is robust regardless of the sequence of data. Regardless of the size of the dataset, this architecture shows robust behavior. The different patterns and structures of data are adapted properly by processing from numerous sides.

(h) **Lesser Data Ambiguity:** The ambiguity and uncertainty of sequential data are handled properly by the BiLSTM architecture. Multiple interpretations are taken into account to avoid data ambiguity. Especially in the case of sentiment analysis and machine translation, these properties allow more flexibility.

4. **Bidirectional Gated Recurrent Unit:** The Bidirectional Gated Recurrent Unit, or BiGRU for short, enables computers to comprehend data sequences, such as sounds or sentences, by examining them from start to finish. This method works well for jobs where context is crucial, such as summarizing lengthy articles, identifying speech, or determining the mood of a document. Typically used technologies like TensorFlow or PyTorch are basically sophisticated software that helps design and train these models, and they are necessary to make a BiGRU perform at its best. When BiGRU finds a good balance, it performs admirably. It isn't overly complicated, wasting little time or processing resources, but it is also intelligent enough to pick up on minute details in the data it examines. Table 3.10 reflects the parametric details of the BiGRU and BiLSTM architecture. Although there are similarities between BiLSTM and BiGRU, significant amount of differences are observed. Some of the points are mentioned below:

(a) **Difference in Memory Management:** While BiLSTM has more complex mechanism of managing memory, BiGRU manages this operation properly. The main difference is that, BiGRU has fewer parameters with only two gates. The gates are known as Reset gate and Update gate.

(b) **Complexity in Computation:** As BiLSTM has more trainable parameters, that is why it requires more computations. Because of the complex architecture, BiLSTM shows such patterns. On the contrary, BiGRU is efficienct in such cases. Deployment of this model is easier due to low amount fo parameters. They reflect great performance in resource-constrianed environments.

(c) **Handling Dependancy:** Although both BiLSTM and BiGRU has almost same capability to capture long sequence by processing input in both ways. Here, both ways refer to forward and backward direction. Experimental result exhibits, BiLSTM has slightly better sequence capturing capability. The primary reason is that, BiLSTM keeps separate memory cells for longer periods.

(d) **Capturing Complex Representation:** As BiLSTM maintains sophisticated memory management that is why they can capture ambiguous and complex patterns from dataset. BiGRU also can capture similar complex representation also.

(e) **Computation Flexibility:** In the case of both BiLSTM and BiGRU, computation plays a vital role. As BiGRU has lower trainable parameters that is why BiGRU is computationally efficient. So, when there is a necessity of a system where computation cost is a priority then BiGRU is more preferred.

(f) **Performance on Generalization:** While performing modeling tasks, BiLSTM and BiGRU shows strong generalization performance. In this case, data must be trained with sufficient data. BiLSTM shows slight better performance when generalization technique is important.

(g) **Robustness:** Both architecture is robust on small and large dataset. Despite of language barrier, these two models have shown suitable performance in every aspects. The long range dependencies are maintained in both architectures. Sentiment analysis and Machine translation are similarly effective in both architectures.

Table 3.10: Parametric details of BiGRU and BiLSTM

| **Hyperparameter Name** | **Value** |
|---|---|
| Number of epoch | 20 |
| Activation function | Softmax |
| Learning rate | 0.03 |
| Optimizer | Adam |
| Recurrent Dropout | 0.25 |

### 3.4.8 BERT Architecture

Transformer models are widely popular for multiple tasks. One of the successful transformer models is BERT. The elaboration of BERT stands for Bidirectional Encoder Representation from Transformers. Primarily, this model is successfully utilized for NLP tasks. The main advantage of these models is they can perform sequence-to-sequence tasks very well. The working procedure of the BERT model is explained below in 3.5

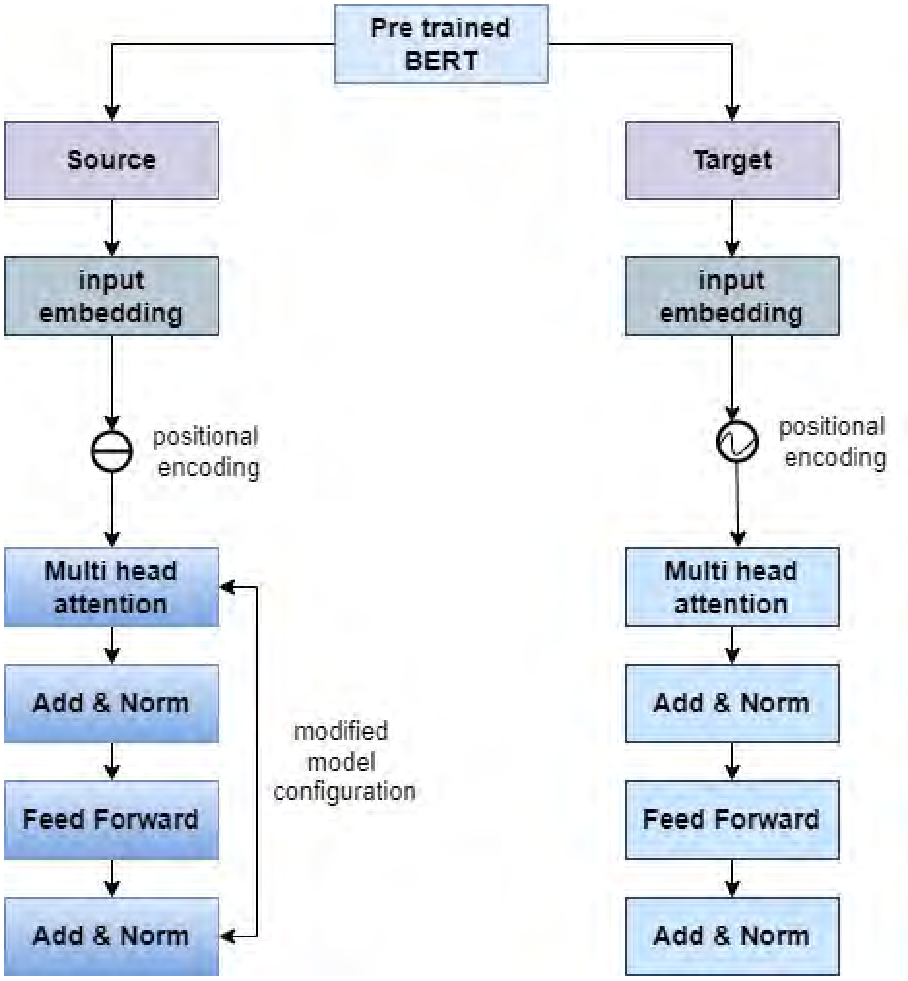



Figure 3.5: Methodology of Proposed Research Work

1. **Understanding the Contextual Representation:** For most of the symbolic language models, word embeddings are performed in a unidirectional way. Examples of this type of model are Word2Vec and GloVe. GloVe is typically used for deep learning architectures. The first bi-directional approach was taken by BERT, where sequence information was captured from both directions. The task is performed simultaneously from both sides. As a result, the context is understood precisely and properly, hence providing an accurate representation.

2. **Availability of Transformer Architecture:** A transformer is basically a Deep learning architecture that was introduced by the researchers based on the self-attention mechanism. This model is highly effective for sequential data.

BERT understands the contextual information with the aid of self-attention mechanism. Here, contextual information refers to the embeddings represented by the words.

3. **Pretrained Architecture:** The main advantage of BERT is that, it is pretrained on large-scale corpus. Among two popular learning techniques, unsupervised learning is utlized here. During the training phase, BERT usually predicts the masked words. Masked words are referred as hidden words within a sentence. For downstream tasks, BERT is usually fine-tuned properly.

4. **Contextual Embeddings:** The vector representation of a word is adjusted dynamically based on the context. Another important property of BERT is that, it can understand multiple meaning of a word.

5. **Attention Mechanism:** Usually self attention is very important as the significance of each word is dynamically understood. The importance is given based on the current word. Using this mechanism, BERT can capture sequence information properly. Despite of small amount of labeled data, BERT can be fine-tuned on downstream tasks. Fine-tuning allows the model to be more flexible and be optimized properly. This open source architecture is easily accessible.

For the BERT model, the language-specific corpus is built. The model is deployed in numerous for each language. One of the most popular architectures in Bengali is Bangla BERT, which is deployed over the BERT architecture.
Transformer models are performing really well, especially in NLP domain-related tasks. For most languages, BERT is mostly utilised. In most cases, BERT does not perform significantly well in Bengali. There are some shortcomings available in such cases. With more power, BERT also possesses some disadvantages. Some of the disadvantages are stated below:

1. **Hardware Resource:** Training the BERT model is hectic when sufficient hardware support is unavailable. Hardware resources are needed to overcome computational complexity. Fine-tuning BERT architecture is also another difficult task. A substantial amount of memory and storage is also needed to train properly.

2. **Massive Training Time:** Depending on the size of the dataset, training the BERT architecture from scratch can be time-consuming. That is why hardware resources are needed to train it in a shorter period of time.

3. **Size of the Model:** Depending on the tasks inside the BERT architecture, there can be billions of parameters. So, deploying the BERT model in numerous systems is very difficult.

4. **Adoption in the Domain:** One of the major issues with BERT is that it can not provide similar results for all domains. For example, BERT is very efficient for English; conversely, the result is not as effective for Bengali.

5. **The issue with Tokenization:** BERT owns its tokenisation technique, which is called WordPiece tokenisation. This tokenisation sometimes affects the performance of the model.

6. **Dealing With Out of Vocabulary tokens:** In most cases, the BERT tokeniser has a fixed vocabulary size. Out-of-vocabulary tokens are represented using subword units. Handling out-of-vocabulary tokens effectively can be a tough task.

7. **Interpretability:** The inner working patterns of BERT are not understood properly. Decision-making can be challenging based on interpretation. Even explainable AI models sometimes remain unable to interpret the model properly.

8. **Dependencies on Data:** As a pre-trained model, BERT depends on scalable labelled datasets for numerous tasks. A sufficient number of data is required to pre-train the model; in most cases, finding such datasets is difficult.

9. **Bias:** Bias available in the train data can be inherited by the BERT model. Ensuring fairness and unbiased behaviour requires a significant amount of dataset cleaning.

### 3.4.9 Proposed Bangla BERT model

BERT or transformer-based architectures perform well in resource-based languages such as English. The primary purpose of using this architecture is that it is trained on more than 40GB of Bengali text data. The model is equally impressive for binary and multi-class classification. The authors focus on this research's multiclass classification of riders' sentiments. To optimize this architecture, we have gone through several steps.
The primary purpose of using Bangla BERT is to solve the shortcomings of BERT architecture in the Bengali language. In contrast to other languages, Bengali has distinct patterns that can be challenging for computers to comprehend. Numerous elements, including morphological complexity, semantic ambiguity, training data limitations, and domain specificity, are present in these patterns. The writers of this study concentrate on outlining the main difficulties in reading Bengali in this part. Issue with Bengali BERT are stated below:

1. **Challenges with Bengali Language:** Morphological Complexities: Bengali is an inflected language that can take on the shape of many different grammatical elements, including person, gender, tense, and mood. These characteristics add layers and more complications to make it more interpretable for NLP.

   Limited Training Data Availability: It is challenging for ML and DL architectures to become acclimated to the Bengali language because of the lack of high-quality training data. Additionally, compared to languages like English, Chinese, or Spanish, Bengali inherits very few resources for training ML and DL models. The scarcity of data hampered the creation of reliable NLP systems.

   Semantic Ambiguity: Bengali has ambiguous semantics, meaning a single word or phrase can have several interpretations. The gender-neutral behaviour in these sentences is another challenge. Domain Specificity: Bengali is supported by specialised vocabulary in several fields, including law, finance, and medicine. Bengali datasets are less prevalent than those for languages that are commonly spoken, making using this dataset for domain-specific tasks challenging.

The BERT architecture primarily handles the input value in three ways: Embedding, Tokenization, and Mapping. In the very first stage, after providing the dataset, the words are converted into tokens and mapped into a unique number from the corpus. Hence, these distinct values are utilized to map the words further. As a result, we obtain an embedding of unique numbers for every sentence. An embedded dimension is also found in the model, which is utilized for backpropagation. A significant advantage of BERT architecture is that it can learn positional encoding. Generally, a matrix represents the sequence of texts; based on a word's position, it can capture the sequence correctly. This positional encoding allows the model to be flexible and adequately capture information. Additionally, the encoder block combines an attention layer and a feed-forward network. Inside a Bangla BERT model, multiple encoder blocks are utilized. Furthermore, the architecture is mainly rotated around the muti-head attention. The attention mechanism's capabilities aim to improve the capability of the attention mechanism to be applied simultaneously multiple times. A different linear projection is made each time based on the input embedding. Here, attention denotes focusing on a specific part of the input sequence. Different weights are utilized for different parts of the input.
In the original Bangla BERT architecture, 12 encoder blocks were used, whereas the research authors applied 8 encoder blocks, leading to fewer trainable parameters. In total 8 attention heads are attached in the optimized model. For normalization purposes, the original denominator value is used. The dropout layer has been set to 0.2. Adam optimizer function is used, and the learning rate is set at 0.0015. The whole architecture is evaluated on the Kaggle platform with the support of the Graphical Processing Unit.
Sub-word tokenization is adopted to build vocabulary and tackle Out of Vocabulary (OoV). All the encountered sub-words are divided into words recognized by the architecture. The preprocessing techniques include Punctuation, stopwords, and noise removal. No need to remove URL and emojis as we don't have this inside the dataset. Finally, Table 3.11 shows the necessary changes made to the original Bangla BERT architecture. [1]

Table 3.11: Parametric details of the proposed model

| **Parameter Name** | **Value** |
|---|---|
| Number of encoder block | 8 |
| Multihead attention | Yes (8) |
| Dropout | 0.2 |
| Learning rate | 0.0015 |
| Recurrent Dropout | 0.20 |

### 3.4.10 Common Explainable AI Models

In computer science, the black box model refers to this model that does not reveal any insightful information regarding generating the result. As a result, it becomes difficult to understand whether the result should be trusted or not. Although the results are not properly explained, these black box models have great computation

[1]Available online: https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=9852438

capability that is helpful for performing tasks that are required for the betterment of society.
Interpreting the black box model is important as most of the deep learning architectures are difficult to understand. Hence. Explainable AI is utilized to understand the results of the models. XAI maintains certain principles to comprehend the result. The cardinal purpose of using XAI is to explain the result. It is important in this research to understand what is happening inside the proposed transformer model. Understanding the optimization process and learning ability of the model is also important. To understand all of these, XAI is required. The meaningful explanation assists the researchers in further optimising the result. Another important property of XAI is that it can detect whether the model is affected by bias or not. Usually, in this research, the XAI models show us which portion of the sentence is dominated by the model while capturing sequence information. The primary advantages of XAI are stated below.

1. **Enhanced Decision Making:** XAI can play a crucial role in understanding insight information. It helps researchers to make important decisions to support further optimization. In the case of this research, this XAI model provides relevant and influential information. The words from the Bengali sentence are detected by the XAI models.

2. **Gaining Trust:** XAI models tell us what is happening inside the models. Without XAI models, we can see results but not explanations. Hence, gaining trust is not easily achievable without the aid of XAI models. With the explanation provided by XAI models, the results from DL models are more accepted.

3. **Minimizing the Amount of Risk:** AI models can be dangerous if not properly driven with an ethical approach. That is why if the model properly understands the results from each layer, it reduces the risk generated by black box models.

4. **Providing Transparency:** The main objective of XAI models is to provide transparency that would be accepted by the users of black box models. In the case of domain-specific tasks, the result must be accepted by all kinds of stakeholders.

5. **Accountability:** The ethical considerations must be taken into account by the XAI models. XAI is responsible for comprehending the result. Hence, all ethical measures must be taken into account to provide a guideline to the users.

As Bengali is a complex language and the target is to let the model learn with training data, that is why XAI models are integrated. After training the model, several XAI models are applied to the proposed framework to observe which section of the sentence is detected by the black box model. Future research must understand how the model captures sequence information from the provided data. Furthermore, to compare the results, three Explainable models are integrated here, namely SHAP, LIME and Grad-CAM. Although these XAI models are very popular in the English language, their validity in the Bengali language has yet to be examined. Apart from this, in Sentiment analysis, it is important to understand what the most valuable insight is.

1. **Local Interpretable Model-agnostic Explanations (LIME):** Local Interpretable Model-agnostic Explanations (LIME) is a methodology developed to demystify the decision-making process of advanced predictive models by offering explanations for individual predictions. This approach generates a simplified, understandable model that mirrors the complex model's reasoning on a specific instance, facilitating insight into the factors influencing a particular outcome. LIME elucidates the model's behavior by introducing slight variations to the data point and monitoring the resultant prediction variations. Through repeated modifications and applying an interpretable model to these altered data points, LIME captures the intricate decision-making process near the instance of interest. The essence of LIME lies in its ability to produce interpretable models, such as linear regressions or decision trees, which approximate and thereby unveil the rationale behind the predictions of otherwise opaque models within a narrowly defined scope around the data point under examination. Despite its utility in enhancing the transparency of model predictions at an individual level, LIME's approach is not without its challenges. The explanations it provides are local, focusing on specific instances rather than the model's overall logic, and the selection of the local interpretive model and its neighbourhood can significantly influence the accuracy and relevance of these explanations. In summary, LIME stands as a valuable instrument for shedding light on the predictive mechanisms of complex models, fostering a greater understanding and trust in their outputs by translating them into a more accessible and interpretable form.

$$\text{Equation: } \theta_{\text{next step}} = \theta_{\text{current}} - \eta \cdot \nabla_\theta J(\theta; \mathbf{x}^{(i)}, y^{(i)}) \tag{3.6}$$

   - $\xi(x)$ represents the explanation model for instance $x$.
   - f denotes the complex model being explained.
   - g symbolizes the simple, interpretable model chosen to approximate f locally, where g belongs to a family of models G.
   - L is the loss function quantifying the discrepancy between f and g in the vicinity of x., which defines this locality.
   - $\Omega(g)$ measures the model $g$'s complexity, promoting simplicity in the explanation.

2. **SHapley Addictive Explanations:** The SHAP method breaks down how individual parts of a predictive model influence its decisions. In SHAP's approach, every data in the model is examined to see how much it affects the outcome. This involves comparing what the model predicts with a particular data piece included versus what it would predict if that piece were absent across all possible combinations of data pieces. This detailed process ensures the accuracy of SHAP values, which precisely measure the impact of changing one

data piece while keeping others constant. These values are valuable because they explain each data piece's role in the model's predictions, making complex models more transparent and easier to understand. However, calculating SHAP values requires significant computational resources. This is because it must account for every potential combination of data pieces, which becomes exponentially more demanding as the number of data pieces increases. This makes the SHAP method resource-intensive, posing challenges when working with models that include many variables.

$$\mathrm{SHAP}_j(f, x) = \phi_0(f) + \sum_{i=1}^{M} \phi_i(f, x_j) \tag{3.7}$$

- $\mathrm{SHAP}_j(f, x)$ represents the SHAP value for feature $j$ in the prediction function $f$ for input $x$.
- $\phi_0(f)$ is the base value, which is the expected output of the model $f$ when no features are present.
- $\phi_i(f, x_j)$ represents the contribution of feature $i$ to the prediction for input $x_j$.

### 3.4.11 Gradient-Weighted Class Activation Mapping:

Gradient Weighted Class Activation Mapping (GradCAM) is mainly utilized for the purpose of computer vision. It highlights the regions inside a picture to find the crucial decisions of a model. It weighs the value of all the pixels and reveals insights into the detected regions. The advantages of using GradCAM is elaborated below:

1. **Improved Transparency:** Grad-CAM enhances CNN's transparency by highlighting the specific regions in the input image that contribute to the model's decision. This visualization helps users understand why and how a model has made a particular prediction.
2. **Debugging Tool:** The Grad Cam model is a very powerful tool when it comes to debugging. Irrelevant parts or predictions for images the model developers can detect and make it correct the problem arises. By deburring, the performance of this model can be improved.
3. **Model Trust and Validation:** Very crucial applications, such as medical imaging, can be used to work on GRAD Cam because of its trustworthy nature in model validation. The judgment of the decision-making process gives it an advantage.
4. **Easy Integration:** GRAD CAM can work significantly better with existing CNN models without needing to modify the network. This criteria makes it easier for developers who implement the model to work.
5. **Works with Any CNN-Based Architecture:** Versatility is a significant criterion for GRAD CAM model architectures as it can be applied to any CNN-based model and used across various domains where CNNs are employed. This includes fields like medical imaging, autonomous driving, security surveillance,

and more. This broad applicability ensures practitioners in different fields can utilize Grad-CAM to understand and improve their models.

6. **Requires No Retraining:** No training can greatly improve this model's time efficiency by saving a significant amount of data. Eliminating session times also reduces costs, which leads to resource conservation. This model is also very simple to use and consistently performs.

The flaws of GradCAM is also stated here so that it can be understood why this XAI model is not integrated in this research.

1. **Limited to Visual Data:** Nodes where images are processed it can be applied for this model GRAD CAM for image extraction tasks. However, it restricted the scope of the application. Data modalities for even structured data are excluded. There should be alternative approaches for model transparency. Moreover, this algorithm works to make it a better visual bias, which can be overlooked by the complexity and richness of this model.

2. **Dependence on Network Architecture:** The effectiveness of Grad CAM is influenced by the underlying CNN architecture. Grad CAM does not provide accurate or meaningful visualizations in some networks, especially those with complex or non-standard layers.

3. **Potential for Misleading Interpretations:** Users may misread the heatmaps produced, giving them false hope concerning what the simulation is doing. For instance, heatmaps that appear to be right can emphasize unrelated aspects that just so happen to be associated with the objective.

4. **Focus on High-Level Features:** Grad CAM may miss small but significant characteristics that previous layers may have captured since it favours emphasising only the most prominent aspects in the final convolutional layers. Higher-order components of the network's structure identify more abstract and sophisticated patterns in the input, which are referred to as high-level features in CNN contexts. Higher-level, abstract aspects are the model's main focus; finer, lower-level details, such as textures or particular edges, are frequently overlooked. Typically acquired in the initial layers of a CNN, these lower-level features.

5. **Granularity:** Even while the heatmaps have a high resolution, they frequently only show a broad area of interest rather than the granularity needed to identify the particular pixels impacting the choice.

### 3.4.12 Performance Metrics

Although numerous performance metrics have been adopted, including sustainability, authors have focused on precision, recall, accuracy, F1-score, data loss, and trainable parameters. As DL and transformer-based architectures can not be judged based on the first four criteria, other techniques have also been considered. The formulas for precision, recall, accuracy, and F1-score are below.

$$\text{Precision} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Positives}} \tag{3.8}$$

$$\text{Recall} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}} \tag{3.9}$$

$$\text{Accuracy} = \frac{\text{True Positive} + \text{True Negative}}{\text{True Positive} + \text{True Negative} + \text{False Positive} + \text{False Negative}} \tag{3.10}$$

$$\text{F1 score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{3.11}$$

### 3.4.13 Training Set Up

Table 3.12: Hardware Requirements for Training Bangla BERT

| **Component** | **Recommendation** |
|---|---|
| GPU | High-performance NVIDIA TESLA P100 GPU's |
| CUDA and cuDNN | CUDA toolkit 12.4 |
| RAM | 16 GB RAM |
| CPU | CORE I5 Gen 11 CPU |
| Storage | 512 GB |
| Cluster/Cloud | Not required |
| Monitoring Tools | NVIDIA System Management Interface occasionally |

As we are proposing an optimized Bangla BERT model, we have to use some high-configuration hardware components. While running the architecture on a Jupyter Notebook, it got disconnected without Graphical Processing Unit (GPU) support. Later, the researchers focused on running the architecture with GPU support. The latest version of Compute Unified Device Architecture (CUDA) runs the project. Finally, the whole thing runs on a CORE i5 GEN 12 computer. Table 3.12 represents the detailed hardware requirement for running the whole project.
The GPU helps us process all the work smoothly. As a result, the architecture runs faster without any decline. Previously, the architecture was attempted to be executed without the aid of the GPU, but the kernel got disconnected many times. Monitoring performance and power consumption is the primary task of the NVIDIA system management interface. With the help of this, authors have monitored all the slightest improvements made by the architecture.

# Chapter 4

# Experimental Result Analysis

## 4.1 Result Analysis of Machine Learning Models

At first, the Machine Learning models have been implemented by providing this dataset. The authors focused on applying Unigram, Bigram, and Trigram token systems to observe the results of the ML models. At first, the Unigram tokenization system was utilized. So, models are trained based on a single time at one time. From there, it can be observed the Logistic Regression has shown the best result with an F1-score of 78.46%. Although the Linear SVM has shown the highest recall, other sector results were not top-notch. Logistic regression showed higher efficiency in all three sectors, producing a higher F1 score. Figure 4.1 reflects the total summary after applying the unigram feature selection process on ML algorithms.

Here, the accuracy, precision, and recall are the average of every class available in the dataset. The original dataset has three polarities: positive, negative, and neutral. Finally, the harmonic mean is taken by the F1-score. Hence, it provided a proper result for the three models. The primary purpose of using these models is, in the state-of-the-art literature survey, that these models are only utilized for extracting features from the Bengali sentences. For extracting results from the Random Forest model, the authors have applied 100 decision trees. Later, the authors focused on applying the bigram feature selection process. In this case, two tokens

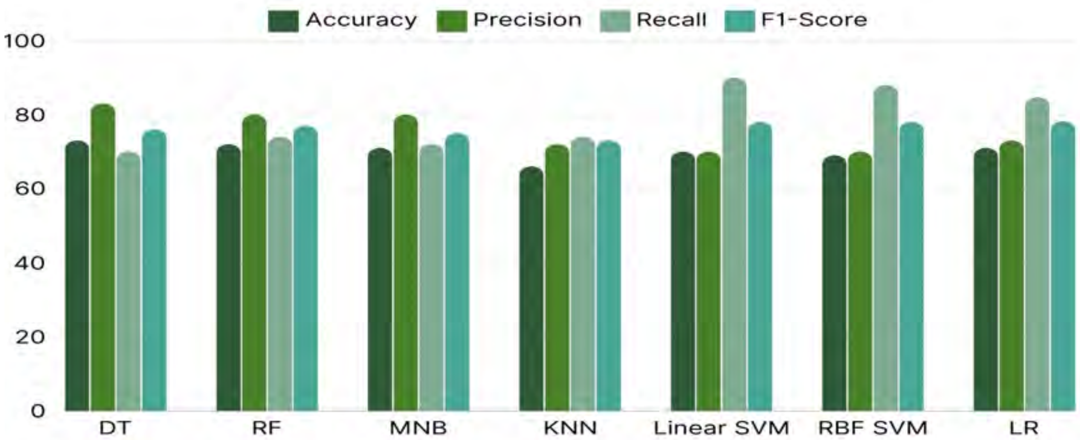



Figure 4.1: Result of ML models using Unigram Feature Selection Process

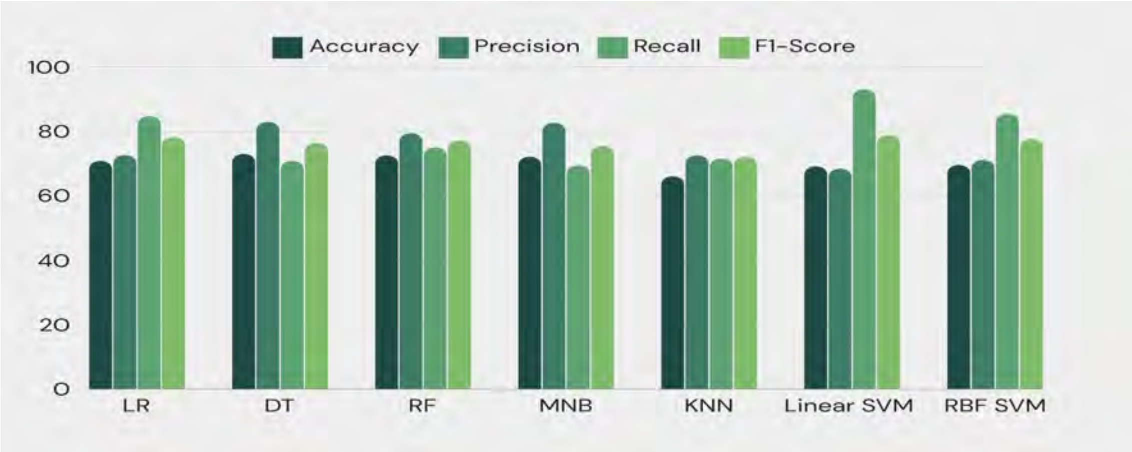



Figure 4.2: Result of ML models using Bigram Feature Selection Process

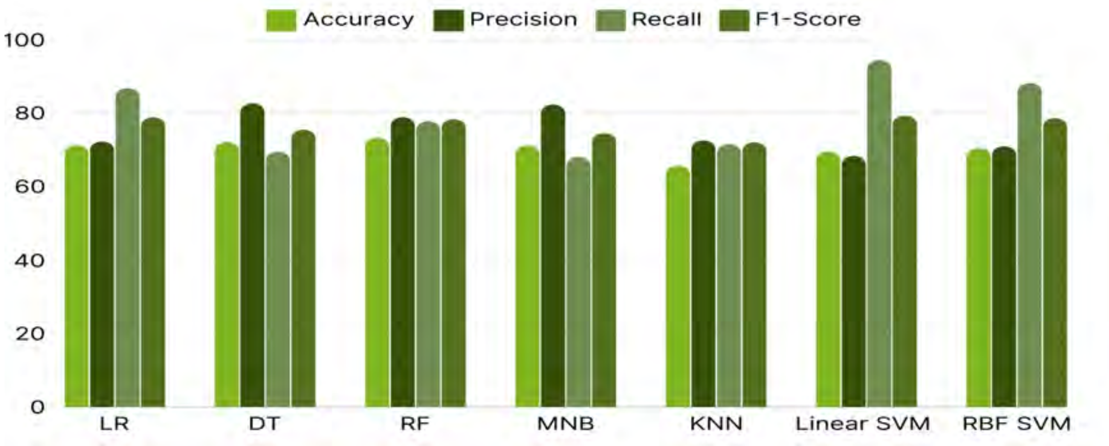



Figure 4.3: Result of ML models using Trigram Feature Selection Process

are considered for training purposes. The superiority of the Bigram model over the unigram is that it can capture local context precisely and aid in reducing sparsity. On the contrary, the unigram model is more robust to noise than the Bigram model. In the case of the Bigram feature selection process, the result is more or less the same, whereas Support Vector has shown better results than other models. The closest model that exhibits an almost similar F1-score is the Logistic regression. The detailed performance summary is reflected in Figure 4.2.

Finally, the authors focus on the trigram selection process to search whether any other model performs better than these models. The trigram token selection procedure considers three tokens simultaneously to capture semantic information. Figure 4.3 demonstrates that only Linear Regression and Support Vector perform better than other models, whereas Random Forest has also shown more impressive results than before. Almost 100 Decision trees are considered for building the Random Forest.

Figure 4.4 and Figure 4.5 show the details of the F1-score and accuracy comparison

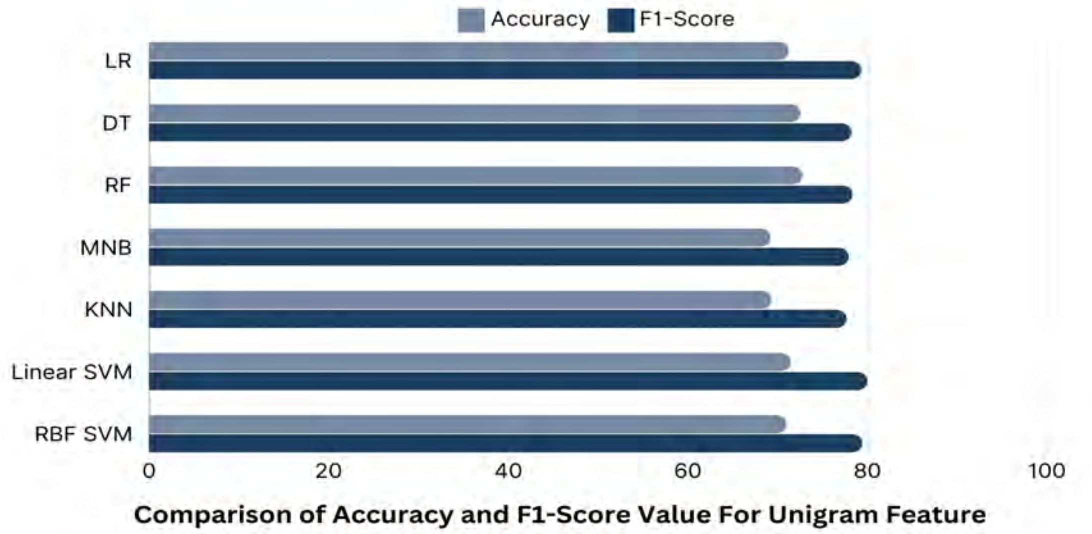



Figure 4.4: Accuracy and F1-Score Bar Diagram for Unigram Token Selection Process

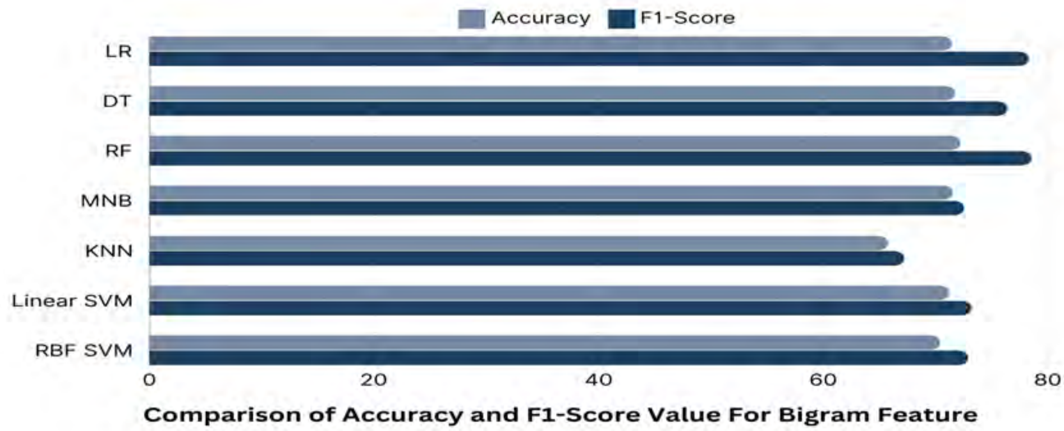



Figure 4.5: Accuracy and F1-Score Bar Diagram for Bigram Token Selection Process

of the models.

### 4.1.1 Result Analysis of the Proposed Architecture

Before understanding the result shown by the proposed model, we need to understand the result shown by the original Bangla BERT model. In total 10 epochs have been run. The primary reason for running 10 epochs are:

1. **Hardware Constraints**: Bangla BERT is a pre-trained model with enormous amount of trainable parameters. As a result, the model requires extensive memory and cloud services so that calculation can be made parallel.

2. **Convergence:** Although, the model has reached its optimal result after 10 epochs because of the size of the dataset is small. The model has understood

most the required words to understand the sentiments precisely. That is why similar type of performance metrics was observed after running more than 10 epochs. The performance of the Bangla BERT model is shown in 4.1.

Table 4.1: Training Phase Details of the Bangla BERT Model

| Epoch | Avg Precision | Avg Recall | Avg Accuracy | Avg F1-Score |
|---|---|---|---|---|
| 1 | 72.14 | 58.13 | 46.70 | 49.05 |
| 2 | 75.79 | 73.77 | 65.95 | 68.16 |
| 3 | 75.14 | 62.73 | 69.22 | 68.26 |
| 4 | 76.01 | 62.58 | 69.64 | 69.74 |
| 5 | 76.21 | 62.63 | 70.28 | 70.29 |
| 6 | 76.73 | 63.67 | 73.61 | 71.53 |
| 7 | 76.67 | 62.50 | 76.10 | 71.26 |
| 8 | 77.17 | 63.68 | 73.73 | 72.39 |
| 9 | 77.04 | 63.27 | 77.18 | 72.22 |
| 10 | 77.27 | 63.04 | 75.82 | 73.16 |

Later, the authors turned their attention to a transformer-based proposed model where a modified Bangla BERT model has been proposed. From there, a detailed analysis of the results is described in this section.
In total, 10 epochs were selected, as the result did not significantly change after 10 epochs. The results were closely observed and compared with the original Bangla BERT model. It has been found that the model did not perform well from the very first epoch. The average F1 score was only 46.41%, but it exhibited good precision of approximately 80%. The model started on the validation set with a good F1 Score of more than 69.4%.

The details of the modified Bangla BERT's result are shown in Table 4.2.

Table 4.2: Training Phase Details of the Modified Bangla BERT Model

| Epoch | Avg Precision | Avg Recall | Avg Accuracy | Avg F1-Score |
|---|---|---|---|---|
| 1 | 79.05 | 62.19 | 54.51 | 68.45 |
| 2 | 82.95 | 66.49 | 67.47 | 71.22 |
| 3 | 83.57 | 68.81 | 72.48 | 73.32 |
| 4 | 85.61 | 69.57 | 73.88 | 74.56 |
| 5 | 85.40 | 68.66 | 78.47 | 76.82 |
| 6 | 85.83 | 71.44 | 77.42 | 77.45 |
| 7 | 84.80 | 72.57 | 82.47 | 78.23 |
| 8 | 86.87 | 73.74 | 80.53 | 78.45 |
| 9 | 87.25 | 73.29 | 81.66 | 80.25 |
| 10 | 87.21 | 73.83 | 81.37 | 80.63 |

The table shows that the average accuracy was initially very low, but with each increment of epoch, the percentage started moving higher. After 8 epochs, the highest precision was found to be 86.87%, whereas the highest recall was found in the same epoch. Furthermore, the average F1 Score was found to be 78.45%.

Although these results are similar to ML models, we have shifted our focus towards the validation set.

Table 4.3: Validation Phase Details of the Modified Bangla BERT Model

| Epoch | Avg Precision | Avg Recall | Avg Accuracy | Avg F1-Score |
|---|---|---|---|---|
| 1 | 82.55 | 68.54 | 69.54 | 70.43 |
| 2 | 86.44 | 69.77 | 72.84 | 72.64 |
| 3 | 88.85 | 69.85 | 76.49 | 73.35 |
| 4 | 87.72 | 72.51 | 81.55 | 76.82 |
| 5 | 88.89 | 73.45 | 82.45 | 78.85 |
| 6 | 89.01 | 74.77 | 84.86 | 79.24 |
| 7 | 88.84 | 74.85 | 85.12 | 82.67 |
| 8 | 88.75 | 74.81 | 85.99 | 82.26 |
| 9 | 89.53 | 75.12 | 86.53 | 83.01 |
| 10 | 89.58 | 76.01 | 87.13 | 84.53 |

The validation result summary is provided in Table 4.3, where it is found that the highest F1-score is found in Epoch number seven, where the F1-score is 82.67%. Now, the factor is that precision and recall remain much higher than in the training set. So, the proposed model performs well with a validation set greater than the state-of-the-art models.

Now, focusing on the data loss, we can find the model performs really well as the dataset is smaller. Table4.4 states the training and validation data loss. Here, it can be viewed that the training loss was really higher at the beginning of the epoch, whereas the validation data loss is not that much. The fewest data loss for the training phase is found in epoch number 8. On the contrary, the same thing is found for the validation set in epoch number 7.

Furthermore, the authors are focused on the Recurrent Neural Network models, which were performed over 20 epochs. The cardinal factors are LSTM, GRU, BiLSTM, and BiGRU, which do not tend to perform well in the Bengali language rather than in the English language.

Table 4.4: Training and Validation Data Loss Comparison

| Epoch | Training Data Loss | Validation Data Loss |
|---|---|---|
| 1 | 72.35 | 33.54 |
| 2 | 44.15 | 29.73 |
| 3 | 31.53 | 12.54 |
| 4 | 15.71 | 9.86 |
| 5 | 11.63 | 4.30 |
| 6 | 8.84 | 4.19 |
| 7 | 4.95 | 3.99 |
| 8 | 4.56 | 5.65 |
| 9 | 4.51 | 5.32 |
| 10 | 4.49 | 5.15 |

Table 4.5: Performance Metrics of the RNN Architectures

| Model Name | Precision | Recall | Accuracy | F1-score |
|---|---|---|---|---|
| LSTM | 78.84% | 65.32% | 74.15% | 75.78% |
| GRU | 75.21% | 68.45% | 75.93% | 74.36% |
| BiLSTM | 72.69% | 63.87% | 72.47% | 72.15% |
| BiGRU | 77.33% | 69.91% | 76.28% | 78.02% |

The table 4.5 provides a detailed comparison of the performance metrics across four RNN models, LSTM, BiLSTM, BiGRU, and GRU, employed in a classification task. These tables compare four classification metrics.

Among the models, LSTM stands out with the highest precision score of 78.84%, showcasing its capability to classify positive instances accurately. However, this model appears to sacrifice some recall, resulting in a relatively lower F1 score. On the other hand, GRU exhibits marginally lower precision but compensates with a higher recall, leading to a comparable F1 Score.

BiLSTM demonstrates a balanced performance, with precision and recall closely aligned, resulting in a moderate F1 score. In contrast, BiGRU emerges as the top performer, boasting the highest F1 Score of 78.02%, indicating its proficiency in simultaneously achieving high precision and recall.

## 4.2 Model Interpretation Using Explainable AI

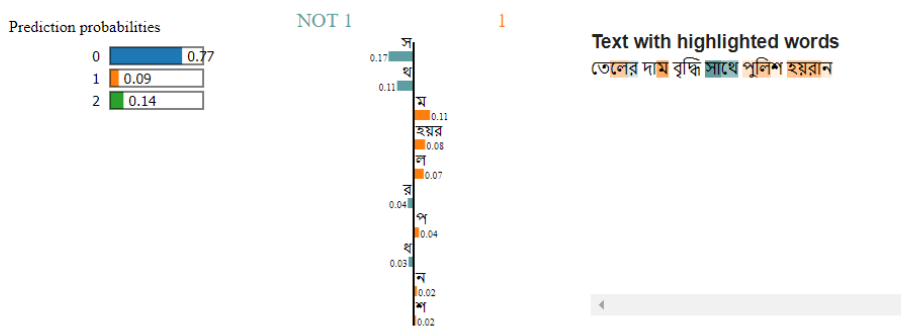



Figure 4.6: Model Interpretation Using Explainable AI

The authors focused on the testing set to understand the modified Bangla BERT model's workflow. Later, the whole process is understood with the aid of Explainable AI. In Figure 4.6, the proposed model can capture the semantics of the sentence and classify it correctly. The maximum prediction probability is assigned to the negative class, which is right in the case of the provided sentence. Furthermore, the authors have observed the performance of SHAP till now. Unfortunately, LIME also provides similar results. Although the model provides correct results, the interpretation is not as strong as LIME and SHAP, which are not built for the Bengali language. LIME and SHAP have excellent capturing capability in English, but in the case of other languages, these interpretation models fail to interpret the model properly.

In the case of GradCAM, the XAI model is suitable for working problems that is integrated with Computer Vision. As the research is mainly focused on textual data, the XAI model is eradicated from the applied XAI model list.

Additionally, the writers have been tracking SHAP's performance up to this point. Regretfully, LIME yields comparable outcomes as well. The interpretation is not as powerful as LIME and SHAP, which are not designed with the Bengali language in mind, even though the model yields accurate findings. While LIME and SHAP perform exceptionally well when capturing English, these interpretation models fall short when interpreting other languages.

### 4.2.1 Reasons Behind this Result

Primary, authors have applied the trail and error procedure in order to achieve the better result. Furthermore, to understand the working procedure behind this result. Reducing the number of encoder layers in BERT can lead to both enhanced efficiency and, under certain conditions, improved accuracy. Here's a detailed mathematical analysis explaining how and why this can happen.

## Reducing the Number of Encoder Blocks

Bangla BERT uses the BERT as its base layer. BERT is mainly focused on the number of encoder layers ($L$) available inside the model. Encoder layer can be divided into two major parts. These are:

#### Self-Attention Mechanism

One of the major tasks of the encoder blocks is to work with the self-attention blocks. Self-attention blocks allows the model to be precise while putting weight to different words in sentiment analysis tasks that is similar to this research. The complexity of the self-attention mechanism is defined by:

$$O(n^2 d)$$

where $n$ is the sequence length and $d$ is the hidden size.

#### Understanding the Feed-Forward Network

Another important part of BERT encoder block is the token representation. The token representation is done with the help of Feed-Forward Network. Two linear transformations are done using a ReLU activation function. The feed-forward network has a complexity of:

$$O(nd^2)$$

#### Overall Complexity in Each Encoder Layer

If needed, we can calculate the overall complexity:

$$O(n^2 d + nd^2)$$

### Final Complexity for All Layers

For a model with 8 layers, the total complexity becomes:

$$O(8 \cdot (n^2d + nd^2))$$

whereas in the original Bangla BERT model, the number of encoder layers were 12. So, the overall complexity in each situation can be described as:

$$O(12 \cdot (n^2d + nd^2))$$

Hence, reducing the number of encoder layers allows the model to be efficient in terms of computational efficiency.

## Explanation of Improved Accuracy

Furthermore, authors are focused on understanding the reasons behind improved accuracy. Primarily, two main reasons are identified by the authors for better accuracy.

### Reducing Overfitting issue

With fewer encoder blocks, the overfitting issue might be reduced. Now, the factor is that while using 7 encoder blocks fails to capture contextual information, the authors found out that 8 encoder blocks are precise in finding the improved result. Reducing $L$ can help prevent overfitting, leading to better generalization on unseen data.

### Stability of Training

Because there are fewer blocks, a straightforward gradient flow is shown between the transformer blocks. As a result, convergence is possible in less time and with reduced computation.

## Analyzing the Part Mathematically

As there are 8 layer, we can say 8 to 12 (with $12 < 8$).
The ratio of computational complexity between the original and reduced models is understood by efficiency gain:

$$\frac{\text{Complexity of Reduced Model}}{\text{Complexity of Original Model}} = \frac{12 \cdot (n^2d + nd^2)}{8 \cdot (n^2d + nd^2)} = \frac{12}{8}$$

So, the efficiency gain is provided with the number of lesser layers. However, if we apply fewer layers than 8, the model fails to capture contextual information. Apart from that, if we look more intensely then, Let $\mathcal{L}$ represent the loss function. For a model with 12 layers, the optimization objective is:

$$\min_{\theta} \infty\in(X, y; \theta, 12)$$

where $\theta$ represents the model parameters, $X$ is the input data, and $y$ are the labels.

For a reduced model with 8 layers, the objective is:

$$\min_{\theta'} \forall(X, y; \theta', 8)$$

where $\theta'$ are the parameters of the reduced model.
The generalization error of the first model will be $E_{\text{gen}}$ will be higher than that of the reduced model:

$$E_{\text{gen}}(12) > E_{\text{gen}}(8)$$

Hence, the reduced number of blocks allows for a generalization error. Finally, 8 encoder blocks lead to lesser computational costs along with more accuracy. Fine-tuning also improves the overall performance of the model.

# Chapter 5

# Conclusion and Future Work

The research primarily focuses on understanding ride-sharing platforms' socio-economic impact. Later, the perspective of drivers is taken into account. As a result, related data regarding this sector has been gathered. The dataset has been processed correctly and is ready to be utilized for future research purposes. Furthermore, an extensive exploratory data analysis was performed to find some insightful information. Later, a modified transformer-based model is proposed with greater accuracy than the proposed state-of-the-art architecture. Explainable AI models have been used to interpret the model, and the researchers observe the attempted result. The research is helpful from both computational and social aspects. It is essential and mandatory to explore the big market so that the financial sector of Bangladesh can stay well organized.

As a developing country, Bangladesh's economy is not stable yet. Ride-sharing platforms contribute significantly on this sector. Analyzing the perspective of riders will not only provide an economic analysis but reflect a certain community's behavioural pattern. From the economic point of view, the dataset reveals that most of the words are related to economics. The dataset is also annotated precisely for further exploration in the concerned sector. The authors also proposed an optimized Bangla-BERT model where fewer trainable parameters are required because of fewer attention layers. The result analysis reflects the model is computationally efficient after observing the classification report and other performance metrics.

It is observed the proposed model outperforms all the state-of-the-art architecture with at most 82.86% average accuracy. On the other hand, the average precision is found to be 89.01%. High precision means most of the positive predictions made in the model are appropriate. On the contrary, if we focus on the RNN-based architectures, we can observe the best result shown by the BiGRU model. The F1-score was 78.02%. The self-attention mechanism and hyperparameter tuning allow the model to be more flexible. 70% of the gathered data were utilized for training purposes.

## 5.1 Limitation of the Research

Although this research is vital in numerous aspects, multiple shortcomings have yet to be resolved. Some of the shortcomings are discussed below:

1. Understanding the user's or passengers' perspective is also important. Discussing with two parties would have allowed the research to be more fruitful.

2. Transformer models demand enormous storage and computationally efficient hardware settings. As a developing country, it is difficult to utilize the full capacity of such models.

3. The research focuses on certain perspectives where other factors such as subsidy, reasons for price hikes and unfair behaviour of riders are not considered.

4. Although the primary focus of this research was to develop an optimized transformer-based model to detect sentiments properly. However, to further upgrade the model, a substantial number of privacy measures should be taken, such as data encryption and compliance, and they should be strictly followed.

## 5.2 Future Work

The authors are focused on improving the research in certain ways. Ride-sharing apps operate in a dynamic technological landscape where advancements continually reshape the industry. Understanding socio-economic factors is important to understanding a certain population's behavioural shift. After analyzing socio-economic factors, crimes, employment opportunities, economic growth, and community development can be understood.

1. A larger dataset will be built, integrating the users' sentiments. If we observe precisely, the proposed model will be more fruitful with increasing data points. The focus is to make the model pre-trained on one of the largest datasets and save weight. Later, those weights will be useful to understand a different dataset. Hence, the improvement in numerous performance metrics will be observed properly.

2. Many factors will be incorporated, and cloud platforms will be utilized for training and data storage. Training the transformer models requires extensive data storage and hardware support. So, utilizing the cloud source is the best example here. The authors will focus on training the architecture on a suitable cloud platform such as AWS.

3. As Bengali is an ambiguous language, the length of the sentence is a vital factor. In the future, large sentences will be considered for training and validation. More focus will be placed on that sector, such as understanding the cultural language pattern more precisely. Training the model on speech signals is also another priority. In such cases, the model should be trained on Bengali speech data.

4. Future work is necessary to adapt analysis methodologies to evolving technologies, such as improvements in GPS tracking and mobile computing capabilities. This ensures that analyses remain relevant and effective in understanding the changing dynamics of ride-sharing services.

5. To understand the socio-economical factors properly, other vehicle drivers, such as CNG drivers, Rickshaw pullers, and Bus drivers, should also be integrated. Due to the recent introduction of Metro Rail in Bangladesh, bus drivers are facing some ups and downs in their likelihood. These things can be incorporated to make a larger dataset. Hence, the dataset will be analyzed based on numerous aspects.

# Bibliography


[1] N. Agatz, "Sustainable passenger transportation: Dynamic ride-sharing," vol. 10, Feb. 2010. DOI: https://doi.org/10.1007/s40890-022-00160-9. [Online]. Available: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1568676.

[2] N. H. Rafiul Islam, "Mathematical modeling and projecting population of bangladesh by age and sex from 2002 to 2031," 1, vol. 13, 2014, pp. 53–60. DOI: https://doi.org/10.1007/978-94-017-8990-5_5. [Online]. Available: https://link.springer.com/chapter/10.1007/978-94-017-8990-5_5.

[3] E. Shmueli, I. Mazeh, L. Radaelli, A. Pentland, and Y. Altshuler, "Ride sharing: A network perspective," in *Social Computing, Behavioral-Cultural Modeling, and Prediction: 8th International Conference, SBP 2015, Washington, DC, USA, March 31-April 3, 2015. Proceedings 8*, Springer International Publishing, 2015, pp. 434–439.

[4] P. Goel, L. Kulik, and K. Ramamohanarao, "Optimal pick up point selection for effective ride sharing," *IEEE Transactions on Big Data*, vol. 3, no. 2, pp. 154–168, 2016.

[5] A. Baj-Rogowska, "Sentiment analysis of facebook posts: The uber case," in *2017 Eighth International Conference on Intelligent Computing and Information Systems (ICICIS)*, IEEE, Dec. 2017, pp. 391–395.

[6] C. Lee, M. Rahafrooz, and O. K. D. Lee, "What are the concerns of using a ride-sharing service?: An investigation of uber," 2017.

[7] A. Surie, "Are ola and uber drivers entrepreneurs or exploited workers," *Economic and Political Weekly*, vol. 53, no. 24, pp. 1–7, 2018.

[8] E. Sthapit and P. Björk, "Sources of value co-destruction: Uber customer perspectives," *Tourism Review*, vol. 74, no. 4, pp. 780–794, 2019.

[9] M. O. R. Tarek and S. A. J. I. D. Amit, "A closer look at the sharing economy, global ridesharing and ridesharing in bangladesh," *ULAB Thought Leadership Article, Centre for Enterprise and Society*, 2019.

[10] Y. Wang, J. Gu, S. Wang, and J. Wang, "Understanding consumers' willingness to use ride-sharing services: The roles of perceived value and perceived risk," *Transportation Research Part C: Emerging Technologies*, vol. 105, pp. 504–519, 2019.

[11] E. Yunus, D. Susilo, S. Riyadi, M. Indrasari, and T. D. Putranto, "The effectiveness marketing strategy for ride-sharing transportation: Intersecting social media, technology, and innovation," *Entrepreneurship and Sustainability Issues*, vol. 7, no. 2, p. 1424, 2019.

[12] R. Karim, “Influence of e-service quality on customer satisfaction & word of mouth in app-based service industry: A case on pathao, bangladesh,” *Journal of Technology Management and Business*, vol. 7, no. 1, pp. 36–49, 2020.

[13] J. Shahid, S. N. Ahsana, S. A. Suma, M. I. Jahan, M. A. Raihan, and M. Hadiuzzaman, “A structural equation modeling approach to analyze perceived service quality of ride sharing service in bangladesh: A pathao case study,” in *1st International Conference on Transportation Research*, Nov. 2020.

[14] X. Yan, X. Liu, and X. Zhao, “Using machine learning for direct demand modeling of ridesourcing services in chicago,” *Journal of Transport Geography*, vol. 83, p. 102 661, 2020.

[15] R. Chowdhury, “Impact of marketing communicators, service quality and price promotion on the brand image: A case study on pathao,” 2021.

[16] T. Dey, M. A. Salam, and T. Saha, “Evaluation and analysis of user satisfaction of ride-sharing service: An assurance and empathy in bangladesh perspective,” *Can. J. Bus. Inf. Stud*, vol. 3, no. 2, pp. 22–28, 2021.

[17] A. K. Lambros Mitropoulos and G. Ayfantopoulou, “A systematic literature review of ride-sharing platforms, user factors and barriers,” vol. 10, Jul. 2021. DOI: https://doi.org/10.1186/s12544-021-00522-1. [Online]. Available: https://etrr.springeropen.com/articles/10.1186/s12544-021-00522-1.

[18] N. F. Sadia Afroj, “The who, when and why of uber trips in dhaka: A study from users’ perspective,” vol. 8(25), Jun. 2022. DOI: https://doi.org/10.1007/s40890-022-00160-9. [Online]. Available: https://link.springer.com/article/10.1007/s40890-022-00160-9.

[19] F. H. Sharin, I. Sentosa, and R. K. Perumal, “A chaotic or orderly digitalization?: Malaysia’s resilient model for sustainable rural tourism,” in *Digital Transformation in Aviation, Tourism and Hospitality in Southeast Asia*, Routledge, 2022, pp. 107–129.

[20] S. A. R. Tonmoy Paul Rohit Chakraborty, “Impact of covid-19 on mode choice behavior: A case study for dhaka, bangladesh. transportation research interdisciplinary perspectives,” vol. 15, Sep. 2022, pp. 22–29. DOI: https://doi.org/10.1016/j.trip.2022.100665. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2590198222001257.

[21] S. S. Umesh Kumar Chun Ta-Li and S. K. Pani, “Design and implementation of an ml and iot based adaptive traffic-management system for smart cities,” vol. 22(8), Apr. 2022. DOI: https://doi.org/10.3390/s22082908. [Online]. Available: https://www.mdpi.com/1424-8220/22/8/2908.

[22] S. Feng, “Job satisfaction, management sentiment, and financial performance: Text analysis with job reviews from indeed.com,” vol. 22(8), Apr. 2023. DOI: https://doi.org/10.3390/s22082908. [Online]. Available: https://www.mdpi.com/1424-8220/22/8/2908.

[23] T. Islam, A. Kundu, R. J. Lima, *et al.*, “Review analysis of ride-sharing applications using machine learning approaches: Bangladesh perspective,” in *Computational Statistical Methodologies and Modeling for Artificial Intelligence*, CRC Press, 2023, pp. 99–122.

[24] K. M. H. Khadija Akter Lima and S. Azam, "A novel data and model centric artificial intelligence based approach in developing high-performance named entity recognition for bengali language," vol. 10, Sep. 2023. DOI: https://doi.org/10.1371/journal.pone.0287818. [Online]. Available: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0287818.

[25] R. mitra.b and Masud, "Assessing demographic and economic vulnerabilities to sea level rise in bangladesh via a nighttime light-based cellular automata model," *Scienctific Reports*, vol. 13, no. 1, pp. 124–129, 2023. DOI: https://doi.org/10.1038/s41598-023-40329-9. [Online]. Available: https://www.nature.com/articles/s41598-023-40329-9.

[26] M. Z. Priasa, "A case study on pathao: Scaling technology-based solutions and social impact in urban bangladesh," *Journal of Management & Educational Research Innovation*, vol. 1, no. 4, pp. 1–14, 2023.

[27] M. M. Shaharuddin, "Controlling traffic congestion in urbanised city: A framework using agent-based modelling and simulation approach.," vol. 12, Jul. 2023. DOI: https://doi.org/10.3390/ijgi12060226. [Online]. Available: https://www.mdpi.com/2220-9964/12/6/226.

[28] T. A. Syed Abdullah Al-Nahid Taskin Jamal, "Additive linear modelling and genetic algorithm based electric vehicle outlook and policy formulation for decarbonizing the future transport sector of bangladesh.," vol. 136, Jun. 2023, pp. 21–46. DOI: https://doi.org/10.1016/j.tranpol.2023.02.005. [Online]. Available: https://www.sciencedirect.com/science/article/abs/pii/S0967070X23000306.

[29] Y. Z. Yuhan Guo and Y. Boulaksil, "Modelling and analysis of online ride-sharing platforms – a sustainability perspective," Jan. 2023, pp. 577–595. DOI: https://doi.org/10.1016/j.ejor.2022.04.035.

[30] N. Siddiqua and M. B. A., "Identification of the factors affecting consumers' usage of smartphone-based delivery service apps: An empirical study in bangladesh,"