FACULTY OF COMPUTING AND INFORMATICS
DEPARTMENT OF INFORMATICS

Predicting International Tourist Arrivals in Namibia Using Machine Learning

BY
SELMA SHIVUTE
200517597

Thesis submitted in partial fulfilment of the requirement for the degree of Master of Data Science

METADATA

STUDY TITLE:	Predicting International Tourist Arrivals in Namibia Using Machine Learning Models
STUDENT NAME:	200517597
SUPERVISOR:	Dr Richard Maliwatu
DEPARTMENT:	Informatics
QUALIFICATION:	Master of Data Science
SPECIALISATION:	Data Science
MAIN KNOWLEDGE AREA:	Machine Learning
KEYWORDS:	**Namibia, Tourist forecast, SARIMA, Random Forest, Machine Learning**
TYPE OF RESEARCH:	Applied Research
METHODOLOGY:	Quantitative
STATUS:	Thesis
SITE:	Namibia University of Science and Technology (NUST)
DOCUMENT DATE:	August 2024

Declaration

The work in the full thesis, **Predicting International Tourist Arrivals in Namibia using Machine Learning Models**, is my original work, and I, Selma Shivute, hereby declare that I have not previously submitted it in its entirety to any other university or higher education institution to receive a degree. I attest that any instances in the work where information has been taken from outside sources have been noted.

Signature……Date…… 08 OCTOBER 2024 ……

Abstract

The lack of accurate and timely predictions for international tourist arrivals in Namibia remains an open problem, leading to inefficiencies in tourism planning and resource allocation. Traditional methods, primarily based on seasonal trends and historical data, dominate the forecasting landscape. Although traditional approaches are good at capturing seasonal patterns, there is always a lack of accounting for more dynamic and non-linear interactions between the predictive variables and more reliable results.

In recent years, machine learning models like Seasonal Autoregressive Integrated Moving Average (SARIMA), Random Forest, and Prophet have gained increasing support for their ability to handle complex, non-linear data and provide more accurate tourist arrival forecasts. SARIMA is particularly good at modelling seasonal time series data, Random Forest excels in capturing non-linear relationships, and Prophet is designed to handle time series data as well as irregular and missing data. However, attempts to implement these three models in predicting international tourist arrivals in the Namibian context have exposed limitations such as a constant prediction with Random Forest and the need for extensive tuning in SARIMA and Prophet, which may result in their prediction accuracy showing little improvements in Namibia's tourism sector.

This study aimed to develop and test the three models: SARIMA, Random Forest, and Prophet, to predict international tourist arrivals in Namibia more accurately. Accurate forecasts can improve decision-making within the tourism sector, infrastructure planning and resource allocation. The methodology used involved data preparation and data exploratory strategies to determine the relationship between exploratory variables and dependent variables. It also included training and validating these models on historical data obtained from the Ministry of Home Affairs, Immigration, Safety and Security. The models were hyper-tuned to overcome the limitations of accuracy by improving accurate predictions.

It is expected that these models will overcome the limitations of non-accuracy predictions of tourist arrivals. The effectiveness of these models' accuracy was evaluated using Root Mean Square Error and Mean Absolute Error and comparing their performance against each other to determine the preferred model for Namibia. The results indicated that SARIMA achieved the most accurate prediction, followed by Random Forest, and Prophet performed poorly in predicting international

visitor arrivals in Namibia. The two models are anticipated to contribute significantly to more efficient tourism management and planning in Namibia.

Dedication

This work is devoted to my precious children, Selma and Daniel. Your unwavering support, endless patience, and the joy you bring into my life have been my constant motivation throughout this journey. This work is a tribute to you and a promise to continue striving for a better future for our family. You always have a soft spot in my heart. I love you!!!!

Acknowledgements

The completion of this study was made possible by the grace and wisdom of Almighty God, who provided me with the strength to persevere. I extend my heartfelt thanks to Dr Richard Maliwatu and Dr Simon Chiutsi for their generosity and the time they dedicated to reviewing my study, despite their busy schedules. Your contributions have greatly enhanced the quality of my work.

My data collection through the Ministry of Home Affairs was a remarkable one. My sincere thanks go to Ms Leilah Nkandi from the Ministry of Home Affairs. Special gratitude goes to the Namibia University of Science and Technology, Faculty of Computing and Informatics lecturers who made the course Master of Data Science a success.

May God Bless You

Selma Shivute

Table of Contents

List of Figures

List of Tables

List of Abbreviations

AI	Artificial Intelligent
GDP	Gross Domestic Product
GIZ	Deutsche Gesellschaft für Internationale Zusammenarbeit
LSTM	Long Short-Term Memory
MAE	Mean Absolute Error
MHAISS	Ministry of Home Affairs, Immigration, Safety and Security
ML	Machine Learning
NCRST	National Commission on Research, Science and Technology
RF	Random Forest
RMSE	Root Mean Square Error
SAE LSTM	Stacked Autoencoder LSTM
SARIMA	Seasonal Auto-Regressive Integrated Moving Average.
SII	Spatial Interaction Index
WTTC	World Travel and Tourism Council

Chapter One: Introduction

1.1. Background of the study

Tourism is a major foreign exchange earner globally and it has since the early 2000s surpassed products such as petroleum, motor vehicles, textiles, and telecommunications equipment. This industry is also known for being labour-intensive, providing jobs for around 100 million people globally, equivalent to 8.3 per cent of world employment (Botha & Saayman, 2022). In 2005, the tourism sector accounted for approximately 10% of the world's Gross Domestic Product (GDP), according to the World Travel and Tourism Council (WTTC). The significance of tourism to economies lies in its ability to generate revenue that can finance infrastructure and development projects and promote cross-cultural understanding. According to the most recent Economic Impact Report released by the World Travel and Tourism Council (WTTC, 2020), the travel and tourism industry supported 330 million jobs globally and generated 10.4% of the world's GDP. But in 2021 and 2022, the COVID-19 pandemic had a major effect on the sector, resulting in a 49.1% decline in the GDP of travel and tourism, which caused many job losses and economic downturns. In Namibia, the tourism sector is of significant importance. In 2018, the industry contributed approximately N$5.2 billion, representing 14.7% of the GDP, and supported 44,700 permanent employment opportunities, encompassing 2,900 positions within community conservation initiatives (GIZ, 2022).

Between 2014 and 2019, Namibia experienced an annual influx of over one million tourists. To anticipate and cater to this demand, tourist sites and suppliers in the region depend on four seasonal forecasts. Th se present study utilised a tourism demand model to predict forthcoming tourist arrivals by analysing historical data. The tourism industry contributes significantly to Namibia's GDP, accounting for 15.4% of total employment. Despite the significant contributions outlined above, Woyo and Amadhila (2018) observed that many African nations fail to fully exploit their tourism potential. Hence, enhancing resource management becomes imperative, potentially catalysing economic expansion. By prioritising the tourism sector, African countries such as Namibia stand to achieve enhanced development and economic growth potential.

Literature is replete with factors that affect the number of tourists visiting a destination, such as income levels, prices of other places, transportation costs, weather conditions, factors related to the tourism industry itself, health concerns, political stability, and security issues such as terrorism and crime.

Despite relying on mineral resources, agriculture, and fishing for growth and development, both before and after gaining independence in 1990, Namibia continues to face high unemployment rates. In recent years, the tourism industry has emerged as a promising sector for job creation and economic growth. Recognising this, the Namibian government has designated tourism as a priority sector in both *Vision 2030, the Harambe Prosperity Plan and the National Development Plans* (National Planning Commission, 2023). Vision 2030 is a long-term national development framework that reflects the aspirations and goals of the Namibian people. At the heart of this vision is the aim to improve the standard of living and quality of life for Namibians, to achieve a standard of living on par with developed nations. Developing the tourism sector is seen as a critical component of the country's Broad-Based Economic Empowerment initiative.

The current study predicted the number of international visitors to Namibia using machine learning techniques based on the given discussion. To give precise and trustworthy projections of foreign visitor visits in Namibia, this research aimed to use data-driven methodologies and predictive modelling tools. Furthermore, by aiding in the planning and administration of tourist-related activities, the study hoped to support Namibia's tourism industry's sustainable growth.

Machine Learning is a powerful tool for making predictions in a wide range of applications. Essentially, machine learning algorithms use statistical techniques to automatically learn patterns and relationships in data, which can then be used to make predictions about new data. One of the most common applications of machine learning for prediction is in the field of supervised learning, where the goal is to train a model to predict an output value based on a set of input features. For example, a machine learning model could be trained to predict whether an email is spam or not based on the email's text and metadata (Sarker, 2022).

Another application of machine learning for prediction is in the field of unsupervised learning,

where the goal is to find patterns or structures in data without having explicit output labels. This can be useful for tasks such as clustering similar items together or identifying anomalous data points. Machine learning is also increasingly being used in combination with other technologies such as big data and the Internet of Things (IoT), to make predictions about complex systems in real-time. For example, machine learning algorithms can be used to predict equipment failure in industrial settings or to optimise energy usage in smart buildings. Overall, the use of machine learning for prediction is a rapidly growing area with numerous applications across many industries, including healthcare, finance, transportation, tourism and more (Takyar, 2019).

Although machine learning models are widely utilised as forecasting tools, there is a scarcity of research studies that employ these models for predicting tourism demand. Previous research has solely employed multilayer perceptron neural network models (Kayral et al., 2023). The present study investigate dthe use of different machine learning models to forecast traveller demand. In particular, the time series forecast and the supervised machine learning were taken into consideration: SARIMA, Prophet and Random Forest.

1.1.1. Tourism

Tourism refers to the concept where people temporarily remain or travel outside of their usual residence for reasons other than employment or work-related activities that are paid for in the place they are visiting (Christianingrum et al., 2022). As the importance of tourism as an economic activity has grown, it has received increased attention from government entities, as well as organisations in both the public and private sectors and scholars. Tourism is a complex activity that encompasses various economic, social, cultural, and environmental factors. Despite being commonly referred to as an industry, it lacks a conventional production function and measurable output like agriculture. Moreover, there is no standard structure that can universally characterise the tourism industry across different countries. As tourism continues to evolve, it demands a multi-faceted approach that considers its diverse aspects and challenges. Tourist attractions vary between countries, as seen in the cases of France, Italy, and Russia, where restaurants and shopping centres play a significant role for tourists in the former two, but not in the latter (Uula et al., 2024). Likewise, the core components of the tourism industry, such as accommodation and transportation can also differ. In the UK, bed and breakfasts in private residences are a popular option, but such

facilities are uncommon in Thailand. Transportation preferences also vary, with most tourists in Western Europe and the USA using cars or buses due to developed road networks and high car ownership, while air travel is the preferred option in India and Indonesia (Žunić et al., 2023).

1.1.2. Machine Learning Models

Machine Learning (ML) plays a crucial role in predicting tourist arrivals by leveraging its ability to analyse vast amounts of data and identify complex patterns and relationships. Traditional forecasting methods often struggle to capture the intricate dynamics and nonlinearities inherent in tourism data. However, machine learning algorithms excel in handling such complexities, allowing for more accurate and reliable predictions. By training models on historical data, encompassing factors such as economic indicators, weather patterns, travel advisories, and demographic information, machine learning algorithms can uncover hidden patterns and generate forecasts that take into account multiple variables simultaneously. This empowers stakeholders in the tourism industry to make well-informed decisions concerning resource allocation, marketing strategies, infrastructure planning, and risk management. Moreover, as machine learning models continuously learn and adapt to new data, they have the potential to enhance their prediction capabilities over time, ensuring that the forecasts remain relevant and up-to-date. Overall, machine learning's relevance in predicting tourist arrival lies in its capacity to leverage data-driven insights, improve forecasting accuracy, and support evidence-based decision-making in the dynamic and ever-evolving tourism industry (Şeker, 2023).

1.2. Problem statement

Namibia's tourism sector contributes significantly to the country's economy, and foreign visitors are essential to the sector's continued existence. Precise estimation of foreign visitor arrivals is essential for efficient tourism administration and planning. Predicting international tourist arrivals can assist tourism stakeholders to develop effective marketing strategies, manage resources efficiently, and enhance the overall visitor experience. However, accurately predicting these arrivals can be a challenging task due to various factors that influence travel decisions such as economic conditions, global events, and changes in travel regulations.

Moreover, traditional statistical forecasting methods may not capture the complex interactions among various factors that influence tourism demand. Machine learning models have shown promising results in predicting tourism demand in other countries, but their effectiveness in predicting international tourist arrivals in Namibia is not well understood. In this study, we aim to develop machine learning models to forecast the number of foreign visitors who will arrive in Namibia. By analysing historical tourist data, and other relevant factors, we will build and evaluate several models to determine the most accurate and reliable method of forecasting tourist arrivals. The results of this study can be used to inform decision-making in the Namibian tourism industry and contribute to its long-term sustainability.

According to Bravo et al. (2023), Machine Learning offers significant advantages over traditional methods for predicting tourist arrivals. It provides higher accuracy by capturing complex, non-linear relationships in data while handling huge datasets from multiple sources. ML models can analyse data in real-time, automatically adapt to changing trends, and incorporate a wide range of factors, including economic indicators and traveller sentiment, which traditional approaches might overlook. Moreover, ML improves predictive power, automates the forecasting process, and reduces manual intervention, thereby making it a more efficient and adaptable solution for tourism forecasting in the Namibia context.

1.3. Objectives of the study

The main objective of this research is to explore historical data for predicting monthly tourist arrivals in Namibia.

To achieve the above-stated objective, the study met the following sub-objective by combining data analysis and literature review:

Intermediate-objectives:

a. To establish the present pattern of tourist arrivals in Namibia;
b. To identify seasonal patterns that influence the number of international tourist arrivals in

Namibia based on their country of origin;

c. To assess the performance of machine learning methods in predicting tourist arrivals in Namibia;
d. To improve prediction accuracy through enhanced pre-processing techniques and feature engineering; and
e. To investigate and propose practical strategies for the Namibian tourism industry to effectively utilise emerging technologies, such as AI and ML.

1.4. Research questions

The following research questions guided this study:

Main question: Can we explore historical data for predicting tourist arrivals in Namibia?

Sub questions

a. What is the present pattern of tourist arrivals in Namibia?
b. How do seasonal patterns influence the number of international tourist arrivals in Namibia based on their country of origin?
c. How do commonly used machine learning methods for predicting international tourist arrivals perform on Namibian data?
d. How can the accuracy of machine learning models to forecast the number of foreign visitors who will arrive in Namibia be improved?
e. How can the Namibian tourism industry leverage emerging technologies like artificial intelligence and Machine Learning to enhance tourism forecasting and decision-making?

1.5. Significance of study

The study on predicting international tourist arrivals in Namibia using machine learning models holds great significance for the country's tourism industry. Namibia heavily relies on tourism as a major contributor to its economy, making accurate predictions of tourist arrivals crucial for effective planning and resource allocation. By harnessing machine learning techniques, this research provides valuable insights into the industry's performance, allowing stakeholders to make informed decisions, allocate resources efficiently, and develop appropriate tourism strategies.

Moreover, the study aids in targeted marketing and promotions by identifying patterns and trends in tourist arrivals, enabling marketers to tailor campaigns to specific markets and align promotional activities with peak travel seasons. Furthermore, economists and policymakers can use this study to estimate the economic impact and assess the benefits and plans for sustainable tourism development. Additionally, machine learning models can help identify potential risks and vulnerabilities, enabling stakeholders to develop contingency plans and mitigate negative impacts. Ultimately, this study serves as a decision-support tool for policymakers, government agencies, and industry stakeholders, providing them with quantitative insights to make data-driven decisions related to tourism infrastructure, investment, and policy formulation.

1.6 Limitation and delimitation of the study

The study was limited to foreign tourists and did not make in-depth predictions about domestic tourists. Not all the Machine Learning models were trained, thus the study only focused on SARIMA, Random Forest and Prophet among others. The accuracy of the model depends on the quality and amount of the previous data it was trained on. The prediction is limited to the past historical arrival data and does not analyse the Namibia tourist internet data. The prediction made in this study may not generalize for other years before 2014 and is not covered by this study due to the unique characteristics of the tourism industry.

1.7 Assumptions

The assumption was that MHAISS data is complete and accurate. The study did not consider proprietary data or real-time data sources that may have additional insights.

1.8. Organisation of the thesis

The rest of the dissertation is organised as follows: Chapter Two discusses previous research on tourist arrival prediction, provides an overview of tourism in Namibia, discusses the literature and models and outlines the case study of the successful implementation of the ML model in some countries. It has also highlighted the benefits and barriers of using machine learning for prediction. Chapter Three outlines the theoretical foundation underpinning the study, and Chapter Four

outlines the methodology employed in this study, followed by the research model which covers the research philosophy, research design, and statistical tests used to analyse the data. The statistical theory behind the model, interpretation of the results, model implementation, model testing results and model evaluation and accuracy results are presented in Chapter Five. Finally, the summary of findings, conclusion and recommendations for future research are given in Chapter Six.

Chapter Two: Literature Review

The literature review covers the overview of the Tourism industry in general and, the factors that influence tourists' choice of destination. This chapter discusses previous work on tourist arrival prediction. This chapter provides a broad overview of the tourism business, examines the factors that affect travellers' choice of location, and addresses the many methods employed to forecast visitor arrivals.

2.1. Overview of tourism industry

The majority of research on tourism demand focuses on two key topics: comprehending the factors that affect demand and efficiently projecting demand. Business managers and policymakers can benefit from effective forecasting since it helps with resource allocation and investment decision-making. However, research on the elements influencing tourism demand offers a better understanding of the variables influencing travellers' decisions. On the other hand, not much is still understood about these variables. In today's competitive world, businesses and policymakers need to understand the behaviour of tourists. Enhancing these elements helps a nation draw in more travellers, which raises the tourism sector's economic contribution to the country. The African continent has the chance to increase the amount that tourism contributes to the GDP of its member nations (Chipumuro & Chikobvu, 2022).

2.1.1. The factors determining tourism demand

In line with microeconomic theory, individuals seek to maximise their utility by choosing travel destinations that align with their preferences. Economic factors such as income, prices of goods and services, substitutes, and social influences like family and friends are known to influence consumer demand (Siwek, 2023). Similarly, tourists consider these factors when deciding which destination country to visit. The demand for tourism is analysed using several key variables, such as income, pricing, the cost of transportation, exchange rates, marketing expenses, weather, historical and cultural characteristics, population size, trends, and supply-side elements including the effectiveness of tourism services and infrastructure (Faridi, 2024). On the other hand, circumstances like political unrest, terrorism, criminal activity, natural disasters, and health hazards can have a detrimental impact on the demand for tourism (Khan et al., 2020).

Income

The income of tourists plays a crucial role in their travel decisions, and it is a commonly used variable in tourism studies. Income has long been recognised by researchers as a key factor influencing traveller demand (Siwek, 2023; Khan et al., 2020). Despite this consensus, there are variations in how researchers express income within their studies. International tourism necessitates financial resources as travellers must allocate a considerable portion of their budget to transportation and accommodation. However, not everyone is interested in exploring foreign countries, and recreational travel does not always require significant expenditures. Therefore, it would be reasonable to consider discretionary income, which represents the remaining amount after necessary expenses are covered, as a factor influencing travellers' demand (Faridi, 2024).

Price

When investigating potential factors of tourism demand, pricing emerges as another crucial factor to take into account after income. Relative costs also referred to as tourism prices, are challenging to measure accurately due to the variety of products and services that travellers typically buy. The two main components of tourism prices are the cost of living at the destination and the cost of transportation. According to Khan et al. (2020), the costs incurred for local travel are included in the entire expense of living at the tourist destination.

Exchange rate

The demand for tourism is heavily influenced by the currency rate, which is a critical factor in determining travel selections. Travellers typically consider fluctuations in currency rates more carefully than shifts in comparable costs when choosing their locations. Travellers usually take annual vacations, mainly in the summer, and deduct travel costs from their annual spending plan. Their purchasing decisions are influenced by both relative exchange rates and relative travel service pricing, which may cause them to choose between travelling overseas and staying home instead of going somewhere cheaper. When making travel plans, tourists may more accurately determine the exchange rates when compared to the prices at their target destination because they are regularly released in a variety of media outlets. On the other hand, travellers frequently rely on their memory of the destination's price range from prior visits and are unaware of price changes beforehand. When there is a low exchange rate, the tourist's home currency is stronger than the

currency of the destination. This can lower the cost of travel for visitors to nations with depreciating currencies (Chipumuro & Chikobvu, 2022).

The fluctuation of exchange rates has a significant impact on the number of tourists visiting a particular country, with varying effects depending on whether the change is favourable or unfavourable. According to Panasiuk (2023), there are several positive effects of exchange rate fluctuations, such as tourists spending more money on items they would have bought otherwise, making more purchases, and drawing in new visitors and cross-border buyers. On the other hand, the study by Montes-Rojas (2020) emphasised the opposite impacts brought about by unfavourable fluctuations in exchange rates. These effects involve tourists traveling less abroad, altering their destination choice, reducing spending on goods and services at the destination, shortening their stay duration, postponing trips, using different modes of transportation, and reducing business travel expenditures. Empirical research has used several definitions of the exchange rate variable, much like they did with income and prices.

Population size

Another factor influencing the demand for tourism is the size of the country of origin's population. Rather than concentrating only on the effects of the overall population, current research has investigated the impact of various population segments on the demand for tourism. The consumption habits of people in different age groups change dramatically. Population ageing refers to the progressive increase over the past ten years in industrialized countries in the number of older people at the expense of younger age groups. The percentage of people above retirement age can be used to quantify this, as it has been rising since the post-World War II baby boom due to longer life expectancies and lower fertility rates. The growing trend of population ageing has led to a significant market of retirees with more time and financial resources to spend on travel, thereby boosting the demand for tourism. Almeida (2020) refers to this segment as "third-age tourism" and notes the emergence of specialised companies catering to the travel needs of seniors.

Country attractiveness

Preferences vary among individuals and are influenced by various socioeconomic factors such as age, gender, marital status, and education level. These factors contribute to the diversity of tastes observed in the population, which can also change over time due to factors like improved living standards, advertising, and industry innovations. Consequently, measuring a single variable to represent tastes becomes difficult. Additionally, capturing changes in destination preferences or the popularity of specific destinations over time can be achieved by incorporating a time trend into the analysis. This enables the examination of temporal patterns and trends in travellers' choices and preferences (Chipumuro & Chikobvu, 2022).

Seasonality

The demand for tourism can be significantly impacted by the scheduling of events within a year, such as a certain season or a school holiday. Twelve seasonal dummy variables are usually included in analytical models with monthly data, whereas four seasonal dummy variables are usually included in models with quarterly data (Chen et al., 2017).

Culture

In the past, people frequently thought of tourism and culture as two different things. Cultural resources were mostly viewed as components of the regional or national legacy, serving to preserve cultural identities and educate the local populace. On the other hand, tourism was seen as a distinct pastime that was unrelated to daily living or the distinctive cultural features of the area. But as the importance of local traditions in drawing foreign travellers and setting cities apart became increasingly clear toward the latter of the century, this view progressively changed. This expanding relationship between tourism and culture was influenced by several causes (Caldwell, 2023).

Country stability

A significant event often incorporated into demand models as a dummy variable is the occurrence of a terrorist attack within a specific year. Sadly, because the tourism sector can receive international media attention, it frequently draws the interest of foreign terrorist organizations. Transportation networks, governmental buildings, populated areas, and military bases are examples of target areas (Chipumuro & Chikobvu, 2022). Terrorist incidents severely affect traveller's ability to make decisions and seriously reduce demand for travel worldwide. Increased security procedures cause delays in transportation, making tourists fearful for their safety and reluctant to visit. Nevertheless, this fear of travelling is usually fleeting. According to Tovmasyan (2023), the short-term effects of a terrorist act on tourism are the most noticeable and have little bearing over the long run. As an example, the aftermath of the September 11, 2001 attacks is noteworthy and has been the subject of much analysis in the last ten years. Due to its extraordinary scope, which stunned everyone, its significant impact on travel demand was not limited to the United States but rather had an impact on the entire world.

2.2 Tourist arrival forecasting/Prediction methods

Empirical studies on tourism demand frequently utilize various models, particularly when it comes to forecasting. These models aim to capture the complex dynamics of tourism demand and provide insights into future trends. ARIMA (autoregressive integrated moving average) models are some commonly used time series models that analyse historical data to detect patterns and make predictions (Ke, 2024). Another approach is the econometric regression model, which assesses the relationship between tourism demand and key determinants, such as income, prices, and marketing expenditure (Tovmasyan, 2023). Additionally, machine learning models, including decision trees, random forests, and neural networks, have grown in status recently due to their capability to capture non-linear relationships and handle large amounts of data. These models can provide accurate forecasts by learning from historical patterns and adjusting to new information. In the end, the selection of a model depends upon the particular study goals, the accessibility/availability of data, and the complexity of the tourism demand issue being studied.

2.2.1 Econometric models

Econometric models are frequently used by researchers to analyse tourism demand and establish

the causal relationships between different elements. Studies on tourism demand typically make use of panel data models, a system of equation models, and single equation estimations (Tovmasyan, 2021). Single equation models are commonly used in econometric models, according to Montes-Rojas, (2020) study. The autoregressive distributed lag model (ADLM), the error correction model (ECM), and the time-varying parameter (TVP) model are a few examples of these models.

2.2.2 Error correctional model

Researchers have developed a keen interest in the Error Correction Model (ECM). It is used to explain the dependent variable's current values based on historical values, independent variables, errors, and current values of the dependent variable (Tovmasyan, 2021). What sets the ECM apart from other econometric approaches like the Vector Autoregressive (VAR) Model is its ability to capture both short and long-term dynamics in cases of cointegration (Gligorić et al., 2019; Jammazi & Aloui, 2019). In some studies, one can see the performance of the Arima model compared to ECM, where Arima performed very well on the prediction of tourism in the long run and ECM performed well in the short run.

2.2.3 Autoregressive Distributed Lag Model

The goal of the Autoregressive Distributed Lag Model (ARDL) is to represent both the short- and long-term correlations between one or more independent variables (predictors) and a dependent variable (tourist arrivals). The main advantage of the ARDL model is its ability to handle both stationary and non-stationary time series data in the same framework, making it suitable for many real-world applications. The ADLM is recognised as a dynamic model as it includes both lagged and current explanatory variables, thereby effectively accounting for the temporal aspect of tourists' decision-making processes (Qureshi & Destek, 2019).

2.2.4 Time-varying Parameter Model

The Time-varying Parameter (TVP) model has found widespread application in various tourism demand studies, leading to relatively accurate modelling and forecasting results (Ramirez-Correa et al., 2017). It is especially useful when working with data that has structural changes or when constant coefficients are thought to be excessively limiting (Juan et al., 2024). It has been noted

that the main benefit of the TVP model lies in its dynamic nature, allowing coefficients to change over time, a departure from traditional econometric models where coefficients remain fixed (Akanbi & Madu, 2018). As a result, recent data hold a more substantial influence on TVP estimation compared to data from the more distant past.

2.2.5 Vector Autoregressive Model

According to Fu et al.'s (2015) study, Vector Autoregressive Model (VAR) models are particularly useful when dealing with multiple interrelated time series variables, as they allow us to analyse the dynamic interactions between these variables over time. In the context of predicting tourist arrivals, a VAR model would involve considering not only the historical data of tourist arrivals but also other relevant variables that may influence tourism, such as exchange rates, GDP, seasonality factors, marketing efforts, and other economic indicators. When employing a VAR model, every parameter is considered an endogenous parameter and the model regresses present values against their respective past values, including all previously used variables. This form of the VAR model is commonly known as an unrestricted VAR model (Hassan & Bornmann, 2016).

The model comes with numerous advantages and a few disadvantages. One of its strengths is its relative ease in forecasting since it eliminates the need to forecast each variable individually (Fu et al., 2015). Moreover, the VAR model requires minimal theoretical knowledge to be implemented. However, this simplicity can also be a drawback as it leads to more complex interpretations of the results. Another limitation of the VAR model is the prerequisite for data stationarity before estimation. To achieve stationarity, data often need to be differenced, but this process results in the loss of essential information about long-run relationships between variables.

3. Machine Learning Model adoption

3.1 Cases of successful Machine Learning model adoption in various countries

Countries that have implemented the Machine Learning Model to predict tourists' arrival among them are Singapore, Thailand, Japan, Australia, Indonesia and the United Kingdom.

Singapore

The Singapore Tourism Board has been actively using data analytics and machine learning to predict tourism demand and arrivals. They use various data sources, including historical travel data and social media trends, to forecast visitor numbers (Wai et al., 2021). Tourism is a major industry in Singapore, and the number of tourists is growing fast. Accurate forecasting of tourism demand is essential for planning and decision-making. However, accurate forecasting of tourism flows has been a challenge due to its stochastic and nonlinear nature (Bouhaddour et al., 2023).

The study done by Bouhaddour et al. (2023) used two AI methods, SARIMA and PROPHET, to forecast the number of air tourists in Vietnam. SARIMA and PROPHET are two models that can be used for time series analysis. PROPHET is an open-source software package that enables time series forecasting. It can be used to predict seasonal and non-linear trends. Time series data can be forecasted using the statistical model SARIMA. The study found that both SARIMA and PROPHET are effective in forecasting tourism demand. PROPHET is more robust to missing data and can handle mixed data without manual effort. However, the study concluded that SARIMA is more accurate in predicting seasonal trends (Bouhaddour et al., 2023).

Thailand

The Tourism Authority of Thailand (TAT) has also employed machine learning and data analytics to analyse past tourist patterns and predict future arrivals. They use these predictions to make more informed decisions about tourism promotions and infrastructure development (Trang, 2019). The study was done by Yotsawat et al. (2016), who extensively employed data clustering to analyse homogeneous tourist groups and understand their characteristics. The study focused on domestic travellers visiting Phranakhon Si Ayutthaya province in Thailand and proposed the development of tourist segment models. By using demographic and behavioural variables, the authors employed a two-step cluster analysis, which is considered the most robust method. The results revealed four well-defined subject groups: "senior tourists with organisation trips," "elderly tourists with family," "employees," and "tourism lovers."

To predict new tourist behaviour among Thai domestic tourists, tourism organisations can use the distance between cluster centroids or create a Bayes Network classifier based on demographic and

behavioural characteristics. This prediction process utilises the predictor variables to assign tourists to specific segments. By leveraging these predictions, service providers can better tailor their products and services to cater to the specific needs of their customers (Yotsawat t al., 2016).

Japan

Japan has been known for implementing advanced technologies in various industries, including tourism. Some tourism boards and organisations in Japan have explored the use of machine learning algorithms to forecast tourist arrivals and optimize marketing strategies (Derdouri & Osaragi, 2021).

The work of Derdouri and Osaragi (2021) introduced a new method for classifying tourists and residents, departing from previous heuristic and probabilistic approaches. The paper employed machine learning (ML) algorithms while considering various parameters like weather, mobility, and photo content to explain the differences between the two groups. This methodology was applied in the analysis of geotagged photos taken between July 2008 and December 2019 in Tokyo's 23 special wards by Flickr members. According to the findings, five supervised-learning algorithms - gradient boosting machine (GBM), generalised linear model (GLM), distributed random forest (DRF), deep learning (DL), and highly randomized trees (XRT) - perform worse than the stacked ensemble (SE) models. Notably, temporal entropy (TEN), movement during the workday, and frequent visits to crowded places and amusement parks all affect the classification of users (Derdouri & Osaragi, 2021).

Australia

The Australian Tourism Data Warehouse (ATDW) has been working on predictive analytics and machine learning models to gain insights into tourist behaviour and predict arrivals during different seasons and events (Ma et al., 2016).

Indonesia

Various tourist destinations and cities in Indonesia have experimented with machine learning models to predict visitor numbers and better manage resources and facilities accordingly. The study of prediction model accuracy in terms of RMSE and MAE revealed interesting findings.

Among the 36 models studied, those utilising multisource Internet data consistently outperform the models relying on single or no Internet data predictors. Furthermore, when comparing training results between data compositions and prediction models, data composition three consistently yields the best results in terms of RMSE and MAE. Overall, the Random Forest model, which incorporates all predictors and is trained using data composition came out with the highest prediction accuracy (Andariesta & Wasesa, 2022)

United Kingdom

Tourism organisations and local authorities in the UK have looked into machine learning applications to predict tourism demand and create tailored marketing campaigns (Yu & Chen, 2022). The study addressed the challenges posed by excessive human flow, and accurate prediction of tourist flow is essential. This involves forecasting tourism volumes in advance, improving the carrying capacity of tourist destinations, and implementing effective preventive measures to ensure a balanced distribution of visitors and resource utilisation. Consequently, the tourism industry has shifted its focus to enhancing forecast accuracy, which holds value for both the market and academic perspectives and has garnered attention from various stakeholders (Yu & Chen, 2022).

According to Bi et al. (2019), traditionally, statistical methods, such as linear regression, were utilised for tourism demand forecasting and achieved satisfactory results. However, these methods disregarded the cyclical variations in tourism demand, leading to limitations in prediction accuracy. Attempts to overcome this limitation involved incorporating techniques like moving average and exponential smoothing, but these approaches remained essentially linear models, revealing inherent constraints. In recent years, the advancement of neural network research has given rise to a new approach for tourism demand forecasting. The neural network-based model, as a nonlinear method, not only captures cyclical patterns in tourism demand but also adapts to time-varying travel demand, resulting in robust and reliable forecast outcomes (Yu & Chen, 2022).

In consideration of the appropriateness of recurrent networks for time series data modelling, the studies by Yu and Chen (2022) and Bi et al. (2020) introduced a novel approach of deeply stacking Long Short-Term Memory (LSTM) based autoencoders. The proposed method involves using a

layered greedy pretraining technique to replace the deep network. The resulting SAE-LSTM prediction model combines this pretraining stage with a fine-tuning network, and it utilises a suggested random weight initialisation method. The primary objective is to enhance the performance of the deep learning model, resulting in superior prediction outcomes.

4. Challenges of implementing Machine Learning to predict tourist demand

The traditional models that are now in use have trouble predicting travel demand since search intensity indices have been overused as indicators of tourism demand. Despite encountering challenges reported by practitioners when incorporating SII (Spatial Interaction Index) data into conventional prediction models, these data remain indispensable for achieving accurate tourism demand prediction. Two primary practical obstacles often arise: (i) concerns regarding feature selection. As noted by Rosselló-Nadal and He (2020), numerous factors, including exchange rates, travel expenses, tourism prices, and diverse SII data, are deemed potential predictors of tourist demand. The expanding array of these potentially influential predictors leads to a reduction in available training datasets within the feature space. Consequently, there is an insufficient data pool to effectively construct precise models (Rosselló-Nadal & He 2020). (ii) The second issue revolves around the choice of lag order. Despite the widespread use of SII data in various tourism demand prediction methods, limited attention has been given to elucidating the significant relationship within time series data. Only a few existing studies have delved into the unpredictability hypothesis, employing techniques such as the Granger causality test or Pearson correlation coefficients. The study was to explore the hypothesis of lesser predictability by evaluating the degree of correlation between a factor's lagged values and the number of visitors arrivals (Akhmet et al., 2021).

Chapter Three: Theoretical Foundation

Chapter three outlines the theoretical foundation, thereby setting the context for the study by discussing relevant concepts. The chapter outlines the theoretical framework that guided the study. The chapter also explains the relationships among the variables and key concepts explored in the study. Later in the chapter, the research approaches are outlined to achieve the research objectives. The data collection method used, pre-processing of data, model selection, training, and evaluation of the proposed ML model will be discussed.

3.1. Tourism and its significance for economies and businesses

The tourism industry is a significant driver of the Namibian economy, drawing in more than 2 million international tourists in 2018 (Ministry of Environment, Forestry and Tourism, 2018). With the rapid expansion of the social economy, there has been a sharp uptick in tourist numbers. The accurate prediction of tourism trends has become especially critical due to the substantial impact of this industry on the economy (Lionetti et al., 2021). Planning processes heavily rely on precise forecasts to minimise uncertainties in decision-making. However, accurately foreseeing patterns in tourism has remained a challenge for decades due to its unpredictable and complex nature. Artificial Intelligence (AI) and Machine Learning (ML) methods have emerged as a promising solution for achieving enhanced forecasting accuracy (He et al., 2021).

3.2. Tourism demand theories:

Based on the microeconomic theory, it is widely recognised that individuals seek to optimise their satisfaction, thus leading them to opt for travel destinations that align with this principle. Economic theory posits that various factors play into consumer preferences, including income, prices of goods and services, costs of alternatives, individual tastes, and social connections (Juan et al., 2024). Similarly, when tourists decide on a foreign country to visit, they often take into account similar determinants. Explanatory variables that are often employed include income, price points, costs of transportation, exchange rates, marketing expenditures, climate, historical and cultural significance, population size, continuous trends, categorical variables, historical data, and supply-

side components like the degree of infrastructure at a destination and the effectiveness of tourism services (Hassan & Bornmann,2016). Negative effects on traveller demand include things like crime, terrorism, political unrest, natural disasters, and health hazards. The majority of empirical investigations addressing tourism demand centre around dependent variables such as the influx and departure of tourists, financial outlays or earnings from tourism, and, in limited cases, the duration of a traveller's sojourn in a particular destination.

3.3. Data-driven tourism

A study conducted by Bi et al. (2019) explored how data-driven approaches, including machine learning, have significantly gained attraction in predicting and understanding tourist behaviour over the past few years. These approaches leverage the power of data collection, analysis, and pattern recognition to provide valuable insights into various aspects of tourism, ranging from destination selection and travel preferences to on-site behaviour and post-trip activities

These methodologies harness extensive data sources such as online searches, social media interactions, and historical travel records to deliver tailored recommendations to individuals, enhancing destination selection and personalising travel experiences. Moreover, these approaches extend to demand prediction, aiding airlines, hotels, and tour operators in optimizing pricing and marketing strategies by analysing historical booking data, seasonal trends, and economic indicators (Bravo et al., 2023). Beyond personalised suggestions, these techniques facilitate destination management through the analysis of tourist movement patterns, enabling local authorities to enhance infrastructure, manage crowds, and refine marketing strategies. They also play a role in sentiment analysis of online content, ensuring that businesses align their offerings with customer preferences and improving post-trip analysis, economic impact assessment, and security measures (Bravo et al., 2023; Koushik et al., 2020). However, a study by Koushik et al. (2020) noted that it is imperative to prioritize ethical considerations and data privacy while implementing these approaches within the tourism industry.

3.4. Machine Learning in tourism

With clear advantages over traditional approaches, predictive modelling utilising machine learning has provided a transformative way to comprehend and forecast tourist behaviour. Machine learning can handle complicated data sources like online searches and social media interactions far more easily than traditional methodologies, uncovering hidden patterns and delivering accurate predictions (Núñez et al., 2024). Its real-time adaptability stands out, as it continuously learns from new data, allowing the industry to respond swiftly to evolving trends and circumstances. Additionally, machine learning's ability to capture intricate nonlinear relationships enhances its predictive accuracy, providing a nuanced understanding of the multifaceted factors influencing tourist behaviour. This shift to machine learning-driven predictive modelling enhances the precision of predictions, personalizes recommendations, and fosters adaptable strategies, ultimately elevating the tourist experience and industry performance (Liu et al., 2020).

3.4.1 Theory behind the SARIMA model

Farsi et al. (2021) describe the theory behind the SARIMA model by explaining that SARIMA explicitly supports univariate time series data that display seasonality. It integrates both seasonal and non-seasonal factors into the time series forecasting model. It is particularly suited for data with periodic fluctuations. The model has four components namely: Seasonal Component(s): Which takes periodic seasonal effects into account; Autoregressive (AR): Which depends on the series' prior values; Integrated (I): which attains stationarity and differentiates the data, and lastly, the Moving Average (MA) which takes historical forecast errors into account.

Under SARIMA, the model is represented by these denote letters (p, d, q) (P, D, Q)_s, where p, d, q are the non-seasonal orders of MA, AR, and differencing. The seasonal cycle length is denoted by s. In contrast, the seasonal orders of AR, differencing, and MA are P, D, and Q. During the application process, one has to identify proper values for p, d, q, P, D, Q and s using data analysis. During the fitting, one employs past data to compute model parameters with Maximum Likelihood Estimation as a technique. Once the model is fitted then project the future values. A residual analysis will be performed to ensure residual errors (the differences between the actual and

predicted values) resemble white noise, which indicates a good fit. Lastly, the model performs a validation using cross-validation to verify the model's ability to predict outcomes (Artley< 2022).

3.4.2 Theory behind the Random Forest model

The Random Forest model is a strong and reliable ensemble learning technique for classification and regression tasks. It comprises many separate decision trees that work together as an ensemble. Every decision tree in the forest makes a prediction, and the ultimate result is obtained either by majority voting (for classification) or by averaging (for regression). The model divides the data into groups according to the values of certain features. Multiple trees are combined in ensemble learning to reduce overfitting and improve prediction stability. To lower variance, each tree is trained using a random sample of the data. To guarantee variation among trees, a random collection of characteristics is used for each split. The model performance is measured based on data that were not part of the training set (Schonlau & Zou, 2020).

3.4.3 Theory behind the Prophet model

According to Liço et al. (2021), the Prophet model is a powerful forecasting tool for time series data with multiple seasons of historical data and strong seasonal patterns. It is appropriate for real-world applications since it is resilient to outliers and missing data. Prophet breaks down the time series into three primary parts using an additive model: trend, seasonality, and holidays.

The trend part measures the data's long-term upsurge or reduction. Both logistic and linear growth models are supported. The seasonality part supports annual, monthly, and daily patterns and models periodic fluctuation. Lastly, the holiday effects take into consideration the unique occurrences that have a big influence on the data.

Model Equation:

$y(t)=g(t)+s(t)+h(t)+\epsilon t$

Where: g(t) is the trend, s(t) is the seasonal part, h(t) represents holiday effects and ϵt is the error term.

Khare (2024) explains that in the Prophet model application , information is gathered such as values and timestamps including any applicable holidays. The model uses maximum likelihood estimation to fit the trend, seasonal, and holiday components to the historical data. By repeating the seasonal and holiday patterns and generalizing the trend, the model predicts future values. During the uncertainty intervals, predictions will be based on the variability of past data.

3.5. Studies that have employed machine learning to predict tourism-related outcomes

Recent studies have embraced machine learning techniques to predict various tourism-related outcomes, offering insights into travellers' behaviour, demand forecasting, and destination management. One notable study conducted by Liu et al. (2020) focused on predicting hotel occupancy rates. Employing an (LSTM) neural network model, the researchers used historical booking data, online reviews, and social media interactions as input features. The study demonstrated that the LSTM model outperformed traditional methods, yielding more accurate occupancy predictions. However, limitations included the need for extensive data preprocessing and potential challenges in interpreting the model's inner workings.

In another study by Zhang et al. (2021), machine learning was employed to predict tourists' travel intentions. The researchers utilised Support Vector Machines (SVM) and Random Forest algorithms to analyse data from online reviews and social media sentiment. The findings indicated that sentiment analysis could effectively predict travel intentions, aiding businesses in strategic planning. Nonetheless, this approach's dependency on subjective user-generated content and the need for continuous model refinement were acknowledged limitations.

A study by Chen et al. (2022) focused on predicting tourist flow patterns within a city using a Recurrent Neural Network (RNN) model. By analysing location-based data from mobile apps, the model accurately forecasted visitor movement, helping local authorities manage traffic and allocate resources efficiently. However, privacy concerns related to location data collection and potential biases in the training data were recognised as challenges.

While these recent studies showcase the potential of machine learning in predicting tourism-related outcomes, common limitations include the necessity for high-quality, diverse datasets, potential bias in training data, model interpretability challenges, and ethical considerations regarding data privacy. As the field advances, addressing these limitations and ensuring responsible implementation will be crucial for realising the full potential of machine learning in predicting and enhancing various aspects of the tourism industry (Chen et al., 2022; Zhang et al., 2021; Liu et al., 2020).

3.6 Theory of planned behaviour

The theory of planned behaviour can be used to predict various behaviours among other tourist travel behaviours. A study done by Pahrudin et al (2021) explained the theory of planned behaviour in terms of tourist prediction as tourist intent to travel for a specific destination is a stimulus by behaviour or attitudes. With attitude behaviour, if the tourist has a strong attitude toward a certain destination, the tourist is likely to travel to that specific destination. The facts that encourage strong attitudes are culture, safety and affordability among others. According to Eom and Han (2019), the component of behaviour in the theory of planned behaviour framework, reflects tourists' perceptions of their ability to control and execute the behaviour of visiting a destination. The facts that perceived this behaviour are time, money and accessibility options. Faber et al. (2023) believes that information obtained from the theory of planned behaviour framework can then be used to develop targeted marketing strategies, improve destination infrastructure, and address any perceived barriers to travel, ultimately increasing tourist arrivals.

Chapter Four: Research Methodology

The goal of this chapter is to outline the process to accomplish the study's objectives. It outlines the research strategy, the empirical procedures used, and the philosophical presumptions that guided the work. It serves as a blueprint for how the research was conducted, providing a clear and structured framework for the entire study. This chapter draws from Saunders' six-layer research onion model (Saunders et al., 2019).

4.1. Conceptualisation

The study employed an application research framework comprising four key components. The first component involved data collection from secondary sources provided by the Ministry of Home Affairs, Immigration, Safety, and Security (MHAISS). The study utilised monthly data on tourist arrivals over nine years from 2014 to 2023. Following data acquisition, data preparation was undertaken by pre-processing the data through feature extraction to derive valuable and highly accurate information from the datasets. The third phase involved the development of the forecast model, which is followed by performance assessment as the final step.

4.2. Research philosophy

The research philosophy guiding this study is positivism. As noted by Saunders et al. (2019), positivism asserts that only "factual" information obtained through sensory observation, including measurement, is dependable. In positivist research, the researcher's role is confined to the objective collection and analysis of data. This approach was well-suited for this study, as it relies on actual data from secondary sources without making assumptions about predicting tourism arrivals in Namibia.

4.3 Research paradigm

The study adopts a data-driven research paradigm, utilising an exploratory approach to analyse data and extract insights through various analytical techniques. Furthermore, the study employs post-positivist methodologies. As Saunders et al. (2019) explain, post-positivism represents a traditional form of research, where assumptions are more applicable to quantitative research than to qualitative research.

4.4 Research strategy

This study used both analytical and predictive research methods. Analytical research entails using critical thinking skills and evaluating relevant facts and information (Pandey et al., 2023). An analytical approach was used in this study because the data were already available from a secondary source (MHAISS). The data were analysed, and a model was created to predict international tourist arrivals in Namibia.

4.5 Study choice

Given that this study applied data science to solve a problem, quantitative research was the preferred approach. Quantitative data, defined as numerical measures or counts, are suitable for this method (Pandey et al., 2023). The study selected the quantitative approach because the available data from the MHAISS on tourist arrivals and departures are numeric.

4.6 Research time horizon

The study adopted a cross-sectional approach, which involved simultaneously measuring both the outcomes and exposures of the study participants. This methodology was applies to the study as participant data are selected based on predetermined inclusion and exclusion criteria, encompassing data from 2014 to 2023.

4.7 Research techniques, tools and procedures

The study employed an applied research framework comprising four integral components. The initial stage involved data collection from a secondary source, specifically the MHAISS, spanning nine years from 2014 to 2023. Subsequently, the collected data underwent preparation through pre-processing techniques, including feature extraction, aimed at extracting valuable and precise information from the dataset. The third phase entailed the development of a forecast model, followed by model evaluation and testing as the final step. Consistent with common practice, 80% of the data was utilised for training the model, while the remaining 20% was allocated for testing purposes. An overview of the research approach is illustrated in Figure 4.1. Model evaluation and comparison were conducted using mean square error. Python, an interpreted object-oriented language, was utilised during the experimentation phase of the methodology. The study leveraged Python libraries, such as Pandas and NumPy, which support statistical analysis.

Theoretical Framework for Predicting International Tourist Arrivals using Machine Learning:

i. Data Collection and Pre-processing

- Gathered historical data on international tourist arrivals from MHAISS, which may include variables such as arrival date, nationality, year, border entry, and exit.

- Pre-processing the data by handling missing values, and outliers, and ensuring data consistency.

ii. Feature Selection and Engineering:

- Identify relevant features that could impact tourist arrivals such as the total entry of tourists a year that contribute to the prediction level.

- Engineer new features such as total entry per year to capture temporal patterns and aid in future prediction using past years.

iii. Model Selection and Development:

- The study chose appropriate machine learning algorithms for the prediction of tourists' arrival, such as regression models (random forests), and time series models (SARIMA and PROPHET).

SARIMA model development

In time series analysis, the SARIMA model has increasingly gained distinction due to its integration of a seasonal variation component, which has become particularly relevant in the tourism industry ((Wu et al., 2020)). Accordingly, this study employed the SARIMA model with a seasonal parameter (s = 12 for monthly data) to forecast tourist arrivals, using the following notion and equation below:

SARIMA(p,d,p)X(P,D,Q)S

Where p: the order of the non-seasonal AR component; d: the order of the non-seasonal differencing, q: the order of the non-seasonal MA component; P: the order of the seasonal AR component; D: the order of the seasonal differencing; Q: the order of the seasonal MA component; and S: The number of periods in each season. The mathematical equation of SARIMA is:

$$\Phi P(Bm)\phi p(B)(1-B)d(1-Bm)Dyt=\Theta Q(Bm)\theta q(B)\epsilon t \qquad \text{eqn. 3.1}$$

Where: yty_tyt is the time series data, BBB is the backward shift operator, ϵt\epsilon_tϵt is the white noise or error term, ϕp(B)\phi_p(B)ϕp(B) is the non-seasonal AR polynomial of order ppp, θq(B)\theta_q(B)θq(B) is the non-seasonal MA polynomial of order qqq, ΦP(Bm)\Phi_P(B^m)ΦP (Bm) is the seasonal AR polynomial of order PPP and ΘQ(Bm)\Theta_Q(B^m)ΘQ(Bm) is the seasonal MA polynomial of order QQQ.

Random Forest model development

The Random Forest model integrates the classification and regression tree algorithm with the bagging technique to enhance predictive accuracy. First, random subsets are selected from the training dataset. Next, decision trees are randomly generated and trained using these subsets. The parent node then splits into two child nodes, and the change in information impurity resulting from this split can be expressed as follows:

$$\Delta g(N) = g(N) - P_L g(N_L) - P_R g(N_R) \qquad \text{eqn. 3.2}$$

where g(N) represents the Gini impurity measure at node N (p_L) denotes the population proportion of the left child node, and (p_R) denotes the population proportion of the right child node.

Each tree makes predictions on the testing dataset, and the predictions from all trees are averaged to produce the final output for forecasting tourist arrivals. The final output of the Random Forest model is as follows; where yˆ is the final output, N tree is the number of trees and yi is the result of a single tree.

$$\hat{y} = \frac{1}{N_{trees}} \sum_{i=1}^{N_{trees}} y_i \qquad \text{eqn 3.3}$$

Prophet Model development

The Prophet model is capable of making accurate time series predictions using straightforward parameters. One of the Prophet's key advantages is its support for incorporating seasonality and irregular components.

The Prophet model equation:

$$y(t)=g(t)+s(t)+h(t)+\epsilon t \qquad \text{eqn 3.4}$$

Where: g(t) is the trend, s(t) is the seasonal part, h(t) represents holiday effects and ϵt is the error term.

- The study Split the dataset into training 80% and testing 20% sets to evaluate model performance.

iv. Model Training and Evaluation:

- Trained the chosen model on the training dataset using the selected features (Total arrivals).

- Tune hyperparameters to optimize model performance using techniques like (Grid SearchCV and Auto-Arima).

- The Mean Absolute Error (MAE) is one of the measures used in the study to assess the model's performance.

v. Interpretation and Validation:

- Interpret the model to understand the impact of each feature on the prediction.

- Validate the model's predictions on the testing dataset and assess its generalization to new data.

vi. Predictions and Insights:

- After the model is validated, the predictions on future international tourist arrivals are made based on new data inputs.

- Extract insights from the model's predictions to understand the factors influencing tourist arrivals.

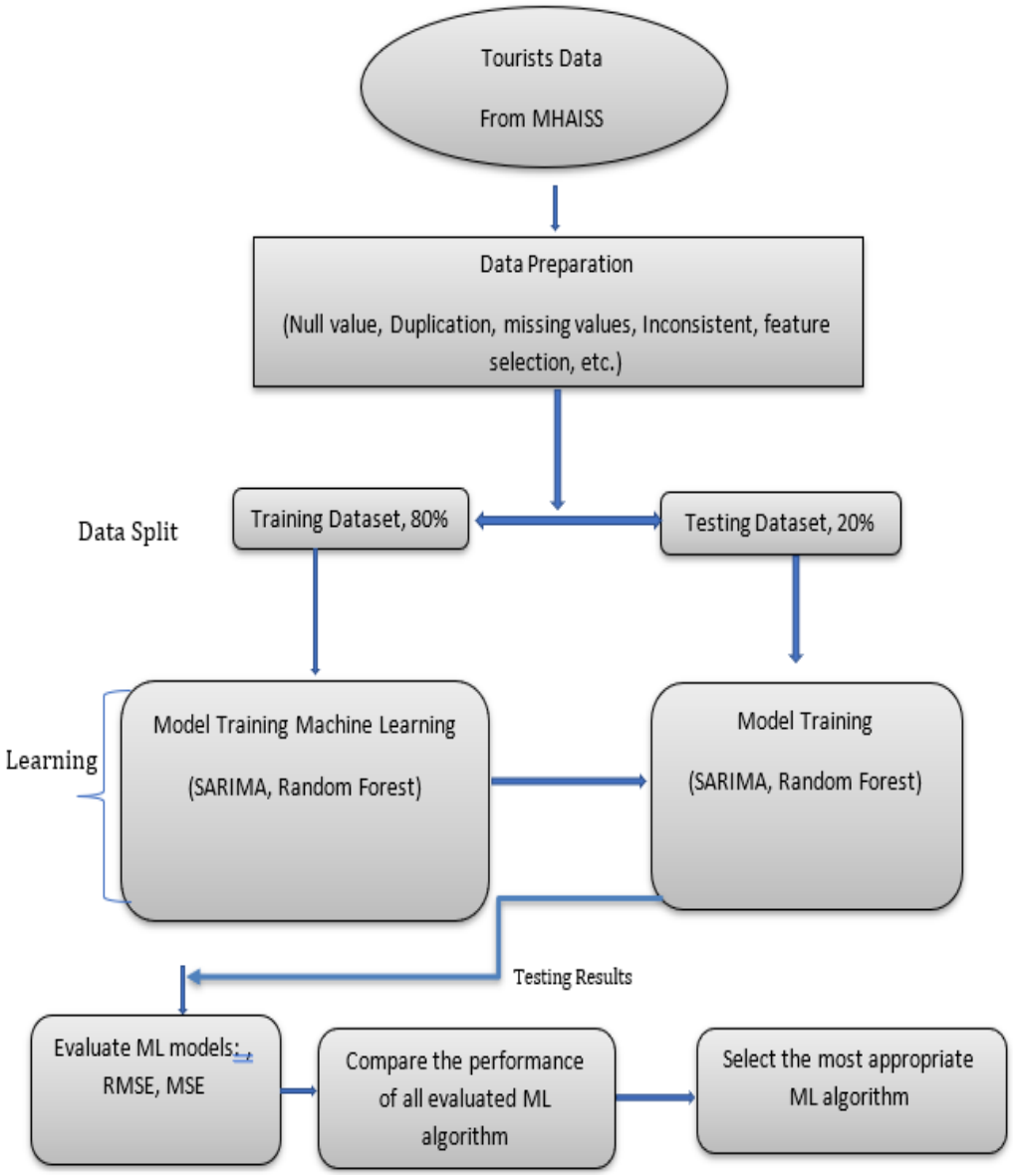

Figure 4. 1 Process diagram for model implementation

4.8 Ethics and confidentiality

Before commencing with the data collection, ethical clearance was obtained from the University which was submitted to the National Commission on Research, Science and Technology (NCRST). The NCRST approved the collection of data from the MHAISS. The MHAISS was informed of the purpose of the study from the onset and ensured that the gathered data would be exclusively utilised for its intended research purpose. The data was treated with the utmost confidentiality and the study ensured completely anonymous in all the fields. No individual information was conveyed and data of a personal nature was anonymised. Besides NUST, the research findings will be shared with the MHAISS which is the custodian of the data, as well as the NCRST which promotes and coordinates research, science, and technology in the country.

Chapter Five: Implementation, Data Analysis and Interpretation of Results

This chapter presents the results and discusses the findings of the study. It gives detailed statistical and analytical methods used to analyse the data. It presents the results of the analysis and discusses the statistical tests performed. This chapter discusses the significance of the results and how they relate to the research questions. The comparison of the findings with previous research is also highlighted.

5.1 Overview

The study analysed international tourist arrivals to Namibia using archived data from the MHAISS for the period 2014 to 2023. Python was used to carry out the analysis and interpretation of the results.

	Exits	Arrivals	year
count	113.000000	113.000000	113.000000
mean	52662.283186	55094.203540	2018.221239
std	29266.706234	28453.875163	2.737747
min	2147.000000	1459.000000	2014.000000
25%	30248.000000	37464.000000	2016.000000
50%	46956.000000	53830.000000	2018.000000
75%	81711.000000	80773.000000	2021.000000
max	102517.000000	100584.000000	2023.000000

Figure 5. 1 Summary of the descriptive statistics

Figure 5.1 offers a comprehensive overview of descriptive statistics of the datasets, particularly focusing on yearly international tourist arrivals. The years included a range from 2019 to May 2023. During this period, the average number of tourists visiting Namibia stands at 55,904. The data reveals that the lowest recorded tourist arrivals occurred in 2014, with only 1,459 tourists visiting Namibia, while the highest number of arrivals was observed in 2023, reaching 100,584.

Standard deviation, a measure of variance from the mean, is calculated at 28,453, indicating significant daily fluctuations in tourist arrivals compared to the average. This suggests that the daily arrival numbers deviate considerably from the mean.

The quartile values (25%, 50%, and 75%) represent the distribution of tourist arrivals across the first three quarters of each year. Specifically, 37,464 tourists arrived in the first quarter, 53,830 in the second quarter, and 86,773 in the third quarter on average during the specified period.

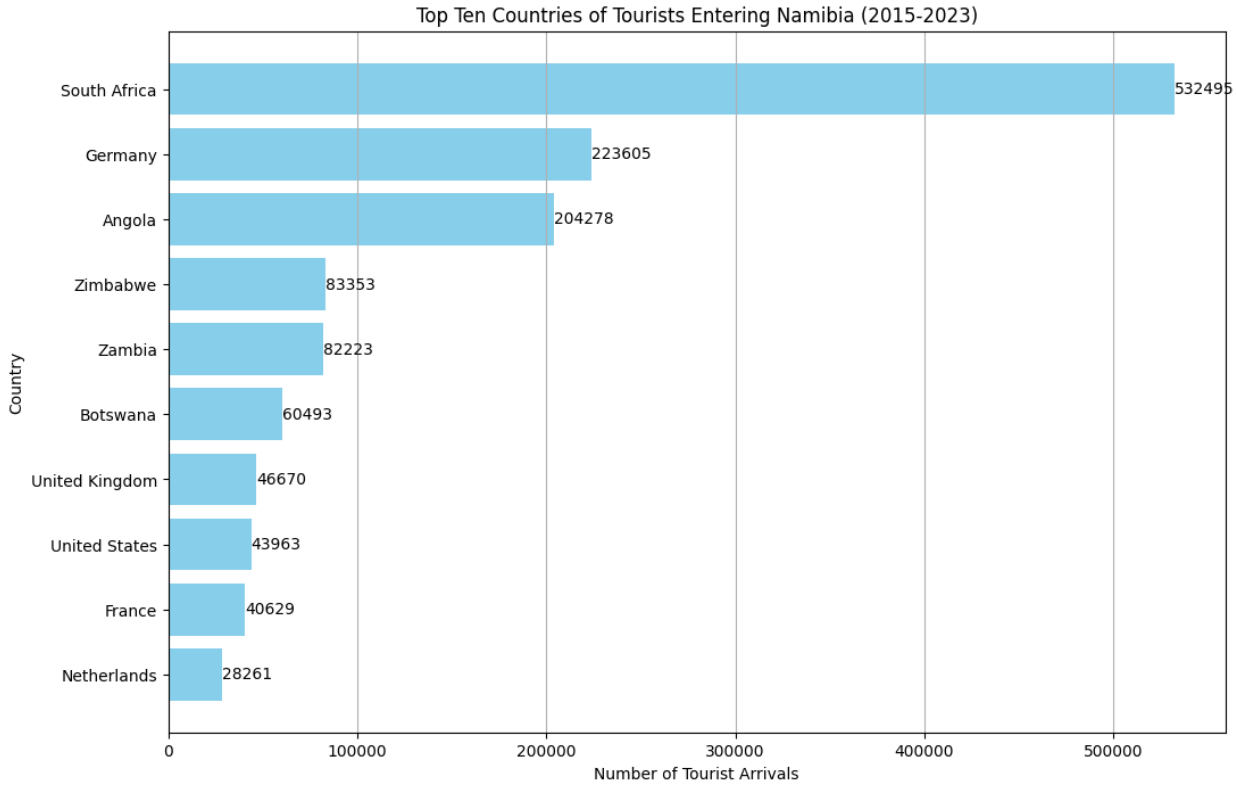

Figure 5. 2 Number of tourist arrival per country

The bar result presented in Figure 5.2 illustrates the number of tourist arrivals per country. The leading country in tourist arrivals in Namibia is South Africa, with a total of 532,495 visitors. This is followed by Germany, which accounts for 223,605 arrivals, and Angola with 204,278 tourists.

Following these top three countries, the next highest numbers of tourist arrivals are seen as coming from Zimbabwe, Zambia, and Botswana. Among Western countries, the United Kingdom and the United States also contribute significantly to tourist arrivals, with France recording 40,629 visitors. Notably, the Netherlands recorded a lower figure of 28,261 tourist arrivals in Namibia.

Overall, the data indicate a strong preference for South Africa as a main tourist origin for Namibia, with Germany and Angola also serving as significant sources of tourists. The consistent arrivals from Western countries such as the UK, the US, and France further underscore the diverse origins of tourists in the region.

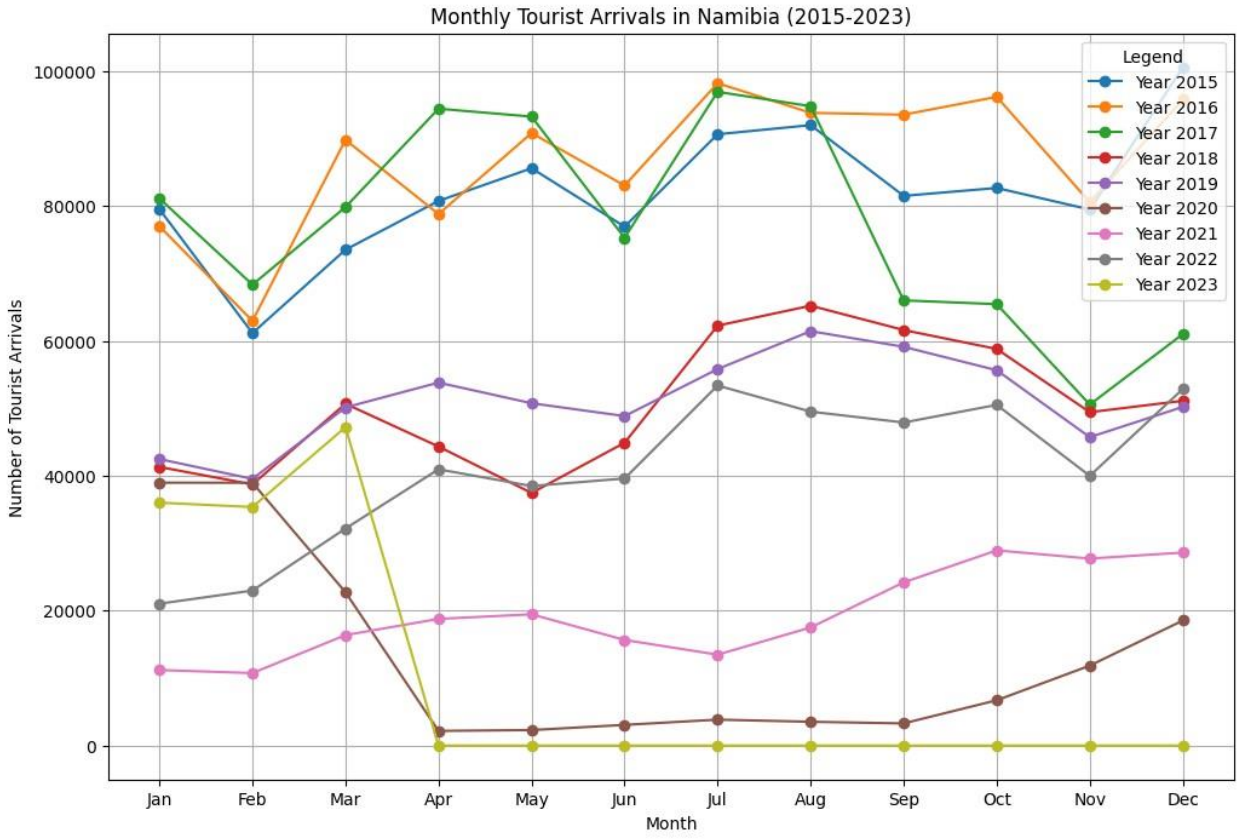

Figure 5. 3 Monthly tourist arrival 2015 to May 2023

Figure 5.3 displays the monthly total arrival between January 2014 and May 2023. In 2017, there was a notable peak in tourist arrivals during the first two quarters, followed by a significant decline in the last quarter. Interestingly, in 2016, the highest number of tourist arrivals was recorded during the last quarter of the same period. The data also indicates a complete absence of tourist arrivals from May to December 2023, attributed to a correction in the data that commenced in early May 2023. Equally, a decline in tourist arrivals was consistently observed between August and October annually.

However, an exception to this pattern was noted in the years 2021 and 2022, during which the period from August to October experienced an increase in tourist arrivals. Furthermore, there was a noticeable decrease in tourist arrivals from April to September 2020, likely due to travel restrictions imposed amidst the COVID-19 pandemic. In 2019, the tourist arrivals remained relatively consistent throughout the year, displaying an average pattern. These results fluctuations suggest the presence of both seasonal and cyclical patterns, alongside random variations stemming from nonlinear dynamics.

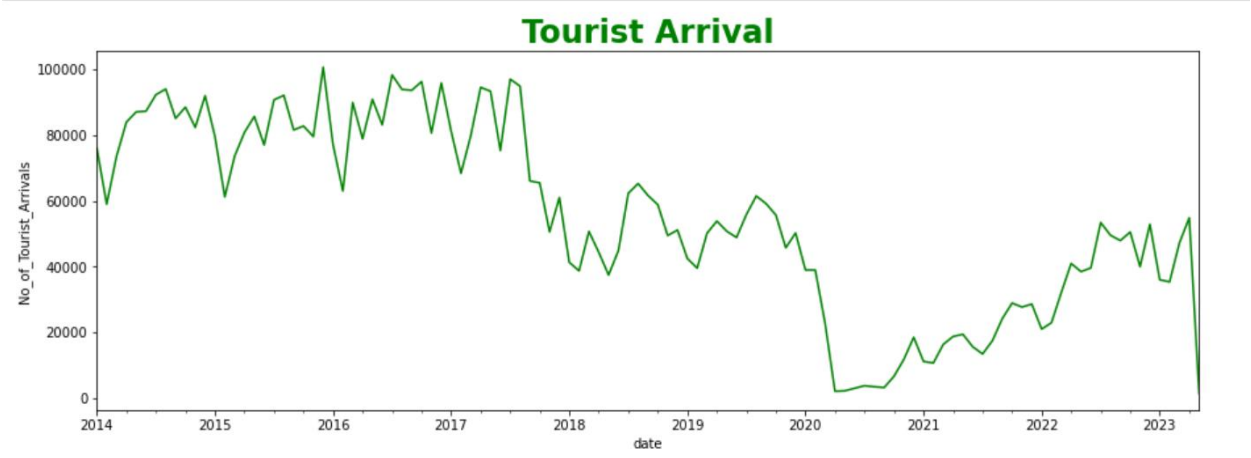

Figure 5. 4 Linear annual total tourist arrival

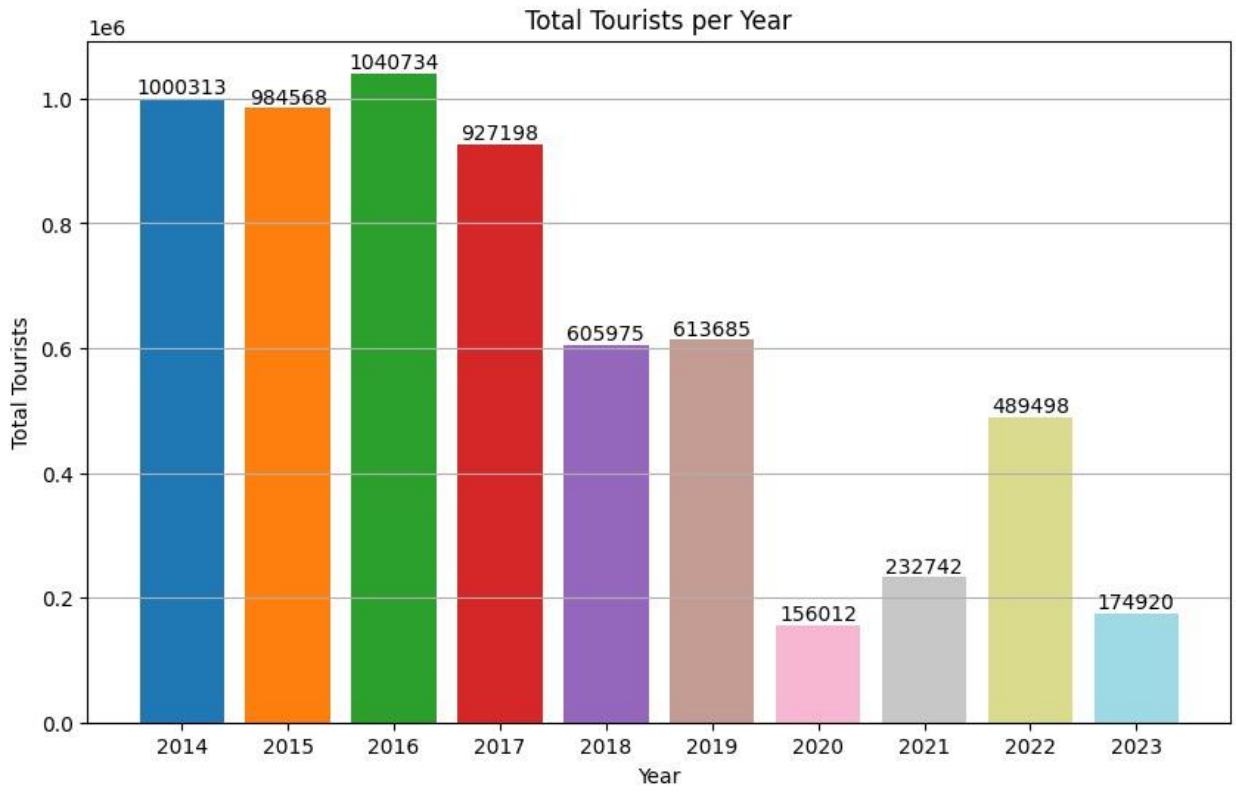

Figure 5. 5 Annual count of total tourist arrival

Figures 5.4 and 5.5 illustrate the yearly trends in tourist arrivals. It is evident that in both 2020 and 2022, there was a decline in tourist numbers due to the imposition of COVID-19 travel restrictions. However, there appears to be a gradual uptick in tourist activity in 2022. Notably, in 2018, the number of tourists dropped by over 30% from previous years. Several factors may have contributed to this, which is not known in this study.

In Figure 5.3, it is observed that there was an increase in tourist numbers during the first quarter of 2017. Despite this, Figure 5.5 indicates that 2016 still recorded the highest total number of tourists annually during the period under study.

5.3 Development of the model

The complete dataset is divided into two segments: training and testing datasets. The data were split into 80% of training and 20% of testing. The purpose of splitting the dataset is to train the machine learning model using a bigger portion of the dataset and use the least data to evaluate the model performance. It is quite useful in evaluating how the model simplifies unseen data. In most cases, the split of datasets to such proportion aids in avoiding overfitting. Moreover, to forecast Namibia's daily tourist arrivals, SARIMA, Random Forest and Prophet have been evaluated. The SARIMA model was chosen because it has a good fitting ability and high short-term prediction accuracy. In addition, since tourists are seasonal in Namibia, SARIMA would be the suitable model to capture time-dependent patterns and forecast time-series data.

However, the Random Forest (RF) model was chosen because it captures non-linear and complex interactions between features, especially to determine if tourist arrivals are being influenced by factors such as destination, exchange rate, culture, etc. RF is less sensitive to outliers in comparison to other models. Furthermore, in comparison to SARIMA, RF is quite useful to integrate other factors which SARIMA might struggle with. In this study, RF was chosen to resolve bias in the model that might exist in certain years or prediction lengths.

Like the SARIMA model, the Prophet model was chosen as it handles seasonality. The Prophet model is designed to model data with strong seasonal effects. Tourist arrivals often exhibit seasonal patterns, such as peaks during holiday seasons or certain months of the year, and Prophet is well-suited to capture these patterns accurately. Moreover, Prophet is robust to missing data and outliers, which are common in real-world data sets like tourist arrivals. This robustness ensures that the model remains reliable even when the data quality is not perfect. Equally, most of the studies compare SARIMA and Prophet in the prediction of tourist arrival hence the models were found fit for this study.

5.2.1 Data pre-processing

Data pre-processing and feature extraction were performed during this phase. During data pre-processing, the study summed all tourists within one month into a total monthly format and had one monthly total arrival. The arrival time was also converted to a date stamp to remove the hours,

minutes and seconds from the time and only have the arrival time with the date. The raw data was in YYYY-MM-DD, HH-MM-SS and was converted to YYYY-MM-DD, for consistency and simplification as the data were aggregated daily and not at specific times of the day. The conversion was also possible for data reduction to reduce the granularity of the data. It makes data easily managed and easier for the type of analyses used for this study.

The following features extractions were performed: The removal of all data with the nationality of Namibia, the dropping of all rows with arrival date which is null, Nal or blank space, checking for a duplicate of tourist details was performed and rows were dropped, and the countries' names with commas were removed.

Data standardization was performed as the last part of the data pre-processing step. To do this, the study performed data standardisation using the equation below:

$$X_{transformed} = (X - \bar{X}) / \sigma,$$

where X stands for the initial value, σ for the standard deviation, and $\bar{X}$ for the mean.

To handle time series data with machine learning methods, this study involved the extraction of two variables, namely, the month and the year and total arrival. Figure 5.6 shows all variables used in the prediction model.

	Month	Exits	Arrivals	year
0	2014-01-01	72029	75960	2014
1	2014-01-02	57135	58961	2014
2	2014-01-03	68315	73499	2014
3	2014-01-04	80527	83877	2014
4	2014-01-05	88192	86997	2014
...	...	...	...	...
108	2023-01-01	43393	36025	2023
109	2023-01-02	32656	35372	2023
110	2023-01-03	42229	47213	2023
111	2023-01-04	53724	54851	2023
112	2023-01-05	2147	1459	2023

113 rows × 4 columns

Figure 5. 6 Prediction model variables

SARIMAX Results

Dep. Variable:	Arrivals	**No. Observations:**	90
Model:	SARIMAX(2, 1, 0)x(0, 1, [1], 12)	**Log Likelihood**	-818.274
Date:	Fri, 06 Oct 2023	**AIC**	1644.547
Time:	09:33:48	**BIC**	1653.922
Sample:	01-01-2014	**HQIC**	1648.297
	- 06-01-2021		
Covariance Type:	opg		

	coef	std err	z	P>|z|	[0.025	0.975]
ar.L1	-0.0497	0.102	-0.485	0.627	-0.251	0.151
ar.L2	-0.1167	0.142	-0.820	0.412	-0.396	0.162
ma.S.L12	-0.3373	0.071	-4.754	0.000	-0.476	-0.198
sigma2	1.12e+08	3.57e-10	3.13e+17	0.000	1.12e+08	1.12e+08

Ljung-Box (L1) (Q):	0.21	**Jarque-Bera (JB):**	8.12
Prob(Q):	0.64	**Prob(JB):**	0.02
Heteroskedasticity (H):	0.98	**Skew:**	-0.60

Figure 5. 7 SARIMA model development

Figure 5.7 presents the results of SARIMA model development, where the number of observations was set to 90, reflecting the historical data points available for analysis and model training. The log-likelihood value of -818.274 measures the likelihood of observing the given data under the SARIMA model. Higher log-likelihood values suggest better model fit, thus indicating that the model adequately captures the underlying patterns in the data.

The coefficient of -0.0497 represents the estimated parameter in the SARIMA model, with a standard error of 0.102 indicating the uncertainty associated with this estimate. The variance of the error term (Sigma2) in the SARIMA model is estimated to be 1.12e+08, signifying the variability of the residuals. A higher Sigma2 suggests greater variability in the model's residuals.

Furthermore, the probability (Q) value of 0.64 indicates the likelihood that the residuals of the SARIMA model are uncorrelated. A higher Q value, closer to 1, implies a better model fit with less autocorrelation in the residuals. Conversely, the heteroskedasticity value of 0.98 denotes the degree of variance in the errors of the SARIMA model. A value nearing 1 suggests less heteroskedasticity, indicating more consistent error variance across observations, as depicted in Figure 5.7.

5.3 Model evaluation

The outcomes demonstrate that the performance of the SARIMA model is better than the Random Forest and Prophet model. Two measures were used in the study to evaluate the prediction performance: RMSE and MAE. Regression analysis and time series forecasting prediction accuracy are measured by RSME. This RSME metric measures how far a set of data points' actual and anticipated values diverge from one another. The model performs better if the RMSE is smaller. The higher the RMSE the poorer the model performance. Contrarily, the MAE metrics quantify the variation between the expected and observed values within a dataset. The MAE is a clear-cut and simple-to-understand indicator of a prediction model's accuracy. When the MAE is low, the model's predictions are generally within a reasonable range of the observed values. The model's predictive accuracy for the target variable increases with a decreasing MAE. When the model's predictions have a high MAE relative to the actual values, the errors are bigger on average. The accuracy of the model decreases with increasing MAE.

RMSE and MAE are calculated as shown below, where "yi" represents the observed or real value,

while "byi" denotes the forecasted or predicted value for tourist arrivals and n is the number of data points.

$$RMSE = \sqrt{[(\Sigma(yi - yi)^2) / n]} \qquad \text{eqn 5.1}$$

$$MAE = (1/n)\ \Sigma(i=1 \text{ to } n)\ |y\_i - \hat{y}\_i| \qquad \text{eqn 5.2}$$

Model	RMSE	MSE	MAE
SARIMA	20781.71	431879362.11	17930.75
Random Forest	23004.62	529212603.3	19970.03
Prophet	43532.75	1895100582.97	38776.2

Table 5.1 Performance of SARIMA vs Random Forest vs Prophet

RMSE is a crucial measure for assessing the accuracy of predictive models. It provides a quantitative indication of how closely the model's predictions match the actual observed data. The RMSE values for both the SARIMA, Prophet and Random Forest models are presented in Table 5.1 above, representing annual predictions spanning from 2014 to May 2023.

SARIMA Model's RMSE score is 20781.71. This has proven how close the model's predictions match the actual observed values. Since the RMSE is moderately low, it indicates the model prediction of the actual observation is somehow on average. As seen in figure 5.8 the prediction of the last quarter of 2022 is almost close to the actual value of the same quarter, which indicates that the SARIMA model was good for predicting tourist arrivals in 2022 last quarter. The Prophet RMSE score is 43532.75 which is a bit higher compared to the SARIMA and Random Forest at 23004.62 and always shows how far the prediction results are from the actual values.

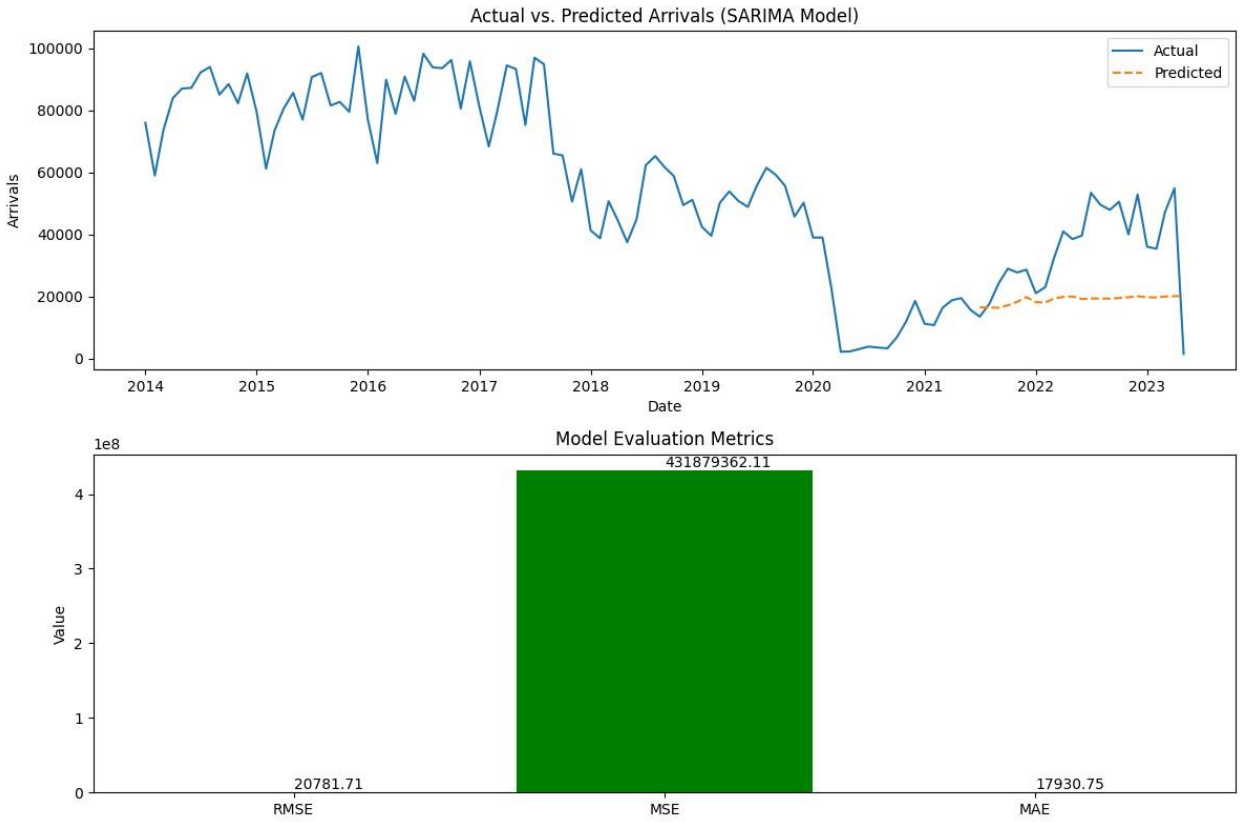

Figure 5. 8 SARIMA predicted vs expected values and SARIMA RMSE performance

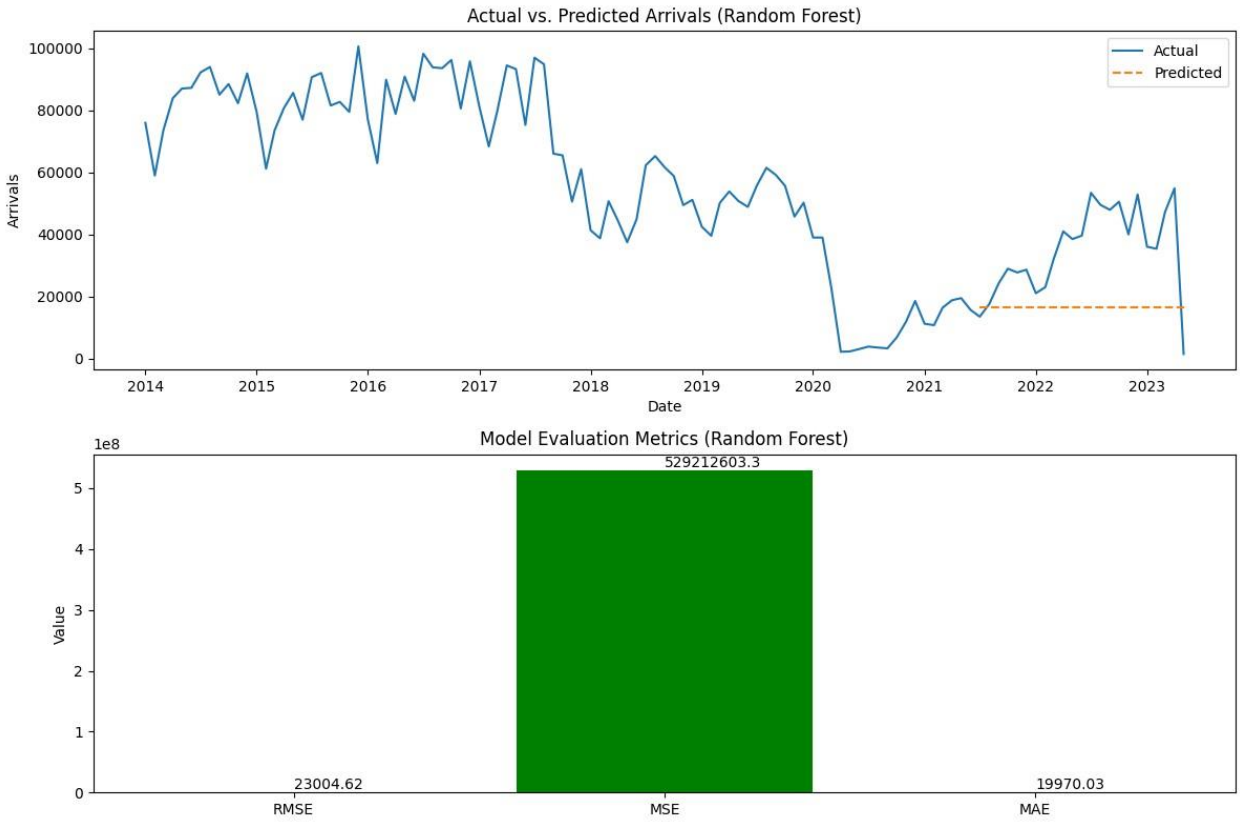

Figure 5. 9 Random Forest predicted vs expected values and Random Forest RMSE performance

The comparison between the RSME values shown in Figures 5.8 and 5.9, covering the period from 2015 to May 2023, is noteworthy. Specifically, the RSME for the Random Forest model stands at 23004 whereas for the SARIMA model, it is considerably lower at 20781. This significant difference in RSME values highlights the SARIMA model's superior accuracy in predicting tourist arrivals.

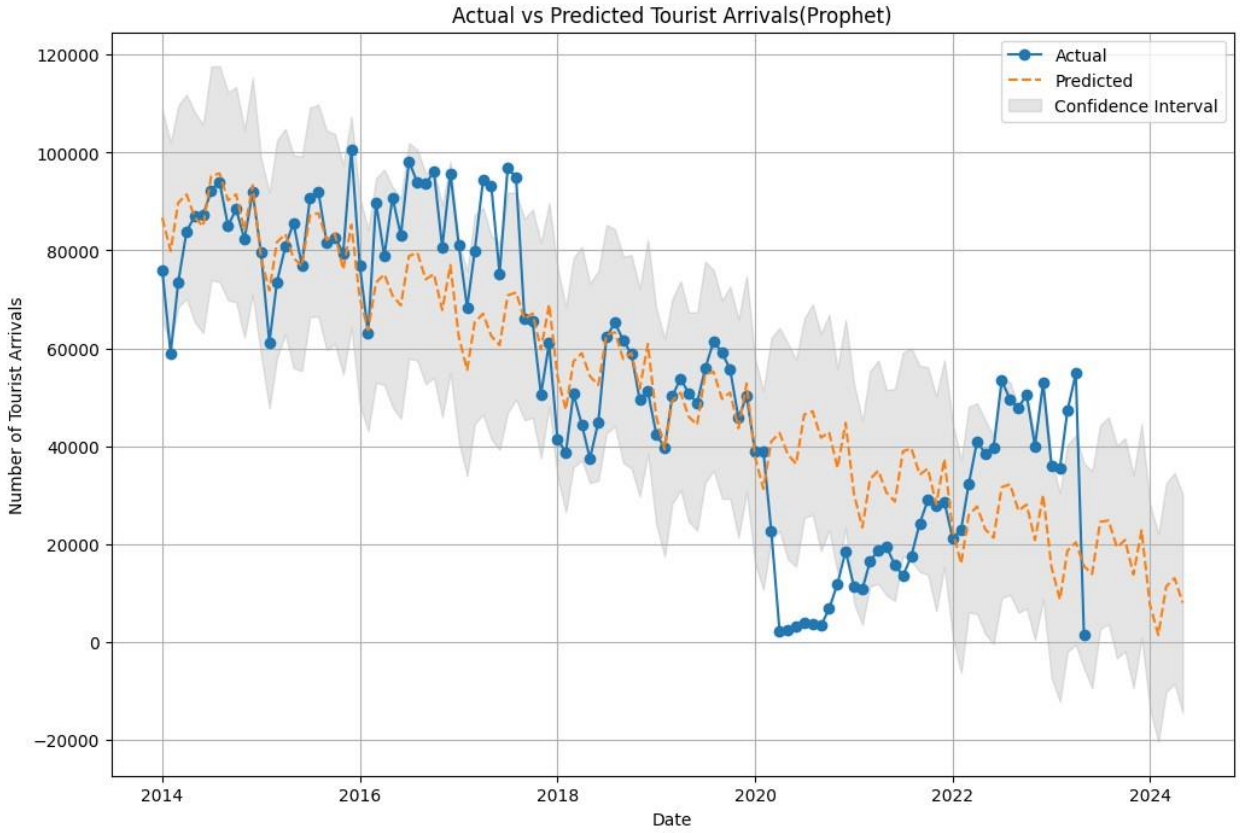

Figure 5. 10 Actual arrivals vs Prophet prediction

Figure 5.10 illustrates the performance of the Prophet model in predicting tourist arrivals. The actual data points are represented by dots or marks that are linked together. According to the Prophet model, the predicted values indicate the expected number of tourist arrivals. Comparing these predictions to the actual data helps to assess the model's accuracy.

The result shows accurate predictions in the last quarter of 2014 to 2015, some months in 2016, 2019, and some months in 2022. Despite the close similarity between actual and predicted values, the Prophet model has the highest RMSE of 43,532.75, compared to SARIMA (20,781.71) and Random Forest (23,004.62).

During the last quarter of 2021 and the first quarter of 2022, the actual data significantly deviates from the predicted values, falling outside the grey area, which indicates higher uncertainty. This deviation is likely due to the impact of COVID-19 on tourist arrivals, which affected the model's performance.

The confidence interval, shaded in grey, represents the uncertainty around the predicted values. It typically shows the range within which the true values are expected to fall with a certain probability, usually a 95% confidence interval. Figure 5.10 indicates a narrow confidence interval, which suggests higher confidence in the predictions. Most of the actual and predicted values fall within this grey area, thus demonstrating that the model's predictions are accurate and reliable.

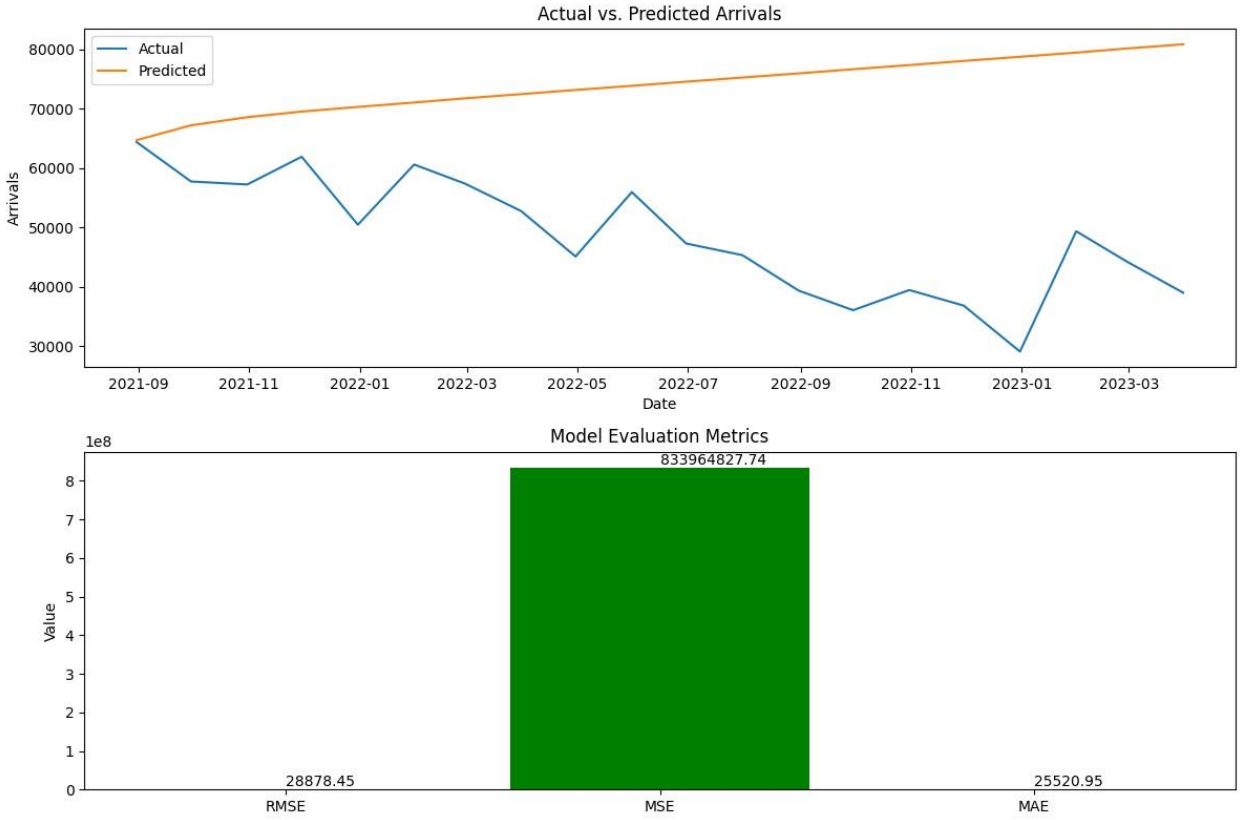

Figure 5. 11 SARIMA Model monthly prediction from September 2021 to March 2023

The graphical representation in Figure 5.11 shows the SARIMA model prediction versus the actual data between September 2021 and March 2023. It is evident that the SARIMA model accurately predicts tourist arrivals only once, notably in September 2021. From September 2021 to June 2022, the model shows closely aligned prediction values, indicating a relatively consistent performance during this period. However, beyond June 2022, the margin of prediction widens substantially until

March 2023. This may be attributed to the fact that the SARIMA model is designed to capture seasonal patterns in data, so 2022 was still a peak of COVID-19, which shifts in seasonal tourist behaviour. This might lead to tourist arrivals that are not captured well by the model, it can lead to wider margins of error.

This widening margin of prediction is verified by the MSE metric, which registers the highest values. High MSE metrics indicate significant discrepancies between the squared differences between the actual and predicted values, which infer that the model's predictions are consistently off from the real values. Such elevated MSE metrics signify considerable inconsistencies between the squared differences between predicted and actual values, thus indicating persistent deviations of the model's forecasts from the true values.

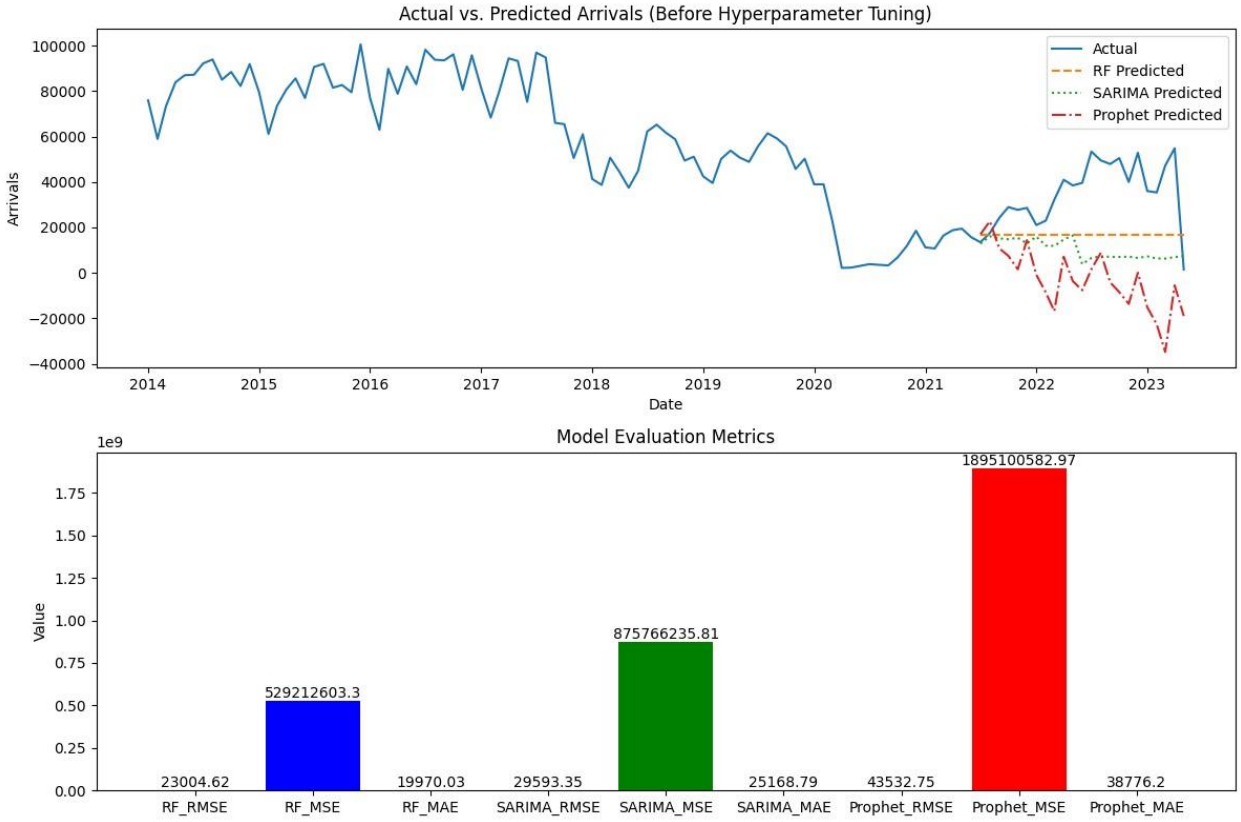

Figure 5. 12 SARIMA, Random Forest and Prophet combined models

Figure 5.12 illustrates the combined predictions of the three models used in the study, tested over the last 20 months (July 2021 to March 2023) to evaluate their performance as tourist numbers began to recover post-COVID-19. The graph indicates that the Random Forest model maintains a constant prediction trend. In contrast, the SARIMA model closely follows the actual pattern of tourist arrivals, particularly from November 2021 to March 2022. However, the Prophet model's predictions diverge significantly from the actual values, often predicting a decrease in tourist numbers when there is an increase and vice versa.

The results also reveal that both the SARIMA and Random Forest models exhibit closely similar prediction trends from August 2022 to February 2022 and April to May 2023. However, between May 2022 and May 2023, both models demonstrate a notable tendency to underestimate actual arrivals. Conversely, a higher level of accuracy is observed in the predictions of both models from September 2021 to February 2022, as shown on the graph.

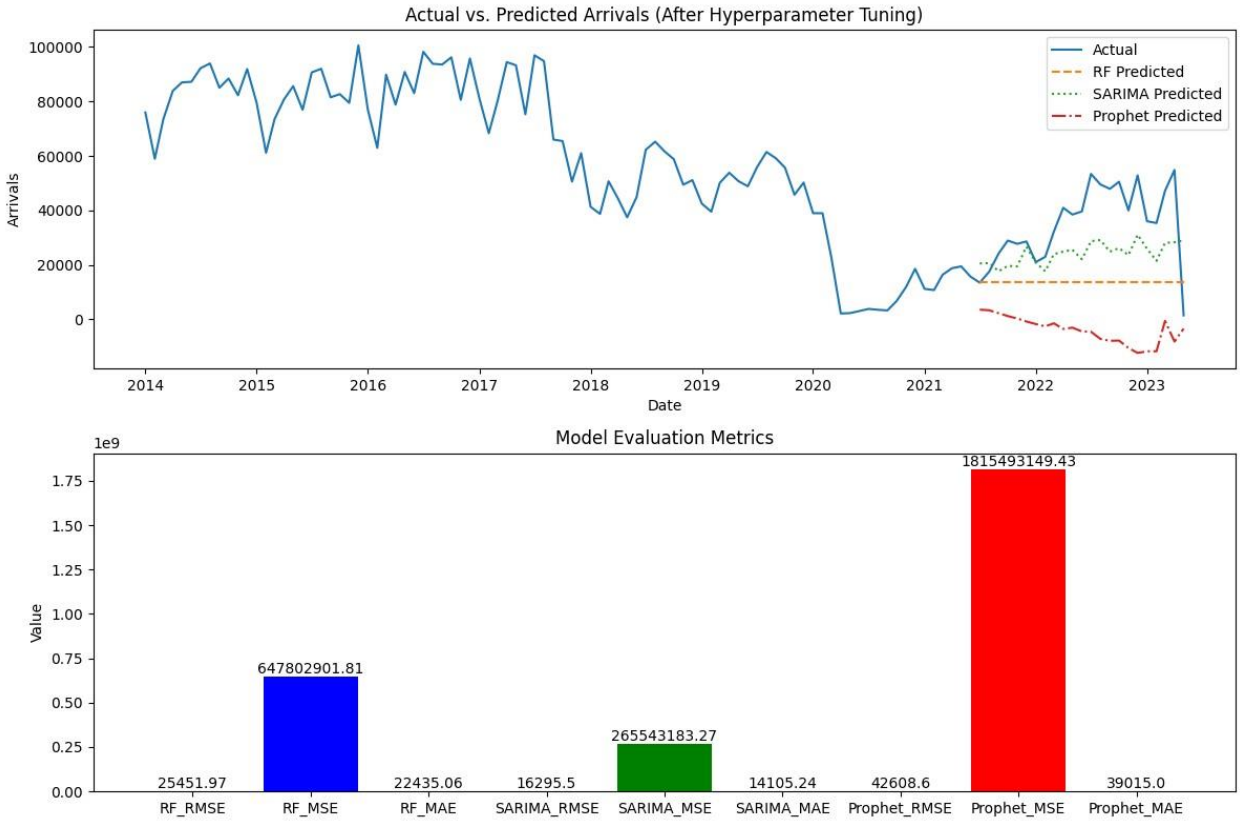

Figure 5. 13 SARIMA, Random Forest and Prophet combined models after hyper tuning

Figure 5.13 illustrates the performance of three models after hyperparameter tuning. The tuning was conducted to assess potential improvements in accuracy, reduction in overfitting, and enhanced generalization to unseen data. The new RMSE values are as follows: Random Forest (RF) at 25,451.97, SARIMA at 16,295.5, and Prophet at 42,608.6.

The graph indicates that the SARIMA model's predictions have moved closer to the actual values, resulting in an accuracy improvement with the RMSE decreasing from 20,781.71 to 16,295.5. This signifies a substantial enhancement in the model's predictive performance.

Before hyperparameter tuning, the Prophet model's predictions deviated significantly from the actual values. Post-tuning, the results show a nearly constant prediction line that further diverges from the actual values. Despite the apparent discrepancy between the predicted and actual values in the graph, there was a slight improvement in the RMSE, from 43,532.75 to 42,608.6, indicating a marginal accuracy enhancement.

For the Random Forest model, hyperparameter tuning led to a degradation in performance. The RMSE increased from 23,004.62 to 25,451.97, indicating a decline in accuracy. The prediction line moved further away from the actual values after tuning.

The hyperparameter tuning methods employed for these models are as follows: Grid Search for Random Forest, Auto-Arima for SARIMA, and manual variable tuning for Prophet. Grid Search was selected for its thoroughness, simplicity, and reliability in systematically exploring the hyperparameter space to find the optimal configuration for the Random Forest model. Auto-Arima was chosen due to its effectiveness in handling complex seasonal data, automatically detecting and incorporating seasonal parameters, thereby simplifying the model-building process for SARIMA. Manual variable tuning was applied to the Prophet model due to the complexity of the dataset.

Further evaluation using Mean Square Error reveals that the SARIMA model, with an MSE of 265,543,183.27 outperforms the RF model, which registers an MSE of 647,802,901.81 and Prophet with an MSE of 1,815,493,149.43. This disparity underscores the SARIMA model's enhanced predictive accuracy. Specifically, the SARIMA model demonstrates a smaller average squared difference between predicted and actual values, highlighting its superior precision in forecasting tourist arrivals relative to RF and Prophet, as per the MSE metric.

A closer examination of the bar graph indicates that RF and SARIMA models yield comparable RMSE, suggesting similar performance in forecasting tourist arrivals. RMSE serves as a key metric for assessing the average discrepancy between predicted and actual values, with lower values indicative of superior predictive accuracy. Consequently, it can be inferred from the graph that both RF and SARIMA models are adequately suited to the data at hand. However, the SARIMA model exhibits a marginally superior performance owing to its lower RMSE compared to RF.

Chapter Six: Conclusion

This chapter summarises the key conclusions and findings and offers a concise response to our study's principal goals and questions. The chapter also details the theoretical and practical results of our findings, thus emphasising their importance to the entire dataset. Furthermore, future study topics and suggestions were provided, hence highlighting areas of concern relating to additional investigation. These findings suggest improvements in research techniques for addressing open-ended issues. The significance of our findings and their potential to influence future research in our field are echoed in the final sentence of this chapter, which is an important conclusion statement.

6.1. Summary of findings

This study uses historical data and machine learning algorithms namely SARIMA, Prophet, and Random Forest to forecast the number of foreign visitors to Namibia monthly. The results demonstrate the SARIMA exhibits the least RMSE, which indicates a higher level of accuracy. To build the model, the dataset was divided into two sections: 80% training data (2014-01-01 to 2021-05-01) and 20% test data (2021-07-01 2023-05-23). Notably, the pandemic crisis is covered in both training and data sets in the most detail as the percentage proportion has divided the COVID-19 pandemic equally. The prediction model's accuracy and learning process were, therefore, enhanced by employing more training sets that include the relevant occurrence. In conclusion, when compared to the RF and Prophet models, the SARIMA model yielded the lowest RMSE.

The tourist arrival and arrival date were the key variables used in this study's model to forecast the arrival of tourists. Towards these variables, the study adopted the Machine Learning model, the SARIMA, Random Forest and Prophet. The next sections go into great depth on the results from the previous chapter and talk about their practical ramifications.

The research questions that guided this study are as follows:

Questions

a. What is the present pattern of tourist arrival in Namibia?

The current trend of tourists in Namibia has shown fluctuating figures between 2014 and 2018, then a continuous drop between 2019-2021 which eventually started picking up again. The dropping in tourist arrivals from 2019 through 2021 was due to COVID-19 restrictions on travelling. Since the data were collected in May 2023, it does not show whether there is a pickup in tourist trends after the years of COVID-19. Arrivals of tourists follow a similar pattern between February and April, June and August, and November and December each year.

b. How do seasonal patterns influence the number of international tourist arrivals in Namibia based on their country of origin?

Figure 5.3 indicates the increase in tourist arrivals between February and April. Namibia has 4 seasons and June to August is considered the dry season in Namibia. During the dry season, this period is the most popular for international tourists due to ideal weather conditions and the heightened visibility of wildlife. Tourists from countries with colder climates such as those in Europe (e.g., Germany), often travel to Namibia during their summer holidays to escape the heat and experience Namibia's natural beauty. This season coincides with school holidays in many European countries and South Africa, thus further boosting arrival numbers.

c. How do commonly used machine learning methods for predicting international tourist arrival perform on Namibian data

The performance of commonly used machine learning methods to forecast the number of foreign visitors who will arrive in Namibia was evaluated using SARIMA, Random Forest, and Prophet models. These models were chosen for their diverse approaches to time series forecasting and machine learning.

The study revealed that the SARIMA model performed better than both the Random Forest and Prophet models on the Namibian data. Specifically, the SARIMA model demonstrated superior predictive accuracy, with an MAE of 14,105.24 and a RMSE of 16,295.5. This suggests that the SARIMA model does a good job of identifying the underlying trends in the data on visitor arrivals.

In comparison, the Random Forest and Prophet models did not perform well, suggesting that SARIMA is more suitable for this specific dataset and context. Better performance was probably a result of the SARIMA model's capacity to take trends and seasonality in the data into account.

d. How can the accuracy of machine learning models to forecast the number of foreign visitors who will arrive in Namibia be improved?

To improve the accuracy of machine learning models to forecast the number of foreign visitors, such as SARIMA and Random Forest, feature engineering, incorporating additional relevant predictors, fine-tuning model hyperparameters, and optimising the training process should be considered to strike a balance between capturing time series patterns (in the case of SARIMA) and leveraging Random Forest's flexibility for handling non-linear relationships. The study used features engineering by extracting data from the main dataset to generate a new variable (the date and the arrival column) that is relevant to this study. In addition, methods of hyperparameter tuning such as Grid Search and Auto-Arima were used.

e. How can the Namibian tourism industry leverage emerging technologies such as artificial intelligence and Machine Learning to enhance tourism forecasting and decision-making?

The accuracy of the machine learning was tested using RMSE. The error values of 16295.5 for the SARIMA model and 25451.97 for Random Forest were obtained, with SARIMA being relatively low and thus indicating the model is making accurate predictions. Conclusively, the Namibia tourism industry can leverage the use of ML to enhance tourist prediction and adequately use the data for future prediction.

Therefore, currently, the ML that can be appropriate for the Namibia tourism industry for prediction and decision-making is SARIMA as its average magnitude of the errors between the

predicted values and the actual values, is close to the true values, which presents the high accuracy of the model. This high accuracy can help the tourism industry anticipate demand more effectively.

During data analysis, the study identified that certain data, specifically those influencing tourists' travel decisions, are not included in the MHAISS dataset. To leverage emerging technologies for predicting tourist arrivals in the tourism industry, data integration is necessary. This involves combining various data sources such as historical arrival data, economic indicators, weather patterns, and marketing efforts, to enrich the dataset and enhance the accuracy of predictions.

Conclusively, the Namibian tourism industry can significantly benefit from integrating AI and ML technologies, particularly by adopting models like SARIMA, which has demonstrated high predictive accuracy. These technologies can enhance the tourism industry's ability to forecast tourist arrivals, optimise operations, and make data-driven decisions.

6.2. Recommendations for the tourism industry

The tourism sector should realise that in this era of modern technology, it is supremely important to assess the digital opportunity that technology can deliver to improve efficiency and effectiveness within the sector. Implementing Artificial Intelligence and Machine Learning results in tourism that is both higher in quality, safer, and more responsive to the tourists' needs and, at the same time, improves planning within the sector. To address the problems of relying on seasonality and past trends to predict future tourist arrival, the industry can explore the use of ML for monthly prediction as they are more accurate because they can recognize complex patterns and non-linear correlations. There are practical studies made on the topics that are more convincing as seen under the literature section that using ML does not require a lot of manual intervention and, that machine learning models can make predictions in real-time and adjust to changing circumstances, which is why is highly recommended for any sector including tourism sector to utilise ML models in any area includes a future prediction of tourists arrivals.

In summary, all things considered, machine learning models are a useful tool for forecasting tourist arrivals because of their benefits for managing complex data, their accuracy, scalability, adaptability, and handling vast and diverse datasets with various variables and non-linear correlations. For simpler cases or when interpretability is a top concern, traditional forecasting techniques could still be helpful.

6.4. Recommendations for further research

To improve the estimations of the models, future studies may employ a longer sample period of more than 10 years. The present study only explored three models, thus future researchers can look into more Machine Learning Models and more Time Series models. In addition, the current study solely made use of two data variables (Total number of tourist Arrivals per month and Date) that were taken from the dataset. Other columns were not relevant to this study. Future studies can explore the relationship of other variables in the same dataset. As this study focused solely on historical data, future research could consider developing a prediction model that incorporates additional data sources such as social media platforms (Twitter, FB) and additional digital forums, to enhance the training dataset and improve forecast accuracy.

Besides historical data, future research should consider incorporating multi-source data such as online travel forums and Google search trends to predict the arrival of tourists in Namibia. Interestingly, future researchers should explore the Theory of Planned Behaviour to explain the decision of international tourists' intentions to visit Namibia.

To enhance the predictive capabilities of these models, future research could integrate additional data such as weather conditions, holidays, and marketing campaigns, which may further improve the accuracy of the forecasts and benefit Namibia's tourism sector.

In conclusion, this study advances the rapidly expanding subject of tourist prediction, emphasises the promise of machine learning approaches in this area, and points to areas that still need to be explored and improved upon in predictive models for foreign tourism in Namibia.

References

Akanbi, O. A., & Madu, C. N. (2018). An empirical analysis of tourism demand in Africa using the autoregressive distributed lag (ARDL) model. *Tourism Economics, 24*(7), 779-795.

Akhmet, M., Fen, M. O., & Alejaily, E. M. (2021, June). *Interdisciplinary Journal of Discontinuity, Nonlinearity, and Complexity, 10*(2), 173-184. https://doi.org/10.5890/dnc.2021.06.001

Almeida, F. (2020). Exploring the impact of socio-demographic dimensions in choosing a city touristic destination. *Journal of Tourism and Heritage Research, 4*(4), pp. 120– 142).

Andariesta, D. T., & Wasesa, M. (2022). Machine learning models for predicting international tourist arrivals in Indonesia during the COVID-19 pandemic: a multisource Internet data approach. *Journal of Tourism Futures*. https://doi.org/10.1108/jtf-10-2021-0239

Artley, B. (2022,). *Time Series Forecasting with ARIMA, SARIMA and SARIMAX*. Medium. https://towardsdatascience.com/time-series-forecasting-with-arima-sarima-and-sarimax-ee61099e78f6

Athanasopoulos, G., Hyndman, R., Song, H., & Wu, D. (2009). The tourism forecasting competition. *International Journal of Forecasting, 27*(3), 822 - 844. https://doi.org/10.1016/j.ijforecast.2010.04.009

Bi, J., Liu, Y., & Li, H. (2020). Daily tourism volume forecasting for tourist attractions. *Annals of Tourism Research, 83,* 102923. https://doi.org/10.1016/j .annals.2020.102923

Bi, J. W., Liu, Y., Fan, Z. P., & Zhang, J. (2019). Wisdom of crowds: Conducting importance-performance analysis (IPA) through online reviews. *Tourism Management, 70*, 460-478. https://doi.org/10.1016/j.tourman.2018.09.010

Botha, I., & Saayman, A. (2022). Forecasting tourism demand cycles: A Markov switching approach. *International Journal of Tourism Research, 24*(6), 759–774. https://doi.org/10.1002/jtr.2543

Bouhaddour, S., Saadi, C., Bouabdallaoui, I., Guerouate, F., & Sbihi, M. (2023). Tourism in Singapore, prediction model using SARIMA and PROPHET. *AIP Conference Proceedings*. https://doi.org/10.1063/5.013128

Bravo, J., Alarcón, R., Valdivia, C., & Serquén, O. (2023). *Application of Machine Learning techniques to predict visitors to the tourist attractions of the Moche Route in Peru*. MDPI. https://doi.org/10.3390/su15118967

Caldwell, R. (2023, June 28). *Finding Balance: Cultural Preservation and Tourism*. Chemonics International. https://chemonics.com/blog/finding-balance-cultural-preservation-tourism/

Chen, J. L., Li, G., Wu, D. C., & Shen, S. (2017). Forecasting seasonal tourism demand using a multiseries structural time series method. *Journal of Travel Research, 58*(1), 92-103. https://doi.org/10.1177/0047287517737191

Chen, X., Cao, J., Li, Y., & Chen, G. (2022). Predicting urban tourist flow patterns using a Recurrent Neural Network (RNN) model. *Tourism Management, 99,* 103889.

Chipumuro, M. & Chikobvu, D. (2022). Modelling Tourist Arrivals in South Africa To Assess The Impact of the COVID-19 Pandemic on the Tourism Sector. *African Journal of Hospitality, Tourism and Leisure, 11*(4),1381-1394. DOI: https://doi.org/10.46222/ajhtl.19770720.297

Christianingrum, C., Zukhri, N., & Rudianto, N. A. R. (2022). Place branding approach as an effort to optimize the image of the regency of Belitung and their implications for the decision to visit tourism destinations in lengkuas island. *Ijbe (Integrated Journal of Business and Economics), 6*(1), 1. Https://doi.org/10.33019/ijbe.v6i1.399

Claveria, Oscar & Monte, Enric & Torra, Salvador. (2015). Tourism demand forecasting with

neural network models: Different ways of treating information. *International Journal of Tourism Research, 17*, 492-500. 10.1002/jtr.2016

Derdouri A, Osaragi T, (2021). A machine learning-based approach for classifying tourists and locals using geotagged photos: the case of Tokyo. *Inf Technol Tourism, 23*(4), 575-609. Epub Sep 18. PMCID: PMC8449224. https://doi.org/10.1007/s40558-021-00208-3

Eom, T., Han, H., (2019). Community-based tourism (TourDure) experience program: a theoretical approach. *J. Trav. Tourism Market. 36* (8), 956-968. https://www.researchgate.net/journal/Journal-of-Travel-Tourism-Marketing-1540-7306

Faber, K., Kingham, S., Conrow, L., & Van Lierop, D. (2023). Differences in Active Travel Between Immigrants in an Active and Less Active Mobility Culture. *Urban Planning, 8*(4). https://doi.org/10.17645/up.v8i4.6977

Faridi, V. a. P. B. R. (2024). *Main determinants of tourism demand: Rashid's Blog: Portal for Inquisitive Learners*. https://rashidfaridi.com/2018/05/09/main-determinants-of-tourism-demand

Farsi, M., Hosahalli, D., Manjunatha, B., Gad, I., Atlam, E. S., Ahmed, A., Elmarhomy, G., Elmarhoumy, M., & Ghoneim, O. A. (2021). Parallel genetic algorithms for optimizing the SARIMA model for better forecasting of the NCDC weather data. *Alexandria Engineering Journal /Alexandria Engineering Journal, 60*(1), 1299-1316. https://doi.org/10.1016/j.aej.2020.10.052

Forsyth, P. & Dwyer, I. (2009). *The travel and tourism competitiveness report*

2009. http://www.weforum.org/pdf/ttcr09/Chapter%201.6.pdf, Fu, X., He, J., Zhang, L., & Cui, Y. (2015). Modelling and forecasting international tourism demand using time series methods: The case of Egypt. *International Journal of Tourism Research, 17*(2), 169-176. https://doi.org/10.1002/jtr.1972

GIZ (2022). Sector brief Namibia: Tourism. In *Sector Brief Namibia: Tourism*. https://www.giz.de/en/downloads/giz-2022-en-sector-brief-namibia-tourism.pdf

Gligorić, M., Božić, S., & Kovačević, M. (2019). Tourism and economic growth: comparative analysis by ARDL approach for selected countries. *Tourism: An International Interdisciplinary Journal, 67*(4), 419-434. https://doi.org/10.1177/13548166231219638

Hassan, S., & Bornmann, L. (2016). Time series methods for modelling and predicting international tourism demand: a case study of Egypt. *Tourism Management, 52*, 49-62. https://doi.org/10.1016/j.tourman.2015.06.016

Jammazi, R., & Aloui, C. (2019). Modelling and forecasting international tourist flows to Egypt using an autoregressive distributed lag model. *Tourism Management Perspectives, 31,* 212-222. https://doi.org/10.1016/j.tourman.2007.07.016

J. Rosselló-Nadal and J. He (2020). Tourist arrivals versus tourist expenditures in modelling tourism demand. *Tourism Economics, 26*(8), 1311-1326. https://doi.org/10.1177/1354816619867810

Juan, S., Fady, A., Giorgio, D. N., Anthony, P., & Philippe, R. (2024). *Design and implementation of the Luenberger observer for estimating the voltage response of a PEM electrolyzer during supply current variations.* https://doi.org/10.1109/access.2023.0322000

Kayral, H. E., Sarı, T., & Aktepe, N. a. T. (2023). Forecasting the tourist arrival volumes and tourism income with combined ANN architecture in the Post COVID-19 period: The Case of Turkey. *Sustainability, 15*(22), 15924. https://doi.org/10.3390/su152215924

Ke, W. (2024). Tourism demand forecast and future market trend research. *Advances in Politics and Economics, 7*(2), 202. https://doi.org/10.22158/ape.v7n2p202

Khan, N., Hassan, A. U., Fahad, S., & Naushad, M. (2020). Factors affecting tourism industry and its impacts on the global economy of the world. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3559353

Khare, P. (2024). *Understanding FB Prophet: A Time Series Forecasting Algorithm. Medium.* https://medium.com/illumination/understanding-fb-prophet-a-time-series-forecasting-algorithm-c998bc52ca10

He, K., Ji L., Wai C., Wu D., K. Fai, & Tso, G. (2021). Journal of hospitality and tourism management using the SARIMA – CNN – LSTM approach to forecasting daily

tourism demand. *J. Hosp. Tour. Manag. 49*, 25-33. https://doi.org/10.1016/j.jhtm.2021.08.022

Koushik, A. N., Manoj, M., & Nezamuddin, N. (2020). Machine learning applications in activity-travel behaviour research: A review. *Transport Reviews, 40*(3), 288-311. https://doi.org/10.1080/01441647.2019.1704307

Liço L., Enesi I. & Jaiswal H. (2021). Predicting customer behaviour using prophet algorithm in a real-time series dataset. *European Scientific Journal, ESJ, 17*(25), 10. https://doi.org/10.19044/esj.2021.v17n25p10

Lionetti, S., Pf, D., & Der Br, T.V. (2021). *Tourism Forec et ast with Weather, Event, and Cross-industry Data.* Proceedings of the 13th International Conference on Agents and Artificial Intelligence, 2, 1097-1104. doi: 10.5220/0010323010971104. https://www.scitepress.org/Papers/2021/103230/103230.pdf

Liu, Y., Wang, D., Li, X., & Bao, J. (2020). Predicting hotel occupancy rates using a Long Short-Term Memory (LSTM) neural network model. *Tourism Management, 81,* 104152. https://doi.org/10.3390/app112110291

Ma, E., Liu, Y., Li, J., & Chen, S. (2016). Anticipating Chinese tourists arrivals in Australia: A time series analysis. *Tourism Management Perspectives, 17*, 50-58. https://doi.org/10.1016/j.tmp.2015.12.004

Ministry of Environment, Forestry and Tourism. (2018). Tourist Statistical Report 2018. https://www.meft.gov.na/files/downloads/5c9_TOURIST%20STATISTICAL%20REPORT%202018%20No%20Write%20Up%20(1).pdf

Montes-Rojas, G. (2020). What are the empirical determinants of international tourist arrivals and expenditures? An empirical application to the case of São Tomé and Príncipe.

Policy Research Working Paper. World Bank. https://documents1.worldbank.org/curated/en/312691584467066838/pdf/What-are-the-Empirical-Determinants-of-International-Tourist-Arrivals-and-Expenditures-An-Empirical-Application-to-the-Case-of-Sao-Tome-and-Principe.pdf

National Planning Commission. (2023). *National plans*. https://www.npc.gov.na/national-plans/

Núñez, J. C. S., Gómez-Pulido, J. A., & Ramírez, R. R. (2024). Machine learning applied to tourism: A systematic review. *Wiley Interdisciplinary Reviews Data Mining and Knowledge Discovery*. https://doi.org/10.1002/widm.1549

Pahrudin, P., Chen, C. T., & Liu, L. W. (2021). A modified theory of planned behavioural: A case of tourist intention to visit a destination post-pandemic Covid-19 in Indonesia. *Heliyon, 7*(10), e08230. https://doi.org/10.1016/j.heliyon.2021.e08230

Panasiuk, A. (2023). *Tourism economics*. MDPI. https://www.econstor.eu/bitstream/10419/279848/1/1870158040.pdf

Pandey, P., Madhusudhan, M., & Singh, B. P. (2023). Quantitative Research Approach and its Applications in Library and Information Science Research. *International Journal of Nepal Library Association, 2*(01), 77-90. https://doi.org/10.3126/access.v2i01.58895

Qureshi, M. I., & Destek, M. A. (2019). Tourism Demand Forecasting Using ARDL Model: A Case of Singapore. *Journal of Travel & Tourism Marketing, 36*(1), 82-95. https://doi.org/10.1177/1354816618812588

Ramirez-Correa, P., Weiermair, K., & Santibañez-Gonzalez, O. (2017). Forecasting international tourism demand to Chile: the role of ARDL models in prediction under data uncertainty. *Journal of Travel Research, 56*(7), 971-984. https://doi.org/10.1016/j.tourman.2014.04.005

Russell, S., & Norvig, P. (2010). *Artificial intelligence: A modern approach* (3rd ed.). Pearson.

Sarker, I. H. (2022). AI-based modeling: Techniques, applications and research issues towards automation, intelligent and smart systems. *SN Computer Science, 3*(2). https://doi.org/10.1007/s42979-022-01043-x

Saunders, N. K., Lewis, P., & Thornhill, A. (2019). *Research methods for business students,* (8th ed.). Pearson Education.

Schonlau, M., & Zou, R. Y. (2020). The random forest algorithm for statistical learning. *The Stata Journal Promoting Communications on Statistics and Stata*, *20*(1), 3-29. https://doi.org/10.1177/1536867x20909688

Şeker, F. (2023). *Evolution of machine learning in tourism: A comprehensive review of seminal research*. ResearchGate. https://www.researchgate.net/publication/376548599_Evolution_of_Machine_Learning_in_Tourism_A_Comprehensive_Review_of_Seminal_Research

Siwek, M. (2023). Economic and social factors shaping consumer behaviour in era of the Covid-19 pandemic. *Journal of Applied Economic Sciences (JAES), 18*(16), 153. https://doi.org/10.57017/jaes.v18.3(81).02

Takyar, A. (2019). *AI use cases & applications across major industries*. LeewayHertz - AI Development Company. https://www.leewayhertz.com/ai-use-cases-and-applications/

Tovmasyan, G. (2023). Factors that influence domestic tourism demand: Evidence from Armenia. *Economics & Sociology, 16*(2), 11–25. https://doi.org/10.14254/2071-789x.2023/16-2/5

Tovmasyan, G. (2021). Forecasting the number of incoming tourists using Arima model: Case study from Armenia. *Marketing and Management of Innovations, 5*(3), 139–148. https://doi.org/10.21272/mmi.2021.3-12

Trang, H.L.T. (2019). *Inbound tourism market segmentation of the Andaman cluster, Thailand.* (Msc. Thesis, Prince of Songkla University). https://kb.psu.ac.th/psukb/bitstream/2010/5524/1/313716.pdf

Uula, M. M., Maulida, S., & Rusydiana, A. S. (2024). Tourism sector development and economic growth in OIC Countries. *Halal Tourism and Pilgrimage, 3* https://doi.org/10.58968/htp.v3i1.343

Wai, K. He, L. Ji, D. Wu, K. Fai, & Tso, G. (2021). *Journal of Hospitality and Tourism Management Using SARIMA – CNN – LSTM approach to forecasting daily tourism demand*. *J. Hosp. Tour. Manag. 49,* 25-33. doi: 10.1016/j.jhtm.2021.08.022. https://doi.org/10.1016/j.jhtm.2021.08.022

Woyo, E., & Amadhila, E. (2018). Desert tourists experiences in Namibia: A netnographic approach. *African Journal of Hospitality, Tourism and Leisure, 7*(3), 1-13.

Wu, D. C. W., Ji, L., He, K., & Tso, K. F. G. (2020). Forecasting tourist daily arrivals with a hybrid Sarima–Lstm approach. *Journal of Hospitality & Tourism Research, 45*(1), 52-67. https://doi.org/10.1177/1096348020934046

WWTC. (2020). https://wttc.org/research/economic-impact-2020

Yao, Y., & Cao, Y.m & Ding, X., & Zhai, J., & Liu, J., Luo, Y., Ma, S., & Zou, D. (2018). A paired neural network model for tourist arrival forecasting. *Expert Systems with Applications, 114*. 10.1016/j.eswa.2018.08.025.

Yotsawat, W., & Srivihok, A. (2016). *Thai domestic tourists clustering model using machine learning techniques: Case study of Phranakhon si Ayutthaya province, Thailand*. . https://www.researchgate.net/publication/301556438

Yu, N., & Chen, J. (2022, June 8). *Design of Machine Learning Algorithm for Tourism Demand Prediction. PubMed Central* (PMC). https://doi.org/10.1155/2022/6352381

Zhang, H., Li, Y., & Yang, Y. (2021). Predicting tourists' travel intentions using sentiment analysis and machine learning algorithms. *Journal of Travel Research,* 00472875211033338. https://www.econstor.eu/bitstream/10419/243089/1

Zhang, L., & Zhang, L. (2021). Tourism demand forecasting and tourists' search behaviour: evidence from segmented Baidu search volume. *Tourism Management, 86*, 102638. https://doi.org/10.1016/j.dsm.2021.10.002

Žunić, L., Pivac, T., & Košić, K. (2023). *The main components of the tourism infrastructure development*. In 7th International Thematic Monograph: Modern Management Tools and Economy of Tourism Sector in Present Era. https://doi.org/10.31410/tmt.2022-2023.393

Appendix A: NUST Ethical Clearance Letter

NAMIBIA UNIVERSITY
OF SCIENCE AND TECHNOLOGY

FACULTY RESEARCH ETHICS COMMITTEE (F-REC)

DECISION/FEEDBACK ON RESEARCH PROPOSAL

Dear Selma Shivute (200517597)

RESEARCH TOPIC: PREDICTING INTERNATIONAL TOURIST ARRIVALS IN NAMIBIA USING MACHINE LEARNING MODELS

Supervisor (if applicable): Dr Richard Maliwatu

Qualification registered for (if applicable): Master of Data Science

(Reference number of applications: **FACULTY RESEARCH ETHICS COMMITTEE REGISTRATION NUMBER: FREC-08/23)**

Re: Ethical screening application No: **FREC-08/23**

The Faculty of **Computing and Informatics** Ethics Screening Committee of the Namibia University of Science and Technology reviewed your application for the above-mentioned research. The research as set out in the application has been:

Approved	X

(Indicate with an X, and N/A if not applicable and proceed)

We would like to point out that you, as a researcher, are obliged to maintain the ethical integrity of your research, adhere to the ethical guidelines of NUST, and remain within the scope of your research proposal and supporting evidence as submitted to the F-REC. Should any aspect of your research change from the information as presented to the F-REC, which could affect the possibility of harm to any research subject, you are under the obligation to report it immediately to your supervisor or F-REC as applicable in writing. Should there be any uncertainty in this regard, you must consult with the F-REC.

We wish you success with your research, and trust that it will make a positive contribution to the quest for knowledge at NUST.

Any ethical issues that need to be highlighted?	Why are these issues important?	What must/could be done to minimize the ethical risk?
No	N/A	N/A

Recommendation: The application is approved.

Sincerely,

Dr Suama L Hamunyela
Chair: Faculty Ethics Screening Committee
Tel: +264-61-207-2922

CC: Co-supervisor: Dr Gloria Iyawa

NAMIBIA UNIVERSITY OF SCIENCE AND TECHNOLOGY
P/Bag 13388 Windhoek NAMIBIA
2023 -03- 30
OFFICE OF THE DEAN
FACULTY OF COMPUTING & INFORMATICS

Appendix B: NCRST Approval Letter

NCRST
NATIONAL COMMISSION ON RESEARCH SCIENCE & TECHNOLOGY

AUTHORIZATION OF RESEARCH PROJECTS

Authorization is hereby granted in terms of Section 21 of the RST Act No. 23 of 2004, to:

Name: Selma Shivute

Address: P.O. Box 26902, Windhoek,
Namibia

Coworkers: N/A

Certificate Number (if applicable): RCIV00032018 **Authorization No:** 202304000

Type of Research:
Non- Commercial research and the use of resources be limited to what is in the proposal.

Title of Research Authorized:
Predicting International Tourist Arrivals in Namibia Using Machine Learning Models.

Locality:
Ministry of Home Affairs, Immigration, safety and Security and Ministry of Environment, Tourism and Forestry.

Duration: 04 April 2023 - 30 April 2024

Research / Sample Collection Conditions:
See research conditions on the next page.

Yours sincerely,

National Commission on Research Science & Technology
04 APR 2023
P/Bag 13253
Windhoek, Namibia

Prof. Anicia Peters
Chief Executive Officer

Head Office:
Cnr. Louis Raymond & Grant Webster Street Olympia, Windhoek
Private Bag 13253, Windhoek
T +264 61 431 7000
F +264 61 216 531
www.ncrst.na
info@ncrst.na
Ncrst @NCRST_Namibia @ncrst.na

Innovation Hub:
Cnr Louis Raymond & Grant Webster Street Olympia, Windhoek
T +264 61 431 7099
F +264 61 235 788

RESEARCH/SAMPLE COLLECTION CONDITIONS

1. The applicant should share the final findings of the study with the Namibia University of Science and Technology, and National Commission on Science, Research and Technology.
2. The results emanating from the study should only be strictly used for the purpose outlined in the research proposal.

Appendix C: Acceptance Letter from MHAISS

REPUBLIC OF NAMIBIA

MINISTRY OF HOME AFFAIRS, IMMIGRATION, SAFETY AND SECURITY

Ministerial Head Quarters
Private Bag 13200
Windhoek

Tel: (061) 2922111
Fax: (061) 2922185

ENQUIRIES: ***Ms. S. Shapaka***
Ext. 2036 | Cellphone: 0813082077| E-mail Address: Selma.Shapaka@mha.gov.na

Our Ref.: S/5/1/1

Human Resources Matters/Confidential

Ms. Selma Shivute
P. O Box 26902
Windhoek

Dear Ms. Shivute

SUBJECT: DATA COLLECTION FOR RESEARCH PURPOSES AT THE MINISTRY OF HOME AFFAIRS, IMMIGRATION, SAFETY AND SECURITY

1. I hereby acknowledge receipt of your letter dated 06 March 2023 on the above subject matter.
2. Permission is granted for you to do research in the Ministry from the period 24 April 2023 to 5 May 2023.
3. Please note that the information collected will only be used for the academic purpose and it should be shared with the Ministry.
4. For any enquiries do not hesitate to contact Ms. Selma Shapaka, Learning and Development Officer at 0819510111 or Selma.Shapaka@mha.gov.na

Yours Sincerely

EXECUTIVE DIRECTOR 2023 -04- 14 Ministry of Home Affairs, Immigration, Safety & Security

.....
ETIENNE MARITZ
EXECUTIVE DIRECTOR

All official correspondence must be addressed to the Executive Director

My NAMIBIA MY COUNTRY, MY PRIDE