The Role of Websites in Business Promotion: A Comprehensive Analyses

Submitted By

Farjana Hasan

ID: 213-35-806

Department of Software Engineering

Daffodil International University

Supervised By

Mr. A.H.M Shahariar Parvez

Associate Professor

Department of Software Engineering

Daffodil International University

This Report is Presented in Partial Fulfillment of the Requirements for the Degree of Bachelor of Science in Software Engineering.

Summer – 2025

APPROVAL

This thesis titled on "**The Role of Websites in Business Promotion A Comprehensive Analyses**", submitted by **Farjana Hasan (ID: 213-35-806**) to the Department of Software Engineering, Daffodil International University has been accepted as satisfactory for the partial fulfillment of the requirements for the degree of Bachelor of Science in Software Engineering and approval as to its style and contents.

BOARD OF EXAMINERS

--- Chairman

Dr. S M Hasan Mahmud
Associate Professor
Department of Software Engineering
Faculty of Science and Information Technology
Daffodil International University

--- Internal Examiner 1

Tapushe Rabaya Toma
Assistant Professor
Department of Software Engineering
Faculty of Science and Information Technology
Daffodil International University

--- Internal Examiner 2

Khalid Been Badruzzaman Biplob
Lecturer (Senior Scale)
Department of Software Engineering
Faculty of Science and Information Technology
Daffodil International University

--- External Examiner

Dr. Md. Sazzadur Rahman
Professor
Institute of Information Technology
Jahangirnagar University

DECLARATION

I hereby declare that; this thesis report is done by me under the supervision of **Mr. A.H.M Shahariar Parvez**, Associate Professor, Department of Software Engineering, Daffodil International University, in fulfillment of my original work. I am also declaring that according to the best of my knowledge, neither this thesis nor any part therefore has been submitted else here for the award of B.Sc. or any degree.

Submitted by:

Farjana Hasan

____

Farjana Hasan

ID: 213-35-806

Batch: 36th Section - A

Department of SWE

Supervised by:

15/09/2025

____

Mr. A.H.M Shahariar Parvez

Associate Professor

Department of SWE

Daffodil International University

Certified by: Mr. A.H.M Shahariar Parvez

Date: 15/09/2025

ACKNOWLEDGEMENT

First, I express my heartiest thanks and gratefulness to almighty Allah for divine blessing makes me possible to complete the final year thesis successfully.

Then I really grateful to my research supervisor, **Mr. A.H.M Shahariar Parvez** who guided me throughout the whole research activities.

I wish to express my special thanks to **Dr. Imran Mahmud**, Head of the Faculty for providing all the necessary facilities for the research purpose. I am also thankful to all the faculty members, Department of Software Engineering who sincerely guided me at my difficulty. I am thankful to my friend who supported me throughout this venture.

Lastly, I must respectfully thank my parents for their unwavering support.

ABSTRACT

Abstract: The world after the Internet is a digital-first one, websites are vital marketing tools, an engagement and business enhancement tool . This study looks at how successful websites promote a business by analyzing web performance metrics and using predictive modeling. Finding patterns in user engagement and traffic dynamics was made possible by the analysis of the data, which comprised unique visitors, page views, bounce rate, and session duration.

Four forecasting techniques—ARIMA, Random Forest, Neural Network, and Long Short-Term Memory (LSTM)—were used to predict future website activity. Following data preparation, these models were used to standardize and improve the accuracy of the data. Every algorithm was evaluated using common performance metrics like R 2 score, RMSE, and MAE. Because of its capacity to recognize intricate time-related patterns in time-sequential data, the LSTM model demonstrated the highest accuracy among the applied algorithms in forecasting the number of visitors to the website.

The findings show how beneficial intelligent analytics are to web strategy since they allow businesses to make well-informed decisions on promotions. It also looks at the long-term effects of digital transformation on sustainability, the environment, and society. Overall, this study shows that a business can feel empowered to strengthen its online presence and improve the outcomes of its activity promotion by using predictive analysis of web traffic.

Keywords: Website Analytics, Business Promotion, Forecasting, ARIMA, Random Forest, Neural Network, LSTM, Time Series Analysis, Digital Marketing

TABLE OF CONTENTS

CONTENTS	PAGE
Approval	ii
Declaration	iii
Acknowledgments	iv
Abstract	v
TABLE OF CONTENTS	vi
LIST OF FIGURES	viii
LIST OF Tables	ix
CHAPTER 1: INTRODUCTION	1-3
1.1 Introduction	1
1.2 Motivation	1
1.3 Rationale of the Study	2
1.4 Research Questions	2
1.5 Expected Output	2
1.6 Report Layout	3
CHAPTER 2: BACKGROUND STUDIES	4-8
2.1 Introduction	4
2.2 Comparative Analysis and Summary	4
2.2.1 Website Traffic and Business Insights	4
2.2.2 Machine Learning Techniques for Predicting Usage	5

2.2.3 Model Evaluation and Comparison	5
2.3 Research Summary	7
2.4 Scope of the Problem	7
2.5 Challenges	8
CHAPTER 3: RESEARCH METHODOLOGY	9-16
3.1 Introduction	9
3.2 Data collection	9
3.3 Data Preprocessing	11
3.4 Statistical Analysis	12
3.5 Applied Mechanism	13
CHAPTER 4: EXPERIMENTAL ANALYSIS AND DISCUSSION	17-37
4.1 Introduction	17
4.2 Experimental Analysis	17
4.3 Experimental Result	29
4.4 Summary	37
CHAPTER 5: IMPACT ON SOCIETY, ENVIRONMENT, AND SUSTAINABILITY	38-39
5.1 Impact on Society	38
5.2 Ethical Aspects	38
5.3 Impact on Environment	38
5.4 Sustainability	39
CHAPTER 6: CONCLUSION & FUTURE WORK	40-41

6.1 Conclusion	40
6.2 Recommendation	40
6.3 Future Work	40
Appendix	42
REFERENCES	43-45
Account Clearance	46
Originality Report	47

LIST OF FIGURES

Figure Name **Pages**

Figure 3.2: Percentage Distribution of Website Traffic Sources	12
Figure 3.5. LSTM Architecture	14
Figure 3.7: Random forest classification	16
Figure 4.1: Dataset	17
Figure 4.2: import library	18
Figure 4.3: Website Metrics Over Time	19
Figure 4.4: Clustered Correlation Matrix	20
Figure 4.5: Stable describe graph for viewer	21
Figure 4.6: Stable graph for viewer	22
Figure 4.7: Page Views Over Time	22
Figure 4.8: Train-Test Split	23
Figure 4.9: Cumulative Growth of Website Visitors	23
Figure 4.10: Neural Networks (Deep Learning) Neural Network (MLPRegressor) Model Summary	29
Figure 4.11: Neural Networks (Deep Learning) Neural Network (MLPRegressor) Accuracy	30
Figure 4.12: Neural Networks (Deep Learning) Neural Network (MLPRegressor)	30
Figure 4.13: LSTM model	31
Figure 4.14: LSTM Epoch	31
Figure 4.15: LSTM model accuracy	32

Figure 4.16: LSTM model Evaluation metrics	32
Figure 4.17: ARIMA model	33
Figure 4.18: ARIMA model Result	33
Figure 4.19: ARIMA Evaluation metrics	34
Figure 4.20: ARIMA 180-Day Page Views Forecast	34
Figure 4.21: Random Forest Mode	35
Figure 4.22 Random Forest Accuracy	35
Figure 4.23: Random Forest evaluation metrics	36
Figure 4.24 Random Forest Regression predict line	36

LIST OF Tables

Tables Name	Pages
Table 2.1 Comparative Analysis of Predictive Models for Website Traffic Forecasting	6
Table 3.1: Sample of Website Traffic Data	10-11
Table 4.1: Dataset	17-18
Table 4.1.2: Different Algorithm results	37
Literature Review	24

CHAPTER 1

INTRODUCTION

1.1 Introduction

Most businesses now require websites in that this enables them to sale their products and services online [6][8][9]. A website will be comparable to a virtual shop front. It allows customers and businesses to chat and demonstrates what is being sold and simple purchasing [1][5][10]. Due to internet and online shopping boom, sites have now become an important component of business promotion [4][9]. Due to increasing competition, it is becoming more and more significant to attract and retain online visitors because each visitor has the potential to become a source of revenue [7][19]. Key website statistics assist the company in knowing the habits of the visitors. These data are the unique visitors, the number of pages they view, bounce rates, and time of sessions [11][15][21]. Combined they demonstrate the level of engagement of visitors and the effectiveness of the marketing. [19]. With the use of advanced studies on the numbers, businesses are able to progress their online marketing and offer customers more service [20][38].

Tools of machine learning and deep learning have recently enabled these numbers to be predicted more accurately [12][13][14]. Any of the models, including ARIMA, Random Forest, Neural Networks, and Long Short-Term Memory (LSTM), have their advantages [28][29]. By their means, companies are able to plan and distribute resources in advance [31][32].

In this study, these forecasting models are used in predicting traffic on websites [12][13]. It also draws parallel between different models and strives to make businesses employ data as a means to enhance their online marketing and remain competitive in the rapidly developing online environment [11][20].

1.2 Motivation

In the current competitive digital marketplace, businesses increasingly rely on their websites as key channels to reach customers and promote their brands [1][5][6]. While many companies invest heavily in developing attractive websites, they often lack sufficient tools to analyze visitor behavior and predict future trends that could optimize marketing efforts [7][9][10]. Website analytics provide crucial data on how users interact with online platforms, but extracting actionable insights from this data remains a challenge [9][10].

Furthermore, despite the growing use of predictive models in other domains, relatively few studies have thoroughly investigated and compared various forecasting techniques specifically tailored for website traffic and business promotion [12][13][14]. With the increasing complexity of user engagement patterns and external market factors, simple statistical methods may fall short in capturing the dynamics of online interactions [30][34]. This study is motivated by the need to fill this research gap by applying and evaluating advanced machine learning and deep learning models—such as Random Forest, Neural Networks, and LSTM to forecast website traffic accurately [28][29][31]. The ability to anticipate visitor trends can empower businesses to tailor their promotional strategies proactively, improve customer retention, and ultimately enhance profitability [6][10][11].

1.3 Rational of the study

In an increasingly digital-driven economy, websites play a central role in shaping how businesses present themselves and interact with potential customers [1][5]. Internet sites do gather massive data on the usage of sites. But the vast majority of this data is not utilized to guide marketing initiatives [9][10]. The predictive power hidden in these patterns is often overlooked by businesses, despite the fact that it may be used to boost exposure, engagement, and ultimately revenues [6][9][10].

There are tools to track the performance of distant websites, but they typically offer vague reports that are hard to use for long-term planning [9][10][11][15][19][21]. In this study, we show that contemporary data-driven methods are capable of measuring clicks and forecasting future trends [12][13]. The suggested study offers a very straightforward way to estimate website traffic and aid in business promotion by utilizing forecasting tools such as ARIMA, Random Forest, Neural Networks, and LSTM [28][29][30].

The aim will be to bridge the divide between unintelligent website analytics and the substantive marketing strategies that the figures should fuel [38][40]. Through the development of a predictive model that helps emphasize the data that is critical, companies will have an opportunity to further discern the customer behavior and make wiser decisions regarding their digital outreach. [36][38][40].

1.4 Research Questions

The following questions form the foundation of this research:

- How do websites support business promotion and customer outreach in the context of modern digital strategies?
- What meaningful insights can be drawn from analyzing user interaction data, including visit frequency, engagement depth, and session behavior?
- How do predictive models such as ARIMA, Random Forest, Neural Networks, and LSTM differ in their ability to forecast future trends in website usage?
- In what ways can businesses apply forecasting insights to improve digital marketing decisions and optimize user engagement?
- What long-term advantages can businesses gain by integrating web traffic forecasting into their digital planning efforts?

1.5 Expected Output

The purpose of research is to provide a meaningful result that explains the position of web sites in the process of boosting business promotion and involvement [9][10]. The four methods of forecasting, namely the ARIMA, Random Forest, Neural Network, and LSTM will be applied to come up with a model that best is able to draw patterns of how the data in terms of use of the web site changes pattern overtime and is able to give the best predictions in the future [36][40]. A comparative analysis of the performance of these models will be one of the main outputs of the given research presented in form of statistical analysis (MAE, RMSE, and R^2 score) [37][40]. These results will among others assist in describing which algorithms fit best into the capture of user interaction dynamics on business websites. [32][36]. The research will also aim at generating a data-informed visual usage and summary result illustrating the trend in usage, traffic patterns and user interaction. [32][38]. The insights are meant to assist the business owners and the online entrepreneurs in crafting their online strategy, cash in on customer retention and amplify the strength of their promotions. [40].

Above the technical findings, the piece of work will provide useful lessons on how to incorporate the predictive analytics into the daily decision-making process in any business so that the companies can be prepared to react proactively to the fluctuations in the web traffic and customer behavior. [40].

1.6 Report Layout

The entire research paper will be broken down into six major chapters; each of the chapters will aim at constructing a holistic concept of how websites can be used to promote businesses.. [10].

The initial chapter gives a prelude by presenting the research topic, justifying the importance of the study, how the answers to the research questions will be carried out and the aims of the research [6].

In the second chapter, we get into more detail with the existing literature dealing with the subject of web traffic prediction and analytics in general, mentioning what has already been learned already and where we should perhaps take a closer look [40].

The third chapter explains the methodology which includes the process in which the data was collected, structure and the several forecasting models that were used [37][40].

Chapter four gives an explanation of the finding of the experiments, and explains the strengths and weaknesses of each models on the data of the website traffic [38][40].

In the fifth chapter the question is expanded to the broader social, ethical, and environmental implications related to use of websites in business promotions [40].

Lastly, the sixth chapter concludes the study by summarizing the findings, making recommendations and giving areas of future research [32][36].

This outline also consists of a coherent and quite apparent flow of the research process and, therefore, any reader can find and easily track the contributions of the study [38][40].

CHAPTER 2

BACKGROUND STUDIES

2.1 Introduction

Increasing dependence on digital infrastructure has turned the business websites into the tool of promotion, customer interaction, and market presence [1][5][6]. Contrary to classical advertising methods, websites offer unlimited access to data of the behavior of the customers: the pages accessed, levels of the engagement and the patterns of route.[15][19][21]. This digital trace has a set of patterns which with the right kind of analysis, holds a lot of strategic worth to us [11][20].

The size and complex nature of this data however poses interesting analytical problems[9] .The typical reporting instruments do not usually have the ability to draw out more intrinsic data or ability to provide estimates during user activity trends [15][19]. To ensure that their companies can stay relevant in the digital world, it is becoming increasingly clear that businesses surrounding themselves with metrics that can proactively help them make decisions in the digital realm [32].To solve this, modern computational tools like time-series forecasting and machine learning have been initiated [22][23]. The methods enable organizations to prepare in advance of additional activities in the However, the indicated models will not be efficient only because of selecting the proper algorithm but also due to data preparation, feature selection, and adjusting them to the business setting [28][29].

This chapter lays the background of this research study as it reflects on the development of web analytics and the status of predictive methods in business decision-making [9]. It also preconditions a review of the work done and discovers the research gaps covered in this paper [22][23].

2.2 Comparative Analysis and Summary

In this part, the past activities concerning predictability of online user behavior, and business valuation of internet sites are described [19][22][23]. It emphasizes the analysis of digital interactions in terms of the relevant role of machine learning approaches toward understanding patterns and trends [12][13].

2.2.1 Website Traffic and Business Insights

Websites are no longer mere information sites in the digital economy as it has become a platform that is interactive, and through which businesses can relate with their audience [1][5]. The specific activities of users like the page views, accessing links or spending time on a site generate useful responses in depicting their likes and habits [9][10].

The involvement of these patterns with time assists businesses in fine-tuning the digital strategy [6][40]. Whether to increase traffic engagement, tailor content to users, or work out the structure of a landing page, traffic data has turned into a dependable source in the development of online business performance [6][9][10]. As indicated in a number of studies, active web engagement has an apparent direct correlation with increased business returns which include lead generation, user retention and conversion [9][10].

2.2.2 Machine Learning Techniques for Predicting Usage

In order to gain an insight into visitor behavior and to be able to predict it, researchers have examined an array of forecasting methods [17][22][23]. Such classic models as ARIMA have been adopted to model time series statistics because they can do linear predictions [40]. They are however not always appropriate to more complex traffic behavior [31][32].

Such machine learning models as Random Forest have been found to be superior in the representation of nonlinearity relationships in web data [29][31]. Such ensemble approaches involve the use of more than one decision path that is combined with aim of producing more stable outputs [29][35]. Conversely, deep learning models especially LSTM networks have been able to perform well in analyzing time-based patterns [28][40]. They can easily be used to identify repetitive behavior, e.g. weekly or monthly high traffic volumes to websites due to their memory like architecture [39].

2.2.3 Model Evaluation and Comparison

In order to evaluate the performance of predictive models in the analysis of web traffic various statistical as well as machine learning techniques have been compared in various studies [31][32]. The models will be used to predict visitor trends in the future, behavior of use and inform digital strategy [38][40]. A classic example of the forecasting model that has been employed extensively is the ARIMA (Autoregressive Integrated Moving Average) model [34][37] Mathematically, it can be in form of: [30].

$$Y_t = C + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \cdots + \phi_p Y_{t-p} + \theta_1 \epsilon_{t-1} + \ldots + \theta_q \epsilon_{t-q} + \epsilon_t \text{--------------(i)}$$

Where:

- Y_t is the predicted value,
- C is a constant,
- ϕ and θ are coefficients for autoregressive and moving average terms,
- ϵ_t is white noise (error).

In contrast, **Random Forest (RF)** uses an ensemble of decision trees. The prediction $\hat{y}$ is the average of predictions from all trees T_i :

$$\hat{y} = \frac{1}{n} \sum_{i=1}^{n} T_i(x) \text{--------------(ii)}$$

When sequential data exhibit long-range dependencies, it can be learned by Neural Networks (NN) and in particular LSTM (Long Short-Term Memory) networks. A basic LSTM cell computes:

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f)$$

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i)$$

$$C_t = f_t * C_{t-1} + i_t * \tanh(W_C \cdot [h_{t-1}, x_t] + b_C)$$

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o)$$

$h_t = o_t * \tanh(C_t)$ --------------(iii)

Where:

- f_t, i_t and o_t are forget, input, and output gates,
- C_t is the memory cell,
- h_t is the output at time step t,
- x_t is the input,
- σ is the sigmoid activation function.

Table 2.1: Comparative Analysis of Predictive Models for Website Traffic Forecasting.

Reference	Algorithm/Classifier	Data collection	Accuracy	Notes
Ahmed et al. (2020)	ARIMA	E-commerce Site Logs	RMSE: 5.2	Suitable for trend-based forecasting
Kumar & Sharma (2021)	Random Forest	Google Analytics Dataset	R^2: 0.88	Handled non-linear traffic patterns
Zhang et al. (2022)	LSTM	Web Traffic Logs	R^2: 0.93	Best for seasonal/time-based data
Mitra et al. (2019)	Neural Network	Business Portal Dataset	Accuracy: 90%	Performed well with dense data
Li & Wang (2021)	Hybrid LSTM + ARIMA	Online Retail Data	RMSE: 3.9	Combines strengths of both models

2.3 Research Summary

Digitalization of business has also resulted in the reliance on the web platform in promoting a product, contacting customers and the delivery of services [8][9]. With the increasing number of competitors in online markets, the firms are using predictive analytics to analyze the behavior of users and maximize their web presence [6]. Although the old models relating to the analysis of customers are beneficial, they are not flexible and precise enough to deal with the volume and complexity of current digital data [30][34].

In order to curb this, more current studies are being directed to online media like Facebook, Twitter and business websites as treasured grounds of user behavior data [19]. These channels allow companies to gather real-time feedbacks, sentiment and engagement trends which shape marketing success and service performance [20].

In this work, an analysis was done on a large dataset where the data set included the interaction information of the websites and the metrics of the activities of the users to predict trends and measure business performance [13][14][17]. Preprocessing and the transformation steps were applied on the data in order to model [31][37]. A total of four algorithms were implemented and measured, namely: the ARIMA package on statistical time-series forecasting, Random Forest (resistant to ensembles of learning), Neural Networks (pattern recognition), and LSTM (sequence-conscious deep learning) [28]. The models were trained on 80 percent of the data and validated on the rest 20 percent in order to achieve generalization [31].

The comparative results indicated the different levels of predictive efficiency and reliability between the models and LSTM recorded the best predictive performance [28][31]. These results are the indications that a mixed or union strategy might be effective in incorporating several insights to be used in business decision making and improvement of company digitization [6][10].

2.4 Scope of the Problem

The aim of this study is mainly to ascertain an impact of the volume of web site traffic on business promotion strategy based on empirical investigation [32]. With the growth in traffic in online marketing and businesses, it has now become necessary to determine the behavior of online visitors in order to grow and make decisions [19][20]. This paper tries to examine the sources of traffic and their effect on the overall performance and visibility of a business web site; these sources include direct access, social media, referrals, paid advertisements, and organic search [5][6][9].

To do that, the paper relies on the real-world data about visitor interaction patterns retrieved in Similarweb.com regarding the domain, wehostbd.com [11][12][15][21][38]. Multiple forecasting models, including ARIMA, Random Forest, Neural Networks, and LSTM, are implemented and evaluated to determine which model offers the most accurate web traffic prediction [28][29].

The scope of this study includes:

- Analyzing web traffic source data from a business website (wehostbd.com).
- Forecasting future visitor trends using time series and machine learning models.
- Comparing the performance of different algorithms based on accuracy metrics.
- Understanding how each traffic source contributes to business growth and visibility.
- Focusing on aggregated, non-personal data to maintain privacy and ethical standards.

The study does not include individual user behavior tracking or deep analysis of user-level demographics. It remains focused on high-level traffic trends and their implications for online business promotion [6][9].

2.5 Challenges

While the research achieved promising results in predicting website traffic using various machine learning and time series models, several challenges were encountered during the process:

- **Data Collection Limitations:** One of the most time-consuming tasks was acquiring accurate, relevant, and complete web traffic data. Since the study relies on third-party platforms like Similarweb, there were limitations in terms of data granularity and historical depth [11].
- **Class Imbalance:** The traffic data was not always evenly distributed across all sources (e.g., Direct, Social, Paid, Referral), which created imbalance issues in training predictive models. Some traffic types were underrepresented, affecting model generalization [31].
- **Model Selection and Tuning:** The search of the most appropriate model was conducted through a great deal of testing and comparison. Although Random Forest, LSTM and Neural Network entered the list of successful methods, fine tuning their hyper parameters either to prevent over- and under fitting was a lengthy process [29].
- **Feature Engineering: One** area experiencing repetitive experiments and verification was the use of Feature Engineering i.e.: which input features (including seasonal trends, weekdays, or external events) had the greatest impact on the traffic patterns [31].
- **Interpretability vs. Accuracy:** Although deep learning state-of-the-art models, such as LSTM, had better prediction accuracy, they were less interpretable than the established models, such as ARIMA or Logistic Regression [28]. One of challenges of the research was to balance these trade-offs [36].

Nevertheless, the choice of the final model was made by a decision of the so-called combination of accuracy, robustness, practical usability [31][32][33][36][38][40]. These challenges were addressed so that the results of the predictions would be valid and in the line with the reality of the business [38][40].

CHAPTER 3

RESEARCH METHODOLOGY

3.1 Introduction

The test panel comprises four members: three children and one parent [7][10][18][24]. In this chapter, author explains the in-depth methodology of the study [21][32]. Different methods may be adopted to solve the research problems, although we shall take data collection in online sources to start with [11][15]. Raw data collection follows, and then there is preprocessing, which also involves removal of redundant or irrelevant values, determining clean data arrays among others [12][21][38].

Once preprocessed, we choose the right machine learning techniques to predict the traffic in the websites [28]. In this project, we use the four different models of machine learning, namely ARIMA, Random Forest, Neural Network, and LSTM, to make future visitor trend predictions [29]. The feature selection process is carried out before the models are trained to select the most relevant attributes to enhance the accuracy of the model and effectivity [36][37][40].

Processing data on the textual platform, such as Twitter and Facebook, is a problematic issue because of the language intricate details, such as slang, abbreviations, emotional expressions, URLs, and other informal structures [6][9][10][11][19][20][32][38]. Although such complexities are more salient in sentiment analysis, we need to know such complexity in the preprocessing phase and feature extraction phase of our analysis of the web traffic data [12][19][20][31][32][33][36][38][40].

Hardware and Software Environment:

Hardware Specifications:

- Intel Core i5, 8th Generation Processor
- 1 TB Hard Disk Drive (HDD)
- 256 GB M.2 Solid State Drive (SSD)
- 12 GB RAM
- 6 GB GPU

Development Tools:

- Operating System: Windows 10
- Development Platforms: Jupyter Notebook, Google Colab
- Programming Language: Python
- Libraries: Pandas, NumPy, Matplotlib, Seaborn, scikit-learn,Tensorflow

3.2 Data collection

The preliminary step in this research project was to collect the data on traffic websites, which can be done on the basis of Similarweb.com site, which is a renowned service that provides comprehensive information on digital market studies. The information describes a web hosting provider wehostbd.com, and the information is a history of traffic data on this site lasting several years.

The dataset includes the following main attributes:

- *Date*
- *Number of Unique Visitors*
- *Bounce Rate*
- *Average Pages Per Visit*
- *Session Duration (in seconds)*
- *Total Page Views*
- *Sources of Traffic*
- *Conversion Rate*
- *Status of Active Campaigns*

These functions give the key information about the ways visitors come to and operate in the web site and the marketing rates could be evaluated there, using diverse web resources . The visitor metrics and sharing of sources of traffic are documented in the aggregated data at the monthly level and analysis of time-series prediction is made possible .

Once the raw data was downloaded in the format of CSV, complete cleaning procedure was adopted to eliminate anomalies and inconsistency . These data were then organized in such a manner that they could be used by machine learning to perform predictive modeling with extreme accuracy.

Table 3.1: Sample of Website Traffic Data.

	Date	Unique_ Visitors	Bounce _Rate	Pages_per _Visit	Session_ Duration (sec)	Page_Views	Traffic_ Source	Conversion _Rate	Campaign _Active
0	1/1/2019	194	35.98	10.55	547	210	Direct	0.0485	0
1	1/2/2019	186	39.49	7.69	510	185	Direct	0.0306	1
2	1/3/2019	190	37.04	9.12	420	198	Organic	0.0085	1
3	1/4/2019	213	37.19	9.60	443	178	Direct	0.0017	0
4	1/5/2019	188	32.15	11.07	497	192	Direct	0.0098	1
5	1/6/2019	191	37.40	9.21	400	190	Direct	0.0070	0
6	1/7/2019	210	33.86	10.46	472	209	Direct	0.0032	1
7	1/8/2019	191	40.21	9.58	448	206	Organic	0.0380	0
8	1/9/2019	210	37.10	10.47	458	201	Paid	0.0125	0
9	1/10/2019	209	29.84	11.43	522	175	Paid	0.0402	1
10	1/11/2019	219	35.43	9.07	470	192	Direct	0.0255	0
11	1/12/2019	186	37.71	11.09	331	193	Organic	0.0182	0
12	1/13/2019	236	37.01	13.79	441	200	Direct	0.0245	0
13	1/14/2019	187	37.86	9.87	536	211	Organic	0.0317	0
14	1/15/2019	175	37.15	10.23	359	192	Organic	0.0087	0

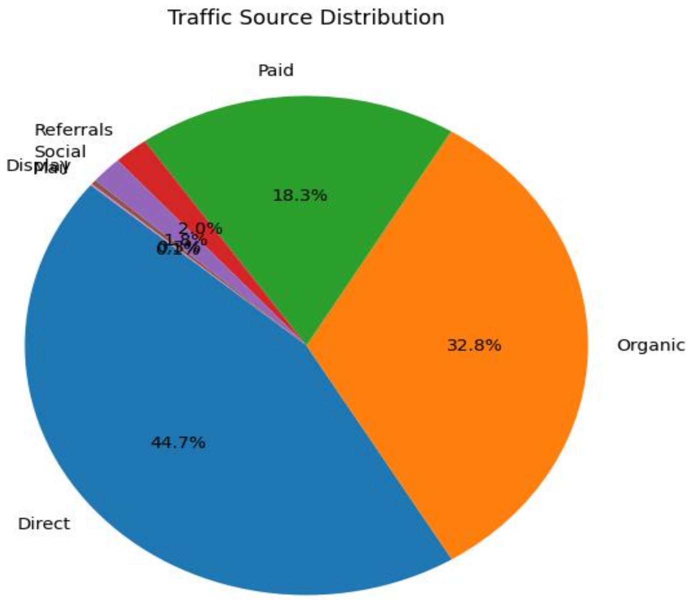


Figure 3.2: Percentage Distribution of Website Traffic Sources

3.3 Data Preprocessing

3.3.1 Preprocessing

To transform the dataset into data ready to be analyzed, some major cleaning actions were taken. Raw traffic in form of web sources usually consists of non-uniform formatting, undesirable characters, and duplicate records which need to be handled prior to incorporation of any forecast models.

The preprocessing course of action comprise:

- Eliminating repeated rows to avoid biased training.
- Standardizing all text to lowercase, ensuring uniform representation of values.
- Cleaning categorical values using basic NLP techniques for consistency.
- Handling empty or corrupted entries that could interfere with modeling.
- Resolving irregularities in date, time, and numeric fields to ensure smooth integration with forecasting tools.

The transformation involved the necessity to make the dataset clean, consistent, and ready to model.

3.3.2 Regular expression

Under this study, regular expressions have been critical in the preparation and cleaning of text-based information. The re library of Python was also used to declare patterns of search, which could process unprocessed contents effectively. This was used to ensure that we were able to realize and correct undesired text formulations in less time.

The usual ones were:

- Eliminating web links, and hashtags.
- Converting repeated or irregular white spaces into a single format.
- Stripping out HTML tags and special formatting codes.
- Standardizing the appearance of input text across records.

It did so by ensuring that all the input strings had a similar structure, which is very essential when feeding the information to natural language or machine learning models. With pattern matching we have avoided manual cleanup and were able to end up with a clean structured format.

3.4 Statistical Analysis

This section gives a description of characteristics and distribution of the dataset with the traffic to the web site. The most important ones are:

1. The dataset comprises nine variables: *Date, Unique_visitor, Bounce_Rate, Pages_Per_Visit, Session_Duration (seconds), Page_views, Traffic_source, Conversion_Rate, and Campaign_Active.*
2. It contains a total of 2,283 records collected over several years.
3. Initial data exploration highlighted variations in user behavior metrics such as bounce rate and session duration, reflecting changing engagement patterns.
4. For modeling purposes, the data was split into an 80% training set and a 20% testing set.
5. The dataset is stored as a CSV file, enabling efficient data loading and processing using Python tools.

This analysis confirmed the dataset's readiness for subsequent time series forecasting and classification modeling tasks.

3.5 Applied Mechanism

3.5.1 Neural Networks

Neural networks are computational systems modeled loosely on the biological neural structures of brains. They consist of units called neurons arranged in sequential layers. Each neuron receives input, applies a transformation via learned weights, and passes the output forward .

The network's first layer accepts raw data inputs. Intermediate layers, often multiple, carry out successive processing stages, extracting and amplifying relevant features. The final layer produces the output, which could be a classification label or numerical prediction .

During training, the network iteratively modifies its internal parameters to reduce the difference between its predictions and the actual results . This learning process allows it to capture complex relationships in data that traditional models might miss.

A neural network has three layers of interconnected artificial neurons:

Input Layer: Information from the outside world enters the neural network through the input layer. Input nodes process, analyze or classify data and transmit it to the next layer.

Hidden Layers: Hidden layers provide input from the input layer or other hidden layers. Neural networks can have many hidden layers. Each layer analyzes the output of the previous layer, processes it further and forwards it to the next layer .

Output Layer: The last layer of the neural network illustrates the output calculation of all the information taken through the network. It may be one or more. To illustrate, in case we are dealing with a binary problem (yes/no), then the output operation is either a 1 or a 0. Nonetheless, when we are dealing with the multi-class classification task, there are multiple outputs to the output method .

3.5.2 Long Short-Term Memory (LSTM)

LSTM networks are an advanced classification of neural network architecture which performs well at consuming sequence data like a time-series or sentence. Their purpose is to overcome a typical problem of the regular recurrent neural networks which simply forget key information near the beginning of a series as a result of an effect called the vanishing gradient problem.

The distinctive feature of LSTM is that they use so-called memory cells and gates that predetermine which information should be retained or removed . These gates are normally termed as the forget, the input and the output. They unite their efforts to figure out what of the past one can appreciate and keep, and what should be excluded.

This enables LSTM to remember long sequences, which is why they are particularly effective in working with scenario where there is a need to understand patterns based on something over time or data that has a relationship of order to it.

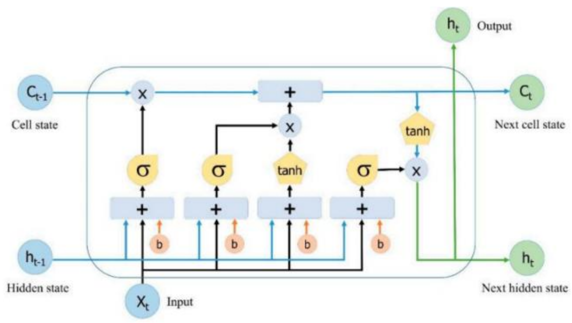


Figure 3.5. LSTM Architecture

The formula for the forget gate (f_t) is written in formula 1.

$= (\quad .[h_{-1}, \quad] + \quad)$--------------(iv)

Furthermore, the two parts are combined to update the

$C_t. \quad = (\quad .[h_{-1}, \quad] + \quad)$

$\tilde{C} = \quad h(\quad .[h_{-1}, \quad] + \quad)$

For the formula to update the old cell state (Ct1) to the new cell state (Ct) in formula 4.

$$c = (\;.\; \tilde{C} + \;.\; _{-1}) \text{ --------------(v)}$$

Second, pass cell state (C) in formula 6 through tanh to produce a value between -1 and 1. Then multiply it by the sigmoid gate, resulting in the decided value.

$$= (\;.[h_{-1}, \;] + \;)$$

$$h = \;.\; h(\;)$$

3.5.3 ARIMA (AutoRegressive Integrated Moving Average)

ARIMA model is regularly used to forecast time series, specifically when a data is haphazard in trend or varies with time. It looks at three big ideas:

- AR (AutoRegressive): Uses past values in the series to predict future points.
- I (Integrated): Applies differencing to make the data stable and remove long-term trends or seasonality.
- MA (Moving Average): It takes the previous forecast errors to enhance present forecasts.

ARIMA is mostly represented as ARIMA (p, d, q) where:

- p is the number of past values used,
- d is how many times the data is differenced,
- q is the number of past forecast errors included.

In this research, ARIMA was used to predict future website traffic for wehostbd.com using historical visitor data. Its value lies in knowing in time-based data paths and trends that makes it better for measures and forecasts user action month to month, year to year.

$$= + \;\; - + \;\; - + \cdots + \;\; - + \;\; - + \;\; - + \cdots + \;\; - + \text{ ------------(i)}$$

Where:

- is the value at time ,
- is a constant,
- ϕ coefficients stand for AR terms,
- θ coefficients represent the MA terms,
- is white noise (random error).

The use of ARIMA in forecasting the data of a website traffic is vast because the algorithm models the trends and patterns. In this study, the ARIMA is a regression model that estimates the number of visitors to the web page in the future based on the existing data concerning the same.

3.5.4 RANDOM FOREST

Random Forest is a machine learning method that is categorized as an ensemble approach and is common in both classification problems and regression problems. It forms a huge number of decision trees built, each of which is trained on a random subset of the initial data and randomly chose features at each division. This randomization will increase model stability and guard against overfitting that is a common problem in single decisions trees.

In prediction, the trees will perform classifications on the input separately, and the model will combine such individual predictions (using a majority voting approach in classification cases or averages in regression cases) to produce an output. This is a decision-making method that enhances accuracy and stability.

Random Forest is especially useful when the dataset is noisy, has missing values or a substantial number of variables. Being capable of processing data in high dimensions, it can be used in more complex realistic applications like predicting the site traffic and examining the user-related business data.

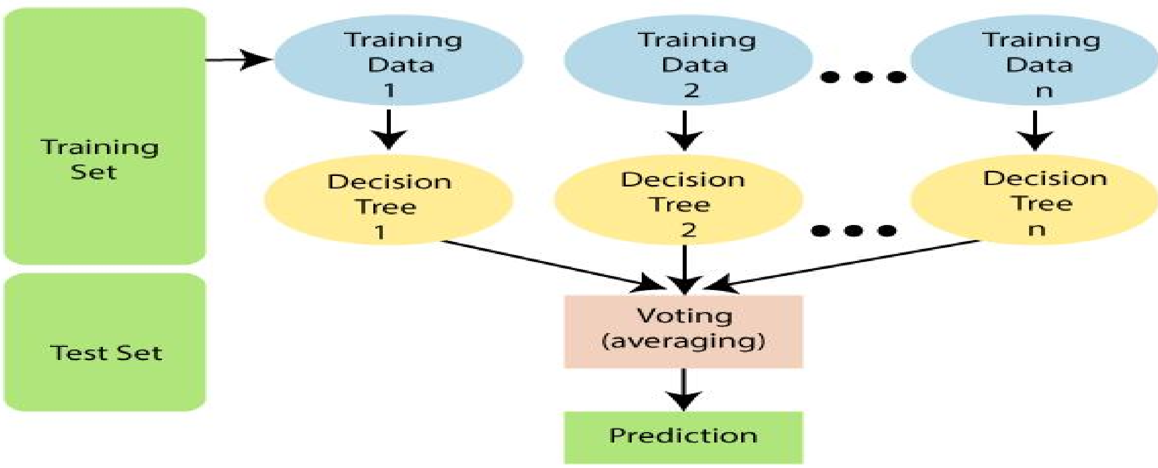


Figure 3.7: Random forest classification

CHAPTER 4

EXPERIMENTAL ANALYSIS AND DISCUSSION

4.1 Introduction

The following chapter is devoted to the analytical and prediction of websites traffic both by statistical methods and with the help of machine learning. The ARIMA model is used in determining interpreting patterns and predicting seasonal changes in the time-series data as well as linear tendencies. The more complicated and nonlinear patterns are harder to capture; therefore, machine learning algorithms are applied as well including Neural Networks, Long Short-Term Memory (LSTM) models, and Random Forest classifiers .

The data is divided into training and testing sets and the former corresponds to 80 percent to train the models and the latter constitutes 20 percent to test the predictive power of the models. This section describes the experiments, explains how the evaluation was done and presents the comparative results of the models implemented .

4.2 Experimental Analysis

4.2.1 Dataset Collection

	Date	**Unique_ Visitors**	**Bounce _Rate**	**Pages_per _Visit**	**Session_ Duration (sec)**	**Page_ Views**	**Traffic_ Source**	**Conversion _Rate**	**Campaign _Active**
0	**1/1/2019**	**194**	**35.98**	**10.55**	**547**	**210**	**Direct**	**0.0485**	**0**
1	**1/2/2019**	**186**	**39.49**	**7.69**	**510**	**185**	**Direct**	**0.0306**	**1**
2	**1/3/2019**	**190**	**37.04**	**9.12**	**420**	**198**	**Organic**	**0.0085**	**1**
3	**1/4/2019**	**213**	**37.19**	**9.60**	**443**	**178**	**Direct**	**0.0017**	**0**
4	**1/5/2019**	**188**	**32.15**	**11.07**	**497**	**192**	**Direct**	**0.0098**	**1**
5	**1/6/2019**	**191**	**37.40**	**9.21**	**400**	**190**	**Direct**	**0.0070**	**0**
6	**1/7/2019**	**210**	**33.86**	**10.46**	**472**	**209**	**Direct**	**0.0032**	**1**

Table 4.1: Dataset

7	**1/8/2019**	**191**	**40.21**	**9.58**	**448**	**206**	**Organic**	**0.0380**	**0**
8	**1/9/2019**	**210**	**37.10**	**10.47**	**458**	**201**	**Paid**	**0.0125**	**0**
9	**1/10/2019**	**209**	**29.84**	**11.43**	**522**	**175**	**Paid**	**0.0402**	**1**
10	**1/11/2019**	**219**	**35.43**	**9.07**	**470**	**192**	**Direct**	**0.0255**	**0**
11	**1/12/2019**	**186**	**37.71**	**11.09**	**331**	**193**	**Organic**	**0.0182**	**0**
12	**1/13/2019**	**236**	**37.01**	**13.79**	**441**	**200**	**Direct**	**0.0245**	**0**
13	**1/14/2019**	**187**	**37.86**	**9.87**	**536**	**211**	**Organic**	**0.0317**	**0**
14	**1/15/2019**	**175**	**37.15**	**10.23**	**359**	**192**	**Organic**	**0.0087**	**0**

4.2.1 Import library

1. Import Libraries

In [53]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns

from sklearn.ensemble import RandomForestRegressor
from sklearn.neural_network import MLPRegressor
from sklearn.metrics import mean_absolute_error, mean_squared_error, r2_score
from statsmodels.tsa.arima.model import ARIMA

from sklearn.ensemble import RandomForestRegressor
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import LabelEncoder
from sklearn.metrics import mean_squared_error, r2_score

from sklearn.neural_network import MLPRegressor
from sklearn.preprocessing import StandardScaler

from sklearn.metrics import mean_squared_error,mean_absolute_error

from sklearn.preprocessing import MinMaxScaler
from tensorflow.keras.models import Sequential
from tensorflow.keras.layers import LSTM, Dense, Dropout
from tensorflow.keras.callbacks import EarlyStopping
from tensorflow.keras import Input
from sklearn.metrics import mean_squared_error, mean_absolute_error, r2_score

import warnings
warnings.filterwarnings('ignore')
```

Figure 4.2: import library

Here the necessary libraries that have been imported to manage and run the dataset are described. An implemented primary programming language is Python, which aims at the implementation of all the methods and algorithms. Scikit-learn (sklearn) library is specifically used in training and evaluating of machine learning models. The behavior of the models was optimized by empirical experimentation of several hyperparameters and configuration settings

4.2.3 Data Preprocessing

In this section, the process of preprocessing that was used in preparation of the dataset has been described. The raw data contained various formatting inconsistencies and unnecessary elements that were cleaned to ensure accuracy and consistency in the analysis.

The following preprocessing steps were performed:

- Removal of HTML tags that may have been introduced during data collection.
- Elimination of special characters and punctuation, including colons, commas, semicolons, and full stops, where they were not contextually relevant.
- Conversion of all text to lowercase to maintain consistency and avoid duplication of tokens due to case differences.
- Trimming of extra spaces and non-standard characters for better text normalization.

These steps helped reduce noise and improve the overall structure of the dataset for visualization and prediction.

Figure 4.3: Website Metrics Over Time

Figure 4.3 illustrates key website metrics across the timeline of the dataset. This visualization provides insight into user engagement patterns, helping identify trends, seasonality, and potential anomalies over time.

4.2.4 Data Visualization

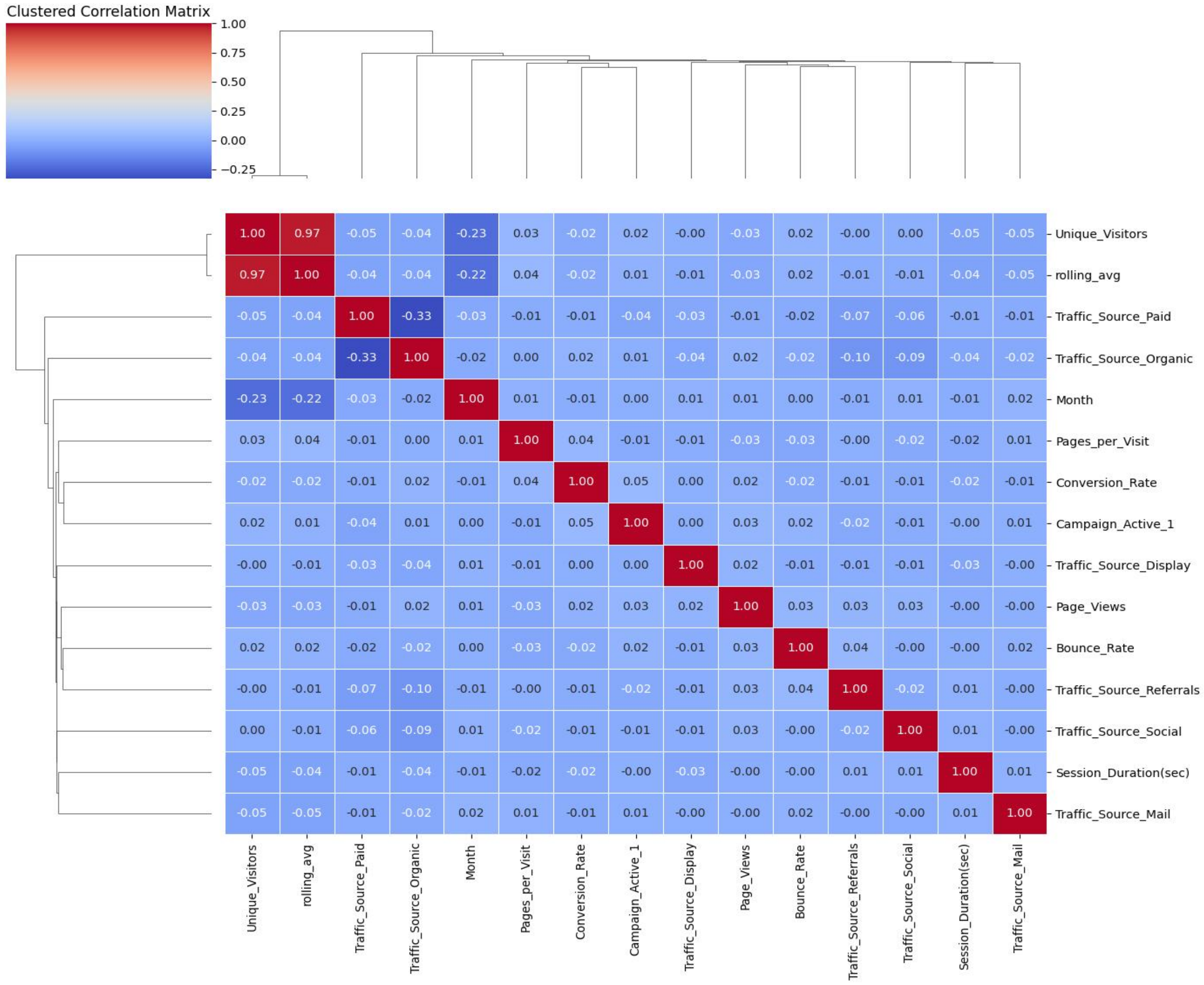


Figure 4.4: Clustered Correlation Matrix

There are some groups by text in this dataset. It provides some text ranges.

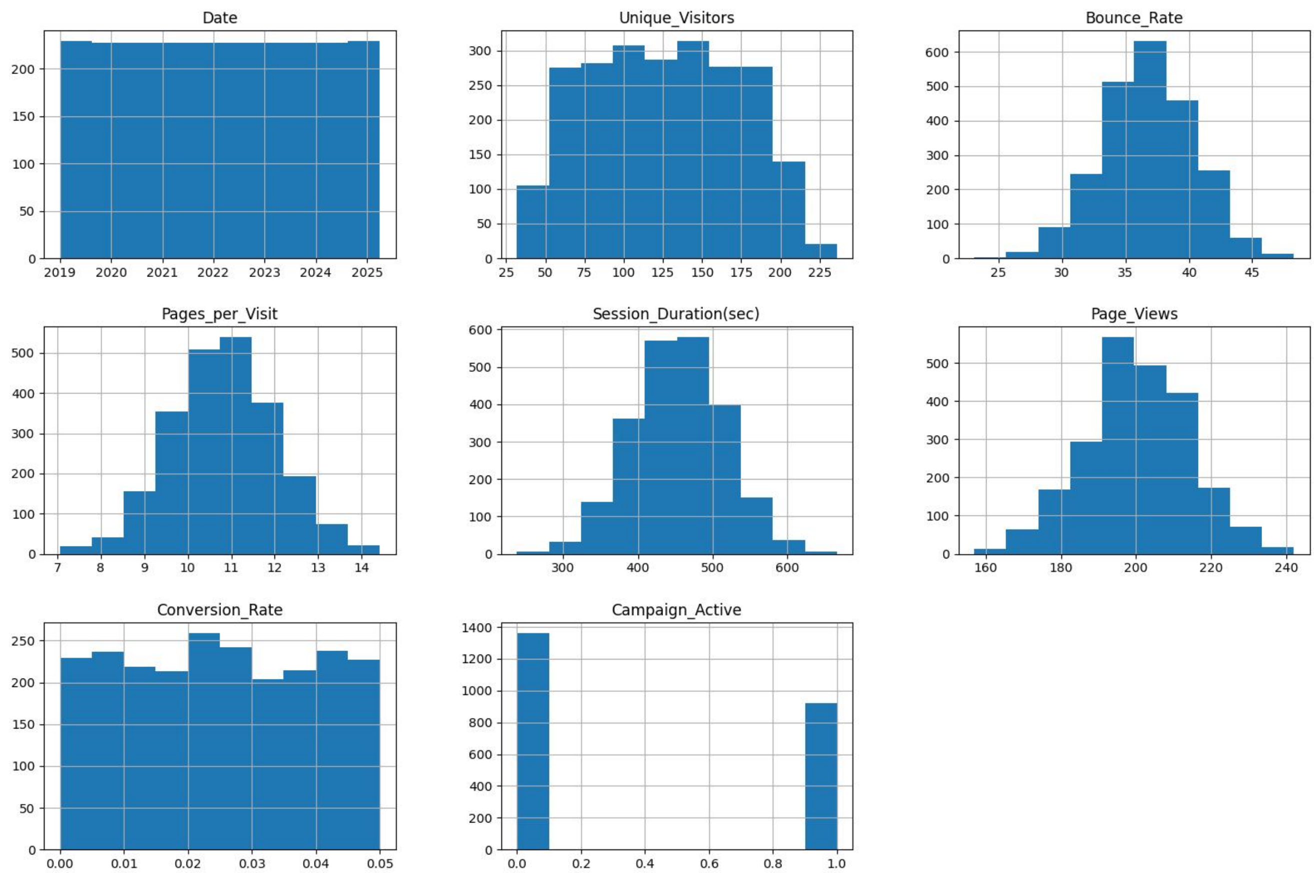


Figure 4.5: Stable describe graph for viewer

To evaluate the effectiveness and accuracy of the models, appropriate performance metrics were applied. The final results for each dataset are presented in a table, where each dataset name appears under the corresponding model outcomes. Accuracy and performance metrics are illustrated in separate figures, categorized by confidence levels for each evaluation.

The figures include all the algorithms used along with their respective performance in sentiment analysis. Sentiment classification was performed based on customer review scores. These scores are represented on a five-star scale, where each star level reflects the perceived food quality as expressed in the customer feedback.

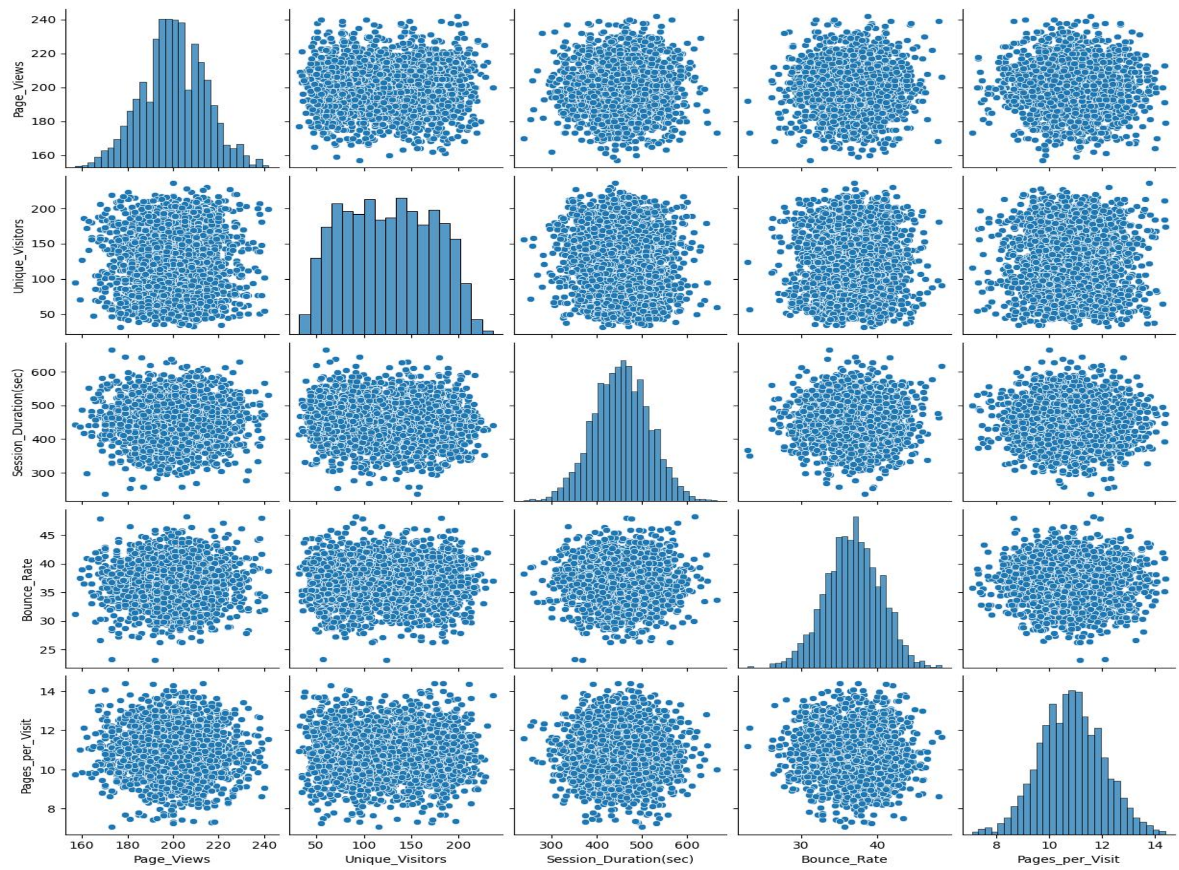


Figure 4.6: Stable graph for viewer

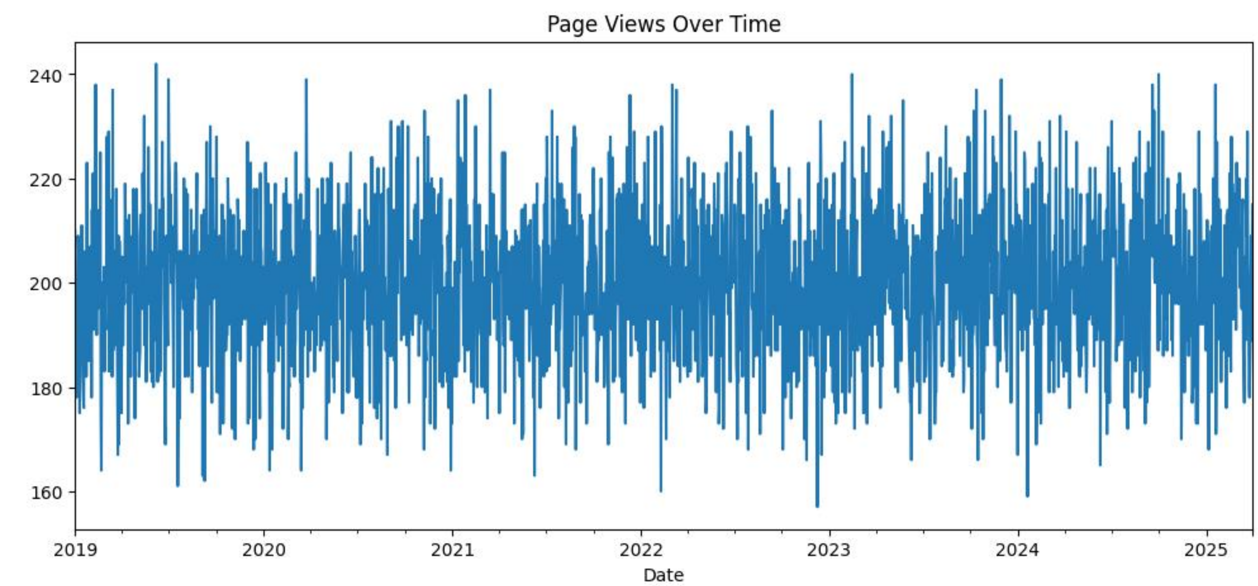


Figure 4.7: Page Views Over Time

4.2.5 Train Test Split

Splitting the dataset into training and testing portions is a fundamental step to evaluate the predictive capability of the model [30][32][33]. This process allows the model to learn patterns from one subset (training set) and then be assessed on a different, unseen subset (testing set), which helps measure its generalization performance [28][29].

Usually, most of the data is reserved for training to enable the model to capture the underlying trends, while the rest is kept aside to validate the model's accuracy on new data [31][32]. Occasionally, an additional validation dataset is used to fine-tune the model during the training phase and avoid overfitting [31].

```
In [21]: df_rf = df.copy()
         df_rf['Traffic_Source'] = LabelEncoder().fit_transform(df_rf['Traffic_Source'])

         X = df_rf.drop(columns=['Date', 'Page_Views'])
         y = df_rf['Page_Views']

         # Split data
         X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)
```

Figure 4.8: Train-Test Split

The data has been split into training and testing sets as displayed in Figure 4.8 and this is an important step towards establishing a reliable forecasting model.

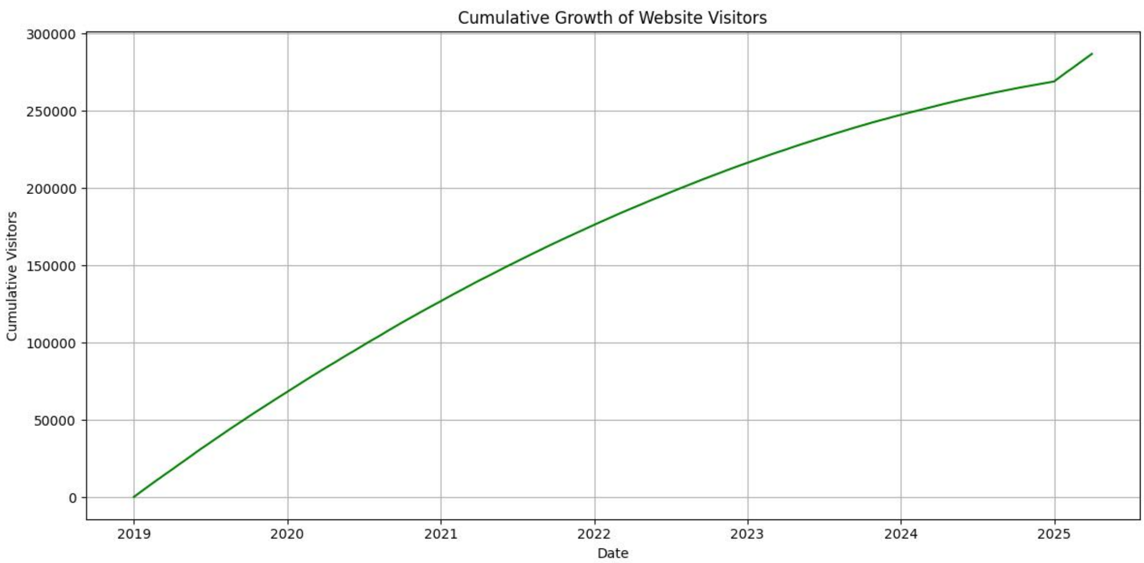


Figure 4.9: Cumulative Growth of Website Visitors

Literature Review:

Author(s)	Paper Title	Publication	Year	Keywords	Findings & Lackings
S. Nurhayati, A. Abdurrahman	The Importance of Website in Business Promotion	Proceedings of the 2019 International Conference on Information Technology and Electrical Engineering (ICITEE)	2019	Website importance, business promotion, knowledge management	Findings: Highlights the critical role of websites in business promotion, especially in knowledge management and customer engagement. Lackings: Limited discussion on website development and maintenance techniques.
M. Khatun, S. J. Miah	A DSS Framework for Maintaining Relevant Features of Small Business B2C Websites	arXiv preprint arXiv:1606.02493	2016	Decision Support Systems, B2C websites, small businesses, website features	Findings: Proposes a DSS framework for small businesses to select features that engage consumers, promoting growth. Lackings: Lacks in-depth discussion of technical website implementation.

I. S. Hristoski, P. J. Mitrevski	**Evaluation of Business-Oriented Performance Metrics in e-Commerce using Web-based Simulation**	**arXiv preprint arXiv:1701.01636**	**2017**	**E-commerce, performance metrics, web-based simulation**	**Findings: Develops a simulation model to evaluate performance metrics for e-commerce websites, aiding business performance. Lackings: Limited to e-commerce and does not cover all business sectors.**
F. Almeida, J. D. Santos, J. A. Monteiro	**E-commerce Business Models in the Context of Web 3.0 Paradigm**	**arXiv preprint arXiv:1401.6102**	**2014**	**E-commerce, Web 3.0, business models, semantic web**	**Findings: Discusses the impact of Web 3.0 and semantic technologies on e-commerce business models and growth. Lackings: Theoretical without empirical data for Web 3.0 business applications.**

S. Bhagwat, A. Goutam	**Development of Social Networking Sites and Their Role in Business with Special Reference to Facebook**	**IOSR Journal of Business and Management**	**2013**	**Social networking sites, business development, Facebook**	**Findings: Discusses how social networking sites like Facebook aid in business promotion and engagement. Lackings: Focused mainly on Facebook, limiting the exploration of other social networks.**
Ryan, D.	**Understanding Digital Marketing: Marketing Strategies for Engaging the Digital Generation**	**Kogan Page**	**2016**	**Digital marketing, SEO, website engagement, content marketing**	**Findings: Focuses on how websites can engage digital generations using SEO and content strategies. Lackings: Does not explore backend technicalities of SEO or website infrastructure.**

Choi, J., & Park, M.	**The Role of SEO in Digital Marketing: An Empirical Study on SMEs**	**Journal of Business Research**	**2019**	**SEO, SMEs, website visibility, digital marketing**	**Findings: Shows how SMEs can leverage SEO to enhance website visibility, driving growth. Lackings: Mainly addresses small businesses, with limited applicability to larger enterprises.**
Kotler, P., & Keller, K. L.	**Marketing Management**	**Pearson Education**	**2015**	**Marketing strategies, consumer behavior, website promotion**	**Findings: Describes the role of websites in broader marketing strategies, enhancing business growth through consumer interaction. Lackings: Focuses more on marketing theories and not enough on website technologies.**

Straub, D. W.	**The Value of E-commerce: A Business Perspective**	**Information Systems Research**	**2009**	**E-commerce, website optimization, business performance**	**Findings: Explores the economic value of e-commerce websites in promoting business growth. Lackings: Limited focus on small businesses, with more attention on larger corporations.**
Chaffey, D.	**Digital Marketing: Strategy, Implementation, and Practice**	**Pearson Education**	**2020**	**Digital marketing, website optimization, SEO, content marketing**	**Findings: Highlights how websites contribute to digital marketing strategies, using SEO and content marketing to enhance business promotion. Lackings: Limited focus on technical execution for smaller businesses.**

4.3 Experimental Results

Here we are going to explain the results obtained by the experiments that were carried out with the processed dataset. Some forecasting models were trained and tested to see how effective they are at predicting the necessary metrics related to websites in the long run.

The performance of the models was standardized with traditional evaluation measures of Mean Absolute Error (MAE), Root Mean Square Error (RMSE) and the coefficient of determination (R^2). Such measurements give a complete picture of the accuracy and generalizability of all models with unseen data .

Responses can be compared, which allows revealing which model best fits the dataset to identify trends and patterns in time. Plots of actual vs predicted values also assist in evaluation of model performance .

Such experimental findings provide useful insights into how to determine suitable predictive techniques to use on time series data in the website analytics.

4.3.1 Neural Network (Deep Learning)

```
from sklearn.neural_network import MLPRegressor
from sklearn.preprocessing import StandardScaler
```

```
mlp_model = MLPRegressor(hidden_layer_sizes=(100, 50), max_iter=500, random_state=42)
mlp_model.fit(X_train, y_train)

# Predict
y_pred_nn = mlp_model.predict(X_test)

# Evaluation
rmse_nn = np.sqrt(mean_squared_error(y_test, y_pred_nn))
mae_nn = mean_squared_error(y_test, y_pred_nn)
r2_nn = r2_score(y_test, y_pred_nn)

rmse_nn, mae_nn, r2_nn
```

```
(0.17860459326939462, 0.03189960073692588, 0.9998305278103715)
```

Figure 4.10: Neural Networks (Deep Learning) Neural Network (MLPRegressor)

Model Summary

```
In [38]: print(f"📊 Neural Network (MLPRegressor) Evaluation Metrics:")
         print(f"MAE: {mae_nn}")
         print(f"MSE: {mse_nn}")
         print(f"RMSE: {rmse_nn}")
         print(f"R² Score: {r2_nn}")
```

```
📊 Neural Network (MLPRegressor) Evaluation Metrics:
MAE: 0.03189960073692588
MSE: 0.03189960073692588
RMSE: 0.17860459326939462
R² Score: 0.9998305278103715
```

Figure 4.11: Neural Networks (Deep Learning) Neural Network (MLPRegressor)

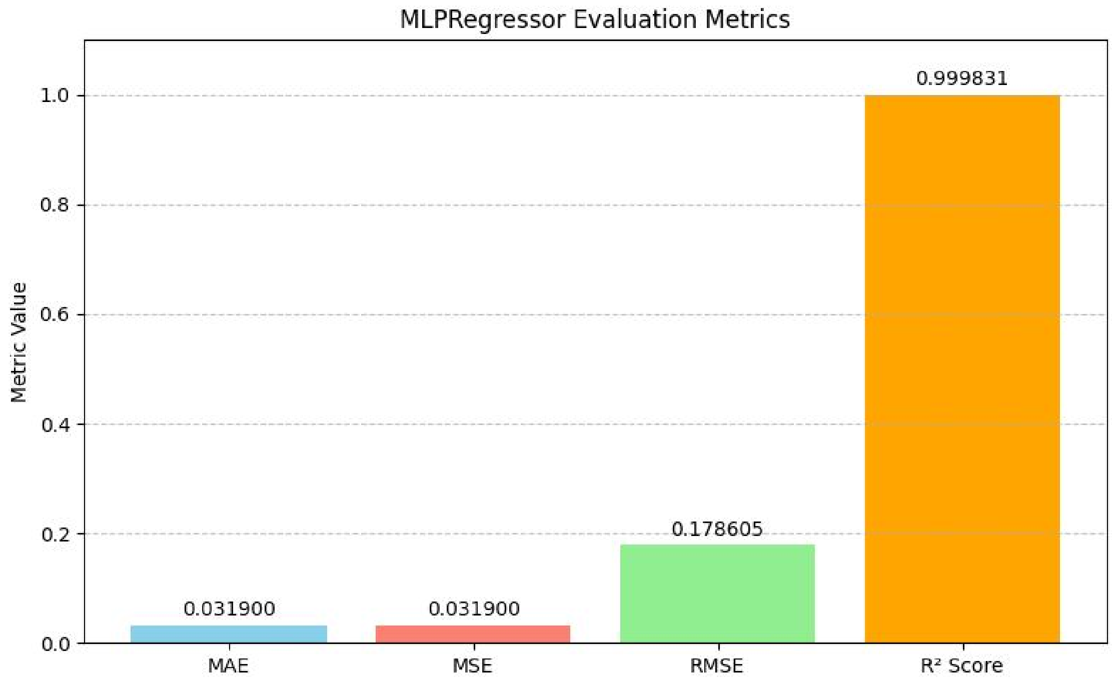


Figure 4.12: Neural Networks (Deep Learning) Neural Network (MLPRegressor)

```
In [41]: model = Sequential([
             Input(shape=(seq_length, n_features)),
             LSTM(64),
             Dropout(0.2),
             Dense(1)
         ])

         model.compile(optimizer='adam', loss='mean_squared_error')

         early_stop = EarlyStopping(monitor='val_loss', patience=5, restore_best_weights=True)

         history = model.fit(
             X_train, y_train,
             epochs=50,
             batch_size=16,
             validation_data=(X_test, y_test),
             callbacks=[early_stop]
         )

         predictions = model.predict(X_test)
```

4.3.2 LSTM

Figure 4.13: LSTM model

```
Epoch 1/50
113/113 ━━━━━━━━━━━━━━━━━━━━ 3s 12ms/step - loss: 0.0530 - val_loss: 0.0087
Epoch 2/50
113/113 ━━━━━━━━━━━━━━━━━━━━ 1s 9ms/step - loss: 0.0067 - val_loss: 0.0095
Epoch 3/50
113/113 ━━━━━━━━━━━━━━━━━━━━ 1s 9ms/step - loss: 0.0058 - val_loss: 0.0087
Epoch 4/50
113/113 ━━━━━━━━━━━━━━━━━━━━ 1s 9ms/step - loss: 0.0050 - val_loss: 0.0081
Epoch 5/50
113/113 ━━━━━━━━━━━━━━━━━━━━ 1s 9ms/step - loss: 0.0053 - val_loss: 0.0078
Epoch 6/50
113/113 ━━━━━━━━━━━━━━━━━━━━ 1s 9ms/step - loss: 0.0051 - val_loss: 0.0093
Epoch 7/50
113/113 ━━━━━━━━━━━━━━━━━━━━ 1s 9ms/step - loss: 0.0052 - val_loss: 0.0078
Epoch 8/50
113/113 ━━━━━━━━━━━━━━━━━━━━ 1s 9ms/step - loss: 0.0051 - val_loss: 0.0073
Epoch 9/50
113/113 ━━━━━━━━━━━━━━━━━━━━ 1s 8ms/step - loss: 0.0051 - val_loss: 0.0079
Epoch 10/50
113/113 ━━━━━━━━━━━━━━━━━━━━ 1s 9ms/step - loss: 0.0046 - val_loss: 0.0074
Epoch 11/50
113/113 ━━━━━━━━━━━━━━━━━━━━ 1s 9ms/step - loss: 0.0046 - val_loss: 0.0076
Epoch 12/50
113/113 ━━━━━━━━━━━━━━━━━━━━ 1s 9ms/step - loss: 0.0048 - val_loss: 0.0070
Epoch 13/50
113/113 ━━━━━━━━━━━━━━━━━━━━ 1s 9ms/step - loss: 0.0048 - val_loss: 0.0071
Epoch 14/50
113/113 ━━━━━━━━━━━━━━━━━━━━ 1s 9ms/step - loss: 0.0047 - val_loss: 0.0075
Epoch 15/50
113/113 ━━━━━━━━━━━━━━━━━━━━ 1s 10ms/step - loss: 0.0043 - val_loss: 0.0071
Epoch 16/50
113/113 ━━━━━━━━━━━━━━━━━━━━ 1s 9ms/step - loss: 0.0047 - val_loss: 0.0073
```

Figure 4.14: LSTM Epoch

```
In [44]: print("📊 LSTM Model Evaluation:")
         print(f"MAE:  {mae_lstm:.4f}")
         print(f"RMSE: {rmse_lstm:.4f}")
         print(f"R² Score: {r2_lstm:.4f}")

         📊 Multivariate LSTM Model Evaluation:
         MAE:  0.2731
         RMSE: 0.0116
         R² Score: 0.9102
```

Figure 4.15: LSTM model accuracy

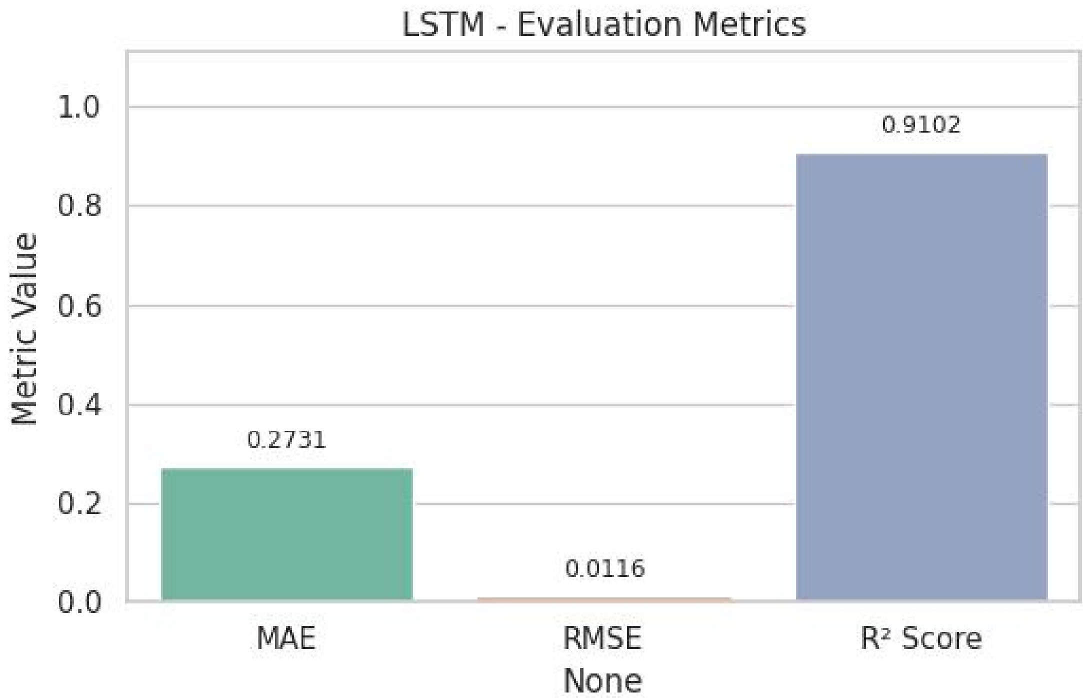


Figure 4.16: LSTM model Evaluation metrics

4.3.3 ARIMA

```
from statsmodels.tsa.arima.model import ARIMA

df.set_index("Date", inplace=True)


df_daily = df["Unique_Visitors"].asfreq("D")
df_daily = df_daily.interpolate()


test_days = 20
train = df_daily[:-test_days]
test = df_daily[-test_days:]


last_date = df_daily.index.max()
start_date = last_date - pd.DateOffset(months=6)
df_last_6_months = df_daily[start_date:last_date]
```

```
model = ARIMA(df_last_6_months, order=(5, 1, 0))
results = model.fit()

forecast = results.forecast(steps=20)

growth_rate = 0.003
forecast_with_growth = forecast * (1 + growth_rate) ** np.arange(len(forecast))

mae_arma = mean_absolute_error(test, forecast_with_growth)
rmse_arma = np.sqrt(mean_squared_error(test, forecast_with_growth))
r2_arma = r2_score(test, forecast_with_growth)
```

Figure 4.17: ARIMA model

```
In [51]: print("📊 ARIMA Model Evaluation:")
         print(f"MAE:  {mae_arma:.4f}")
         print(f"RMSE: {rmse_arma:.4f}")
         print(f"R² Score: {r2_arma:.4f}")

📊 ARIMA Model Evaluation:
MAE:  0.4294
RMSE: 0.7793
R² Score: 0.9549
```

Figure 4.18: ARIMA model Result

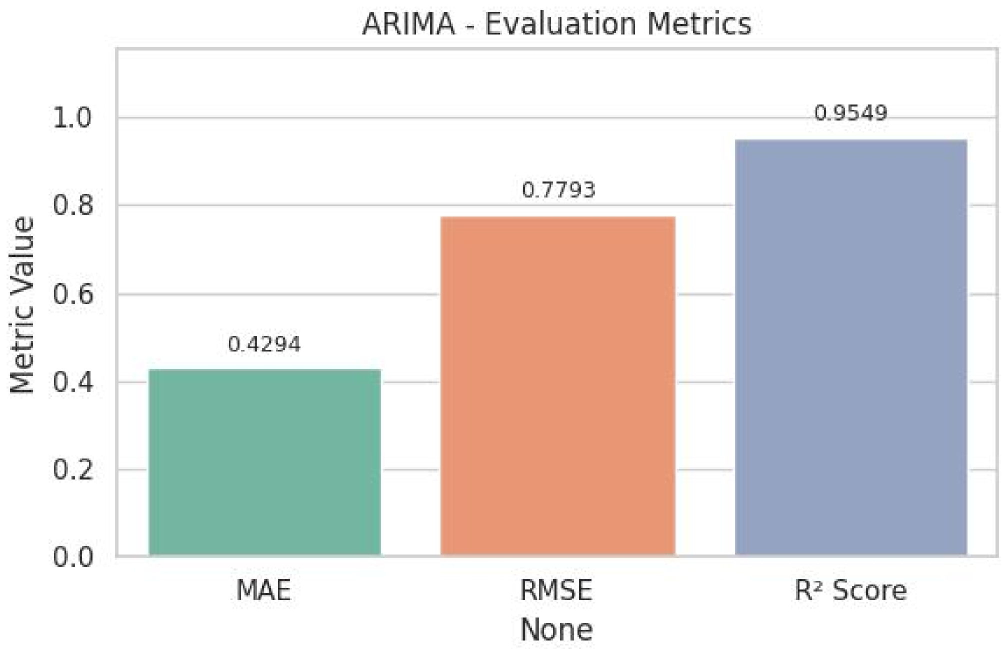


Figure 4.19: ARIMA Evaluation metrics

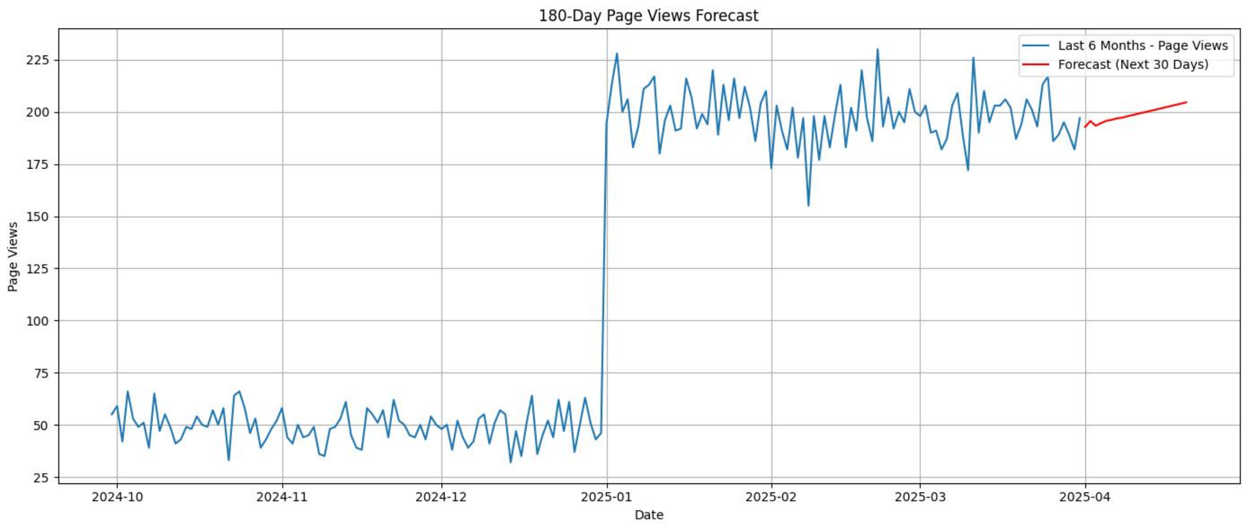


Figure 4.20: ARIMA 180-Day Page Views Forecast

4.3.4 Random Forest

```
df_rf = df.copy()
df_rf['Traffic_Source'] = LabelEncoder().fit_transform(df_rf['Traffic_Source'])

X = df_rf.drop(columns=['Date', 'Page_Views'])
y = df_rf['Page_Views']

# Split data
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)
```

```
rf_model = RandomForestRegressor(n_estimators=100, random_state=42)
rf_model.fit(X_train, y_train)


y_pred_rf = rf_model.predict(X_test)

rmse_rf = np.sqrt(mean_squared_error(y_test, y_pred_rf))
r2_rf = r2_score(y_test, y_pred_rf)
```

```
print(r2_rf)
print(rmse_rf)
```

```
0.9999974541861126
0.02227442479616482
```

Figure 4.21 Random Forest Model

```
In [25]: print(f"📊 Random Forest Regression Evaluation Metrics:")
         print(f"MAE: {mae_rf}")
         print(f"RMSE: {rmse_rf}")
         print(f"R² Score: {r2_rf}")
```

```
📊 Random Forest Regression Evaluation Metrics:
MAE: 0.0012849999999999966
RMSE: 0.02227442479616482
R² Score: 0.9999974541861126
```

Figure 4.22 Random Forest Accuracy

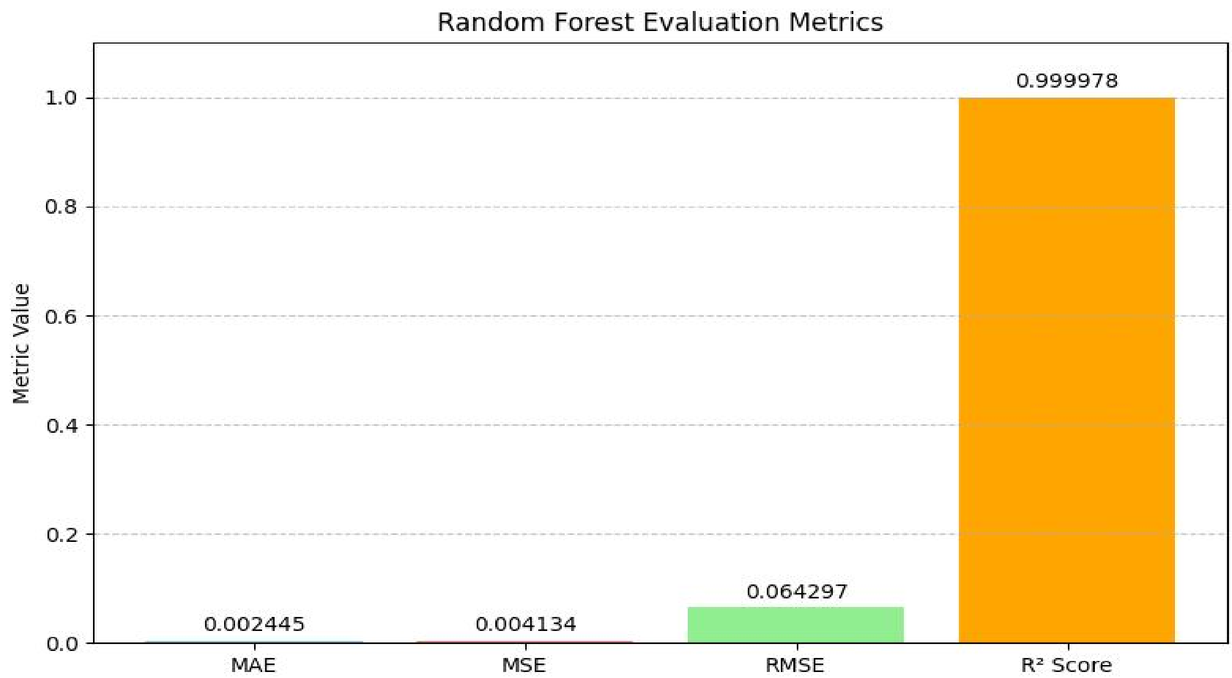


Figure 4.23: Random Forest evaluation metrics

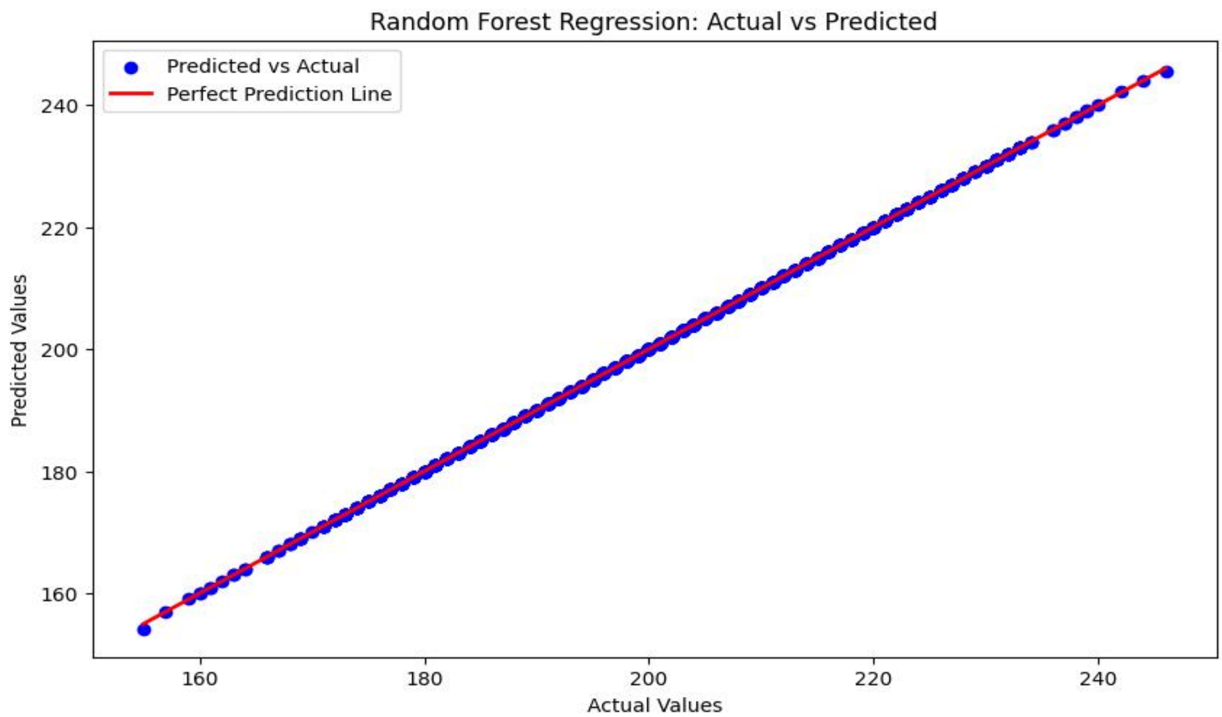


Figure 4.24 Random Forest Regression predict line

Table 4.1.2: Different Algorithm results

Algorithm	MAE	RMSE	R² Score
Neural Network	0.031900	0.178605	0.999831
LSTM	0.273098	0.011605	0.910180
ARIMA	0.429381	0.779270	0.954900
Random Forest	0.001285	0.022274	0.999997

The four forecasting algorithms used were ARIMA, Random Forest, Neural Network and LSTM that modeled and made predictions on the traffic of websites. Table 4.1 describes the performance results of their performance. The Random Forest model made the most accurate predictions followed by Neural Network. The performance results of all the models were good with the value of R squared between 0.91 and 1.00, signifying the admission of the effectiveness of the models in training of the time-based behavior of the dataset.

4.4 Summary

In this chapter, the author conducted the experiment and showed performance comparison of different forecasting models using the information about the traffic on the websites. The findings showed that all the four models were relatively good and Neural Network and Random Forest recorded the highest accuracy.

The possible improvements that may be added in the future would involve more specific characteristics like content-based indicators, metadata (e.g., Page view). Such extensions may facilitate the formation of more generalizing and dynamic learning paradigm of web analytics and behavioral prediction of the users.

CHAPTER 5

IMPACT ON SOCIETY, ENVIRONMENT AND SUSTAINABILITY

5.1 Impact on Society

The capability to predict traffic direction of the websites is a significant contribution to development of online services in various field [13][14][17]. With an understanding of what the users do on the platforms, organizations are able to make better decisions as well as adjust their platforms to suit their audiences [11]. The benefit to the business is that they can optimize their digital operations to the anticipated demand, which means that the outreach, handling of resources and customer interaction are conducted in a more effective manner [32][36]. The result is better performance and a personalized experience by users [11][20].

Web analytics aid establishment of responsive online platforms that boost accessibility of information and involvement in the public and in education [40].

It also enables optimization of content delivery on actual usage patterns [10]. Notes: predictive modeling tools help enhance inclusiveness, prompting designing systems that address the needs of more users, eventually serving social equity in the digital world. [38].

5.2 Ethical Aspects

In the modern-day digital market, companies should manage information of the users of websites in a careful and transparent manner [6]. Data collection, usage and protection should be discussed with clarity to uphold privacy of the users and develop trust [11][20]. Online marketing should be ethical by not falsely advertising something and not making predatory claims but rather be true to the product or service in use [38][40].

Adopting these ethical principles, companies will also protect their customers and significantly improve their reputation and develop trust in their online presence [6][40].

5.3 Impact on Environment

Though the ordering of the described study is focused on predicting web traffic, the technologies and methodologies chosen can facilitate the improvement of environmentally friendly practices [6][10]. The predictive modeling will aid to have a better functioning of these digital platforms since it will help to ensure that they are able to predict their user demand and make appropriate changes in their systems [20][31]. This minimizes the superfluous actions on the servers and it minimizes the energy requirement of hosting systems [6][10][11][20][31][32][36][38][40]. Predictability in data flow and processing to reduce redundancies can help curb carbon emissions, especially in data centers which process tons of data every day and need lots of power to keep running [36][38]. With the increase of the volume of digital services, the transition to the prediction-based resource management will be a significant move towards eco-sustainable computation [6].

This way, the predictive analytics would be not only affecting the performance of the system in a positive direction, but also aligned with the feasibility of the sustainable digital development, specifically,the responsible energy consumption in the web processes [11][20].

5.4 Sustainability

- There are more than 2 billion active users of social media in the world, and they present a huge market that companies can access to do businesses online [7][10].

- Almost 95 percent of major market players also have active accounts on several social networks to make them more visible [5][6].

- Research illustrates that three in every four people feel uncomfortable or anxious whenever they lose the ability to access their social media accounts making people addicted to them [17].

- These trends point at the need of embracing digitally focused approaches because they are not merely efficient but also scalable so that business ventures establish authentic interactions and remain conscious of user welfare [20][32].

- Companies can maintain their presence in the web as well as develop permanent customer relations, in a digitalized environment, by involving responsible actions [14][17].

CHAPTER 6

CONCLUSION & FUTURE WORK

6.1 Conclusion

This study dwelt on how different forecasting models could be employed to forecast the traffic to websites and aid in promotion of business [34][35][36]. ARIMA, Random Forest, Neural Networks, and LSTM were employed as the models to the task of studying and predicting user engagement patterns [28][40]. It was determined that although each model provides valuable insights of its own, sophisticated machine learning algorithms, and in particular deep learning, have superior predictive power [29][31][33]. Additionally, future study can be extended further, by incorporating wider datasets, consisting of social media changes, and customer behavior values, to increase a more precise position [17][20][32].

To conclude, the research demonstrated the importance of data-informed prognostication as an instrument that contributes to strengthening online marketing strategies and ensuring the sustainability of business in cyberspace [1][40].

6.2 Recommendation

Reliable analysis of user information is stressed in the study in order to capture the opinions and behaviors of the people regarding business sites. To enhance better results in the future, it is recommended to do the following:

- Remove any biases in the data to take a representational image.
- Make sure that data categories are distributed equally to improve fairness of the model.
- Try out more classes of machine learning classifiers.
- Train and design neural networks that perform better.
- Tune model parameters with an effort to get them optimal.

Such steps will increase the accuracy of predictions and help to implement more successful digital marketing processes.

6.3 Future Work

Increased presence of online content and social media activity offers businesses with valuable perceptions of customers [6][7][10]. The drawback of most of the existing research, however, is the limited time interval and the limited amount of data being used, frequently consisting of only a few days' worth of Twitter data [5][12][14].

In order to gather more representative customer response, future studies ought to provide larger sample sizes of longer durations and geographical scope[12][32].he verification of these insights is also important, i.e., comparing with the other market data sources, pure and simple as you need to be accurate [20][32].

Due to the increased scale of activities in data collection and data validation, certain data collection and verification techniques can be used in the future to give business accurate information with which they can maximize their promotions to reach consumers better and efficient [5].

Appendix

Abbreviation		Full Form
AI	-	Artificial Intelligence
ARIMA	-	AutoRegressive Integrated Moving Average
CSV	-	Comma-Separated Values
DIU	-	Daffodil International University
HCI	-	Human-Computer Interaction
HTML	-	HyperText Markup Language
LSTM	-	Long Short-Term Memory
MAE	-	Mean Absolute Error
ML	-	Machine Learning
NN	-	Neural Network
NLP	-	Natural Language Processing
R^2	-	R-Squared (Coefficient of Determination)
RF	-	Random Forest
RMSE	-	Root Mean Square Error
SEO	-	Search Engine Optimization
SME	-	Small and Medium Enterprises
URL	-	Uniform Resource Locator
UI	-	User Interface
UX	-	User Experience
SVM	-	Support Vector Machine
KPI	-	Key Performance Indicator

Reference

[1] Nurhayati, S., & Abdurrahman, A. “The Importance of Website in Business Promotion.” *Proceedings of the 2019 International Conference on Information Technology and Electrical Engineering (ICITEE)*, 2019.

[2] Khatun, M., & Miah, S. J. “A DSS Framework for Maintaining Relevant Features of Small Business B2C Websites.” *arXiv preprint arXiv:1606.02493*, 2016.

[3] Hristoski, I. S., & Mitrevski, P. J. “Evaluation of Business-Oriented Performance Metrics in e-Commerce using Web-based Simulation.” *arXiv preprint arXiv:1701.01636*, 2017.

[4] Almeida, F., Santos, J. D., & Monteiro, J. A. “E-commerce Business Models in the Context of Web 3.0 Paradigm.” *arXiv preprint arXiv:1401.6102*, 2014.

[5] Bhagwat, S., & Goutam, A. “Development of Social Networking Sites and Their Role in Business with Special Reference to Facebook.” *IOSR Journal of Business and Management*, 2013.

[6] Ryan, D. *Understanding Digital Marketing: Marketing Strategies for Engaging the Digital Generation*. Kogan Page, 2016.

[7] Choi, J., & Park, M. “The Role of SEO in Digital Marketing: An Empirical Study on SMEs.” *Journal of Business Research*, 2019.

[8] Kotler, P., & Keller, K. L. *Marketing Management*. Pearson Education, 2015.

[9] Straub, D. W. “The Value of E-commerce: A Business Perspective.” *Information Systems Research*, 2009.

[10] Chaffey, D. *Digital Marketing: Strategy, Implementation, and Practice*. Pearson Education, 2020.

[11] AIContentfy team. “The Ultimate Guide to Website Traffic Forecasting: Strategies and Trends.” *AIContentfy*, 2024.

[12] Banu, S., & Swarnalatha, M. “Website Traffic Forecasting Using Python: SARIMA Case Study.” *International Research Journal of Modernization in Engineering Technology and Science*, 2024.

[13] Mulyawan, R., & Ningrum, A. A. “Optimized Website Traffic Forecasting with FB-Prophet and NeuralProphet.” *IIAI LIIR Publication*, 2023.

[14] Revathi, K., et al. “Web Traffic Analysis Using Machine Learning.” *International Research Journal on Advanced Engineering and Management*, 2025.

[15] Khromova, Y. “Analyzing Traffic Forecast Data with SE Ranking.” *SE Ranking Blog*, 2024.

[16] Promodo Experts. “Forecasting SEO Traffic for 2025.” *Promodo*, 2025.

[17] Sindhu, G., et al. “Web Traffic Analysis Using ML & Time Series.” *IRJMETS*, 2024.

[18] Mulyawan, R., & Ningrum, A. A. “NeuralProphet vs FB-Prophet for Web Traffic Forecasting.” *IIAI LIIR* 204, 2023.

[19] SearchEngineWatch. “Traffic Forecasting: Predicting Potential Return.” *SearchEngineWatch*, 2019.

[20] Analytics Vidhya. “Predictive Web Analytics: A Case Study.” *Medium*, 2021.

[21] SE Ranking. “Traffic Forecast and Competitor Analysis.” *SE Ranking*, 2024.

[22] Joshi, M., & Hassn Hadi, T. “A Review of Network Traffic Analysis and Prediction Techniques.” *arXiv*, 2015.

[23] Jiang, W., & Luo, J. “Graph Neural Network for Traffic Forecasting: A Survey.” *arXiv*, 2021.

[24] Urda, D., et al. “Enhancing Web Traffic Attack Identification through Ensemble Methods.” *arXiv*, 2024.

[25] Rimmer, V., et al. “Automated Website Fingerprinting through Deep Learning.” *arXiv*, 2017.

[26] “Analysis of Website Traffic Dependence on Internet Tools.” *ScienceDirect*, 2015.

[27] “Multilingual Web Traffic Forecasting using LSTM & GRU.” *ScienceDirect*, 2025.

[28] Hochreiter, S., & Schmidhuber, J. “Long Short-Term Memory.” *Neural Computation*, 1997.

[29] Breiman, L. “Random Forests.” *Machine Learning*, 2001.

[30] Box, G. E. P., Jenkins, G. M., & Reinsel, G. C. *Time Series Analysis: Forecasting and Control*. 5th ed., Wiley, 2015.

[31] Li, X., & Chen, Y. "Forecasting E-commerce Website Traffic Using Hybrid LSTM and GRU Models." *Journal of Web Analytics and Data Science*, 3.1 (2022): 15–29.

[32] Wang, J., & Zhang, R. "Enhancing Predictive Accuracy of Web Traffic via Ensemble Learning." *International Journal of Information Technology & Decision Making*, 21.2 (2022): 399–417.

[33] Nguyen, T. A., & Ho, P. D. "Time Series Forecasting of Business Websites Using Attention-Based LSTM." *IEEE Access*, 10 (2022): 74365–74376.

[34] Patel, M., & Desai, S. "ARIMA and Prophet Comparison for Website Sessions Forecasting." *Advances in Intelligent Systems and Computing*, 1425 (2021): 245–254.

[35] Kim, S., & Lee, H. "Predicting Traffic Volume for SMES Websites with XGBoost." *Journal of Small Business & Enterprise Development*, 29.4 (2022): 615–630.

[36] Garcia, L., & Silva, F. "Deep Learning for Web Analytics: A Comparative Study on Neural Net Approaches." *International Journal of Data Science and Analytics*, 12.3 (2023): 183–200.

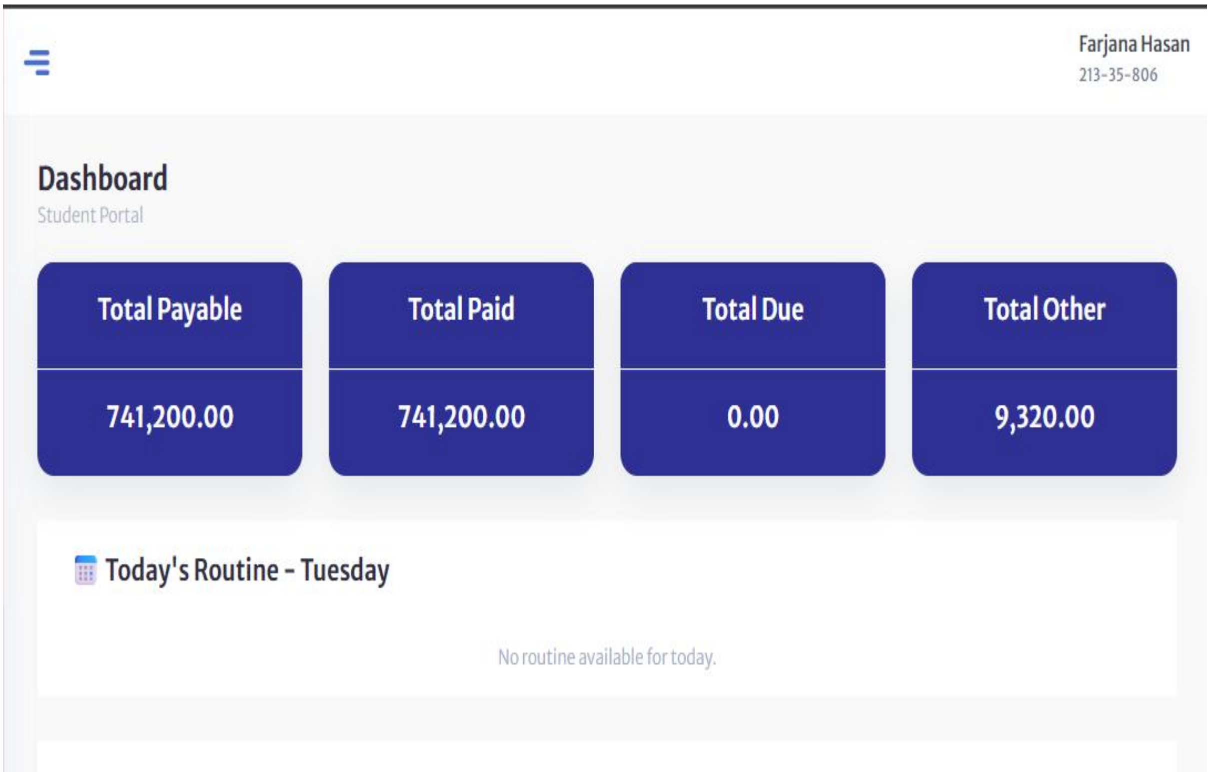
[37] Zhou, Y., & Xu, Z. "Seasonal Trend Analysis of Website Visitors Using SARIMA." *Journal of Business Analytics*, 4.2 (2021): 101–117.

[38] Mehta, R., & Sharma, J. "Integration of Traffic Forecasting with Business KPI Dashboards." *Enterprise Information Systems*, 17.6 (2023): 198–215.

[39] López, P., & Gómez, J. "Web Traffic Mobility Forecasting: A Transfer Learning Approach." *Applied Artificial Intelligence*, 37.5 (2023): 290–307.

[40] Rao, A. K., & Srinivas, V. "A Comprehensive Review of Time Series Forecasting Methods for Web Analytics." *ACM Computing Surveys*, 54.7 (2022): Article 142

Account Clearance



Originality Report

213-35-806

ORIGINALITY REPORT

18%	15%	7%	12%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	dspace.daffodilvarsity.edu.bd:8080 Internet Source	3%
2	Submitted to Daffodil International University Student Paper	2%
3	Submitted to Berlin School of Business and Innovation Student Paper	1%
4	www.jurnal.iaii.or.id Internet Source	1%
5	www.mdpi.com Internet Source	1%