Somasiri, Nalinda ORCID logoORCID: https://orcid.org/0000-0001-6311-2251, Karunarathne, Lakmali ORCID logoORCID: https://orcid.org/0009-0000-7720-7817 and Ganesan, Swathi ORCID logoORCID: https://orcid.org/0000-0002-6278-2090 (2024) Predictive Modeling of Tourist Arrivals in Sri Lanka Using Linear Regression. The 1st International Conference on Advanced Computing Technologies (ICACT 2024), 1 (1). pp. 52-62.

Downloaded from: https://ray.yorksj.ac.uk/id/eprint/12702/

The version presented here may differ from the published version or version of record. If you intend to cite from the work you are advised to consult the publisher's version: https://www.nsbm.ac.lk/event/icact-2024/

Research at York St John (RaY) is an institutional repository. It supports the principles of open access by making the research outputs of the University available in digital form. Copyright of the items stored in RaY reside with the authors and/or other copyright owners. Users may access full text items free of charge, and may download a copy for private study or non-commercial research. For further reuse terms, see licence terms governing individual outputs. Institutional Repositories Policy Statement

RaY

Research at the University of York St John

For more information please contact RaY at ray@yorksj.ac.uk

ResearchGate

See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/385379540

Predictive Modeling of Tourist Arrivals in Sri Lanka Using Linear Regression

Conference Paper · October 2024

CITATIONS
0

READS
222

3 authors:

Lakmali Shashika Karunarathne Mapa Senavige
York St John University
15 PUBLICATIONS 22 CITATIONS
SEE PROFILE

Nalinda Somasiri
York St John University
71 PUBLICATIONS 1,998 CITATIONS
SEE PROFILE

Swathi Ganesan
York St John University
31 PUBLICATIONS 61 CITATIONS
SEE PROFILE

All content following this page was uploaded by Lakmali Shashika Karunarathne Mapa Senavige on 30 October 2024.

The user has requested enhancement of the downloaded file.

ICACT

INTERNATIONAL CONFERENCE ON ADVANCED COMPUTING TECHNOLOGIES | 2024

NSBM GREEN UNIVERSITY TOWN

PROCEEDINGS

EMPOWERING DIGITAL TRANSFORMATION THROUGH ADVANCED COMPUTING INNOVATIONS

OCTOBER 24, 2024

RESEARCH CONFERENCE 2024 - ORGANIZED BY FACULTY OF COMPUTING

THE 01ST INTERNATIONAL CONFERENCE ON ADVANCED COMPUTING TECHNOLOGIES

“EMPOWERING DIGITAL TRANSFORMATION THROUGH ADVANCED COMPUTING INNOVATIONS”

FOCUS AREAS:

- Data Science & Artificial Intelligence
- Computer Networks, Cyber Security, and Computer Systems
- Computer Vision, Image, and Signal Processing
- Information Systems, HCI, and Enterprise Systems
- Smart Living, Digital Technologies and IoT
- Social Media and Web Technologies

Organized by

Faculty of Computing,
NSBM Green University
Pitipana, Homagama,

Disclaimer

The views expressed in the papers published in these proceedings are those of the authors. The statement and opinions stated either in the papers published herein or at the conference and discussions do not necessarily represent the views of the NSBM Green University of Sri Lanka. All papers are formatted using the seventh edition APA referencing style. If an author wants to present a work that has already been published or accepted by the conference, they need to write to the chair of the conference and inform that they are publishing the work elsewhere. However, the authors have the authority to present only a part of the study and not the content in its entirety. There might be some restrictions on how the authors can reuse the material that has been published/accepted.

ISSN 3051-5106
ISBN 978-955-23-0019-3

NSBM GREEN UNIVERSITY TOWN

All rights reserved.
Copyright ©2024 by NSBM Green University.

Conference Agenda

10:00 am	Registrations
11:00 am	Lighting of the Traditional Oil Lamp & Welcome
11:15 am	Welcome Speech Prof. Chaminda Wijesinghe, Dean – Faculty of Computing
11:20 am	Conference Proceedings Presenting
11:30 am	Keynote Speech 1, Prof. Nathan Clarke, Professor of Cyber Security and Digital Forensics
11:50 am	Entertainment Act – Conducted by NSBM Music Club
12:00 pm	Keynote Speech 2, Dr. Malith Jayasinghe, Vice President Research & AI, WSO2.
12:20 pm	Concluding Remarks, Dr. Mohamed Shafraz, Conference Chair
12:30 pm	Lunch
01:30 pm	Commencement of the Conference Parallel Sessions

PREFACE

We are pleased to present the proceedings of ICACT 2024 and extend a warm welcome to all authors and delegates attending this prestigious event. We are confident that the insights within will be both enlightening and transformative. This year marks a special milestone as the Faculty of Computing at NSBM Green University proudly hosts the International Conference on Advanced Computing Technologies (ICACT 2024) for the first time, pushing the frontiers of academic exploration. NSBM continues to lead in higher education by fostering innovation and promoting scholarly exchange.

ICACT 2024, under the theme "Empowering Digital Transformation through Advanced Computing Innovations" boldly ventures into the rapidly evolving technological landscape. In an age where digital advancements are reshaping industries and societies, this conference seeks to ignite discussions on the role of technology in driving meaningful and positive change. Emerging technologies have become essential for organizations and nations, presenting new opportunities and challenges at the intersection of business and technology.

This conference serves as an important platform to unite local and international researchers, academics, and industry leaders in exploring the profound effects of digital transformation on society. The chosen theme underscores the urgent need to harness digital innovation for social progress. Through strategic implementation, businesses, governments, and communities can work together to accelerate positive societal change via emerging technologies.

As we embark on this intellectual journey, we extend our heartfelt gratitude to Prof. E. A. Weerasinghe, Vice-Chancellor of NSBM Green University, and Prof. Chaminda Rathnayaka, Deputy Vice-Chancellor, for their unwavering support and visionary leadership. Special thanks go to Prof. J. Baratha Dodankotuwa, Head of Academic Development and Quality Assurance, for his valuable contributions, and to Prof. Chaminda Wijesinghe, Dean of the Faculty of Computing, for his expert guidance and knowledge.

We sincerely appreciate all the local and international presenters and participants whose contributions enhance the quality of this conference. ICACT 2024's success is a result of the hard work and dedication of our reviewers and the diligent efforts of the Conference Organizing Committee.

In conclusion, we expect ICACT 2024 to spark rich academic discussions and generate innovative solutions that contribute to a digitally transformed and progressive future. We wish all attendees a rewarding and memorable experience.

The Conference Organizing Committee
ICACT 2024

MESSAGE FROM VICE-CHANCELLOR

Prof. E.A. Weerasinghe
Vice Chancellor

It is with immense pleasure and pride that I warmly welcome you to the International Conference on Advanced Computing Technologies (ICACT) 2024, hosted by the Faculty of Computing at NSBM Green University. As we launch this inaugural research conference, we find ourselves at the intersection of academia, innovation, and the transformative potential of digital technologies.

Our theme for ICACT 2024, “Empowering Digital Transformation through Advanced Computing Innovations” highlights the ever-evolving landscape of research and innovation. In today’s fast-paced global environment, the seamless integration of emerging technologies with business processes is not just relevant—it is essential for driving societal progress. ICACT 2024 builds on the success of our previous initiatives, serving as a beacon for scholars, researchers, and industry experts to come together and examine the profound impacts of digital transformation. We stand on the brink of a new era, where the fusion of advanced computing technologies with business operations holds the potential to streamline processes and shape a future that is both sustainable and inclusive.

Universities play a crucial role in driving research and fostering innovation, contributing to economic and societal development. ICACT 2024 exemplifies this commitment by providing a platform for multidisciplinary researchers to share their work and engage in productive discussions on how digital technologies can empower businesses to operate more efficiently and contribute to societal growth. I extend my heartfelt congratulations to the organizing committee for their dedication to crafting a conference that promises to be a wellspring of knowledge, ideas, and innovative solutions. The high-quality papers and expert insights presented at this event are a reflection of our academic community's commitment to advancing the boundaries of knowledge.

A special thanks goes out to all the presenters and delegates whose contributions enrich our conference with insightful presentations and meaningful dialogues. Your participation plays a key role in strengthening our university’s research culture and, more importantly, in supporting broader national development. I eagerly anticipate the stimulating discussions that will take place at ICACT 2024. May this conference serve as a source of inspiration, collaboration, and transformative ideas, propelling us toward a future where digital transformation not only enhances business processes but also accelerates positive societal change.

Thank you!

MESSAGE FROM DEPUTY VICE-CHANCELLOR

Prof. Chaminda Rathnayake
Deputy Vice Chancellor

It is with great pleasure that I extend my warmest greetings to the inaugural International Conference on Advanced Computing Technologies (ICACT 2024), hosted by the Faculty of Computing at NSBM Green University. As the Deputy Vice-Chancellor, I am honored to welcome you to this esteemed event centered around the theme "Empowering Digital Transformation through Advanced Computing Innovations”.

In a time of remarkable technological progress, the influence of digital transformation on various aspects of society is both profound and wide-reaching. ICACT 2024 provides an essential platform for sharing ideas, insights, and research that deepen our collective understanding of the transformative role of digital innovation. At NSBM Green University, we are committed to nurturing an environment that promotes academic excellence, innovation, and collaboration. This conference perfectly aligns with our mission to encourage cutting-edge research and stimulate discussions that contribute to the advancement of business practices and societal well-being.

As we engage in discussions focused on empowering digital transformation through advanced computing, I encourage each of you to actively participate, share your knowledge, and explore opportunities for collaboration. The diverse viewpoints represented here will undoubtedly enhance our understanding of the challenges and opportunities that emerging technologies bring, especially in how they streamline and optimize business processes.

I would like to extend my sincere appreciation to the organizing committee, presenters, reviewers, and all participants for their dedication and efforts in bringing ICACT 2024 to life. Your commitment to advancing knowledge and driving positive change is truly commendable, and I am confident that this conference will foster meaningful developments in the field of business innovation through technology. May your experiences at ICACT 2024 be both intellectually stimulating and personally rewarding. I wish you a highly productive and inspiring conference.

Thank you!

MESSAGE FROM THE HEAD OF ACADEMIC DEVELOPMENT AND QUALITY ASSURANCE

Prof. J Baratha Dodankotuwa

Head Of Academic Development and Quality Assurance

It is my distinct pleasure to extend a warm welcome to all participants, researchers, and scholars who have gathered for the inaugural International Conference on Advanced Computing Technologies (ICACT 2024), proudly hosted by the Faculty of Computing at NSBM Green University. As the Head of Academic Development and Quality Assurance, I am thrilled to witness the assembly of brilliant minds and groundbreaking ideas converging around the central theme of "Empowering Digital Transformation through Advanced Computing Innovations”.

The rapidly evolving digital landscape is not only revolutionizing academia and industry but also fundamentally transforming the way we live and work. ICACT 2024 provides an exceptional platform for us to delve into the many ways in which emerging technologies—such as artificial intelligence, big data analytics, cloud computing, and blockchain—are driving this digital revolution. These technologies hold the potential to reshape business processes, making them more efficient, agile, and adaptable to the challenges of the modern world. The conference encourages us to explore how leveraging such technologies can streamline operations, optimize decision-making, and create new opportunities for growth and innovation.

At NSBM Green University, we are dedicated to upholding the highest standards of academic excellence. Conferences like ICACT play a critical role in fostering a culture of inquiry, collaboration, and continuous improvement. As we embark on this intellectual journey, I urge you to actively engage in the sessions, present your research, and participate in insightful discussions with your peers. The theme of this conference, focusing on digital transformation, emphasizes the need for businesses and societies to adapt swiftly and strategically to the digital era. By embracing advanced computing technologies, businesses can smooth out complex processes, reduce inefficiencies, and accelerate their path to success.

I would also like to take this opportunity to express my heartfelt gratitude to the organizing committee, and all contributors who have worked tirelessly to bring ICACT 2024 to life. Your unwavering dedication to advancing knowledge and fostering innovation is pivotal to the success of this conference. It is our hope that the knowledge shared, and the collaborations formed here will deepen our understanding of the transformative power of digital technologies, while also highlighting their ability to drive business innovation and societal progress. May your experience at ICACT 2024 be intellectually enriching and may the connections you make during this conference lead to fruitful partnerships that further propel the boundaries of technology-driven business and societal advancements.

Thank you!

MESSAGE FROM THE DEAN

Prof. Chaminda Wijesinghe
Dean, Faculty of Computing

Dear Students, Faculty Members, and Esteemed Guests,

It is with immense pride and joy that I extend my heartfelt congratulations to all participants and organizers of the first-time conference on International Conference on Advanced Computing Technologies (ICACT 2024) of the Faculty of Computing at NSBM Green University. Today marks a significant milestone in our academic journey, showcasing the innovative spirit, dedication, and intellectual prowess of our students.

The establishment of this first-ever conference is a testament to our commitment to fostering a robust research culture within the Faculty of Computing. It provides a unique platform for our students to present their pioneering research, engage in scholarly discussions, and collaborate with peers and experts in the field. Your hard work, curiosity, and passion for knowledge have brought us to this remarkable moment, and I am confident that the insights and discoveries presented today will pave the way for future advancements in computing and technology. I would like to express my gratitude to the faculty members, reviewers, and staff whose guidance and support have been instrumental in the success of our students' research endeavors. Your unwavering dedication to nurturing the next generation of innovators is truly commendable. To our students, I commend you for your perseverance and creativity. Your contributions are not only significant to your personal growth but also to the broader field of computing. As you present your research, remember that this symposium is just the beginning of your journey as a researcher and thought leader.

Congratulations once again to all participants for your outstanding achievements. I wish you a productive and inspiring symposium. Let us celebrate this momentous occasion and look forward to many more successful research symposia in the years to come.

Thank you!

MESSAGE FROM THE CONFERENCE CHAIR

Dr. Mohamed Shafraz
Conference Chair, ICACT 2024

I am deeply humbled and honored to share this message for the first-ever International Conference on Advanced Computing Technologies (ICACT) 2024, which will take place on 24th October 2024 at NSBM Green University. As South Asia's first green university, NSBM Green University continuously strives to make meaningful contributions to the nation and the global community through research and development, embodying our essential role in advancing academic endeavors.

The theme of ICACT 2024, "Empowering Digital Transformation through Advanced Computing Innovations," highlights the vital role of digital innovation in today's world. The conference aims to explore how technologies like AI, machine learning, IoT, and cloud computing are reshaping sectors such as healthcare, education, business, and governance. These emerging technologies are no longer confined to IT but drive efficiencies, optimize decision-making, and solve complex societal challenges across industries. This conference serves as a platform to explore the transformative potential of these innovations and how they can accelerate not only business success but also social and economic progress. In making ICACT 2024 a reality, countless individuals have contributed to every phase of the process.

This achievement would not have been possible without the visionary leadership of our Vice-Chancellor, Prof. E.A. Weerasinghe, whose guidance and encouragement have been a constant source of strength. I would also like to extend my heartfelt gratitude to Prof. Chaminda Rathnayake, Deputy Vice-Chancellor, and Prof. Baratha Dodankotuwa, Head of Academic Development and Quality Assurance, and Prof. Chaminda Wijesinghe, Dean of the Faculty of Computing for their immense support in enhancing the research culture at NSBM.

On behalf of the Organizing Committee, I would like to express a special note of gratitude to all local and international speakers, authors, reviewers, researchers, and presenters for their invaluable contributions to ICACT 2024. Your time, efforts, and insights are what make this conference a success. As the Conference Chair, I encourage all participants to actively engage in discussions, share their valuable expertise, and seek collaborative opportunities. The collective knowledge and expertise gathered here have the potential to drive significant change. Let us seize this moment to shape the digital transformation narrative, ensuring it aligns with our shared values, enhances business processes, and propels us toward a future where technology is a powerful force for positive societal change. Thank you.

ORGANIZING COMMITTEE

ADVISORY BOARD:
Prof. E. A. Weerasinghe - Vice Chancellor
Prof. Chaminda Rathnayaka - Deputy Vice-Chancellor
Prof. J. Baratha Dodankotuwa - Head, Academic Development and Quality Assurance
Prof. Chaminda Wijesinghe - Dean, Faculty of Computing
Prof. Noel Fernando - Professor, Faculty of Computing

CONFERENCE COMMITTEE:
Dr. Mohamed Shafraz - Conference Chair
Dr. Pabudi Abeyrathne - Conference Co-Chair
Mr. Chamara Dissanayake - Conference Co-Chair
Ms. Lakni Peiris - Conference Secretary

REVIEW BOARD:
Prof S.Vasanthapriyan
Prof. Nalinda Somasiri
Prof. Noel Fernando
Prof. Chaminda Wijesinghe
Prof Chaminda Wijesinghe
Dr. Dilrukshi Gamage
Dr. Kushani De Silva
Dr. Gihan Chathuranga
Dr. Hiruni Rupasinghe
Dr. Nirosha Wedasinghe
Dr. Shan Jayasinghe
Dr. Manoja Weerasekara
Dr. Pabudi Abeyrathne
Dr. Rasika Ranaweera
Dr. Mohamed Shafraz
Mrs. Udeshika Indumini
Ms. Dilhara Batanarachchi
Mr. Shashika Suren
Ms. Nimasha Arambepola
Mr. Indrajith Jayasuriya
Mr. Liyanage Perera
Ms. Chalani Oruthotaarachchi
Mr. Chamara Dissanayake
Mr. Chamara Attanayake
Ms. Pavithra Subashini
Mr. Isuru Sri Bandara
Ms. Hirushi Dilpriya
Mr. Gayan Perera

COMMITTEE MEMBERS:
Mr. Naji Saravanabavan
Mr. Chamindra Attanayake
Ms. Pavithra Kankanamge
Ms. Dulanjali Wijesekara
Ms. Natasha Chamba
Mr. Madusanka Mithrananda
Ms. Yasanthika Mathotaarachchi
Ms. Isuri Caldera
Mr. Gayan Perera
Ms. Nethmi Weearsinghe
Ms. Hirushi Dilpriya
Mr. Isuru Sri Bandara
Ms. Kavishka Rajapakshe
Ms. Sarangi Attanayake
Ms. Chathurma Wijesinghe
Ms. Chathurika Ranaweera
Ms. Gayathmi Kariyapperuma
Mr. Vimukthi Nilupul
Mr. Poorna Hettiararchchi

EDITORIAL COMMITTEE:
Mr. Anton Jayakody
Ms. Thisarani Wickaramasinghe
Mr. Diluka Wijesinghe
Ms. Sanuli Weerasinghe
Ms. Dharani Rajasinghe
Mr. Supun Gajendrasinghe

KEYNOTE SPEECH I

Prof. Nathan Clarke

Abstract: Cybersecurity continues to be a significant challenge with reports stating cybercrime will cost the global community $10.5 trillion by 2025. Cybercrime is not a new problem, but the scale is escalating to new levels, and the impact cuts across all of society. Artificial intelligence (AI) is seen by many as a potential step change in capability for cyber security - yet the prospect of truly intelligent systems also opens up equal opportunities for the hacker community. This presentation will draw upon Prof Clarke's 20+ years of developing AI and machine learning systems to aid in a variety of cyber security solutions - providing examples of where and how it can be used to provide more effective cyber security. The talk will also address the weaknesses of current approaches and the need for a new generation of explainable and trustworthy AI to ensure we understand how these systems operate to protect us. Finally, Prof Clarke will draw upon some examples of how technology can be misused and the potential dangers that exist for us all now and in the future.

KEYNOTE SPEECH II

Dr. Malith Jayasinghe

Abstract: Software engineering is a fundamental discipline within computer science, focusing on the systematic design, development, testing, and maintenance of software systems. Recent advancements in artificial intelligence (AI) have significantly transformed this field, leading to remarkable improvements in the software development process. AI enhances efficiency by automating tasks such as code generation, debugging, testing, and system maintenance through intelligent tools. However, it also brings forth several challenges, including ensuring accuracy, maintaining acceptable speed, protecting data privacy, addressing security vulnerabilities, and managing the inherent complexities of AI-driven systems. In this talk, I will explore AI's substantial impact on software engineering, the challenges associated with developing AI-enabled tools that foster these innovations, and strategies to overcome those challenges.

Table of Contents

Predictive Modeling of Tourist Arrivals in Sri Lanka Using Linear Regression

Nalinda Somasiri

Dept. of Computer Science & Data Science
York St John University London Campus
London, United Kingdom
n.somasiri@yorksj.ac.uk

Lakmali Karunarathne

Dept. of Computer Science & Data Science
York St John University London Campus
London, United Kingdom
l.karunarathne@yorksj.ac.uk

Swathi Ganesan

Dept. of Computer Science & Data Science
York St John University London Campus
London, United Kingdom
s.ganesam@yorksj.ac.uk

***Abstract* - One of the sectors of a developing nation's economy that directly generates income is tourism. For this reason, predicting the number of visitors is crucial when deciding on policies to upgrade facilities and other relevant aspects of this sector. The data for this paper were gathered from the Corporate website of the Sri Lanka tourism development authority(SLTDA)This paper attempts to forecast tourist arrivals in Sri Lanka using Linear Regression Model. The time span used for this study is from January 2021 to March 2024. The performance of the model was evaluated using metrics such as Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE) and R2 score. This paper forecasts tourist arrivals in Sri Lanka using the Linear Regression Model. While Linear Regression is suitable due to its simplicity and the scope of the dataset, more advanced forecasting models such as ARIMA or deep learning approaches like LSTM could have been considered for a more robust comparison. The study uses data from January 2021 to March 2024, and the model's performance is evaluated using metrics such as Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and R^2 score. The findings suggest that machine learning models can improve the forecasting accuracy of visitor arrivals, providing valuable insights for stakeholders in the tourism industry.**

Keywords: Linear Regression Model; forecast; tourist arrivals; predictions; MAE; RMSE; R2 score

I. INTRODUCTION

Sri Lanka's economy, like many others across the world, depends heavily on the tourism industry. Sri Lanka's tourist sector makes a substantial contribution to the country's GDP, employment, and foreign exchange profits. The island's distinctive wildlife, varied landscapes, and rich cultural legacy draw millions of visitors each year, making it one of the most dynamic areas of the national economy. Precise forecasting of visitor numbers is crucial for several reasons. Such as economic planning, resource management, policy formulation, crisis management and more. Better economic forecasting and planning are made possible by the ability to predict visitor arrivals. It ensures that companies, the government, and infrastructure may adjust resources, infrastructure, and services in response to variations in the number of tourists. Stakeholders can more efficiently manage resources, such as lodging, transportation, and attractions, by knowing predicted visitor inflows. This contributes to upholding excellent service standards and improving the traveler experience. Precise forecasts aid authorities in developing plans to boost travel, upgrade infrastructure, and guarantee environmentally friendly tourism practices. It also aids in the creation of regulations to lessen the negative impacts of tourism, such as deterioration of the environment and dilution of culture. The travel and tourist industry are susceptible to a range of external shocks, including natural disasters, pandemics, and geopolitical crises. By offering early warnings and assisting in the creation of backup plans, predictive modelling can help with crisis management [1], [2], [3].

Predictive modelling using machine learning approaches has demonstrated encouraging results in a number of sectors, including tourism, in recent years. Compared to conventional statistical techniques, machine learning models provide more reliable and accurate predictions by analyzing intricate patterns and correlations inside huge datasets.

Using linear regression, one of the foundational and extensively used machine learning techniques, this study attempts to create a predictive model for tourist arrivals in Sri Lanka. The model endeavors to furnish dependable forecasts that may facilitate strategic deliberation and planning in the tourist domain by using past data and diverse economic factors.

A. Research Objectives

The main goal of this research is to use machine learning techniques, notably linear regression, to create a reliable predictive model for the number of tourists that will arrive in Sri Lanka. The goal of the project is to produce accurate and trustworthy estimates of tourist arrivals by examining historical data on arrivals as well as a range of economic factors. This will enable stakeholders to plan ahead and anticipate changes in the demand for tourism.

Furthermore, in order to provide important insights into the dynamics of the tourism industry, the study attempts to identify and comprehend the major external and economic factors impacting the number of tourists that arrive. Based on the insights obtained from the model, this predictive capability is meant to assist enterprises, tourism authorities, and policymakers in making well-informed decisions about infrastructure development, resource allocation, and strategic planning. Additionally, by giving early warning signals through precise predictions, the study seeks to enhance stakeholders' capacity to manage and reduce the effects of external shocks, such as pandemics and economic downturns. By enabling improved planning and management of tourist intakes, this study ultimately hopes to contribute to the development of sustainable tourism practices by striking a balance between economic rewards and environmental and cultural preservation. By taking an all-encompassing approach, we hope to show how machine learning can be used to forecast tourism trends and promote the sustainable growth and development of Sri Lanka's tourism sector [1], [2].

II. LITERATURE REVIEW

For many nations, like Sri Lanka, where it contributes significantly to GDP, employment, and foreign exchange profits, tourism is an important economic industry. Predicting visitor numbers with precision is essential for economic planning, resource management, and policy development. This study of the literature looks at the situation of tourism in Sri Lanka right now, looks at earlier research on visitor arrivals, and emphasizes the economic factors that affect tourism patterns.

A. Tourism Industry in Sri Lanka

Over the past few decades, Sri Lanka's tourism industry has grown significantly, propelled by the island's unique wildlife, varied landscapes, and rich cultural legacy. The Sri Lanka Tourism Development Authority (SLTDA) reports that there has been a steady growth in the number of tourists arriving, with particular surges in the years leading up to the COVID-19 pandemic. The SLTDA offers thorough data on a range of tourism-related topics, such as visitor demographics, spending trends, and seasonal variations, all of which are essential for comprehending the workings of the sector [1].

The Central Bank of Sri Lanka (CBSL) offers economic indicators that shed more light on the financial aspects of the travel and tourist industry. Important indicators like GDP share, inflation rates, and currency rates have a direct impact on travel-related activities. The economic indicators of the CBSL show how closely related the tourism industry is to other economic sectors, underscoring the necessity of a comprehensive approach to tourism management and policy-making [2].

B. Economic Indicators Influencing Tourism

Arrivals of tourists in Sri Lanka are influenced by a number of economic factors, such as monetary policy, GDP growth, inflation, and exchange rates. The World Bank offers a wealth of information about Sri Lanka's economic circumstances, including insights into long-term patterns and structural shifts in the country's economy. For example, the affordability and appeal of Sri Lanka as a travel destination can be impacted by GDP growth rates and inflation levels. Changes in exchange rates can affect how affordable the resort is for tourists from other countries [3].

Understanding the macroeconomic conditions affecting tourism requires knowledge of the money supply, interest rates, and fiscal policies, all of which are covered in detail in the Central Bank of Sri Lanka's economic indicators report. A shift in the central bank rate or the average weighted fixed deposit rate may have an indirect effect on tourism by influencing investment conditions and general economic stability [2].

C. Previous Studies on Tourist Arrivals in Sri Lanka

In emerging nations, tourism has become a major engine of economic expansion, created jobs and increased foreign exchange reserves. Sri Lanka's post-war economic recovery is heavily dependent on tourism, which makes precise forecasts of foreign visitor arrivals necessary for informing business environments, infrastructure development, and successful governmental decisions.

In their paper "Forecasting international tourist arrivals in formulating tourism strategies and planning: The case of Sri Lanka," the authors create accurate forecasting models for both the top 10 source countries and total international arrivals using the Seasonal Autoregressive Integrated Moving Average (SARIMA) method. The training dataset consisted of monthly data from January 1984 to December 2016, and the accuracy of the model was tested using data from January 2017 to December 2017.

The SARIMA approach was selected because of its resilience in capturing the intricate seasonal patterns present in tourism data, as well as its reputation for handling seasonal changes. The results highlight a number of important insights such as four million tourist arrivals by 2020 is an ambitious goal set by the Sri Lankan government, forecasting accuracy is critical to strategic tourism planning and policy formulation and SARIMA models show efficacy in producing accurate forecasts even in the face of seasonal fluctuations. By illustrating the usefulness of SARIMA models in predicting the number of foreign visitors, this study makes a substantial contribution to the literature on tourism. Governmental organizations and private sector stakeholders engaged in infrastructure planning, advertising campaigns, and policymaking in the tourism sector can benefit from the actionable insights this research offers. Stakeholders can optimize resource allocation and strategic actions to optimize the socio-economic advantages of tourism in Sri Lanka by utilizing these forecasting methodologies [4].

The study "Time Series Patterns of Tourist Arrivals to Sri Lanka" examined monthly time series data from January 2008 to December 2014, with a particular focus on the top four market regions—Asia, Western Europe, Eastern Europe, and the Middle East—for tourist arrivals. To find trends in arrivals in various regions, the study used Time Series plots, Auto-Correlation Functions (ACF), and descriptive statistics. The study's approach involved comparing the mean number of tourists arriving in each of the chosen regions using one-way Analysis of Variance (ANOVA). The results showed that, with 29,361 and 25,982 average arrivals, respectively, Asia and Western Europe had the highest arrival rates, with no discernible difference between the two regions. On the other hand, average arrivals in the Middle East and Eastern Europe were lower, at 4,300 and 5,866, respectively. Important distributional features of tourist arrivals were also emphasized by the study: arrivals from Asia, Western Europe, and Eastern Europe were non-normally distributed and showed positive skewness. Additionally, it was discovered that the data series for all four regions were non-stationary, highlighting the dynamic character of visitor flows to Sri Lanka. Also, this study suggested investigating more complex time series forecasting techniques in the future, including moving averages, exponential smoothing, decomposition methods, linear and non-linear trend models, and circular models. These techniques are suggested in order to improve forecast accuracy and identify the cyclical and seasonal trends present in Sri Lankan tourism arrivals [5].

A recent study by Basnayake et al. (2023) employs a time-series approach with Change Point Analysis (CPA) and Seasonal Autoregressive Integrated Moving Average (SARIMA) models to model and forecast tourist arrivals in Sri Lanka. The analysis makes use of monthly data on visitor arrivals from January 2000 to December 2019, a time that was remarkable for important events like the tsunami and the civil war, both of which had a major impact on tourism numbers. The goal of the authors' SARIMA model is to adequately capture seasonal fluctuations in visitor arrivals by extending standard ARIMA models to integrate seasonality. The best-performing models are chosen using the Akaike Information Criterion (AIC), and their accuracy is assessed using the Normalized Root Mean Squared Error (NRMSE) and Mean Absolute Percentage Error (MAPE) [6].

III. METHODOLOGY

Time series data of tourist arrivals, denoting visits of foreigners to Sri Lanka for the purpose of holidays, business, sight-seeing, excursions, medical treatment, religious pilgrimages, and other purposes that are uniquely different from their usual place of residence, were used to forecast future trends. The data spans from January 2021 to March 2024 , obtained in monthly time intervals from the Sri Lanka Tourism Development Authority. The secondary data for the study was collected from publications, reports, books, journals, and relevant websites that included data on tourist arrivals. Then the data was prepared for analyzing by finding null values, duplicate values, rearranging datatypes at the data pre preparation stage. After that the data set was analyzed by splitting it into training and testing sets and employed the Linear Regression model to get the forecast tourist arrivals in Sri Lanka. The Linear Regression Model was chosen for its simplicity and interpretability. More advanced models like ARIMA or LSTM, which could account for non-linearities in the data, were considered but excluded in favour of Linear Regression due to the linear nature of the dataset and the study’s focus on simplicity. However, these models offer benefits like handling seasonality and long-term dependencies, making them worthy of exploration in future work.

The dataset was pre-processed by addressing missing data using mean imputation and removing duplicates. Handling outliers was done by reviewing statistical distributions, ensuring the cleaned data is suitable for predictive modelling. Expanding the details on how the dataset was pre-processed improves the study’s reproducibility.

A. Data Collection

The study's dataset includes historical information on visitor arrivals to Sri Lanka from January 2021 to March 2024 and the dataset contains 39 rows and 13 columns. The 13 columns contain Date, Head Inflation, Core Inflation, Tourist Arrivals, Covid, GDP, %GDP, Rev(%GDP), Expenditure(%GDP), Money Supply(Annual Change), Money Supply(%GDP), Average Weighted Fixed Deposit Rate (AWFDR) Rate(%) and the Central Bank Rate. The Head Inflation columns represent the overall inflation rate in the economy and normally it measured by the consumer price index (CPI). It reflects the average change in prices paid by consumers for good and services. Same as Head Inflation the Core Inflation reflect the prices of some volatile goods like food and energy excluded. A better picture of the economy's underlying long-term inflation pattern can be obtained by looking at core inflation and it also measured by the CPI. The number of tourists that arrive in Sri Lanka during the mentioned period is shown in Tourist Arrivals column and this is the study's main topic and the target variable for the predictive model. The Covid column, which is a binary variable with a value of 0 for no impact and a value of 1 for impact throughout the time, represents the effect of the COVID-19 pandemic on tourist arrivals. GDP refers to the Gross Domestic Product which is represent the total economic output of Sri Lanka in the mentioned period which is January 2021 to March 2024 and %GDP refers to the proportion of the economic output of tourist arrivals in Sri Lanka relative to the total GDP of this time period. From Rev(%GDP) represent the government revenue as a percentage of GDP and it shows the percentage of the nation's total economic production that the government receives in revenue from tourist arrivals. Expenditure(%GDP) column represent the expenditure of tourist arrivals as a percentage of GDP. This expenditure could be various services or projects which is done for the tourist arrivals in Sri Lanka. Money Supply(Annual Change) column shows the annual change in the money supply and the Money Supply(%GDP) represent the money supply as a percentage of GDP. Average Weighted Fixed Deposit Rate (AWFDR) Rate(%) indicates the average interest rate offered on fixed deposits by financial institutions, weighted by the number of deposits. This reflects the cost of borrowing for tourist arrivals. Central Bank Rate represents the policy interest rate set by the central bank of Sri Lanka [2], [1], [3].

```
#Import the Libraries
import pandas as pd
import dateutil
import seaborn as sns
import matplotlib.pyplot as plt
import numpy as np
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
from sklearn.metrics import mean_squared_error, mean_absolute_error
from sklearn.preprocessing import StandardScaler, MinMaxScaler
from sklearn.metrics import r2_score
```

Figure 2: Import Libraries

After selecting the dataset, which is suitable for this study, first imported the relevant libraries to the Jupitar notebook for the further analysis. From these libraries pandas library is more important to import at first because it helps to load the dataset to further analysis. This Pandas library has functions for analyzing, cleaning, exploring and manipulating data and it has an ability to handle large dataset. Also importing NumPy library is help to working with the arrays.

```
#Load the dataset to the pandas library
data = pd.read_csv('economic_data.csv')
```

Figure 3: Load the dataset

After importing required libraries, then load the dataset to the pandas library for further analysis. Here the pandas library offers the 'pd.read_csv' function to read a comma-separated values (CSV) file which is named as 'economic_data.csv' into a pandas DataFrame and assign it to the variable name 'data'.

```
#Get the first few raws
data.head()
#Get the last few raws
data.tail()
```

Figure 4: Get first and last few rows

As a common step of after loading the dataset to the pandas DataFrame get the first and last few rows using 'data.head()' and 'data.tail()' to check whether the dataset has been properly updated.

```
#Find the shape of the dataset
print('Number of rows in the dataset: ', data.shape[0])
print('Number of columns in the dataset: ', data.shape[1])
```

Figure 5: Get the shape of the dataset

Then using 'data.shape', the researcher can get the idea about the number of rows and columns of the dataset.

```
#Get the columns of the dataset
print(data.columns)
```

Figure 6: Print the column names

Using 'data.columns' can get the information about the column names and using print function the column names can be print to the console.

```
#Get the information of the dataset
data.info()
#Get the description of the dataset
data.describe()
```

Figure 7: Get the information and the description of dataset

Using 'data.info()', all the non-null count and the data type of each column can be printed to the console and the researcher can get an idea about the memory usage to the dataset as well. From 'data.describe()' the count, mean, standard deviation, minimum, 25%, 50%, 75% and the maximum value of each column can be taken.

B. *Data Preprocessing*

a. *Handling Missing Values and Duplicates*

Missing values are a common issue in machine learning, which has formed the foundation for data analysis and extraction. A number of reasons can lead to missing values, such as missing entirely, missing at random, or missing not at random. These could all be the consequence of human error in the pre-processing stage of the data or a system failure during data collecting. However, it is crucial to address missing values before doing any data analysis because doing so could lead to biassed or incorrect conclusions [7, Vol. 8]. Missing values can be handled in a variety of ways, many of which are rather straightforward. A few of the techniques are based on statistical principles such as ignoring, deletion, mean/mode imputation. The first and most straightforward method for handling missing values is to ignore them, as previously mentioned. The analysis proceeds with complete disregard for the missing values. Even though it's a straightforward technique, it can be highly dangerous if the data's missingness percentage is high enough to affect the analysis's conclusion. As stated, the deletion approach involves simply removing the observation data instance or missing variable in order to proceed with the analysis or data mining process. Mean/Mode Imputation method comes as a solution to give a better result, and it solves the missing values problem and the number of data that is expected to be proceed is remain to be the same. While ignoring and deletion did not give a good result of an analysis or data mining process caused by missing values (Ignoring) and less data to be proceed (Deletion), this method works to improve the results. However, this method's flaw is the bias it introduces because so many of the data's values are similar [8]. Inaccurate findings may arise from an analysis that is distorted by duplicate records. To keep data integrity, duplicates must be found and eliminated. The dataset was made ready for additional analysis and modelling by methodically removing duplicates and missing values. This made sure that the predictive model that was created was founded on accurate and clean data.

```
#Check the null values of the dataset
data.isnull().sum()
```

```
print('The number of null values in the dataset: ',data.isnull().sum().sum())
```

Figure 8: Get the null values

From 'data.isnull().sum()' , the result is a Series object, each element of which reflects the quantity of missing data in the relevant DataFrame column. Using data.isnull().sum().sum() can get the total number of null values of the dataset.

```
#Check duplicate values
data.duplicated().sum()
```

Figure 9: Check the duplicates

From data.duplicated().sum() the duplicated values of this dataset can be identified, and this step is essential for the data cleaning and preparing for analysis.

```
#Fill the null value of the 'GDP' column using the mean value of the 'GDP' Column and get the 'GDP'
data['GDP'].fillna(data['GDP'].mean(), inplace=True)
data['GDP']
```

```
#Fill the null values of the 'GDP%', 'Rev(%GDP)', 'Expenditure(%GDP)', 'Money Supply(Annual Change)',
#'Money Supply(%GDP)' columns using the mean value of each Column
data['GDP%'].fillna(data['GDP%'].mean(), inplace=True)
data['Rev(%GDP)'].fillna(data['Rev(%GDP)'].mean(), inplace=True)
data['Expenditure(%GDP)'].fillna(data['Expenditure(%GDP)'].mean(), inplace=True)
data['Money Supply(Annual Change)'].fillna(data['Money Supply(Annual Change)'].mean(), inplace=True)
data['Money Supply(%GDP)'].fillna(data['Money Supply(%GDP)'].mean(), inplace=True)
data
```

Figure 10: Fill the null values using mean values of each column

The next step after finding null values and duplicates is to address for those values if exists. Here the null values which are in the 'GDP', 'GDP%', 'Rev(%GDP)', 'Expenditure(%GDP)', 'Money Supply (Annual Change)', 'Money Supply(%GDP)' filled with the mean value of each column. After filling the null values, using data.isnull().sum().sum() can confirm whether there are any other null values in the dataset.

b. Datatype Conversion

It is necessary to convert each column in the dataset to the proper data type in order to guarantee the precision and effectiveness of data processing and analysis.

```
# Convert date from string to date times
data['Date'] = data['Date'].apply(dateutil.parser.parse, dayfirst=True)
data.head(10)
```

Figure 11: Convert data type

Access the date column using 'data['Date']' and from '.apply(dateutil.parser.parse, dayfirst=True)', apply the function to each element in the column of the 'Date' column in the DataFrame and in the date format, the day comes before the month according to the 'dayfirst=True' parameter.

C. Model Selection and Evaluation

In this study the Linear Regression model in machine learning has been used to get the prediction of tourist arrivals in Sri Lanka. The Linear Regression (LR) approach is a significant machine learning methodology for future data prediction. Predictive modelling, which is used to determine the relationship between input and output variables, is the main application for this technique. The model is expressed as a set of input values, x and y, with two coefficients; the first coefficient denotes the scaling factor, while the second coefficient, sometimes known as the intercept coefficient, denotes the degree of freedom on the classifier line [9].

At first found the correlation which is the quantity of degree in which two variables are associated. After that rearrange the data in order by date.

```
# Plot the line graph
plt.figure(figsize=(10, 6))
plt.plot(data['Date'], data['Head Inflation'], marker='o', linesty
plt.title('Head Inflation Over Time')
plt.xlabel('Date')
plt.ylabel('Head Inflation')
plt.grid(True)
plt.xticks(rotation=45)  # Rotate x-axis labels for better readabi
plt.tight_layout()
plt.show()
```

Figure 12: Head inflation over time

Then plot the line graph about 'Head Inflation Over Time' which is get an idea about the head inflation with respective to the time. Here the figure size is (10,6) and the 'Date' is for X axis and 'Head Inflation' for Y axis.

```
# Plot the line graph
plt.figure(figsize=(10, 6))
plt.plot(data['Date'], data['Average Weighted Fixed Deposit Rate (AWFDR) Rate (% )'], marker='o', linestyle='-', color
plt.title('AWFDR Over Time')
plt.xlabel('Date')
plt.ylabel('Average Weighted Fixed Deposit Rate (AWFDR) Rate (% )')
plt.grid(True)
plt.xticks(rotation=45)  # Rotate x-axis labels for better readability
plt.tight_layout()
plt.show()
```

Figure 13: Average weighted fixed deposit rate

Then plot the 'Average Weighted Fixed Deposit Rate (AWFDR) Rate (%)' over time to get an idea about AWFDR with respective to the 'Date'.

```
# Plot the line graph
plt.figure(figsize=(10, 6))
plt.plot(data['Date'], data['Central Bank Rate'], marker='o', linestyle='-', color='g
plt.title('Central Bank Rate Over Time')
plt.xlabel('Date')
plt.ylabel('Central Bank Rate')
plt.grid(True)
plt.xticks(rotation=45)  # Rotate x-axis labels for better readability
plt.tight_layout()
plt.show()
```

Figure 14: Central bank rate over time

```
# Plot the line graph
plt.figure(figsize=(10, 6))
plt.plot(data['Date'], data['Tourist Arrivals'], marker='o', linestyle='-', color='g')
plt.title('Tourist Arrivals Rate Over Time')
plt.xlabel('Date')
plt.ylabel('Tourist Arrivals')
plt.grid(True)
plt.xticks(rotation=45)  # Rotate x-axis labels for better readability
plt.tight_layout()
plt.show()
```

Figure 15: Tourist arrival rate over time

Same as above 'Central Bank Rate Over Time' and 'Tourist Arrivals Rate Over Time' plotted with respective to the 'Date' to get an idea about each variable changes with the tome.

Then, to work with the Linear Regression model X and Y identified as below.

```
X = data[['Head Inflation', 'Core Inflation', 'Covid',
        'GDP', 'GDP%', 'Rev(%GDP)', 'Expenditure(%GDP)',
        'Money Supply(Annual Change)', 'Money Supply(%GDP)',
        'Average Weighted Fixed Deposit Rate (AWFDR) Rate (% )',
        'Central Bank Rate']]
```

```
y = data[['Tourist Arrivals']]
```

Figure 16: Identifying X and Y

Then split dataset in to training and testing. The training test is 80% of the dataset and testing set is 20% of the dataset. From 'train_test_split' function from 'sklearn.model_selection' module which is in the 'scikit-learn' library. Also, it splits arrays and metrices into random train and test subsets.

```
# Split the data into training and testing sets
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

print(len(X_train),len(X_test))
print(X_train.shape)
print(y_train.shape)
print(X_test.shape)
print(y_test.shape)
```

Figure 17: Split dataset into training and testing

Here X represents the feature data and y represents the target data. 'X_train' is the training set of features, and 'X_test' is the testing set of features. 'y_train' is the training set of target values, and 'y_test' is the test set of target values. This split's objective is to train a machine learning model on a training set of data and assess its performance on a test set of data that it hasn't seen yet. This aids in evaluating the model's ability to generalize to fresh, untested data.

```
# Initialize and train the linear regression model
model = LinearRegression()
model.fit(X_train, y_train)
```

Figure 18: Initialize and train the model

By doing this, a Linear Regression object from the sklearn.linear_model module is created. From 'LinearRegression()' by fitting a linear equation to observable data, it describes the connection between a dependent variable (goal) and one or more independent variables (features).

```
# Make predictions on the testing set
predictions = model.predict(X_test)

# Evaluate the model
mse = mean_squared_error(y_test, predictions)
mae = mean_absolute_error(y_test, predictions)
print(f'Mean Squared Error: {mse}')
print(f'Mean Absolute Error: {mae}')


# Visualize predicted vs. actual tourist arrivals
plt.scatter(y_test, predictions)
plt.xlabel('Actual Tourist Arrivals')
plt.ylabel('Predicted Tourist Arrivals')
plt.title('Actual vs. Predicted Tourist Arrivals')
plt.show()
```

Figure 19: Make predictions and evaluate the model

By utilizing the trained linear regression model to make predictions on the testing set, this code assesses the model's performance by computing the Mean Squared Error (MSE) and Mean Absolute Error (MAE). These metrics represent the average squared and absolute differences between the predicted and actual values. For evaluation, the MSE and MAE are printed. In order to help evaluate the visual accuracy of the model's predictions, the code also plots the relationship between the actual and expected arrivals of tourists on a scatter plot.

```
plt.figure(figsize=(10, 6))
plt.scatter(y_test, predictions, label='Predicted Tourist Arrivals')
plt.scatter(y_test, y_test, label='Actual Tourist Arrivals')
plt.xlabel('Actual Tourist Arrivals')
plt.ylabel('Predicted Tourist Arrivals')
plt.title('Actual vs. Predicted Tourist Arrivals')
plt.legend()  # Add legend to distinguish between predicted and actual val
plt.grid(True)
plt.show()
```

Figure 20: Actual tourist arrivals vs predicted tourist arrivals

After that, a scatter plot with a given figure size of 10 by 6 inches is created by the above code to improve the visualization of the link between actual and expected tourist arrivals. It depicts, with labels to differentiate between them, the actual and predicted numbers of tourists arriving on the same graph. The actual tourist arrivals are shown on the x-axis, while the anticipated tourist arrivals are shown on the y-axis. A grid is enabled to enhance readability, and a legend is given to distinguish between the actual and forecasted numbers. This graphic comparison aids in assessing the regression model's performance and accuracy.

```
# Calculate R2 score
r2 = r2_score(y_test, predictions)

# Print R2 score
print(f'R2 Score: {r2}')
```

Figure 21: Calculate R2 score

At last, the R2 score, a statistical indicator of how closely the regression model's predictions match the actual data, is computed by the code mentioned above. R2_score(y_test, predictions) is used to generate the R^2 score, which has a range of 0 to 1. A value of 1 denotes perfect prediction accuracy, whereas values closer to 0 indicate poor predictive ability. The R2 score, which shows how much of the variance in the dependent variable is predictable from the independent variables, is calculated and then printed to the console with a formatted text to give a clear indication of the model's explanatory ability.

IV. RESULTS AND DISCUSSION

This dataset, which included data on arrivals of tourists and a variety of economic factors, was used to train and assess the linear regression model. Accurately forecasting the number of tourists arriving in Sri Lanka was the main objective. Mean Squared Error (MSE), Mean Absolute Error (MAE), and R-squared (R^2) value were used to evaluate the model's performance. Before approaching to the Linear Regression model, the basic idea of the dataset has been taken such as the size of the dataset with rows and columns, data type, column headings, null values, duplicates and etc.

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 39 entries, 0 to 38
Data columns (total 13 columns):
 #   Column                                          Non-Null Count  Dtype
---  ------                                          --------------  -----
 0   Date                                            39 non-null     object
 1   Head Inflation                                  39 non-null     float64
 2   Core Inflation                                  39 non-null     float64
 3   Tourist Arrivals                                39 non-null     float64
 4   Covid                                           39 non-null     float64
 5   GDP                                             24 non-null     float64
 6   GDP%                                            24 non-null     float64
 7   Rev(%GDP)                                       24 non-null     float64
 8   Expenditure(%GDP)                               24 non-null     float64
 9   Money Supply(Annual Change)                     24 non-null     float64
 10  Money Supply(%GDP)                              24 non-null     float64
 11  Average Weighted Fixed Deposit Rate (AWFDR) Rate (% )  39 non-null     float64
 12  Central Bank Rate                               39 non-null     float64
dtypes: float64(12), object(1)
memory usage: 4.1+ KB
```

Figure 22: Information of the dataset columns

According to the result of 'data.info()' all the columns have the datatype of 'float64' except date column and it has the data type of 'object'.

```
Date                                                     0
Head Inflation                                           0
Core Inflation                                           0
Tourist Arrivals                                         0
Covid                                                    0
GDP                                                     15
GDP%                                                    15
Rev(%GDP)                                               15
Expenditure(%GDP)                                       15
Money Supply(Annual Change)                             15
Money Supply(%GDP)                                      15
Average Weighted Fixed Deposit Rate (AWFDR) Rate (% )    0
Central Bank Rate                                        0
dtype: int64
```

Figure 23: Null values

The null values were observed and 'GDP', 'GDP%', Rev(%GDP), Expenditure(%GDP), Money Supply (Annual Change) and Money Supply (%GDP) has 15 null values for each. So, the total number of null values of the dataset was 90 and there were no duplicates values.

	Date	Head Inflation	Core Inflation	Tourist Arrivals	Covid	GDP	GDP%	Rev(%GDP)	Expenditure(%GDP)	Money Supply(Annual Change)	Money Supply(%GDP)	Average Weighted Fixed Deposit Rate (AWFDR) Rate (%)	Central Bank Rate
0	2024-01	6.5	2.2	208253.0	0.0	1.252789e+13	-2.155	8.325	18.755	11.2	49.12	14.28	14.50
1	2024-03	2.5	3.4	209181.0	0.0	1.252789e+13	-2.155	8.325	18.755	11.2	49.12	13.10	14.00
2	2023-01	53.2	52.0	102545.0	0.0	1.252789e+13	-2.155	8.325	18.755	11.2	49.12	19.15	26.50
3	2023-02	53.6	50.1	107639.0	0.0	1.252789e+13	-2.155	8.325	18.755	11.2	49.12	19.23	26.00
4	2023-03	49.2	44.2	125495.0	0.0	1.252789e+13	-2.155	8.325	18.755	11.2	49.12	19.80	25.00
5	2023-04	33.6	31.8	105498.0	0.0	1.252789e+13	-2.155	8.325	18.755	11.2	49.12	19.72	24.50

Figure 24: First few rows after filling null values

After filling the null values with the mean value of each column the dataset was printed again and recheck the null values of the dataset again and it shows as zero as below.

```
Date                                                     0
Head Inflation                                           0
Core Inflation                                           0
Tourist Arrivals                                         0
Covid                                                    0
GDP                                                      0
GDP%                                                     0
Rev(%GDP)                                                0
Expenditure(%GDP)                                        0
Money Supply(Annual Change)                              0
Money Supply(%GDP)                                       0
Average Weighted Fixed Deposit Rate (AWFDR) Rate (% )    0
Central Bank Rate                                        0
dtype: int64
```

Figure 25: Recheck the null values

	Head Inflation	Core Inflation	Tourist Arrivals	Covid	GDP	GDP%	Rev(%GDP)	Expenditure(%GDP)	Money Supply(Annual Change)	Money Supply(%GDP)	Ave Weig F De (AW Rate
count	39.000000	39.000000	39.000000	39.000000	3.900000e+01	39.000000	39.000000	39.000000	39.000000	39.000000	39.00
mean	23.976923	21.038462	77886.153846	0.307692	1.252789e+13	-2.155000	8.325000	18.755000	11.200000	49.120000	12.09
std	24.858023	22.460288	64367.010722	0.467572	4.053408e+11	4.502085	0.003974	0.965584	1.827855	4.490165	5.60
min	0.800000	0.600000	1497.000000	0.000000	1.201785e+13	-7.820000	8.320000	17.540000	8.900000	43.470000	5.62
25%	5.100000	4.100000	26286.500000	0.000000	1.201785e+13	-7.820000	8.320000	17.540000	8.900000	43.470000	6.12
50%	10.800000	8.800000	82327.000000	0.000000	1.252789e+13	-2.155000	8.325000	18.755000	11.200000	49.120000	13.10
75%	47.250000	40.950000	108419.000000	1.000000	1.303793e+13	3.510000	8.330000	19.970000	13.500000	54.770000	17.41
max	73.700000	64.100000	218350.000000	1.000000	1.303793e+13	3.510000	8.330000	19.970000	13.500000	54.770000	19.84

Figure 26: Description of the dataset

Then the description of dataset has been taken and it shows all the summary of the dataset such as count, mean, standard deviation, minimum, 25%, 50%, 75% and maximum.

	Date	Head Inflation	Core Inflation	Tourist Arrivals	Covid	GDP	GDP%	Rev(%GDP)	Expenditure(%GDP)	Money Supply(Annual Change)	Money Supply(%GDP)	Average Weighted Fixed Deposit Rate (AWFDR) Rate (%)	Centr Ban Rat
0	2024-01-24	6.5	2.2	208253.0	0.0	1.252789e+13	-2.155	8.325	18.755	11.2	49.12	14.28	14.
1	2024-03-24	2.5	3.4	209181.0	0.0	1.252789e+13	-2.155	8.325	18.755	11.2	49.12	13.10	14.
2	2023-01-24	53.2	52.0	102545.0	0.0	1.252789e+13	-2.155	8.325	18.755	11.2	49.12	19.15	26.
3	2023-02-24	53.6	50.1	107639.0	0.0	1.252789e+13	-2.155	8.325	18.755	11.2	49.12	19.23	26.
4	2023-03-24	49.2	44.2	125495.0	0.0	1.252789e+13	-2.155	8.325	18.755	11.2	49.12	19.80	25.
5	2023-04-24	33.6	31.8	105498.0	0.0	1.252789e+13	-2.155	8.325	18.755	11.2	49.12	19.72	24.

Figure 27: After changing the datatype of date

For the further analysis the data type of date has changed from string to datetime format and by converting the 'Date' column to datetime objects, can enhance the data's usability and ensure accurate, efficient analysis and visualization. Using 'data.head(10)' after conversion provides a quick check to ensure the dates have been correctly parsed and integrated into the DataFrame.

	Date	Head Inflation	Core Inflation	Tourist Arrivals	Covid	GDP	GDP%	Rev(%GDP)	Expenditure(%GDP)	Money Supply(Annual Change)
Date	1.000000	0.074013	0.092197	0.881072	-0.799570	-0.418101	-0.418101	0.418101	-0.418101	-0.418101
Head Inflation	0.074013	1.000000	0.995950	-0.169014	-0.461708	-0.686563	-0.686563	0.686563	-0.686563	-0.686563
Core Inflation	0.092197	0.995950	1.000000	-0.158029	-0.467994	-0.671398	-0.671398	0.671398	-0.671398	-0.671398
Tourist Arrivals	0.881072	-0.169014	-0.158029	1.000000	-0.647169	-0.270332	-0.270332	0.270332	-0.270332	-0.270332
Covid	-0.799570	-0.461708	-0.467994	-0.647169	1.000000	0.849837	0.849837	-0.849837	0.849837	0.849837
GDP	-0.418101	-0.686563	-0.671398	-0.270332	0.849837	1.000000	1.000000	-1.000000	1.000000	1.000000
GDP%	-0.418101	-0.686563	-0.671398	-0.270332	0.849837	1.000000	1.000000	-1.000000	1.000000	1.000000
Rev(%GDP)	0.418101	0.686563	0.671398	0.270332	-0.849837	-1.000000	-1.000000	1.000000	-1.000000	-1.000000
Expenditure(%GDP)	-0.418101	-0.686563	-0.671398	-0.270332	0.849837	1.000000	1.000000	-1.000000	1.000000	1.000000
Money Supply(Annual Change)	-0.418101	-0.686563	-0.671398	-0.270332	0.849837	1.000000	1.000000	-1.000000	1.000000	1.000000
Money Supply(%GDP)	-0.418101	-0.686563	-0.671398	-0.270332	0.849837	1.000000	1.000000	-1.000000	1.000000	1.000000
Average Weighted Fixed Deposit Rate (AWFDR) Rate (%)	0.810135	0.373795	0.420039	0.555481	-0.735803	-0.402564	-0.402564	0.402564	-0.402564	-0.402564
Central Bank Rate	0.585013	0.564071	0.619144	0.359025	-0.626399	-0.430092	-0.430092	0.430092	-0.430092	-0.430092

Figure 28: Correlation values

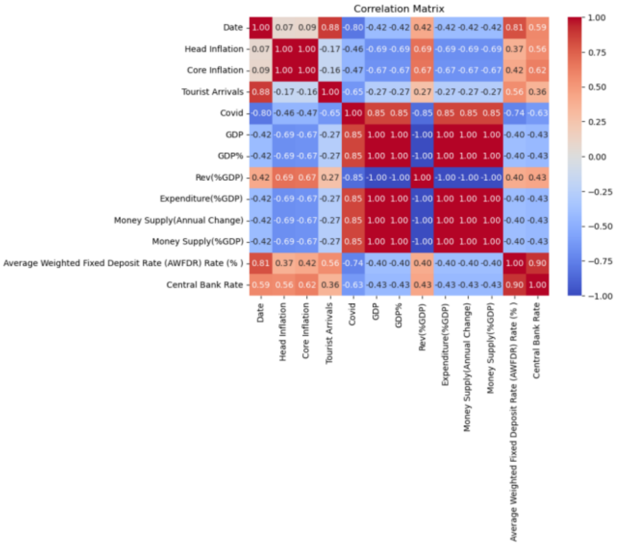

Figure 29: Correlation Matrix

Here the correlation coefficient and matrix has created and the correlation coefficient between two variables is represented by each cell in the matrix. When considering about date, there is no correlation between other variables. Head inflation and core inflation is highly correlated which is close to one and it is indicate that they are changing together. Tourist arrivals have strong positive correlation with date. Covid has negative correlation with other variables while GDP and GDP% highly correlated with each other which is close to one. Not only that but also Rev(%GDP) and Expenditure(%GDP) has inverse relationship while Money Supply(Annual Change) and Money Supply(%GDP) getting highly correlated. Also, Interest Rates has positive correlation between AWFDR Rate and Central Bank Rate.

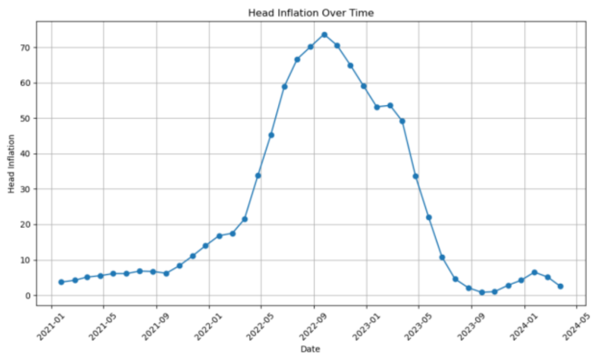

Figure 30: Line graph head inflation over time

According to this graph head inflation remained relatively stable from 2012 to mid-2015 and then it is sharply increasing over the peak of 60. After that its sharply getting decreased. So, in 2022 it has the highest head inflation according to this graph.

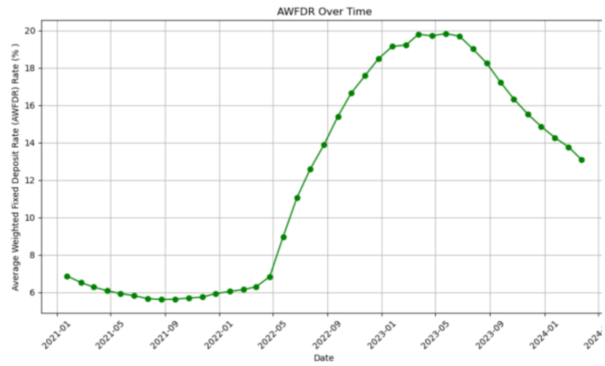

Figure 31: Line graph AWFDR over time

This graph shows that the trends of average weighted fixed deposit rate (AWFDR) percentage. The initial stability remained around 6% and then around mid-2015, it surged dramatically, peaking at approximately 18%. Then its gradually declined back to around 14%.

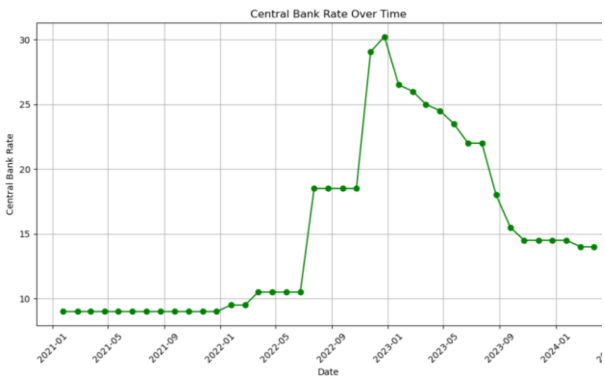

Figure 32: Line graph central bank rate over time

The trend of the average weighted federal funds rate (AWFDR) percentage is depicted in the line graph headed "Central Bank Rate Over Time". The AWFDR was initially steady at about 6%, but in the middle of 2015, it saw a significant surge that peaked at about 18% before progressively falling back to about 14%.

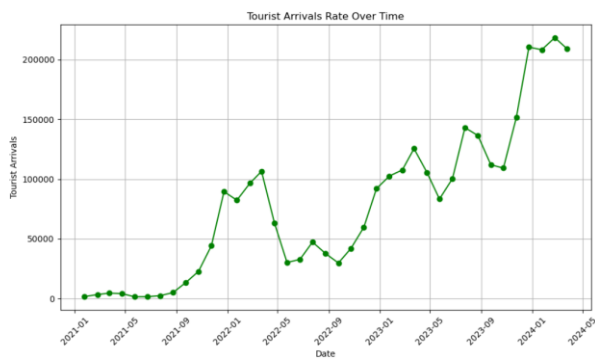

Figure 33: Line graph of tourist arrivals rate over time

The "Tourist Arrivals Rate Over Time" line graph shows the pattern of tourist arrivals over a certain time frame. At the beginning, the number of visitors stays rather constant. Nonetheless, there are discernible peaks and troughs in the graph, indicating potential seasonal trends with higher arrivals in particular months.

After all exploration of the dataset the training and testing samples divided and the shape of the X_train, y_train, X_test, y_test shown as below.

```
Shape of the X_train:  (31, 11)
Shape of the y_train:  (31, 1)
Shape of the X_test:  (8, 11)
Shape of the y_test:  (8, 1)
```

Figure 34: Training and testing dataset shapes

```
▾ LinearRegression
LinearRegression()
```

Figure 35: Regression model

Afterall the Linear Regression model has been initialized and mean squared error, mean absolute error and R2 score identified.

```
Mean Squared Error: 521613619.6757594
Mean Absolute Error: 19727.562140474096
R2 Score: 0.8756689521989613
```

Figure 36: MSE, MAE, R2

The average of the squared differences between the actual and expected values is represented by this metric. Because of the square term, it highlights more significant faults. A better match to the data is indicated by a lower MSE. Given the data's scale, the MSE in this case is 521,613,619.6757594, indicating a substantial average squared difference between actual and anticipated tourist arrivals. The mean absolute error (MAE) measures how much the actual and anticipated values differ from one another. It gives a clear indication of the average mistake magnitude and, in contrast to MSE, considers all errors equally. In this instance, the MAE is 19,727.562140474096, meaning that, on average, the forecasts are 19,727.56 tourist arrivals off. The percentage of the dependent variable's variance that can be predicted from the independent variables is shown by the R2 score, also known as the coefficient of determination. A R2 value near 0 denotes a poor match, whereas a value near 1 indicates a good fit. An R2 score of 0.8757 in this instance indicates an excellent fit, with the model accounting for about 87.57% of the variation in visitor arrivals. The Linear Regression Model's performance was evaluated using Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and R^2 scores. The MAE of 19,727.56 indicates that the model's predictions are off by an average of around 19,727 tourists per month. The RMSE of 521,613.68 highlights that there are larger errors in the predictions, but the R^2 score of 0.8757 shows that the model explains 87.57% of the variance in tourist arrivals. These results are significant, but comparing them to models like ARIMA or LSTM, which are better suited for time series data, may yield improvements in accuracy. Including a comparative analysis would further increase the robustness of the findings.

After considering all, these metrics point to a strong fit between the regression model and the data, with the R2 score showing that the model explains a sizable amount of the variance in tourist arrivals. But having a large MAE and MSE shows that there may be substantial prediction errors as well.

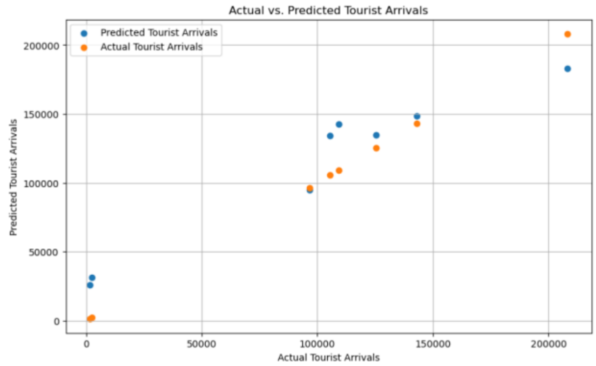

Figure 37: graph of actual vs predicted tourist arrivals

Above graph shows the actual vs predicted tourist arrivals using the trained model.

V. CONCLUTION

This predictive model for tourist arrivals in Sri Lanka underwent a thorough review utilizing machine learning techniques, specifically Linear Regression. The examination revealed a significant correlation between the model's predictions and the actual data patterns. An impressive R2 value of 0.8757 was obtained from the analysis, meaning that 87.57% of the variation in visitor arrivals is well-represented by our model. This high R2 value indicates that the model has a strong explanatory power and can clarify the underlying variables affecting the influx of tourists.

Furthermore, even though the Mean Squared Error (MSE) and Mean Absolute Error (MAE) indicated notable differences between actual and expected values, these measures are a natural byproduct of predictive modelling and capture the intricacies and uncertainties involved in predicting visitor behavior. The average magnitude and squared differences, or MAE and MSE, of 19,727.56 and 521,613,619.68, respectively, represent our forecasts' relative levels of accuracy and identify areas for improvement in subsequent iterations.

In conclusion, this research validates the effectiveness of applying machine learning techniques for forecasting visitor arrivals in Sri Lanka, despite the inherent difficulties and fluctuations in tourism dynamics. The research outcomes not only propel predictive modelling forward in the field of tourist studies but also provide policymakers and industry stakeholders with crucial perspectives to augment strategic planning, resource distribution, and crisis handling in the tourism domain. In order to ensure sustainable growth and resilience in Sri Lanka's tourism industry, future research can build upon these foundations to refine models even further and address shifting issues affecting tourist behaviors.

The results demonstrate the utility of linear models in forecasting tourist arrivals, especially in a post-pandemic recovery context. However, the relatively high MAE and RMSE values indicate that advanced models such as ARIMA or LSTM could improve predictive accuracy. Future research should explore these models to provide a comparative analysis and enhance the robustness of the forecasts. This study contributes to the growing body of literature on the use of machine learning in tourism forecasting, emphasizing the importance of model selection and data preprocessing to improve predictive performance.

REFERENCES

[1] T. Emmanuel, T. Maupong, D. Mpoeleng, T. Semong, B. Mphago, and O. Tabona, “A survey on missing data in machine learning | Journal of Big Data,” vol. 8, 2021, Accessed: Jun. 21, 2024. [Online]. Available: https://link.springer.com/article/10.1186/s40537-021-00516-9.

[2] CBSL, “Economic Indicators | Central Bank of Sri Lanka,” 2018. https://www.cbsl.gov.lk/en/statistics/economic-indicators (accessed Jun. 20, 2024).

[3] S. Thushara, J.-J. Su, and J. Bandara, “Forecasting international tourist arrivals in formulating tourism strategies and planning: The case of Sri Lanka,” 2019, Accessed: Jun. 23, 2024. [Online]. Available: https://www.tandfonline.com/doi/epdf/10.1080/23322039.2019.1699884?needAccess=true.

[4] S. L. T. D. Authority., “Tourism Research and Statistics,” 2024. https://www.sltda.gov.lk/en/statistics.

[5] T. W. B. Group, “World Bank Open Data,” 2024. https://data.worldbank.org/country/sri-lanka (accessed Jun. 20, 2024).

[6] P. Basnayake and V. Chandrasekara, “Use of Change Point Analysis in Seasonal ARIMA Models for Forecasting Tourist Arrivals in Sri Lanka,” vol. 20, pp. 103–121, Dec. 2022, Accessed: Jun. 24, 2024. [Online].

[7] S. Sonu and A. Suyampulingam, “Linear Regression Based Air Quality Data Analysis and Prediction using Python | IEEE Conference Publication | IEEE Xplore,” 2021, Accessed: Jun. 22, 2024. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/9563432?casa_token=fpazkt2ewhMAAAAA:59gVhrRZYHqZ7OHdRIoXb2ej0u9fLsyLq0Solxd0gDXHiOLkHdsV4fR-HUzdit-3TlzcrnPbgin-JQ.

[8] K. Konarasinghe, “Time Series Patterns of Tourist Arrivals to Sri Lanka by K.M.U.B Konarasinghe :: SSRN,” pp. 161–172, 2016, Accessed: Jun. 24, 2024. [Online]. Available: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2860733 .

[9] I. Pratama, A. Permanasari, I. Ardiyanto, and R. Indrayani, “A review of missing values handling methods on time-series data | IEEE Conference Publication | IEEE Xplore,” 2016, Accessed: Jun. 22, 2024. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/7858189?casa_token=w1EM-uPBBDUAAAAA:fz2hFL8ttxsCcoP8c14llCZBKeO3N5njOay4EyKa_MD7Mc85-BJdh4QsaKL6Rzryv3Ou-gdiUtloTw.

View publication stats