SOCIAL MEDIA SENTIMENT ANALYSIS FOR TOURISM PROMOTION IN NAMIBIA

By

Johannes Shoowa

219070490

Submitted in partial fulfilment of the requirements for the degree of

Master of Data Science

In the

Faculty of Computing and Informatics

Department of Informatics

at the

NAMIBIA UNIVERSITY OF SCIENCE AND TECHNOLOGY

Supervisor: Dr Richard Maliwatu

DECLARATION

I Johannes Shoowa hereby declare that the work contained in this thesis presented for the degree of Master of Data Science at the Namibia University of Science and Technology, titled: "***Social Media Sentiment Analysis for Tourism Promotion in Namibia***", is my original work, and that I have not previously, in its entirety or part, submitted it to any other university or higher education institution for the award of a degree. I further declare that I have fully acknowledged the sources of information used for the research by the institution's rules.

Signature: Date:

________________ 08/12/2025

Johannes Shoowa

METADATA

Study Title	Social Media Sentiment Analysis for Tourism Promotion in Namibia
Student Name	Johannes Shoowa
Student Number	219070490
Faculty	Faculty of Computing and Informatics
Department	Department of Informatics
Programme	Master of Data Science (09MADS)
Status	Dissertation
Specialisation	Machine Learning
Main knowledge area	Data Science
Keywords	Social media, Artificial Intelligence (AI), Machine Learning (ML), Sentiment analysis, Tourism
Type of research	Applied research
Methodology	Mixed Methods
Site	Windhoek
Document Date	08/12/2025
Supervisor	Dr Richard Maliwatu
First Co-Supervisor	N/A
Digital Signature of Student	
Digital signature of the Supervisor	
Digital Signature of Programme Coordinator (or equivalent)	
Digital Signature of FCI HDC Representative	

ABSTRACT

The rapid evolution of the Internet and advancements in information and communication technologies (ICT) have drastically changed how consumers share experiences and opinions, particularly in the tourism industry. Specialised social media platforms such as TripAdvisor now serve as primary channels through which tourists express their views via Electronic Word-of-Mouth (eWOM), significantly influencing the decisions of potential visitors. Despite Namibia's growing reputation as a tourist destination, there has been limited exploration into how tourists' opinions shared on specialised social media platforms can be systematically analysed to inform marketing strategies. Existing work in sentiment analysis has proven its value across multiple sectors, but its application within the tourism industry, particularly in Namibia, remains underexplored. This research builds upon previous studies that have utilised artificial intelligence (AI) and machine learning (ML) for sentiment classification and prediction but innovates by applying these techniques specifically to tourism promotion in Namibia.

This study investigates the application of sentiment analysis, utilising AI and machine learning techniques, to enhance tourism promotion in Namibia. Data was scraped from TripAdvisor, a specialised social media review platform, using Instant Data Scraper. Relevant reviews were filtered to retain only those directly related to Namibian tourism. Data pre-processing included cleaning, tokenisation, stopword removal, and standardisation to prepare the text for sentiment classification. Sentiment analysis was then conducted using Natural Language Processing (NLP) tools, namely the Valence Aware Dictionary and sEntiment Reasoner (VADER) and the RoBERTa model from Hugging Face. These tools were employed to classify user-generated content into positive, negative, and neutral sentiment categories to facilitate analysis.

The results from this study reveal insightful trends in tourist perceptions. Positive sentiments frequently highlight Namibia's natural beauty, wildlife, and unique cultural experiences as standout attractions. In contrast, negative sentiments often relate to issues such as inconsistent service quality, poor infrastructure in remote areas, and occasional dissatisfaction with customer service. These insights not only validate the value of applying sentiment analysis in this context but also uncover specific areas that require attention for improvement.

These findings, specifically the positive perceptions of Namibia's natural attractions and the negative feedback on service quality and infrastructure, can help tourism stakeholders craft targeted, data-driven marketing strategies. Additionally, a strategy that addresses areas of concern while amplifying the country's strengths can enhance Namibia's appeal as a top travel destination and contribute to the sustainable growth of its tourism industry.

DEDICATION

To my mother, Helena Shivute, whose strength and love shaped the person I am today; to God, for His unwavering guidance and grace; and to Isabel Nghinamupika, my partner, whose support and belief in me mean more than words can express.

ACKNOWLEDGEMENTS

I want to extend my sincere appreciation to all those who supported and contributed to the completion of this thesis.

I want to sincerely thank my Supervisor, Dr Richard Maliwatu, for his guidance, expertise, and encouragement. His insightful feedback and great encouragement were critical in helping me complete this research.

I sincerely thank the Namibia University of Science and Technology for providing the necessary resources and environment that enabled me to complete this research. Their support has been instrumental in bringing this work to completion. Special thanks also go out to the Faculty of Computing and Informatics and the Department of Informatics for their facilitation during the research.

Finally, I am grateful to all the participants and stakeholders in the tourism sector whose input and data made this research possible. Their contributions have been instrumental in achieving the study's goals. Thank you all for your support and encouragement.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ACRONYMS AND ABREVIATIONS

API	Application Programming Interface
AI	Artificial Intelligence
AUC	Area Under the Curve
BiLSTM	Bidirectional Long Short-Term Memory
BERT	Bidirectional Encoder Representations from Transformers.
CNN	Convolutional Neural Network
DMO	Destination Marketing Organisation
eWOM	Electronic Word-of-Mouth
k-NN	k-Nearest Neighbours
LDA	Latent Dirichlet Allocation
LSTM	Long Short-Term Memory
LLM	Large Language Models
ML	Machine Learning
MNB	Multinomial Naive Bayes
NUST	Namibia University of Science and Technology
NWR	Namibia Wildlife Resort
NLP	Natural Language Processing
NLTK	Natural Language Toolkit
RNN	Recurrent Neural Network
ROC	Receiver Operating Characteristic
UGC	User-generated Content
SA	Sentiment Analysis
SMI	Social Media Influencer
TAM	Technology Acceptance Model
TPB	Theory of Planned Behaviour
SVM	Support Vector Machine
VADER	Valence Aware Dictionary and sEntiment Reasoner

1. INTRODUCTION

In an era where millions of people actively engage with social media daily, these platforms have emerged as powerful channels for connecting with diverse audiences across the globe. Recognising this widespread usage, tourism promoters worldwide have tapped into the vast potential of social media to showcase destinations and attract international visitors (Waramontri, 2020).

This study explores how Namibia's tourism industry can leverage the influence of social media platforms to develop an effective promotional strategy that showcases the country's unique charm and natural beauty. By using sentiment analysis, tourism industry stakeholders can gain valuable insights into visitor experiences and preferences. This enables them to tailor promotional efforts more precisely and effectively (Wassenaar et al., 2013).

A data-driven approach such as this not only strengthens Namibia's appeal as a travel destination but also encourages more engaging and responsive interactions with potential tourists, ultimately enhancing the country's global tourism profile.

1.1 Background of the study

Namibia, a country rich in natural beauty and cultural heritage, is increasingly recognising the significance of social media in shaping travel decisions (Pohamba Shifeta, 2022a) .The diverse offerings of Namibia, from the stunning landscapes of the Namib Desert to the vibrant wildlife of Etosha National Park, serve as compelling attractions for tourists seeking authentic experiences (Marina L, 2013). However, in an era where digital engagement is paramount, traditional marketing strategies alone are insufficient to capture the attention of a global audience (Ministry of Environment and Tourism, 2016; Pohamba Shifeta, 2022).

The rise of tourists' review pages also presents significant opportunities for sentiment analysis Damar et al., (2024), a growing field within Artificial Intelligence (AI) that involves the use of natural language processing, text analysis, and computational linguistics to identify and extract subjective information from user-generated content (George & Ramos, 2024). In essence, sentiment analysis helps determine whether online opinions such as reviews, comments, or posts express positive, negative, or neutral sentiments (Mao et al., 2024a).

In the tourism sector, sentiment analysis plays a crucial role in understanding how travellers perceive destinations, attractions, and services (Flores-Ruiz et al., 2021a). By analysing feedback shared on platforms like TripAdvisor, Facebook, and Instagram, tourism stakeholders can gain deep insights into visitor experiences, expectations, and areas for improvement (Balamurali et al., 2023). This information enables more targeted marketing campaigns and service enhancements, ensuring that promotional strategies resonate with potential tourists.

For Namibia, where tourism is a vital economic sector (Puh & Bagić Babac, 2023a), harnessing sentiment analysis can help tailor messages that align with travellers' desires, enhance destination appeal, and strengthen the country's competitiveness in a fast-evolving global tourism market (Pohamba Shifeta, 2022b).

As the tourism sector plays a vital role in Namibia's economy (Ahmad Khan et al., 2024), stakeholders need to use innovative approaches that harness the power of social media. The ability to create an engaging online presence is crucial for effectively showcasing the country's unique charm and fostering connections with travellers (Pohamba Shifeta, 2022b). Furthermore, fostering a community through user-generated content can amplify word-of-mouth marketing, transforming satisfied visitors into passionate brand ambassadors.

This research, with its practical applications in sentiment analysis, is not just important but crucial for crafting effective marketing campaigns and enhancing the tourism experience. Furthermore, positive and unique experiences inspire visitors to advocate for their destinations (Pop et al., 2022), thereby fostering optimism about the potential of these places. Additionally, technological advancements offer many options for travellers to accurately plan their trips, covering everything from destinations and duration to tourist attractions, transportation, dining, and lodging. Various applications and websites facilitate extensive pre-planning for travellers. Social media communication often reveals individuals' thought processes and sentiments, enabling the classification of people based on their perspectives (Khan et al., 2020).

This study applied AI-based sentiment analysis to social media data to uncover key insights into tourist perceptions of Namibia. By analysing user-generated content, it identified common themes, preferences, and areas of concern among visitors. The findings offer concrete, data-driven guidance for tourism stakeholders to enhance marketing efforts, improve service delivery, and better align with travellers' expectations. In doing so, the study contributes meaningfully to both academic research and practical tourism development in Namibia.

1.2 Problem Statement

In the digital era, social media has become a powerful tool for destination marketing and promotion (Smith, 2020). Tourists increasingly use online platforms to share their travel experiences, shaping public perceptions of destinations and influencing potential visitors. However, accurately tracking and interpreting tourists' sentiments and opinions through social media remains a complex task, due to the unstructured nature of user-generated content, linguistic diversity, and context sensitivity (Flores-Ruiz et al., 2021b).

In the context of Namibia, a growing tourist destination known for its diverse natural and cultural attractions (Pohamba Shifeta, 2022b), there is limited empirical research that uses sentiment analysis to understand tourist experiences and perceptions. This creates a gap in using available online posts data to inform tourism promotion strategies. Without systematically identifying and analysing tourists' sentiments, tourism stakeholders may fail to understand visitors' expectations, address negative perceptions, or promote the aspects of Namibia most valued by tourists (Williams, 2021).

This presents a problem for tourism development in Namibia. The absence of data-driven insights hinders the design of effective marketing campaigns, misaligns offerings with visitor preferences, and limits the ability to build a resilient and competitive tourism sector. Understanding what drives tourists' sentiments is essential for enhancing promotional strategies and ensuring the sustainable growth of tourism (Garcia & Martinez, 2020) in Namibia.

1.3 The Aim of the Study

To enhance tourism promotion efforts in Namibia, it is essential to first analyse the sentiments and perceptions expressed by tourists on social media and web-based platforms. Insights gained from this analysis will inform the development of more effective and targeted marketing strategies. By understanding tourists' experiences and opinions, stakeholders can tailor their promotional efforts to better align with visitor expectations, ultimately strengthening Namibia's image and appeal as a travel destination.

Thus, this study is envisaged to contribute to tourism promotion in Namibia by leveraging sentiment analysis to uncover valuable insights that guide strategic planning and improve stakeholder decision-making.

1.3.1 Research Objective

The primary objective of this study is to apply sentiment analysis to social media posts related to tourism in Namibia to uncover actionable insights that can inform and enhance the country's tourism promotion strategies. This entails analysing tourists' online expressions, both in terms of sentiment (positive, negative, or neutral) and underlying perceptions, to assess how Namibia is currently viewed as a travel destination.

The study not only explores opportunities for improvement in general terms but also seeks to identify specific gaps and emerging trends in existing tourism promotion efforts. By leveraging data-driven insights, the research aims to provide targeted recommendations that align marketing strategies with tourists' expectations and preferences. In doing so, it offers practical guidance for tourism stakeholders seeking to strengthen Namibia's positioning in the global travel market.

Underlying this objective is the hypothesis that positive online sentiment significantly contributes to Namibia's perceived attractiveness as a travel destination. This proposition stems from the understanding that emotional expressions captured in social media reviews play a crucial role in shaping the perceptions and decision-making processes of potential tourists. Consequently, examining these sentiments provides valuable evidence on how online narratives influence destination appeal and can be strategically leveraged to enhance tourism promotion efforts.

This primary objective is further supported by a set of sub-objectives (as detailed in Table 1.1 below), which focus on identifying key sentiment drivers, assessing the effectiveness of current promotional strategies, and developing actionable recommendations for improvement. Together, these objectives form a structured framework that guides the direction and scope of this study.

Table 1.1: Research questions, objectives, and Hypotheses

Research Objectives	**Research Questions**
Main	
Apply sentiment analysis to better understand Namibia's tourism promotion strategies and explore improvement opportunities.	What are the predominant sentiments expressed by tourists in their online review sites posts about their visits to Namibia (Negative, positive, or neutral)?
Sub	
1. Identify the key factors that influence tourists' sentiments and perceptions during their trips to Namibia.	1. What key factors influence tourists' sentiments as reflected in their posts on online review sites during their trips to Namibia?
2. Derive insights from online reviews sentiment analysis to improve tourism promotion strategies in Namibia, ensuring they better align with the preferences and expectations of tourists.	2. How can we improve tourism promotion strategies using insights from online reviews sentiment analysis to better meet the preferences and expectations of tourists?
3. Assess the effectiveness of current tourism promotion strategies in shaping tourists' perceptions and sentiments as expressed on online review sites.	3. How effective are the current tourism promotion strategies in shaping tourists' perceptions and sentiments on online review sites?

1.3.2 Significance of the study

The study offers valuable insights into tourists' sentiments, as expressed in social media and online review site content related to tourism in Namibia, thereby aiding the development of more targeted and effective promotional strategies. Understanding these sentiments allows tourism authorities and marketers to tailor their promotional strategies, increasing positive perceptions and addressing negative ones, to enhance Namibia's overall destination marketing efforts.

In summary, this study demonstrates the value of sentiment analysis as a strategic tool for enhancing Namibia's tourism sector. By uncovering tourists' sentiments and preferences from online content, the research supports data-driven decision-making in policy development, marketing, and tourism product design. These insights not only improve the visitor experience but also contribute to Namibia's global competitiveness, sustainable tourism development, and economic growth.

1.3.3 Scope and Delimitations of the Study

This section outlines the boundaries of the research by distinguishing between its scope and delimitations. The scope of the study involves the analysis of tourist sentiments expressed in English-language social media posts and online reviews specifically related to tourism experiences in Namibia. The primary data source was publicly accessible user-generated content, with TripAdvisor selected as the main platform due to its relevance in the tourism sector and the practical ease of data extraction.

The delimitations refer to the intentional boundaries set to maintain the study's focus and feasibility, explicitly clarifying what the research does not address (Christina Lederle, 2020). Firstly, the analysis was confined to content written in English. Reviews and posts in other languages were excluded due to limitations in the Natural Language Processing methods used and the unreliability of automated translation tools for detecting sentiment nuances (Potter, 2024). Secondly, the data collection was restricted to selected online platforms, primarily TripAdvisor. Other platforms, even those containing relevant English-language content about Namibian tourism, were not included. Thirdly, the study relied solely on user-generated content (UGC) from these platforms, excluding other data sources such as official tourism reports, surveys, interviews, or traditional media. Fourthly, while sentiment analysis was applied, the research did not extend to deep cultural interpretation of language use, recognising a limitation inherent in automated analysis across diverse cultural contexts. Finally, the findings are context-specific and are not intended to be generalised beyond Namibian tourism. No effort was made to apply the results to other countries or tourism settings.

TripAdvisor was selected as the sole data source because it is the most comprehensive and structured tourism-specific review platform for Namibia, offering detailed, experience-rich narratives that are well-suited for sentiment and thematic analysis. Other platforms such as Twitter, Facebook, and Instagram, were excluded due to API access restrictions, high data acquisition costs, and the predominance of short-form or non-tourism-focused content, which limits their usefulness for advanced sentiment modelling. Concentrating on TripAdvisor ensured access to high-quality, context-rich user-generated content necessary for accurate sentiment classification and meaningful comparison across machine learning models.

1.4 Ethical Consideration

To uphold ethical standards, the study received official ethical clearance from the Namibia University of Science and Technology (NUST). Ethical considerations were strictly followed during data collection, ensuring user privacy and adherence to platform usage policies. The study also prioritised fairness, transparency, and accountability to support the responsible use of data and analysis methods.

1.5 Organisation of the Dissertation

The remaining sections of this dissertation are structured as follows:

- **Chapter 2: Literature Review**

Reviews relevant academic and industry literature on sentiment analysis, tourism marketing, and the use of social media data. It highlights key theoretical frameworks, identifies gaps in existing research, and justifies the study's focus and approach.

- **Chapter 3: Methodology**

Describes how this study was carried out, the research design (*Saunders' research onion*), tools, and techniques used to conduct sentiment analysis. It details data collection methods, the selection of platforms (e.g., TripAdvisor), the use of machine learning/NLP models, and ethical considerations.

- **Chapter 4: Implementation**

Explains the practical steps taken to apply sentiment analysis techniques, including data cleaning, pre-processing, model selection, and analysis execution. This chapter outlines the technical processes and challenges encountered during implementation.

- **Chapter 5: Results and Discussion of Findings**

Presents and interprets the findings of the sentiment analysis, including predominant tourist sentiments, key influencing factors, and insights into perceptions of Namibia as a travel destination. The results are discussed in relation to the research objectives and existing literature.

- **Chapter 6: Conclusions and Recommendations**

Summarises the key findings and contributions of the study. It also provides practical recommendations for tourism stakeholders, acknowledges limitations, and suggests directions for future research to support data-driven tourism development in Namibia.

- **References**

Contains all the scholarly sources, articles, and tools cited throughout the study, following the required referencing style.

- **Appendix**

Includes supplementary materials such as sample TripAdvisor reviews, data collection procedures, pre-processing code snippets, and additional results or charts supporting the main analysis.

2. LITERATURE REVIEW

2.1. Introduction

The application of sentiment analysis in tourism, particularly in using social media to promote tourism in Namibia, is an emerging field of study that bridges technology and human emotions (Firdaus & Afdholy, 2023). This chapter explores how AI-powered sentiment analysis can uncover valuable insights from tourists' online expressions and contribute to more targeted tourism strategies.

A systematic literature review was conducted using databases such as IEEE Xplore, NUST Digital Library, and Google Scholar, as well as relevant Namibian sources, including TripAdvisor data. The review prioritised recent publications (2018–2024) and case studies that applied sentiment analysis in tourism contexts.

Articles were selected based on their relevance to the topic, novelty, and data accuracy. The study's inclusion and exclusion criteria are as follows in Table 2.1 below:

Table 2.1: Research Inclusion and Exclusion Criteria

Inclusion Criteria	Exclusion Criteria
Articles and journal papers published between 2018 and 2024.	Publications before 2018, unless they are seminal works directly relevant to the study.
Studies focus on sentiment analysis, particularly those applying social media or online data to promote tourism.	Studies that do not focus on sentiment analysis, AI, ML, or their application in tourism.
Research involving AI, ML, LLM, NLP, SVM, VADER sentiment analyser, and RoBERTa.	The journal does not explicitly discuss sentiment analysis techniques.
Peer-reviewed journal articles, conference papers, reputable books, and authoritative digital repositories.	Sources that are not peer-reviewed or lack academic rigour include non-credible websites and non-reviewed conference abstracts.
Emphasis on practical applications and case studies demonstrating the effectiveness of sentiment analysis in tourism.	Duplicate studies or those offering no new insights beyond what is already covered by other sources.

Adhering to these inclusion and exclusion criteria, the literature review ensures a focused and relevant exploration of the intersection between sentiment analysis, social media, and tourism promotion in Namibia. This systematic approach helped identify the most relevant studies and provided a robust foundation for understanding how AI and ML can enhance tourism through sentiment analysis.

2.2. Theoretical literature

2.2.1. Sentiment Analysis Overview

The theoretical review outlines existing theories or concepts, explores their interrelationships, examines how they have evolved, and discusses the development of new hypotheses based on these advancements. Sentiment analysis, or opinion mining, refers to the computational identification and classification of emotional tone within text, typically categorised as positive, negative, or neutral (Huang, 2021). It uses a combination of NLP and machine learning techniques to extract subjective information from user-generated content such as reviews and social media posts.

Figure 2.2 below provides an overview of sentiment analysis, adapted from Ligthart et al. (2021). A holistic approach to sentiment analysis is necessary, as categorisation or classification alone is insufficient (Haque et al., 2022).

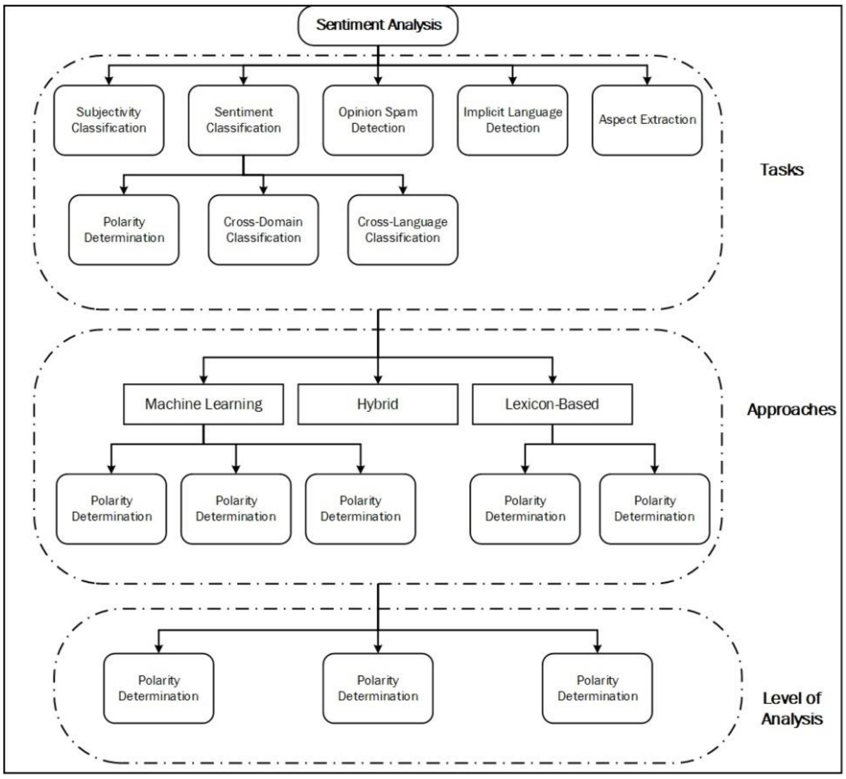


Figure 2.1: Sentiment analysis concept overview

Source: (Ligthart et al., 2021)

In addition, Sentiment scoring involves quantifying sentiment polarity into a numeric score (e.g., -1 to 1), which aids in analysing patterns across large datasets (An & Moon, 2022). In tourism, these metrics are invaluable for assessing visitor satisfaction, identifying areas for improvement, and guiding enhancements.

Similarly, sentiment analysis is defined by Zucco et al. (2020) as the process of extracting and analysing subjective information or opinions from textual data to determine the sentiments or opinions expressed by the author. These sentiments encompass a range of emotions, attitudes, evaluations, or opinions towards an entity, such as a product, service, brand, event, or issue. Sentiment analysis employs natural language processing methods to analyse and identify emotional attitudes contained in texts (Malviya et al., 2020).

However, sentiment and opinion are related constructs, and both are included when referring to either (Ligthart et al., 2021). This research adopts sentiment analysis as a general term encompassing both opinion mining and sentiment analysis.

Moreover, sentiment scoring, a crucial element of sentiment analysis, is the classification of polarity (Huang, 2021), which indicates the general sentiment of a text, phrase, or word. Polarity can be quantified as a "sentiment score," typically represented by a numerical value (An & Moon, 2022). For instance, this score might range from -100 to 100, where 0 signifies neutral sentiment. Sentiment scores can be computed for an entire text or specific segments (Aziz et al., 2023).

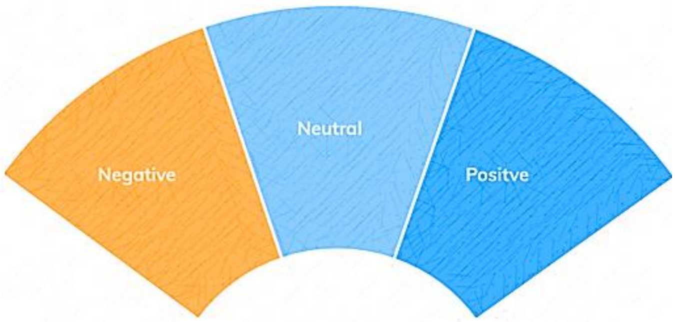


Figure 2.2: Represents a sentiment score expressed in various formats.

Source: Author

The goal is to understand people's feelings, opinions, and responses to a topic, entity, product, or service based on the text they generate, such as reviews, comments, tweets, and other online content.

2.2.2. Social media sentiment analysis

According to Alaei et al. (2019), social media sentiment analysis involves the computational study of opinions, sentiments, and emotions expressed on online platforms such as review websites, Twitter, Facebook, and Instagram. It utilises natural language processing (NLP) techniques to identify and extract subjective information from social media texts (Kang et al., 2020). The primary goal is to determine the overall sentiment (positive, negative, or neutral) towards a particular subject, which can provide valuable insights for various industries, including tourism (Alireza Alaei et al., 2023; Balamurali et al., 2023).

2.3. Sentiment Analysis Methods

Generally, sentiment (or opinion) refers to a way of thinking about something (Zucco et al., 2020). Sentiment analysis involves the computational study of sentiments, emotions, opinions, and attitudes regarding a specific topic, as expressed in texts or other media (Muley & Irfan, 2023).

In the Namibian context, where social media usage for tourism feedback is moderate compared to global trends, structured review platforms like TripAdvisor provide a rich and consistent source of sentiment-laden content. The application of lexicon-based methods, for instance, has proven effective in extracting sentiment (Alaei et al., 2019) from these English-language reviews without requiring extensive labelled training data. These methods can be enhanced by adapting them to local nuances, such as traveller expressions unique to Southern African English or specific to regional attractions.

As mentioned in Chapter 1, one reason for the significant interest in sentiment analysis research is the potential to develop various applications across various domains, such as product and service reviews, tourism, and medicine. Consequently, sentiment analysis research has evolved in parallel across different fields (Zucco et al., 2020), leading to an increased terminology that reflects refined differences in tasks, such as sentiment analysis, opinion mining, opinion extraction, sentiment mining, emotion analysis, review mining, and more.

Sentiment analysis is a broad concept encompassing various tasks, approaches, and types of analysis, which are explained in this chapter.

Sentiment analysis methods fall into three major categories:

1. **Rule-Based Methods**: These rely on lexicons and grammatical rules (Amalia et al., 2018). They are transparent but labour-intensive to design and maintain. Figure 2.3 below presents the rule-based method:

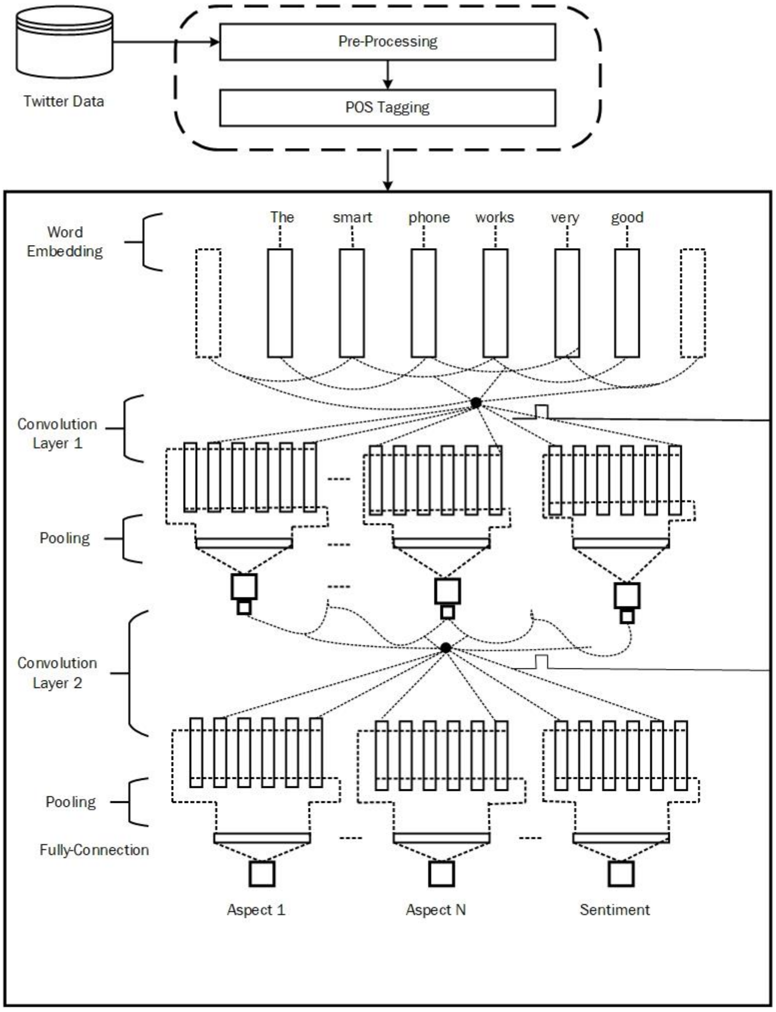


Figure 2.3: Rule-Based Method

Source: (Amalia et al., 2018)

Rule-based sentiment analysis relies on predefined linguistic rules and dictionaries to identify sentiment. This approach typically involves:

- **Lexicon-Based Techniques**: These techniques use sentiment lexicons to assign sentiment scores to words (e.g., positive, negative, neutral). The scores are aggregated to determine the overall sentiment of a text.
- **Grammatical Rules**: Applying syntactic and semantic rules to capture the context in which words appear facilitates understanding of negations and modifiers (Hilario et al., 2021).

Given the limited availability of large, labelled datasets in Namibia, rule-based and lexicon-based approaches (e.g., using VADER or SentiWordNet) can be beneficial for analysing reviews on platforms like TripAdvisor. However, the growing accessibility of pre-trained language models enables the application of transfer learning even in low-resource settings, such as Namibia, supporting more nuanced sentiment classification across regional tourism reviews.

Based on a study conducted by Amalia et al. (2018), the rule-based method in sentiment analysis relies on predefined or heuristic rules to classify texts according to their sentiments. This approach in data analysis utilises established rules or principles to categorise or make decisions about new data. It employs a logical framework where set rules are applied to recognise patterns within the data. These rules may range from simple logical statements to complex sequences of conditions and actions. They can be developed from domain expertise, practical experience, or learned from existing data through rule learning techniques. Implementing rule-based methods can also involve data mining to uncover relevant and significant rules from the data, which are then used to classify or make decisions about new data based on existing criteria (Steenwinckel et al., 2021).

Additionally, rule-based methods are known for their transparency and interpretability, but require significant effort in rule creation. These methods involve developing rules or heuristics to classify text based on its sentiments. The rules are crafted using human knowledge and understanding of language and sentiment (Mao et al., 2024b).

The primary advantage of rule-based methods is their transparency. Since the rules used in sentiment analysis are explicitly defined, the analysis results are explainable and understandable by humans. This enables users or domain experts to interpret and validate. However, there are some limitations to this approach. One major drawback is the substantial effort required to create the rules. Developing accurate rules that cover a wide range of cases demands considerable time and in-depth domain knowledge. Crafting comprehensive and detailed rules is a complex and time-consuming task.

Moreover, according to a study conducted by Krishnamoorthy (2018), rule-based methods may lack flexibility in handling variations and complexities in human language and sentiment. Any change in language or context might necessitate the recreation or adaptation of existing rules. Additionally, these methods may struggle with ambiguity and fail to grasp more complex contexts in human language.

2. **Machine Learning-Based Classification Methods:** These include algorithms such as Support Vector Machines (SVM), Naive Bayes, and decision trees. These models automatically learn patterns from labelled datasets (Mao et al., 2024b).

Machine learning techniques have become increasingly popular in sentiment analysis (Mao et al., 2024a) because they can learn from data and improve over time. Key methods include:

- **Supervised Learning**: Involves training algorithms on labelled datasets where the sentiment is already known. Standard algorithms include Support Vector Machines (SVM), Naive Bayes, and decision trees.
- **Deep Learning**: Techniques such as Long Short-Term Memory (LSTM) networks and Convolutional Neural Networks (CNNs) are used for more complex sentiment analysis tasks, particularly in handling large datasets and capturing contextual nuances (Mao et al., 2024a).

The classification-based method is a sentiment analysis approach that utilises machine learning algorithms to classify text based on the sentiments automatically expressed (Ligthart et al., 2021). This method, a subset of machine learning, classifies data into predetermined categories or classes by identifying patterns within the data. (Atqia et al., 2023)

Classification-based methods involve creating models or classifiers that examine the relationship between features in the data and the predefined class labels (Ligthart et al., 2021). These features can include numeric, categorical, or text attributes. The model learns from the training data to recognise patterns (Birjali et al., 2021a), classifying new, unlabelled data into the appropriate class, each offering distinct advantages and disadvantages depending on the data characteristics and the analysis objectives (Vali et al., 2020).

According to Pratiwi and Adiwijaya (2018), in sentiment analysis, classification-based methods utilise text data, converting it into features to classify the sentiment as positive, negative, or neutral (Mao et al., 2024b). Models trained on data with known sentiments can classify sentiments in new text data, making these methods highly effective for various applications such as document classification, spam detection, and pattern recognition. Their ability to automatically classify data facilitates more efficient and accurate decision-making across domains.

This method employs machine learning techniques to analyse patterns within a training dataset and then applies these learned patterns to classify new, unfamiliar texts (Nguyen et al., 2018). The classification-based approach offers several advantages:

- **Automatic Classification:**

This approach enables sentiment classification to be performed automatically without requiring human intervention. Machine learning algorithms can learn intricate patterns of sentiment, enabling the prediction of sentiment with speed and efficiency.

- **High Accuracy:**

Classification-based methods can achieve high accuracy in sentiment classification by leveraging advanced machine learning techniques. These algorithms can identify significant features that contribute to the development of precise classifiers.

- **Scalability:**

These methods efficiently handle large datasets. Machine learning algorithms trained on extensive datasets can effectively manage and process vast amounts of data.

However, there are also some drawbacks to classification-based methods:

- **Dependence on Training Dataset:**

Accurate classification results depend heavily on having a high-quality and representative training dataset. Gathering and annotating this data can be time-intensive and laborious.

- **Lack of Interpretability:**

Although these methods can provide accurate classification outcomes, the underlying sentiments may be challenging to interpret and understand. The complexity of machine learning algorithms often makes them difficult for humans to interpret directly.

3. **Deep Learning Models**: Models such as CNNs, RNNs, LSTMs, and transformer-based architectures (e.g., BERT, RoBERTa) are capable of capturing complex patterns in language (Damar et al., 2024b).

In all approaches, data pre-processing (tokenisation, stemming, stopword removal) is vital for preparing text for analysis (Nhlabano & Lutu, 2018).

A research paper conducted by Mittal et al. (2018) defined machine learning as an approach to sentiment analysis that uses machine learning techniques to examine and interpret textual data. This method involves training a model on a dataset with texts annotated for sentiment (positive, negative, or neutral). According to Xu et al. (2020), the model then learns to detect patterns and relationships between text features and their corresponding sentiments.

Moreover, classification algorithms such as SVM, Naive Bayes, and Random Forest (RF) are commonly used in machine learning-based methods to assign sentiment labels to text. The process includes extracting features from the text, such as keywords, phrases, or n-grams, and using a classification algorithm to identify sentiment-related patterns (Shahi et al., 2022).

- **Convolutional Neural Network (CNN)**

One notable machine learning model employed in sentiment analysis is the CNN. CNNs are particularly adept at processing spatial information (Puh & Bagić Babac, 2023), making them widely used in image processing. The convolutional layers in CNNs use connections from the preceding layer, where adjacent neurons connect to neurons in the following layer. This structure enables the layer to comprehensively understand the input data, enhancing the model's ability to analyse and interpret the information it receives. Figure 2.4 below provides a visual representation of how a CNN operates.

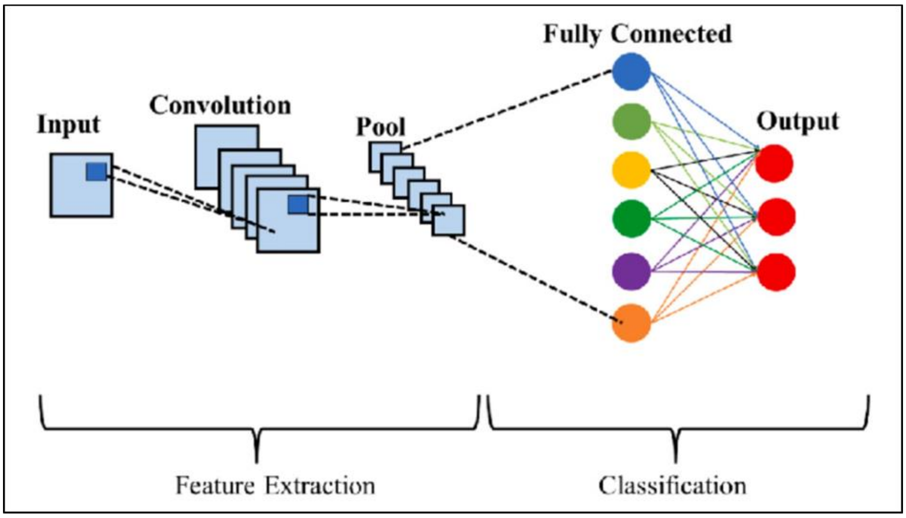


Figure 2.4: Convolutional Neural Network (CNN)

Source: (Damar et al., 2024b)

- **Recurrent Neural Networks (RNNs)**

RNNs, as illustrated in Figure 2.5 below, are utilised in sentiment analysis to incorporate previous computations and exploit sequential data (Xu et al., 2020). RNNs are particularly useful for Natural Language Processing tasks, such as predicting the next word in a sentence (Puh & Bagić Babac, 2023). They are adept at modelling the contextual relationships between words within sentences, making them practical for understanding and analysing textual data (Damar et al., 2024).

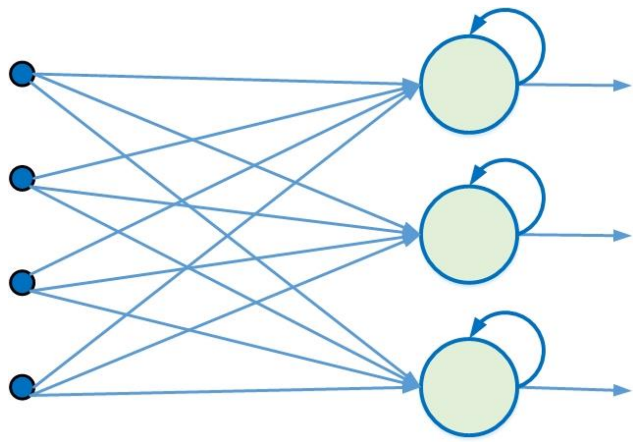


Figure 2.5: Recurrent Neural Networks

Source: (Dobilas, 2022).

- **Long Short-Term Memory (LSTM)**

LSTM networks are a specialised type of RNN created to address the issue of forgetting historical data in lengthy sequences (Balamurali et al., 2023). LSTMs utilise memory gates to preserve important historical information (Dobilas, 2022), preventing the data loss that can occur with inadequately trained RNNs. This capability allows LSTMs to analyse long sequences of data more effectively. Figure 2.6 below presents how an LSTM works.

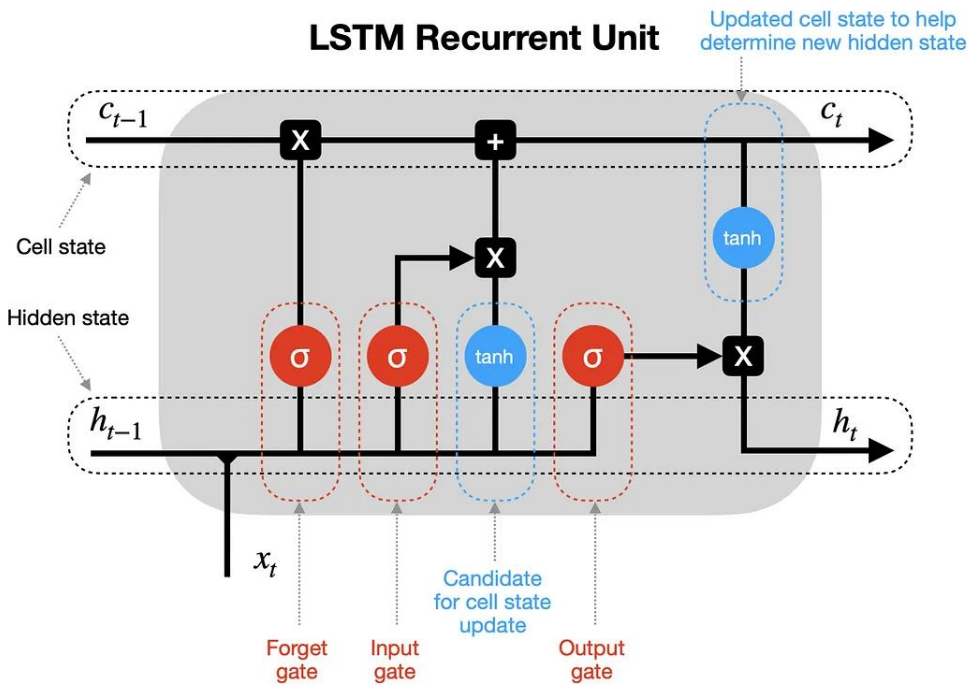


Figure 2.6: Long Short-Term Memory

Source: Dobilas (2022).

- **Bidirectional Long Short-Term Memory (BiLSTM)**

BiLSTM extends its capabilities by enhancing its performance in complex contexts (Puh & Bagić Babac, 2023), such as psychosocial analysis. It integrates forward and backwards hidden states, enabling information to flow in both directions. This bidirectional information flow enables BiLSTM to model better contextual relationships (Suhashini Chaurasia et al., 2021), resulting in improved performance (Puh & Bagić Babac, 2023) in tasks that require understanding both past and future contexts. The architecture of this model is presented in Figure 2.7 below.

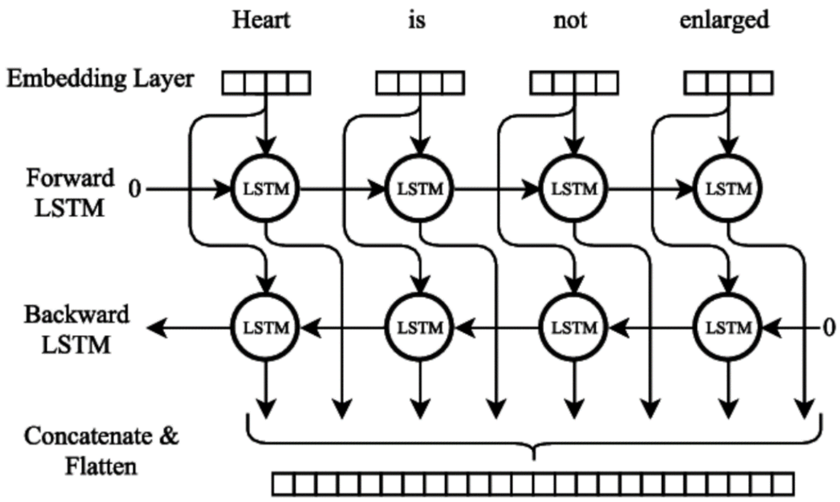


Figure 2.7: Bidirectional Long Short-Term Memory

Source: (Damar et al., 2024a)

Machine learning-based methods utilise a computer model trained on substantial data sets that include both the desired inputs and outputs (Muley & Zainab Irfan DrKirti Muley, 2023). The model then learns patterns in the data and establishes relationships between these inputs and outputs. Once the training process is complete, the model can predict or classify new, unlabelled data (Malviya et al., 2020). These methods offer several benefits for handling data complexity in sentiment analysis, including:

- **Flexibility:**

Machine learning-based methods are well-suited for handling complex data. They can identify intricate, non-linear patterns within textual data, enabling them to recognise subtle sentiments and nuanced contexts.

- **Generalisation:**

These methods can generalise from the training data to new, unseen data. Machine learning algorithms can accurately identify similar sentiments in new data by learning common patterns from the training data.

- **Scalability:**

Machine learning-based methods are suitable for large and complex datasets. These algorithms can handle high-dimensional and voluminous data, making the processing and analysis of complex data efficient.

However, machine learning-based methods also pose specific challenges:

- **Deep Understanding:**

Implementing these methods requires an in-depth understanding of modelling and data processing. Choosing the appropriate algorithm, efficiently processing data, and optimising parameters require substantial knowledge and experience.

- **Complex Data Processing:**

While capable of handling data complexity, these methods demand sophisticated data processing. This includes pre-processing tasks like text cleaning, removing stopwords, vectorising text, and selecting relevant features.

- **Dependence on Data Quantity:**

A significant amount of data is needed to develop an effective and accurate model. Limited training data can negatively impact the model's performance and generalisability (Liu, 2024).

Recent advancements in sentiment analysis have integrated these techniques with other technologies, such as natural language processing (NLP), social network analysis, and social media analytics (Birjali et al., 2021). This integration facilitates more advanced processing and a deeper understanding of user opinions and sentiments. Cutting-edge NLP techniques, particularly those based on transformer models, have enhanced the accuracy of sentiment analysis by comprehending more complex contexts (Rodríguez-Ibánez et al., 2023). Additionally, incorporating social network data allows sentiment analysis to consider social interactions and relationships (Gandhi et al., 2023). Social media analytics uses data from social media platforms to understand opinion distribution and user influence. Moreover, multimodal sentiment analysis, which combines text, images, and videos, has emerged, using deep learning techniques to achieve superior results in sentiment classification (Ligthart et al., 2021). This technological convergence comprehensively and precisely analyses user sentiments across various communication channels and media.

After outlining the foundational sentiment analysis methods, the following section focuses on how these techniques are applied within the tourism sector, particularly in the context of Namibia.

2.3.1. Applications of Sentiment Analysis in Tourism

Sentiment analysis is widely used for:

- Monitoring public perception
- Improving destination marketing
- Guiding infrastructure investment
- Supporting real-time crisis response (e.g., service failures)

Namibia-specific insights reveal that those positive sentiments centre on wildlife, landscapes, and cultural experiences (Marina L, 2013), while negative feedback often points to service quality or road conditions (Ndlovu, 2013).

The tourism industry can apply sentiment analysis in various scenarios. Monitoring social media platforms for tourist reviews and feedback helps understand the overall sentiment toward destinations, accommodations, and services. Rodríguez-Ibánez et al. (2023) Suggests using this information to make informed decisions and enhance tourism services.

A specific application of sentiment analysis is in the development of AI-driven chatbots and models for tourism (Muley & Irfan, 2023). These models can provide real-time information and support to tourists, enhancing their experience by responding appropriately to their emotional states (Balamurali et al., 2023). For example, during transportation disruptions or emergencies, sentiment analysis can help identify and address the concerns of affected tourists.

Tourism heavily relies on word-of-mouth and user-generated content (Pop et al., 2022), making sentiment analysis a crucial tool for understanding traveller perceptions and experiences. Studies have shown that social media sentiment can significantly influence the decisions and perceptions of potential tourists regarding destinations (Lisheng Weng et al., 2020). Social media platforms are rich sources of data for analysing tourist experiences, preferences, and behaviours (Xiang et al., 2017).

2.3.2. The role of social media influencers in tourism promotion is significant.

Sentiment analysis offers a feedback loop to improve promotional strategies. Based on the analysed sentiments, Maada et al. (2022) suggested that focusing on improving infrastructure, promoting cultural exchange programmes, and offering comprehensive travel guides can significantly improve tourists' perceptions and experiences. Moreover, leveraging user-generated content, such as shared pictures or stories on social media, can serve as organic promotional materials to attract more tourists (Sarah Anis et al., 2020).

Social media platforms serve as powerful tools for Destination Marketing Organisations (DMOs) to engage with potential tourists and promote destinations (Mirzaalian & Halpenny, 2021). Analysing and interpreting social media sentiment allows DMOs to tailor their marketing strategies effectively (Pop et al., 2022). Positive sentiments can enhance destination image and attractiveness, while negative sentiments can highlight areas needing improvement (B. Uugantsetseg, 2022).

Social media's ubiquity and integration into daily life have made it an invaluable source of data for tourism research (M. Kavitha et al., 2022). According to Yuan et al. (2022), travellers frequently share their experiences, complaints, and praises on platforms like Instagram, Twitter, and TripAdvisor, making it feasible to gauge the overall sentiment regarding a destination.

According to Lorgeoux and Divakaran (2023), Social Media Influencers (SMIs) play a significant role in shaping travel decisions. Trust in the content generated by SMIs has a positive impact on each phase of the travel decision-making journey, including the desire to travel, information search, evaluation of alternatives, purchase decisions, and sharing of experiences (Pop et al., 2022). Influencers are perceived as credible and trustworthy sources of information, which enhances their ability to influence followers' travel choices (Singh & Singh, 2021).

By understanding the predominant tourist sentiments, tourism authorities and marketers (B. Uugantsetseg, 2022) can tailor their promotional activities to meet the expectations and preferences of potential visitors. For instance, positive feedback can be leveraged to highlight Namibia's strengths, such as its natural attractions and cultural experiences, in marketing campaigns. Conversely, addressing common complaints or negative feedback in promotional materials and service improvements can mitigate potential deterrents for future tourists.

Effective use of sentiment analysis involves continuous monitoring and the implementation of adaptive strategies (Muley & Irfan, 2023). Real-time data from social media can inform immediate marketing responses and long-term strategic planning. For instance, upon detecting a spike in negative sentiments about a specific issue, tourism authorities can promptly address the issue and disseminate the resolution to the public, thereby upholding a positive image (Pop et al., 2022).

Evaluating the effectiveness of current tourism promotion strategies in shaping tourists' perceptions and sentiments on social media is crucial (Liu, 2024). A study by Johnson & White (2018) has shown that there is often a gap between the intended message of tourism campaigns and the actual perceptions formed by tourists. Various factors, such as the authenticity of the promotional content, the medium of promotion, and the alignment with tourists' actual experiences, can account for this discrepancy.

In Namibia, current tourism promotion strategies have been somewhat effective in attracting tourists (Williams, 2021), as evidenced by the positive sentiments expressed in social media posts. However, there are areas for improvement (Johnson & White, 2018). For instance, the emphasis on Namibia's unique natural and cultural attractions has been well-received (Marina L., 2013); however, issues related to infrastructure and service quality remain common themes in negative sentiments (Ndlovu, 2013). Enhancing the effectiveness of these strategies requires a more nuanced understanding (Romeo Asa et al., 2022) of tourists' experiences and a proactive approach to addressing negative feedback.

2.4. Empirical Literature

Empirical literature examines previous studies that have employed data-driven approaches to address specific research questions, providing evidence-based insights that inform and support the current study's objectives. Empirical studies examining the intersection of sentiment analysis and tourism promotion have increased significantly over the past decade, leveraging advancements in natural language processing (NLP), social media mining, and machine learning. These studies provide critical insights into how tourists express perceptions and emotions online, which in turn can inform more targeted and effective tourism strategies.

2.4.1. Sentiment Analysis in Tourism Contexts

Several studies have employed sentiment analysis on user-generated content to evaluate tourist experiences. For instance, Mao et al. (2024) analysed over 10,000 TripAdvisor reviews to assess visitor satisfaction across various tourist destinations in China. Their findings highlighted key sentiment drivers, such as cleanliness, hospitality, and cultural authenticity, and demonstrated how sentiment scores could predict return visit intentions.

In a similar study, Mariani and Borghi (2020) used machine learning algorithms to classify sentiment in hotel reviews and correlate it with service quality indicators. Their work confirmed that sentiment analysis can serve as a proxy for customer satisfaction and be integrated into tourism performance dashboards. Moreover, it underscored the importance of review platform data for real-time monitoring of tourist sentiment.

In the African context, empirical research is emerging but still limited, and social media sentiments toward tourism in East Africa have been investigated using Twitter data. The study employed VADER and Naïve Bayes classifiers to assess sentiment polarity and found that political stability and infrastructure were frequently discussed themes influencing perception.

Beyond classification, recent studies have applied topic modelling techniques, such as Latent Dirichlet Allocation (LDA), to extract dominant themes in tourism-related content. Hasan et al (2020) Applied LDA and sentiment scoring to TripAdvisor reviews of Malaysian attractions, uncovering nuanced patterns in tourist feedback that informed destination marketing priorities.

In terms of methodology, there is a trend toward hybrid models combining lexicon-based and deep learning approaches. For example, Liu (2024) compared VADER, TextBlob, and transformer-based models, such as BERT and RoBERTa, concluding that transformer models yield higher accuracy in context-sensitive sentiment tasks. Their study emphasises the relevance of model selection based on data characteristics and research objectives.

These empirical findings collectively demonstrate the potential of sentiment analysis in capturing real-time tourist perceptions and enhancing promotional strategies. However, a gap remains in applying these techniques to the Namibian tourism context. The current study addresses this gap by implementing both traditional (VADER) and modern (RoBERTa) sentiment analysis methods on TripAdvisor reviews, aiming to inform data-driven tourism promotion efforts in Namibia.

F. Mehraliyev et al., 2021; Xu et al(2020) Emphasise that the emotional aspects of travel have a significant influence on tourist decision-making and post-travel satisfaction. These studies utilised large datasets from review platforms, such as TripAdvisor, to extract sentiment trends, revealing that emotional expression is a strong predictor of destination loyalty and satisfaction.

Empirical studies have demonstrated that online reviews can serve as rich, real-time data sources for analysing tourist sentiment. Yuan et al. (2022a) used machine learning models to evaluate online feedback from hotels and destinations, identifying patterns in positive and negative experiences that helped refine tourism offerings. Similarly, Damar et al. (2024) applied natural language processing techniques to classify and quantify emotions from review data, reinforcing the value of text-based sentiment analysis in tourism analytics.

In the Namibian context, platforms like TripAdvisor host substantial user-generated content on key destinations such as Etosha National Park and Swakopmund. Although studies in Namibia remain limited, the available data offer a valuable empirical basis for understanding local tourism dynamics. Taherdoost & Madanchian.(2023) Emphasise the economic significance of tourism in Namibia and advocate for the strategic use of data analytics to harness user feedback for industry development.

Building on the application of sentiment analysis tools, Balamurali et al. (2023) investigated sentiment-aware chatbots in the tourism sector, demonstrating how real-time emotional detection can enhance user satisfaction and improve service interaction. Their study employed machine learning classifiers to recognise sentiment polarity and emotional cues, providing a functional prototype for AI integration in tourism services.

Additional studies provide insight into tourists' use of social media and review sites as platforms for sharing travel experiences. Flores-Ruiz et al.(2021) Found that online reviews influence consumer decisions and provide critical feedback to service providers. Pearce. P & Oktadiana. H. (2020) Confirmed the significance of such data in shaping tourist choices, while Christina Lederle (2020) emphasised the importance of mining these reviews to identify service strengths and weaknesses. With platforms like TripAdvisor housing over 884 million reviews in 2020 (Bandi Tanner & Hämmerli, 2018), the scope for empirical analysis is immense.

Puh & Bagić Babac, (2023) Applied sentiment classification to tourist-generated content, demonstrating that tourism stakeholders can use this data to improve services, inform targeted campaigns, and adapt offerings to better meet visitor expectations. Mao et al. (2024) further reinforced this by showing how sentiment analysis can guide precision marketing and enhance tourism strategies.

In terms of methodological advances, studies such as those by Gandhi et al. (2023) and Khan et al. (2020) provide the empirical foundation for the application of deep learning and machine learning models in sentiment analysis. These include algorithms such as Naïve Bayes, linear regression, and modern transformer-based models, which improve sentiment prediction accuracy across large text datasets.

Despite its broader application in fields such as marketing and finance (Ibrahim Sabuncu, 2021), sentiment analysis remains relatively underutilised in tourism, particularly in the African context. As Balamurali et al. (2023) suggest, sentiment-aware AI systems can transition from basic data processors to emotionally intelligent entities, capable of enhancing user engagement and satisfaction. This presents a compelling opportunity for empirical research focused on harnessing sentiment analysis for tourism promotion in Namibia.

2.4.2. Case Studies and Applications

A growing body of research has demonstrated the practical applications of online review sites and social media sentiment analysis in tourism promotion. For instance, Puh & Bagić Babac (2023) analysed

Twitter data to gauge sentiment towards major European cities, providing insights into tourist satisfaction and areas for improvement. Another study by F. Mehraliyev et al. (2021) examined TripAdvisor reviews to identify sentiment trends and their impact on tourism demand.

Additionally, Yilmaz (2016) examines how sharing travel experiences on social media platforms, such as blogs and forums, can impact potential tourists' perceptions and decisions. The study emphasises the role of storytelling in enhancing the tourist experience by making it more meaningful and memorable.

Finally, Khan et al. (2020) demonstrated the use of sentiment analysis to assess public sentiment on Twitter in response to government policies during the COVID-19 pandemic. This approach is particularly relevant for tourism, as it can help understand traveller sentiment toward evolving policies, such as travel restrictions or health protocols.

2.4.3. Challenges and limitations of Sentiment Analysis in tourism

Despite its growing application, sentiment analysis in tourism encounters various challenges that impact its effectiveness. One major challenge lies in the complexity of natural language; elements such as sarcasm, slang, idiomatic expressions, and context-dependent meanings often lead to misclassification of sentiments (Kang et al., 2020). Another challenge is the dynamic and evolving nature of social media language, which necessitates continuous retraining and updating of models to ensure (Nhlabano & Lutu, 2018).

In addition to these operational challenges, the study also identified several methodological limitations. Firstly, many sentiment analysis tools rely on predefined lexicons or machine learning models trained on generic datasets (Nguyen et al., 2018), which often struggle to interpret domain-specific terminology and cultural nuances present in tourism discourse (George & Ramos, 2024). Secondly, these tools are typically limited to sentiment polarity classification (positive, negative, or neutral), without providing insights into the emotional intensity or underlying motivations behind user sentiments (Dong & Shirai, 2023). This restricts the depth of analysis and limits the ability of tourism stakeholders to derive actionable insights (Puh & Bagić Babac, 2023b). Lastly, the exclusion of non-English content due to natural language processing constraints further narrows the scope and inclusivity of the findings.

Together, these limitations constrain the analytical value of sentiment analysis and highlight the need for more tailored, multilingual, and context-aware models to support effective tourism strategy and decision-making (George & Ramos, 2024).

2.5. Theoretical frameworks

2.5.1. Destination Image Theory and Tourist Perception Models

Destination Image Theory (Madden et al., 2016) provides a conceptual framework for understanding how travellers form perceptions and emotional associations with destinations. It distinguishes between three core dimensions: the cognitive image (knowledge and beliefs about a destination), the affective image (emotional responses or feelings), and the conative image (intentions or behavioural outcomes such as revisiting or recommending). In the context of this study, tourists' online reviews represent cognitive and affective expressions that collectively shape Namibia's destination image.

Similarly, the Theory of Planned Behaviour (Bosnjak et al., 2020) complements this by linking attitudes, subjective norms, and perceived control to behavioural intentions, such as visiting or revisiting a destination. By applying sentiment analysis (SA), this study interprets tourists' emotions and opinions to understand how online sentiment contributes to destination image formation and travel behaviour. Integrating these theories strengthens the conceptual foundation of the research and positions SA as a valid tool for analysing destination perceptions within the broader field of tourism behaviour.

2.5.2. Tourism Behaviour Theory

Tourism behaviour theory seeks to explain the motivations, attitudes, and decision-making processes that influence travel behaviour (Siri et al., 2020). It suggests that tourists are driven by both internal (psychological, emotional) and external (social, environmental) factors that shape their perceptions of destinations. Online reviews represent an extension of this behaviour, as they capture post-travel reflections and social validation processes. Sentiment analysis provides a data-driven means to interpret these behavioural expressions, revealing emotional drivers such as satisfaction, attachment, and loyalty.

2.5.3. Theory of Planned Behaviour

Bosnjak et al. (2020) Theory of Planned Behaviour (TPB) explains how attitudes, subjective norms, and perceived behavioural control influence individuals' intentions to act, in this case, to travel or recommend a destination. Online reviews can be interpreted as behavioural expressions of attitude and intention. Positive sentiments often translate into strong behavioural intentions (e.g., recommending Namibia as a travel destination), while negative sentiments reflect barriers or dissatisfaction. Sentiment analysis therefore, supports TPB by identifying attitude valence and social influence patterns embedded in tourist-generated content.

2.5.4. Technology Acceptance and Social Influence Frameworks

The rise of online travel platforms has introduced new dimensions to tourist decision-making, as described by the Technology Acceptance Model (TAM) (Schorr, 2023) and the Social Influence Theory (Dinara & Savvas, 2025). Tourists rely on peer-generated content to reduce uncertainty, validating destinations through collective experience. These frameworks suggest that user-generated reviews function as persuasive mechanisms that enhance the perceived usefulness and trustworthiness of destination information. Integrating these perspectives with sentiment analysis allows tourism researchers to map how emotional and informational cues from online platforms influence travel behaviour.

By combining insights from Tourism Behaviour Theory, Destination Image Theory, and the Theory of Planned Behaviour, this study establishes a conceptual foundation linking tourists' emotional expressions with destination perception and promotional effectiveness. These frameworks reinforce the interpretation of sentiment analysis results, ensuring that data-driven insights remain grounded in human behavioural understanding.

2.6. Gaps in the Literature

While extensive research has been conducted on social media sentiment analysis (Mukhopadhyay et al., 2022) and its application in tourism, a significant gap remains in studies that explicitly focus on Namibia. Most existing research centres on global destinations or broader regional contexts, leaving a lack of Namibia-specific sentiment insights derived from user-generated content.

Key gaps identified include:

- Lack of domain-specific sentiment models tailored to Namibian tourism data
- Limited inclusion of non-English reviews, which restricts full representation of tourist perceptions
- Underutilisation of emerging platforms and influencer-driven data, which are increasingly shaping travel decisions
- Insufficient exploration of deep qualitative insights, especially those revealing nuanced expectations and experiential perceptions of tourists

Despite growing international interest in applying AI/ML techniques to tourism analytics, no previous study has systematically applied both lexicon-based and transformer-based sentiment models to Namibia's tourism context. Existing studies either rely on generic sentiment tools, focus on other countries, or employ limited qualitative assessments that do not leverage the predictive and contextual strengths of modern NLP models.

This study addresses these gaps by providing the first comprehensive, AI-driven sentiment analysis of Namibia-specific TripAdvisor reviews. By integrating VADER, SVM, Naive Bayes, and RoBERTa, the research applies both traditional and state-of-the-art sentiment analysis methods to domain-specific data and links the derived insights directly to tourism promotion strategies. This contributes to a more robust, contextualised, and data-driven understanding of tourist perceptions in Namibia, thereby filling a clear gap in the literature.

2.7. Chapter Summary

Various techniques and approaches have been developed and tested in sentiment analysis, each with strengths and limitations. Rule-based methods offer transparency and interpretability, but require significant effort to create and maintain rules. Additionally, classification-based methods enable automatic classification with high accuracy but depend on extensive training datasets. In contrast, machine learning-based methods provide flexibility in handling data complexity but require deep data modelling and processing expertise. Recent advancements have integrated sentiment analysis with NLP, social network analysis, and social media analytics, enabling a more comprehensive and real-time understanding of user sentiments.

Finally, sentiment analysis of social media or online reviews data offers valuable insights for tourism promotion, as it reveals travellers' emotions and attitudes. In Namibia, using sentiment analysis can help tourism stakeholders understand public perception, improve customer experiences, and optimise marketing strategies. As the field evolves, continued advancements in sentiment analysis methodologies will further enhance its applicability and effectiveness in the tourism industry, offering a promising future for tourism promotion.

Table 2.2 below provides a summary of the key literature reviewed, including the methodologies employed and core findings from each study.

Table 2.2: Summary Table: Key Literature Reviewed

#	Author(s) & Year	Focus Area	Methodology	Key Findings
1	Huang (2021); Zucco et al. (2020)	Definition of sentiment analysis	NLP, ML overview	Emphasised polarity scoring and text emotion classification.
2	Kang et al. (2020).	Tourism sentiment complexity	NLP with social media data	Sarcasm and context limit classification accuracy.
3	Balamurali et al. (2023)	Tourism chatbots & sentiment	Deep learning, chatbots	Emotional tone improves tourist satisfaction.
4	Williams (2021); Ndlovu (2013)	Namibia-specific tourism reviews	TripAdvisor sentiment analysis	Identified common praise/complaint themes in Namibian tourism.
5	George & Ramos (2024)	Methodological limitations	Lexicon-based vs. domain models	Generic tools miss cultural nuances in tourism data.
6	Flores-Ruiz et al. (2021)	Social media review challenges	Comparative study	Real-time social data can be misleading without context-aware tools.

This synthesis highlights gaps such as the underrepresentation of cultural context in general-purpose sentiment models, and opportunities like the application of domain-specific tools in analysing Namibian tourism data.

The preceding chapter provided a theoretical and empirical foundation for understanding how sentiment analysis contributes to tourism research. The following chapter builds on this by describing the research methodology adopted to operationalise these concepts within the Namibian tourism context.

3. RESEARCH METHODOLOGY

3.1. Introduction

This chapter outlines the comprehensive methodology used to conduct sentiment analysis for tourism promotion in Namibia, guided by Saunders' Research Onion framework (Saunders et al., 2020), as illustrated in Figure 3.1 below. The methodology spans key elements, including the research philosophy, approach, strategy, data generation methods, and analysis techniques employed in this study.

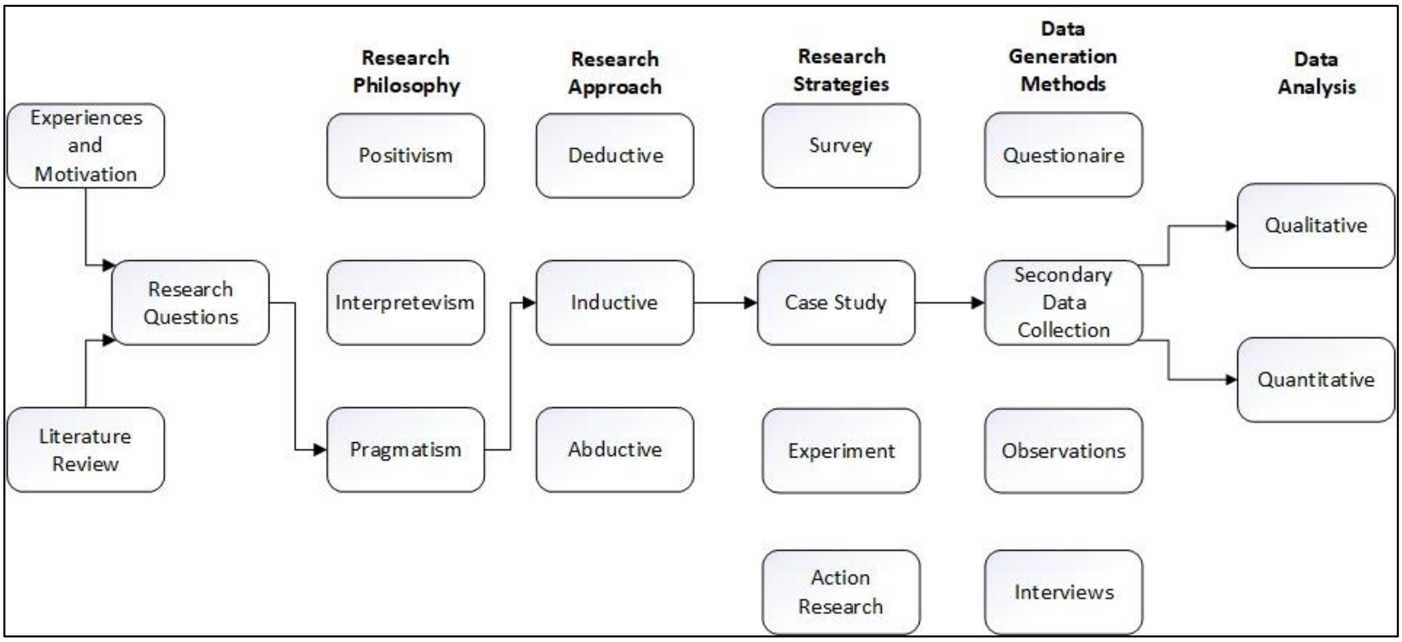


Figure 3.1: Model of the Research Process

Source: Adapted from (Saunders et al., 2020)

In this study, the term "social media" is specifically used to refer to online review platforms such as TripAdvisor. Although TripAdvisor is not a conventional social networking site like Facebook or Twitter, it embodies fundamental social media features such as user-generated content, peer-to-peer interaction, ratings, and comment threads. This clarification is crucial for ensuring the accurate classification of data sources and maintaining conceptual clarity throughout the research.

The study applies Natural Language Processing, lexicon-based techniques, and Machine Learning models to analyse tourist sentiments expressed in TripAdvisor reviews. The primary objective is to derive actionable insights that can enhance tourism marketing and strategic decision-making in Namibia.

The methodology involves multiple components: data collection, data pre-processing, sentiment analysis techniques, model evaluation, and a reflection on the underlying philosophical and strategic research decisions. Each methodological layer has been carefully chosen to align with the research aims and contextual realities of tourism in Namibia.

3.2. Research philosophy and paradigms

This study adopts a pragmatic research philosophy, which is particularly well-suited to applied research aimed at solving real-world problems. Pragmatism allows for the combination of both quantitative and qualitative techniques, offering methodological flexibility guided by the research questions rather than strict adherence to a single epistemological school of thought (Saunders et al., 2020). This philosophy is particularly suitable for tourism sentiment analysis, where numerical data and textual nuances are critical for understanding tourist experiences.

In the context of this study, the pragmatic approach supports the integration of sentiment scoring (quantitative) with contextual interpretation of tourist narratives (qualitative). The hybrid nature of tourist reviews, which comprises both measurable sentiment and expressive, subjective experiences, necessitates this dual methodological stance.

Paradigms in research refer to broader frameworks that guide how researchers view and interpret the world (Tengli, 2020). Aligned with pragmatism, the study adopts a mixed-methods paradigm, blending positivism and interpretivism to provide a more comprehensive understanding of tourist sentiments expressed on social media:

Each paradigm contributes uniquely to the research design and interpretation of findings. Positivism supports the quantitative aspect of the study, particularly the use of sentiment analysis algorithms (such as VADER and RoBERTa) to classify reviews as positive, negative, or neutral. From a positivist standpoint, social phenomena like tourist satisfaction can be measured objectively through empirical observation and statistical analysis. The application of machine learning models and natural language processing aligns with the positivist belief in observable, measurable reality, enabling the study to identify sentiment trends based on large volumes of textual data.

On the other hand, Interpretivism informs the qualitative dimension of the research. Since online reviews are rich in personal expression, context, and meaning, an interpretivist lens is necessary to understand the underlying themes, cultural references, and subjective experiences embedded in the text. This is especially relevant during topic modelling or thematic analysis, where the goal is to uncover the deeper meanings and narratives tourists attach to their experiences. Interpretivism acknowledges that human behaviour is complex and context-dependent, and therefore, the interpretation of tourists' sentiments cannot rely solely on objective measures.

Integrating both paradigms, the study leverages the strengths of each: the objectivity and generalisability of Positivism, and the depth and contextual understanding of Interpretivism. This philosophical pluralism enhances the validity and richness of the findings, making the study more robust in exploring how social media sentiments can inform tourism promotion strategies in Namibia.

3.3. Research Approach, Strategy, and Methodological Choice

3.3.1. Research Approach

This study adopts an inductive research approach, which is appropriate given the exploratory nature of the investigation. Inductive reasoning involves drawing general conclusions from specific observations, particularly when there is limited existing theory in the local context (Ligthart et al., 2021). In this case, the study focuses on user-generated content from a single social media platform, TripAdvisor, to analyse tourists’ sentiments about travel experiences in Namibia. Rather than testing a pre-defined hypothesis, the research seeks to identify emerging patterns, themes, and sentiment trends through qualitative and quantitative techniques such as sentiment classification and topic modelling. The aim is to develop insights into how online reviews reflect tourists' perceptions and how these sentiments might influence tourism promotion efforts. Starting with specific data and allowing the findings to inform broader conclusions, the study aligns with the principles of inductive reasoning, making it suitable for uncovering new understandings in the context of social media and tourism in Namibia.

3.3.2. Research Strategy

The study follows a case study strategy. According to Priya (2021), this strategy is especially valuable when the goal is to gain a deep understanding of a specific issue, event, or phenomenon within its real-world context. In the context of this study, it enables an in-depth, context-rich exploration of tourists' sentiments related to Namibia. The case study strategy is particularly suitable for sentiment analysis because it enables a focused examination of how tourists experience and evaluate Namibia as a destination through online reviews. This strategy supports a detailed investigation of both statistical sentiment trends and the contextual narratives behind tourist feedback.

3.3.3. Methodological Choice

The study employs a mixed-methods methodological choice, integrating both quantitative and qualitative techniques. Quantitative methods include sentiment scoring, classification algorithms, and statistical evaluation metrics such as precision, recall, and F1-score. However, Qualitative techniques are used to interpret emotional tone and thematic patterns in tourist reviews. This dual-method approach is consistent with the pragmatic research philosophy (Feilzer, 2010), adopted in the study, which supports the use of multiple methods to address the research questions most effectively.

The following, Table 3.1, presents a summary of the three key methodological layers, research approach, strategy, and methodological choice, highlighting the specific decisions made in this study and their alignment with the research objectives.

Table 3.1: Summary of the three methodological Layers

Source: Author

Layer	The Study's Choice	Justification
Research Approach	Inductive	Applying existing sentiment analysis models to Namibian tourism data
Strategy	Case Study	Focused, in-depth analysis of Namibia's tourism sentiment
Methodological Choice	Mixed Methods (Quant + Qual)	Combines statistical classification with narrative interpretation

3.4. Techniques and Tools

This section outlines the specific techniques and tools used throughout the research process, including data collection, data extraction, pre-processing, and sentiment analysis. These tools and methods were selected to support accurate, efficient, and contextually appropriate sentiment analysis for tourism-related content in Namibia.

3.4.1. Data Collection and Extraction

Data extraction involves utilising Application Programming Interfaces (APIs) provided by social media platforms (Atqia et al., 2023). For this study, data were collected exclusively from TripAdvisor, a user-driven travel review platform that functions similarly to social media through features such as user-generated content, peer-to-peer interaction, ratings, and narrative-style reviews. TripAdvisor was selected because it offers rich, experience-based textual content that is highly suitable for sentiment analysis and tourism research.

The data collection process involved web scraping using Instant Data Scraper, a browser-based tool that automates the extraction of public content from TripAdvisor review pages. A similar approach was used by (Yuan et al., 2022b) in extracting tourism-related content from platforms such as Instagram and TripAdvisor. In this study, scraping targeted reviews explicitly related to tourism in Namibia, using keywords such as "Namibia travel," "Namibia safari," "Etosha National Park," and other destination-specific search terms.

The dataset comprises reviews posted between January 2006 and April 2025. This extended timeframe was chosen to capture long-term sentiment trends and the evolution of tourist perceptions of Namibia as a destination. Analysing reviews across nearly two decades enables the identification of recurring patterns, shifts influenced by external events, such as the COVID-19 pandemic, and changes driven by tourism-related policy or infrastructure development. Including the most recent reviews ensures that the findings remain current and relevant for stakeholders.

After extraction, the raw dataset was filtered to retain only reviews that directly referenced Namibian destinations, attractions, or travel experiences. This ensured a focused, high-quality dataset aligned with the research objectives. After this filtering, the textual data underwent a series of pre-processing steps designed to enhance consistency and accuracy in sentiment analysis.

Other platforms, such as Twitter, Facebook, and Instagram, were not included because accessing their APIs requires paid subscriptions, and their content tends to be shorter, less descriptive, and less structured than TripAdvisor reviews. Such platforms generally contain brief posts, hashtags, or images rather than detailed narratives, limiting their suitability for advanced sentiment modelling. Given the constraints of accessing alternative platforms and the study's requirement for long-form, experience-rich textual data, TripAdvisor provided the most accessible, reliable, and contextually relevant dataset for this research.

3.4.2. The dependent variables

The sentiment polarity extracted from tourist reviews serves as the dependent variable in this study. It reflects the emotional tone of the reviews and enables the analysis of tourist satisfaction and perception. Following similar approaches in sentiment-based tourism research, categorising sentiment as positive, negative, or neutral provides valuable insights into tourist experiences (Mehraliyev et al., 2022)

3.4.3. Independent Variables

The independent variables consist of key attributes that may influence tourist sentiment. These include tourist type, country of origin, visit date, reviewed location, and the textual content of the review. As outlined in Table 3.2, these factors serve as the main explanatory variables in understanding variations in sentiment polarity.

Table 3.2: List of independent variables used in the study

Source: Author

Variable	Description	Type of Data
tourist_name	Username or identifier of the tourist who left the review	Categorical (String)
tourist_country	Location of the tourist (city and/or country)	Categorical (String)
topic	Title or brief headline of the review	Textual (String)
reviews	Full content of the review written by the tourist	Textual (String)
visit_date	Date when the tourist visited the place	Date
tourist_type	Type of traveller (e.g., Friends, Couples, Family)	Categorical (String)
reviewed_place	Name of the tourist attraction or place being reviewed	Categorical (String)
WeekDay	The day of the week on which the visitor visited the destination.	Categorical
Month	A month during which the tourist visited the destination.	Categorical
Year	Year in which the tourist visited the destination.	Date
PreprocessedReviews	Review text after initial pre-processing (e.g., lowercasing, stopword removal).	Textual (String)
CleanedReviews	Thoroughly cleaned review text ready for sentiment analysis.	Textual (String)

3.5. Data Pre-processing and Feature Engineering

Before conducting sentiment analysis, the collected textual data underwent a series of pre-processing steps to improve quality, remove noise, and ensure consistency in analysis. The following procedures were applied:

- **Text Cleaning**: Unwanted characters were removed from the text to reduce noise and irrelevant data. This included:
  - **Special characters** such as @, #, $, %, &, and *
  - **HTML tags**, emojis, and numeric symbols that did not add semantic value.
  - **Punctuation marks** such as commas, periods, question marks, quotation marks, and exclamation marks were stripped unless necessary for sentiment emphasis.
  - **URLs** and hyperlinks were removed using regular expression patterns (http\S+ and www\S+) to eliminate non-linguistic content.
- **Duplicate Removal**: Duplicate entries were identified and removed based on matching review text content and identical timestamps. Reviews with identical text posted more than once were retained only once in the dataset.

- **Tokenisation**: Each review was broken down into individual words (tokens) using Python's nltk.word_tokenize() function to allow further linguistic processing.
- **Stopword Removal**: Common English stopwords such as *"the," "is," "in,"* and *"of"* were removed using the NLTK stopwords list. This helped retain only the most meaningful content in each review.
- **Lemmatisation**: Words were reduced to their base or dictionary form using WordNet Lemmatiser (nltk.WordNetLemmatizer()). For example, *"travelling"* became *"travel"*, and *"better"* was normalised to *"good"*. Lemmatisation was preferred over stemming to preserve the grammatical structure of the words.

3.6. Sentiment Analysis

To enhance the analytical depth of this study, Latent Dirichlet Allocation (LDA) was incorporated as a topic modelling technique to identify dominant themes within the tourist reviews. This approach complements sentiment classification by revealing contextual drivers behind positive, neutral, and negative sentiments. LDA enabled the extraction of recurring topics such as wildlife, landscapes, accommodation, and infrastructure, providing thematic insight into what tourists discuss most frequently.

The sentiment analysis process employed two primary approaches: lexicon-based methods and machine learning classification techniques. These methods were used to categorise tourist reviews into three sentiment classes, positive, negative, or neutral, and to extract meaningful insights regarding perceptions of tourism in Namibia.

Lexicon-based methods classify sentiment using predefined word lists associated with specific sentiment polarities or emotional categories (Birjali et al., 2021b) In this study, the VADER and NRC Emotion Lexicons were used. VADER is tailored for social media-style and informal text, accounting for lexical features such as valence, punctuation, capitalisation, and degree modifiers to compute a compound sentiment score (Youvan, 2024a)). The NRC Emotion Lexicon maps words to eight fundamental emotions as well as general positive and negative sentiment. Reviews were tokenised, matched to lexicon entries, and assigned sentiment scores based on aggregated polarity values.

In addition to lexicon-based analysis, the study incorporated supervised machine learning models to enable more robust sentiment classification. The selection of VADER, SVM, Naive Bayes, and RoBERTa was guided by their complementary strengths and their relevance in prior tourism sentiment research. VADER served as an appropriate baseline due to its effectiveness in handling social media text. SVM and Naive Bayes were included as established machine learning classifiers well-suited for high-dimensional text data, commonly used in sentiment analysis of online reviews. RoBERTa, a state-of-the-art transformer-based deep learning model, was selected for its superior ability to capture linguistic context and nuanced sentiment, addressing many limitations of traditional models. Together, these techniques allowed for a comprehensive comparison across lexicon-based, classical machine learning, and advanced deep learning approaches to determine the most effective model for analysing Namibia-specific tourism reviews.

To complement the lexicon-based analysis, supervised machine learning models were trained on a manually labelled dataset. These models learned sentiment patterns from the data and were subsequently applied to classify unseen reviews. The classifiers used include:

- **Support Vector Machine (SVM):** A linear kernel SVM was implemented due to its robustness in handling high-dimensional text vectors. TF-IDF was used to convert text into numerical feature representations.
- **Naive Bayes Classifier:** A probabilistic model that estimates the likelihood of sentiment categories based on word frequencies (Puh & Bagić Babac, 2023). Its efficiency and simplicity make it widely used for text categorisation.
- **RoBERTa:** A transformer-based deep learning model fine-tuned for sentiment classification (Rodríguez-Ibánez et al., 2023). RoBERTa leverages contextual embeddings to understand meaning at the sentence level, making it especially effective for longer and more complex reviews.

All machine learning models were evaluated using standard performance metrics, including accuracy, precision, recall, and F1-score, and validated against a manually annotated subset of the dataset to ensure reliability and accuracy of predictions.

3.6.1. Sentiment Scoring

According to Chakraborty et al. (2021), sentiment scoring is the process of assigning numerical values to textual content to quantify its emotional tone, typically along a scale ranging from negative to positive. This technique is commonly used in lexicon-based sentiment analysis, where predefined sentiment dictionaries are used to evaluate individual words within a text.

In this study, sentiment scoring was applied exclusively to the lexicon-based approaches. After data pre-processing, each review was tokenised and its words were matched against sentiment lexicons such as AFINN, NRC Emotion Lexicon, and VADER. Each word carries an associated sentiment value:

- A positive score (e.g., *“amazing”* = +3),
- A negative score (e.g., *“terrible”* = –2), or
- Neutral/no score (e.g., *“view”*).

The overall sentiment score for a review is calculated by summing the values of all scored words. For example, a sentence containing the words *“amazing”*, *“view”*, and *“terrible”* would be scored as:

  - **+3 (amazing)** + **0 (view)** + **–2 (terrible)** = **+1 (slightly positive)**.

It is essential to note that this numeric scoring method is not employed in the machine learning approaches used in this study. Instead, machine learning models such as SVM, Naive Bayes, and RoBERTa classify the sentiment of an entire review based on learned linguistic patterns, without relying on fixed word-level sentiment scores.

Although VADER generates a compound score ranging from –1 to +1, this study opted to use only three primary sentiment categories; positive, negative, and neutral, because these discrete classes are required for training supervised machine learning models and for comparative model evaluation. Compound scores are continuous numerical values that are not directly compatible with multi-class classification frameworks like SVM, Naive Bayes, and RoBERTa. Standardising all models to the same three-class structure ensured consistent evaluation and allowed for direct comparison across all sentiment analysis techniques.

3.7. Model performance evaluation

The performance of the machine learning models used in this study was evaluated using standard classification metrics, including accuracy, precision, recall, and F1-score. These metrics measure the models' ability to correctly classify sentiment categories —positive, negative, or neutral —based on labelled data (Ramasamy et al., 2021). To ensure robustness and minimise the risk of overfitting, k-fold cross-validation was employed during model training and testing.

A manually labelled subset of the dataset was developed to serve as the validation set for assessing the performance of the machine learning classifiers: SVM, Naive Bayes, and RoBERTa. This facilitated the identification of the model that most accurately captured tourist sentiment from text data.

In addition to the machine learning models, the lexicon-based method VADER was also evaluated using the same classification metrics. Although VADER does not require model training, its sentiment predictions were treated as categorical outputs and compared against the manually labelled reference set. This approach enabled a consistent and fair comparison between VADER, and the machine learning models, thereby providing a comprehensive assessment of each method's effectiveness in analysing tourist sentiment.

3.7.1. Area Under the Curve (AUC)

In addition to standard classification accuracy, this study evaluated model performance using Precision, Recall, F1-score, and Area Under the Curve (AUC). The inclusion of AUC, derived from the Receiver Operating Characteristic (ROC) curve, provides a robust measure of a model's ability to discriminate between sentiment classes across different thresholds. AUC complements F1 by assessing the trade-off between true positive and false positive rates, thereby enhancing the reliability of comparative model assessment.

These metrics ensure that model performance is not overestimated by class imbalances, providing a balanced view of each algorithm's predictive capability.

3.7.2. Accuracy

Accuracy was measured as the proportion of correctly classified sentiments in the validation dataset. This metric provides a straightforward evaluation of how often the sentiment classification models, such as TextBlob, Naive Bayes, SVM, and RoBERTa, correctly predicted the sentiment of tourist reviews. The maximum accuracy score is 1 (100% correct predictions), and the minimum is 0 (no correct predictions). Mathematically, accuracy is calculated as:

Equation 3.1: Accuracy

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN}$$

Where TP and TN represent true positives and true negatives, respectively, and FP and FN represent false positives and false negatives.

3.7.3. Precision

Precision measures the proportion of optimistic sentiment predictions that were correct. This is particularly relevant in sentiment analysis when evaluating the model's ability to avoid falsely labelling a review as positive. A high precision indicates that the model produces few false positives and is reliable in identifying positive sentiment:

Equation 3.2: Precision

$$Precision = \frac{TP}{TP + FP}$$

This metric was crucial in assessing the reliability of the model when it comes to capturing favourable tourist experiences in Namibia

3.7.4. Recall

Recall, also known as sensitivity, reflects the model's ability to identify all actual positive cases correctly. In this study, it represents the percentage of truly positive reviews that were correctly identified by the model:

Equation 3.3: Recall

$$Recall = \frac{TP}{TP + FN}$$

A high recall indicates the model's ability to capture most of the positive sentiments expressed by tourists, which is valuable when ensuring that critical feedback about visitor satisfaction is not missed.

3.7.5. F1-Score

The F1-score balances the trade-off between precision and recall by computing their harmonic mean. It is beneficial in cases where the class distribution is imbalanced, such as datasets with more positive than negative reviews:

Equation 3.4: F1-score

$$F1 - score = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

An F1-score close to 1 indicates a strong performance in both identifying true positive sentiments and minimising false classifications, which is critical for extracting reliable insights from user-generated content (Haque et al., 2022).

3.7.6. Comparative Analysis

A comparative analysis was conducted to assess the performance and effectiveness of the lexicon-based approach, VADER, relative to the machine learning models (SVM, Naive Bayes, RoBERTa) used in this study. While machine learning models were quantitatively evaluated using standard metrics such as accuracy, precision, recall, and F1 Score, lexicon-based methods were assessed through a descriptive analysis of their sentiment outputs.

The comparison aimed to identify which approach was more effective in capturing sentiment nuances within tourism-related reviews about Namibia. Particular attention was given to how well each method handled polarity detection, contextual understanding, and consistency in classification across different types of reviews.

The comparative findings provided valuable insights into the trade-offs between interpretability, accuracy, and computational complexity, helping determine the most suitable sentiment analysis approach for tourism applications in resource-constrained or domain-specific contexts (Balamurali et al., 2023).

3.8. Chapter Summary

This chapter presented a detailed account of the research methodology employed to conduct sentiment analysis for tourism promotion in Namibia, guided by Saunders' Research Onion framework. The study adopted a pragmatic research philosophy, enabling the integration of both quantitative and qualitative methods to effectively capture the complexity of tourist sentiment expressed on online review platforms.

A deductive research approach was selected to apply and test existing sentiment analysis models within the specific context of Namibia's tourism sector. The study followed a case study strategy, focusing on TripAdvisor as a data source for understanding tourist experiences in Namibia. The mixed-methods choice enabled the combination of lexicon-based techniques and machine learning models, offering both interpretability and predictive power.

Data was collected through web scraping from TripAdvisor reviews posted between January 2007 and April 2025, ensuring a comprehensive dataset representing both historical and recent tourist sentiments. Pre-processing steps, including text cleaning, tokenisation, lemmatisation, and stopword removal, were systematically applied to prepare the data for analysis.

The implementation phase utilised Python and associated NLP and ML libraries, including NLTK, Scikit-learn, TextBlob, and Hugging Face Transformers. Sentiment classification was conducted using both lexicon-based tools, VADER, and machine learning models (SVM, Naive Bayes, and RoBERTa). The outputs of these models were evaluated using appropriate performance metrics (e.g., accuracy, precision, recall, F1-score), and a comparative analysis was conducted to determine the most effective approach for tourism sentiment analysis.

By following this structured methodology, the study ensures transparency, reproducibility, and analytical rigour. The findings derived from this approach are expected to support tourism stakeholders in making evidence-based decisions, enhancing marketing strategies, and ultimately improving the overall tourist experience in Namibia.

Having established the methodological framework and tools, the next chapter presents the implementation and results, highlighting how sentiment analysis models were applied and evaluated on real-world TripAdvisor data related to Namibian tourism.

The methodological framework presented in this chapter has established the foundation for the study's practical implementation. By combining quantitative and qualitative analytical methods, including lexicon-based, machine learning, and transformer-based approaches, the study ensures both analytical rigour and contextual relevance. The data collection, pre-processing, and sentiment classification methods described here provide a systematic blueprint for examining tourist perceptions of Namibia.

The subsequent chapter, Chapter 4 - Implementation, applies these methodological procedures to the TripAdvisor dataset, presenting the execution of the sentiment analysis models, data visualisation processes, and preliminary findings that inform the broader discussion of results.

4. IMPLEMENTATION

4.1. Introduction

This chapter outlines the practical execution of the research methodology presented in Chapter 3. It provides a step-by-step account of how sentiment analysis was operationalised using real-world data and computational tools. The chapter begins by describing the programming environment, software, and libraries used, followed by a detailed explanation of the data preparation techniques, including data cleaning, normalisation, tokenisation, and handling of duplicates.

Subsequently, the chapter details the model training and implementation process for both lexicon-based and machine learning approaches. This includes how the dataset was split, the application of algorithms such as Support Vector Machine (SVM), Naive Bayes, and RoBERTa, and the specific optimisation techniques used (e.g., cross-validation, hyperparameter tuning) to enhance model performance.

Finally, the chapter presents visual outputs and intermediate results, laying the foundation for the performance evaluation and comparative analysis in the next chapter.

Figure 4.1 below illustrates the structure and key processes covered in this chapter.

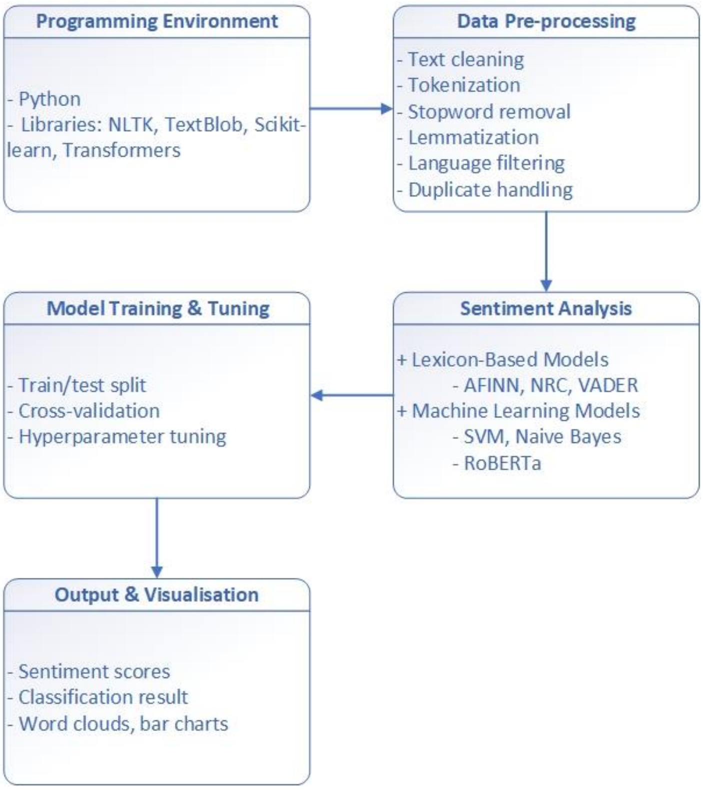


Figure 4.1: Implementation Flow Diagram

Source: Author

4.2. Data Collection

For this study, data were collected using Instant Data Scraper, a browser-based tool that enables the extraction of structured information from websites without requiring advanced programming skills. The dataset was sourced from TripAdvisor, one of the world's leading travel review platforms, to capture authentic tourist sentiments and experiences related to Namibia. Reviews were collected across various attractions, accommodations, and destinations within the country. The data spans a period from 2006 to 2024, providing a rich historical perspective on how tourist perceptions have evolved. This long-term dataset enabled both temporal analysis and the identification of recurring themes and sentiment trends, thereby offering a comprehensive foundation for sentiment analysis and insight generation.

4.3. Data Cleaning and Pre-processing

Data cleaning is the process of detecting, correcting, or removing inaccurate, inconsistent, or irrelevant data elements from a dataset (Pratama, 2024) to enhance its overall quality and reliability for analysis. In the context of sentiment analysis, data cleaning is a critical step that ensures the textual data is well-structured, semantically meaningful, and free from noise (Yonatan & Edwin R, 2024) that could skew model performance or mislead interpretation.

For this study, data cleaning was undertaken to standardise and refine the review texts collected from TripAdvisor. The process involved removing non-informative elements such as special characters, irrelevant punctuation, and duplicate records. These tasks were essential to prepare the data for accurate tokenisation, sentiment scoring, and machine learning classification.

The process was implemented in two phases: manual cleaning in Microsoft Excel, followed by advanced processing in Python.

4.3.1. Excel-Based Cleaning

The initial data cleaning was performed in Microsoft Excel immediately after data extraction. The following actions were taken:

4. **Renaming and Rearranging Columns:** Columns were renamed for clarity and consistency. Only the relevant fields required for sentiment analysis were retained.
5. **Splitting Combined Fields:** As part of the data pre-processing phase, it was necessary to extract the type of trip associated with each review to support further sentiment segmentation and trend analysis. The dataset included a field where the trip date and type were embedded in a single

string, for example, entries like "Oct 2016 â€¢ Couples". This format combined two pieces of distinct information: the visit date and the trip type. To separate them, the implementation process involved identifying the special delimiter "â€¢", which was consistently used to separate the two components. Using Excel, this delimiter was used to split the string into two parts: the left-hand side (e.g., "Oct 2016") was retained as the visit date, while the right-hand side (e.g., "Couples") was assigned to a new column called trip type. After extraction, the trip_type values were further reviewed and standardised into five main categories: solo, couples, family, friends, and business. This categorisation was performed manually to ensure consistency and resolve any typos or variations (e.g., "friends trip" vs "with friends").

6. **Removing Encoding Artefacts:** Artefacts such as "â€¢", caused by character encoding issues, were cleaned using Excel's Find and Replace tool.
7. **Standardising Trip Type Categories:** Values under trip_type were cleaned and categorised into five consistent types: *solo*, *couples*, *friends*, *family*, and *business*.
8. **Cleaning review_date and creating tourist_type:** Unwanted characters were removed from the review_date field, and embedded visitor types were extracted and stored in a new tourist_type column.
9. **Adding reviewed_place:** A new column was introduced to indicate the specific place or attraction that was reviewed.
10. **Checking for Duplicates:** A check confirmed that the dataset contained no duplicate entries.

4.3.2. Python-Based Cleaning

After the manual phase, the dataset was imported into Python for further refinement:

- **Managing the ID Column:** The existing review_id was dropped and replaced with a new sequential ID column ranging from 1 to 3670.
- **Standardising Tourist Countries:** The tourist_country field contained inconsistent formats. A custom dictionary (country_mapping) was used to standardise country names.
- **Handling Missing Values:** 15 missing values were found in the tourist_type column and replaced with the placeholder "Unknown."
- **Date Formatting:** The visit_date column was reformatted into a standard date and time format for consistency.
- **Generating Time-Based Features:** New columns (Weekday, Month, and Year) were created to support temporal analysis.

- **Pre-processing Review Texts:** The reviews column was cleaned by removing emojis and URLs. A new column called PreprocessedReviews was created. Stopwords were retained to maintain semantic context for text classification.
- Table 4.2 below presents a coherent summary of the data cleaning tasks implemented in this study, along with the techniques and tools used at each step.

Table 4.1: Data cleaning implementation details

Step	Action Area	Description
1	Column Renaming	Renamed columns for clarity and dropped irrelevant fields.
2	Splitting Fields	Split combined visit_date and trip_type values.
3	Encoding Artefact Removal	Removed unwanted characters like "â€¢" from text fields.
4	Standardising Trip Types	Grouped trip_type into five consistent categories.
5	Cleaning Review Dates	Removed 698 irrelevant characters from review_date; created tourist_type.
6	Adding Reviewed Place Column	Created reviewed_place to identify reviewed attractions.
7	Duplicate Check	Confirmed no duplicate rows in the dataset.
8	ID Reassignment	Dropped old ID and created a new ID column (1 to 3670).
9	Country Name Standardisation	Used dictionary mapping to clean tourist_country entries.
10	Handling Missing Values	Replaced 15 missing tourist_type entries with "Unknown."
11	Formatting Dates	Reformatted visit_date into proper date and time format.
12	Creating Temporal Features	Generated Weekday, Month, and Year columns.
13	Text Pre-processing	Removed emojis/URLs from reviews and created PreprocessedReviews.

4.4. Sentiment Classification Techniques

This section outlines the dual sentiment analysis approach employed in this study, comprising the lexicon-based VADER method and supervised machine learning classifiers, specifically Naïve Bayes, SVM, and RoBERTa. The integration of these techniques aimed to maximise both interpretability and predictive accuracy in classifying tourist sentiments from TripAdvisor reviews.

4.4.1. Lexicon-Based Method: VADER

The Valence VADER was selected as the lexicon-based tool due to its suitability for social media-style text, such as online reviews (Youvan, 2024b). VADER uses a predefined dictionary that maps lexical features to emotion intensities ((Samuel, 2022) and computes sentiment scores based on both the presence of words and contextual modifiers (e.g., negations, intensifiers).

Implementation Steps:

- The *SentimentIntensityAnalyzer()* class from Python's *nltk. Sentiment. The Vader* module was used.
- Each review text was analysed to compute four scores: *positive*, *neutral*, *negative*, and *compound*.
- The **compound score** (a normalised, weighted score ranging from -1 to +1) was used to assign sentiment classes:
    - **Positive**: compound ≥ 0.05
    - **Negative**: compound ≤ -0.05
    - **Neutral**: otherwise

This unsupervised method enabled the classification of all 3,670 reviews in the dataset without the need for manual labelling.

4.4.2. Supervised Machine Learning Models

To enhance accuracy and enable performance benchmarking, three machine learning models were also implemented: Naïve Bayes, SVM, and RoBERTa. These models required labelled data, necessitating a manual labelling process.

4.4.3. Data Labelling Process

A sample of 1000 reviews was randomly selected and manually labelled based on the overall tone, keywords, and sentiment polarity inferred from each text. Labels were categorised as:

- Positive
- Negative
- Neutral

This labelled dataset was then used to train and evaluate the machine learning models. The process ensured representativeness and reduced bias in model training.

4.4.4. Feature Engineering and Vectorisation

- For Naïve Bayes and SVM, the text data was vectorised using TF-IDF with the following settings:
    - Lowercasing enabled

    - English stopwords removed
    - Minimum and maximum document frequency thresholds set to filter rare and overly common terms
- The dataset was split into **training (80%)** and **testing (20%)** subsets.
- For **RoBERTa**, the roberta-base model from Hugging Face's transformers library was fine-tuned:
    - RobertaTokenizer was used to tokenise text.
    - Training was conducted for four epochs with a learning rate of 2e-5.
    - Early stopping and dropout regularisation were used to prevent overfitting.

4.5. Chapter Summary

This chapter outlined the end-to-end implementation of the sentiment analysis pipeline, focusing exclusively on TripAdvisor as the primary data source. Reviews were collected using the Instant Data Scraper tool, covering the period from 2006 to 2024. The implementation began with extensive data cleaning and pre-processing, including renaming and restructuring columns, handling encoding artefacts, standardising trip type categories, and deriving key variables such as visit_date, trip_type, and reviewed_place. Trip type information was extracted from embedded text in the review metadata, which was cleaned and categorised into five standard groups: solo, couples, family, friends, and business.

Further preparation included formatting date fields, mapping tourist countries for consistency, and generating new variables such as visit month, weekday, and year. A significant portion of the work involved text processing to remove noise (e.g., emojis, URLs) while preserving stopwords to retain context essential for sentiment classification.

The sentiment analysis approach combined lexicon-based and supervised machine learning techniques. VADER was used to produce initial polarity scores, which also informed the labelling of reviews as positive, neutral, or negative for training purposes. This labelled dataset enabled the training and evaluation of three machine learning models: SVM, Multinomial Naive Bayes (MNB), and the transformer-based RoBERTa model. While this chapter described the implementation process, model performance metrics and a detailed comparison are presented in the subsequent chapter.

Throughout the implementation, care was taken to contextualise the process within Namibia's tourism sector, ensuring that methodological decisions aligned with the practical realities of tourist experiences in the country. This chapter thus provides a strong technical foundation for analysing tourist sentiment and sets the stage for the presentation of results and key findings in Chapter 5.

5. RESULTS AND DISCUSSION OF FINDINGS

5.1. Introduction

This chapter presents and interprets the key findings from the sentiment analysis and machine learning techniques applied to TripAdvisor reviews related to tourism in Namibia. The primary aim is to assess how social media sentiment reflects tourist perceptions and to derive insights that can inform data-driven tourism promotion strategies. The results are aligned with the study's research objectives and are supported by visualisations and statistical summaries to enhance interpretability.

5.2. Dataset Description and Descriptive Analysis

5.1.1. Dataset Description

To provide context for the analytical results, this section begins with a summary of the dataset used in the study. The data comprises 3,671 user-generated reviews about tourism experiences in Namibia, collected from **TripAdvisor**. The dataset includes various attributes related to the review content, reviewer demographics, and metadata that inform sentiment classification and thematic analysis.

Table 5.1 below presents a summary of key features extracted from the dataset.

Table 5.1: Summary of key features extracted from the dataset

Feature (s)	Description	Sample Value(s)
tourist_name	Anonymised username of the reviewer	_christiane_domen_cd
tourist_country	Country and city of origin of the reviewer	Belgium.
topic	Title or subject line of the review	*Travel to Namibia*
reviews	Full text of the review describing the tourist experience	"We booked a visit to Kolmanskop..."
visit_date	Date of the tourist's visit (as stated in the review)	23-Aug-19
tourist_type	Type of traveller (e.g., Solo, Couples, Friends, Family)	Couples
reviewed_place	Tourist attraction, destination, or site reviewed	Okaukuejo waterhole

5.1.2. Descriptive analysis of the dataset

To provide context for the analytical results that follow, the structure and general characteristics of the dataset are first explored, as illustrated in Figure 5.1. This initial step involves examining variables such as review dates, tourist types, and countries of origin. In particular, the analysis of numerical variables reveals that the majority of review submissions occurred around 2017. Moreover, the dataset spans a broad timeline, containing reviews from as early as 2007 up to 2024. Notably, out of the 3,671 records, the median review date falls in July 2017, while approximately 75% of the reviews were submitted by April 2019. This temporal distribution highlights a concentration of reviews in the years immediately preceding the COVID-19 pandemic.

	Id	visit_date	Year
count	3671.000000	3671	3671.0
mean	1836.000000	2017-08-07 16:22:13.696540672	2017.074639
min	1.000000	2006-06-06 00:00:00	2006.0
25%	918.500000	2015-10-26 12:00:00	2015.0
50%	1836.000000	2017-05-23 00:00:00	2017.0
75%	2753.500000	2018-12-15 12:00:00	2018.0
max	3671.000000	2024-12-18 00:00:00	2024.0
std	1059.870747	NaN	2.761232

Figure 5.1: Comprehensive view of the reviews dataset

Source: Author

Additionally, a standard deviation of 2.76 years shows moderate variability in the distribution of reviews over time. The Id values, ranging from 1 to 1992 with an average of 996.31, indicate a well-distributed set of unique entries. This large span of both ID and Year underscores the temporal breadth and consistency of the dataset.

5.2.1. Interpretation of VADER Sentiment Polarity Statistics

In this subsection, the sentiment of the reviews was analysed by calculating each review's polarity, and then, based on the polarity, determining whether the review expressed a negative, positive, or neutral sentiment.

Figure 5.2 below summarises descriptive statistics for sentiment scores (*neg, neu, pos, compound*) generated using VADER on 3671 tourist reviews.

	neg	neu	pos	compound
count	3671.000000	3671.000000	3671.000000	3671.000000
mean	0.031726	0.791172	0.177110	0.676222
std	0.042306	0.095213	0.097838	0.422521
min	0.000000	0.354000	0.000000	-0.990800
25%	0.000000	0.733000	0.110000	0.598350
50%	0.017000	0.801000	0.166000	0.859300
75%	0.051000	0.857000	0.235000	0.948700
max	0.367000	1.000000	0.621000	0.999900

Figure 5.2: VADER Sentiment Polarity Statistics

Source: Author

VADER provides three probability scores indicating the likelihood that a text is negative, positive, or neutral (Samuel, 2022). Based on these probabilities, it calculates a compound score: a negative value for negative reviews, a positive value for positive reviews, and a score of 0.000 for neutral reviews.

The descriptive statistics derived from the VADER sentiment analysis provide a comprehensive overview of the emotional tone conveyed in tourist reviews. The results reveal a clear skew towards positivity, with a mean compound score of **0.676** and a median of **0.859**, indicating that over half of the reviews express strong positive sentiment. The average positive sentiment score is **0.177**, while the negative sentiment remains minimal with a mean of just **0.0317**. Interestingly, the neutral sentiment is the most dominant component, averaging **0.791**, which suggests that while many reviews are written in a neutral or factual tone, they still tend to conclude with a positive overall sentiment. The relatively low standard deviation across the polarity scores, especially in the negative and positive dimensions, reflects consistent sentiment patterns across the dataset. These findings imply that tourist experiences in Namibia are generally favourable, with minimal dissatisfaction expressed and a strong tendency toward positive reflections.

5.3. Research Objective 1: Uncovering Sentiments and Perceptions

To address the first research objective, *to identify the key factors that influence tourists' sentiments and perceptions during their trips to Namibia*, it is essential to begin with an exploratory analysis of the review data. This step serves to uncover broad patterns and temporal trends in the sentiments expressed by tourists, which provides contextual grounding for deeper analytical techniques applied later in the study. Furthermore, examining the distribution of sentiments over time and across various attributes such as tourist types or countries of origin, the analysis lays a foundation for identifying possible factors that contribute to tourists' emotional evaluations. The results of this exploratory phase guide the interpretation of more detailed findings, ensuring that the subsequent identification of influencing factors is rooted in the observed data patterns.

5.3.1. Exploratory Data Analysis: Trends and Patterns in Tourist Reviews

Figure 5.3 visually represents the number of reviews categorised by sentiment (Positive, Neutral, Negative) for each year from 2006 to 2024. The y-axis shows the number of reviews, while the x-axis lists the years. The stacked bars for each year display the proportion of positive (typically the most significant segment), neutral, and negative sentiments.

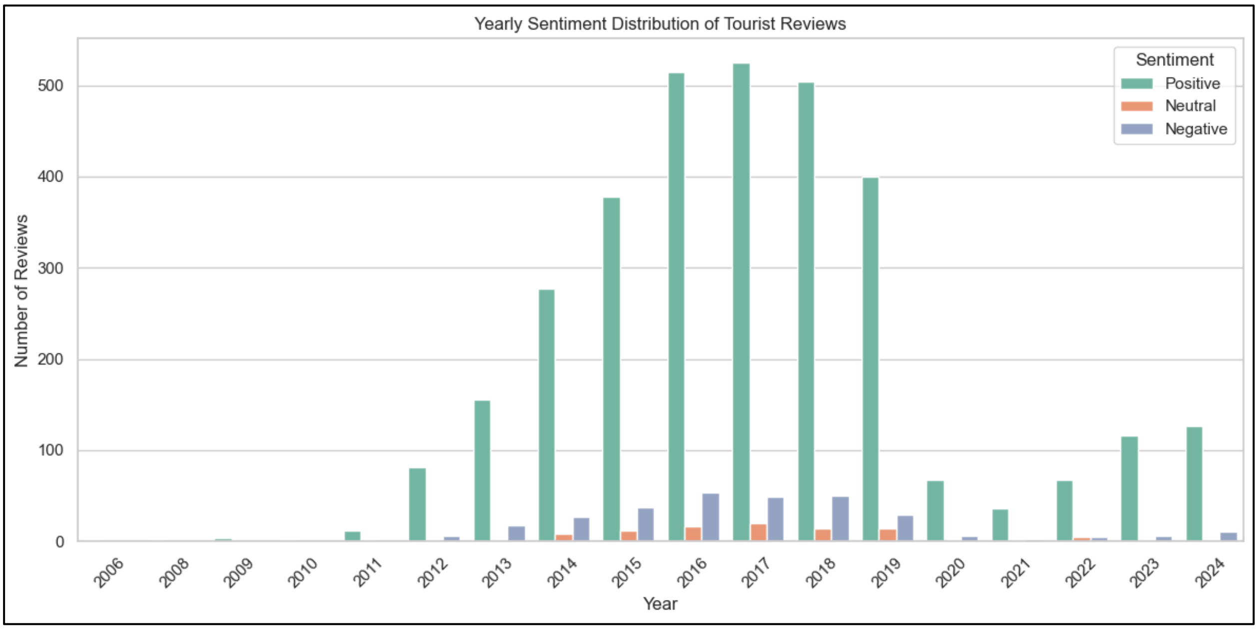


Figure 5.3: Yearly Sentiment Distribution of Tourist Reviews

Source: Author

Key observations:

- **Overall Trend:** Positive sentiments dominate across all years, with neutral and negative sentiments forming smaller portions of the reviews.
- **Peak Activity:** The years 2016–2018 show the highest number of reviews, with counts exceeding 500 in 2016 and 2017.
- **Decline in 2020-2021:** A notable decrease in review volume is observed during 2020 and 2021, likely due to the impact of the COVID-19 pandemic on global travel. (Firdaus & Afdholy, 2023) Similarly, a significant reduction in tourist activity was reported during this period, attributed to travel restrictions imposed across many countries.
- **Recent Recovery**: By 2022 and 2023, the number of reviews begins to rebound, although it has not yet reached pre-pandemic levels.

The plot highlights the resilience of positive sentiment despite external disruptions, suggesting consistent satisfaction among tourists.

Figure 5.10 visually presents the distribution of sentiment categories per recorded year. Table 5.4 below summarises the yearly breakdown of sentiment counts and their respective percentages, offering insight into the evolution of tourist perceptions over time.

Table 5.2: Summary of Sentiment Distribution by Year

Year	Positive Count	Positive %	Neutral Count	Neutral %	Negative Count	Negative %
2006	1.0	100.0%	0.0	0.0%	0.0	0.0%
2008	1.0	100.0%	0.0	0.0%	0.0	0.0%
2009	4.0	100.0%	0.0	0.0%	0.0	0.0%
2010	1.0	100.0%	0.0	0.0%	0.0	0.0%
2011	12.0	85.7%	2.0	14.3%	0.0	0.0%
2012	81.0	91.0%	2.0	2.2%	6.0	6.7%
2013	155.0	88.6%	2.0	1.1%	18.0	10.3%
2014	277.0	88.5%	9.0	2.9%	27.0	8.6%
2015	378.0	88.5%	12.0	2.8%	37.0	8.7%
2016	515.0	87.9%	17.0	2.9%	54.0	9.2%
2017	525.0	88.4%	20.0	3.4%	49.0	8.2%
2018	504.0	88.7%	14.0	2.5%	50.0	8.8%
2019	400.0	90.3%	14.0	3.2%	29.0	6.5%
2020	68.0	90.7%	1.0	1.3%	6.0	8.0%
2021	36.0	90.0%	1.0	2.5%	3.0	7.5%
2022	67.0	87.0%	5.0	6.5%	5.0	6.5%
2023	116.0	93.5%	2.0	1.6%	6.0	4.8%
2024	127.0	91.4%	1.0	0.7%	11.0	7.9%

The data in Table 5.2 reveals a consistently high proportion of positive sentiment across all recorded years, with values generally above **85%** and peaking at **93.5%** in 2023. This indicates a strong and possibly increasing level of satisfaction among tourists visiting Namibia. Negative sentiment has remained low and stable, rarely exceeding **10%**, indicating that dissatisfaction levels have not increased significantly despite the growing volume of reviews over time. Neutral sentiment consistently accounts for a minor share of the total, reinforcing the tendency of tourists to express clear positive or negative experiences. Overall, the results suggest a stable or slightly improving trend in tourist satisfaction, with no significant rise in negative sentiment, thereby reflecting positively on Namibia's tourism offerings over the years.

- **Top 10 Countries Represented in Namibia Reviews**

Figure 5.4 below illustrates the top ten countries by the number of TripAdvisor reviews related to tourism in Namibia. The United Kingdom leads with the highest number of reviews, followed closely by the United States and South Africa. These three countries contribute a substantial share of the total reviews, suggesting that a significant number of reviewers on TripAdvisor who have visited Namibia are based in these regions.

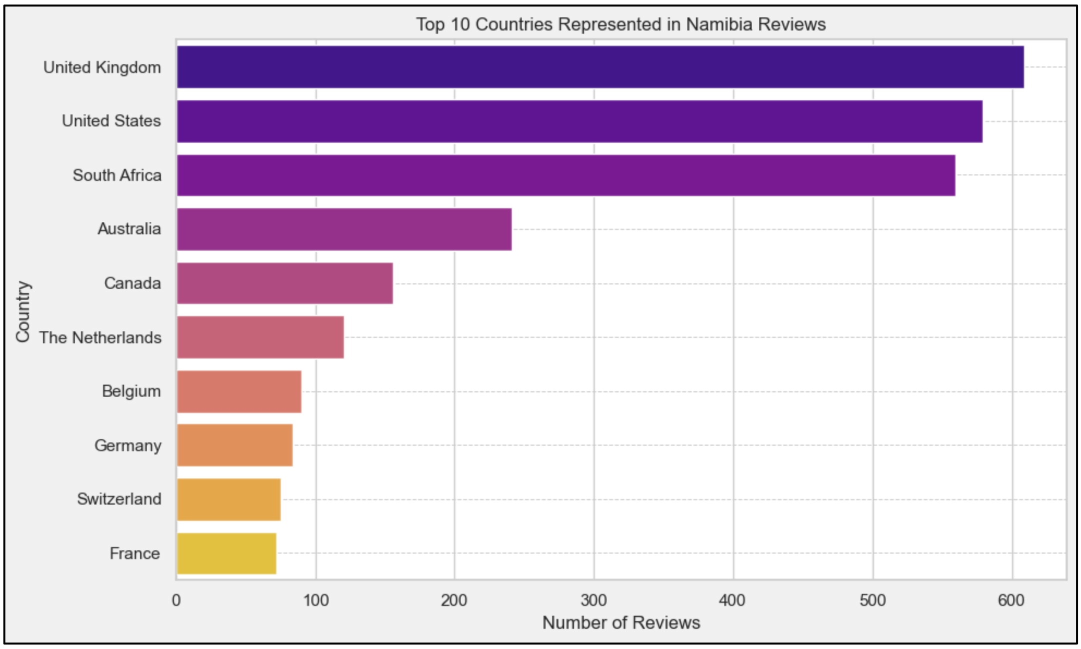


Figure 5.4: Top 10 Countries Represented in Namibia Reviews

Source: Author

The presence of Australia, Canada, and several European countries, including the Netherlands, Belgium, Germany, Switzerland, and France, further demonstrates a broad international interest in Namibia as a travel destination. While these countries show fewer reviews compared to the top three, they still represent meaningful levels of engagement.

A higher number of reviews from a specific country may suggest a relatively higher volume of tourists from that country, or at least a higher level of online engagement in sharing travel experiences. However, this metric does not directly reflect geographic proximity or the total number of visitors. It should be interpreted in the context of online behaviour, TripAdvisor user demographics, and market penetration in those regions.

These insights are valuable for tourism stakeholders in Namibia. Understanding the origin of online reviews can inform destination marketing strategies, enabling more targeted promotional efforts in countries that exhibit sustained or growing engagement. For example, the high review activity from countries such as the UK, the US, and South Africa signals strong outbound interest. At the same time, noticeable participation from Australia, Canada, and Western Europe indicates the presence of additional active tourism markets.

- **Distribution of Reviews Over the Years**

Figure 5.5 below is a bar chart that displays the number of reviews submitted for Namibia from 2006 to 2024.

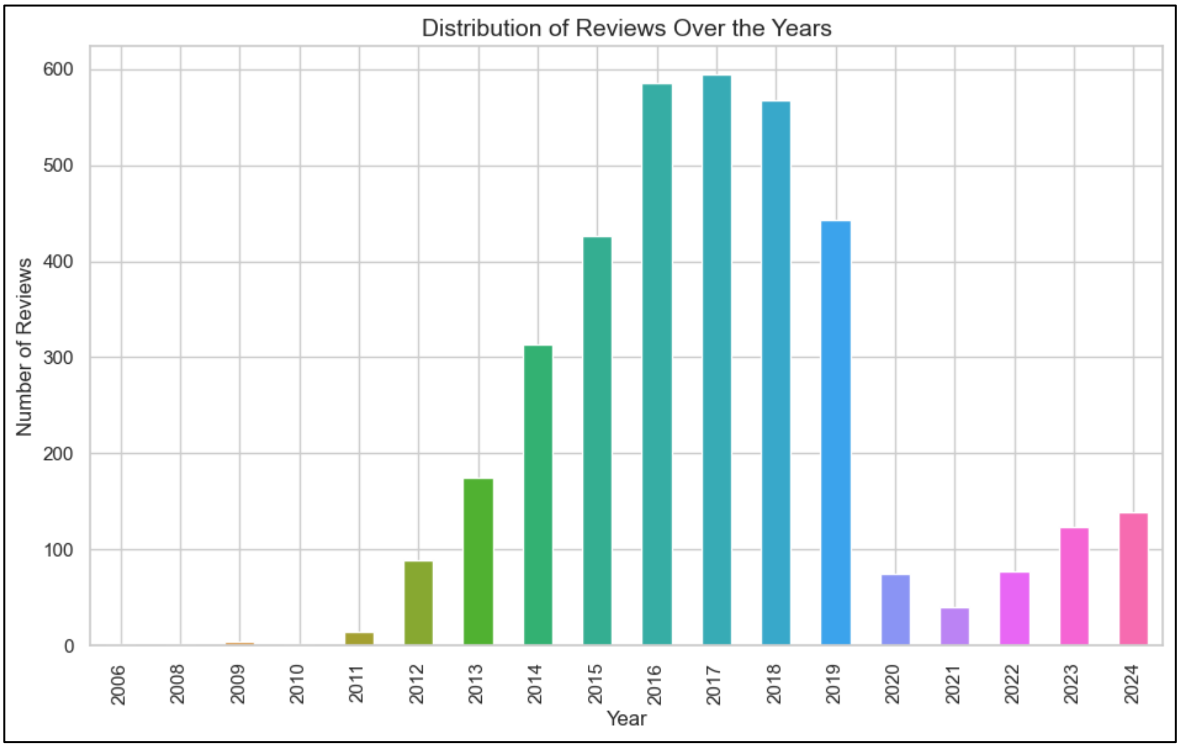


Figure 5.5: Distribution of Reviews Over the Years

Source: Author

The chart reveals a significant increase in reviews starting from 2012, with steady growth peaking between 2016 and 2018, where the number of reviews exceeds 500, with 2017 being the highest. After 2018, there is a noticeable decline in 2019, followed by a steep drop in reviews during 2020, likely due to the global pandemic (COVID-19), and lower activity in 2021.

However, the chart shows a gradual recovery beginning in 2022, with reviews steadily increasing again through 2023 and projected into 2024. This suggests a resurgence in tourism and interest in Namibia following the pandemic. The sharp rise in reviews in the earlier years highlights periods of significant tourism engagement, while the recent increase indicates a revival of interest in the country. This data is critical for understanding tourism trends and can guide future marketing and service improvements to sustain and enhance tourist feedback.

- **Annual Trends of Tourist Reviews from the Top 5 Countries Visiting Namibia**

While we are at it, let us narrow down the exploration to the number of reviews annually from the top 5 countries represented in Namibia. The x-axis represents the years from 2012 to 2024, and the y-axis shows the number of reviews. Each coloured line in Figure 5.6 represents a different country, as indicated in the legend. Below is an interpretation of the trends:

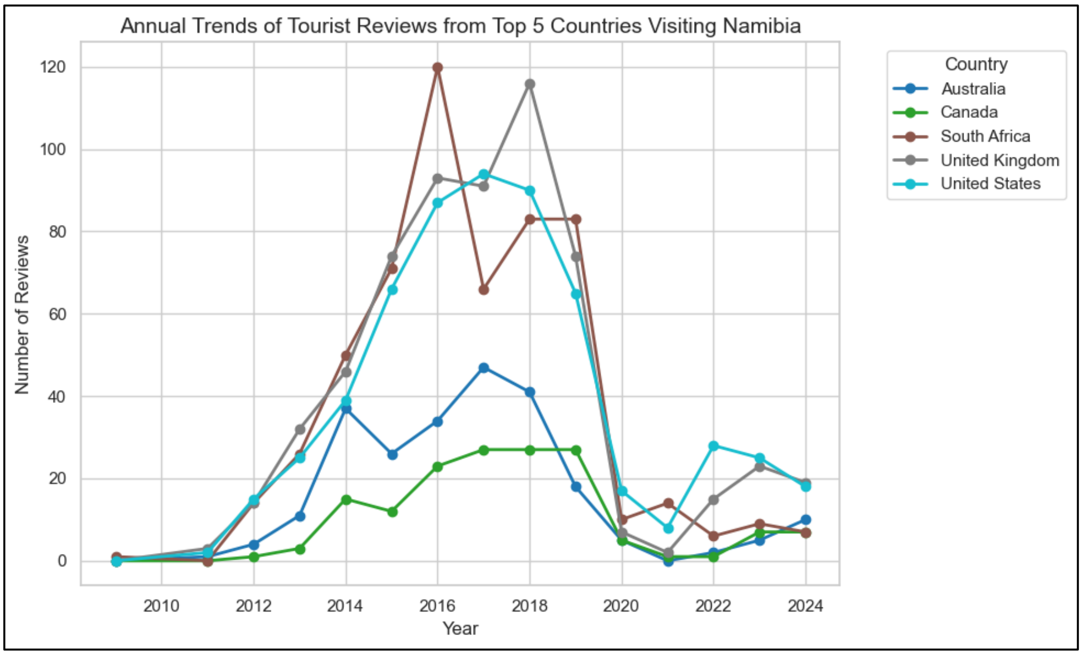


Figure 5.6: Annual Trends of Tourist Reviews from Top 5 Countries Visiting Namibia

Source: Author

The data spans from 2009 to 2024 and reveals several important patterns:

- **Peak Activity (2015–2018)**: All five countries experienced a sharp increase in review activity between 2013 and 2018. The United Kingdom reached its peak around 2018, with over 120 reviews, making it the most active reviewer group in the country. South Africa showed its highest activity in 2016, surpassing 110 reviews. The United States maintained strong and relatively stable growth during this same period, peaking just below 100 reviews around 2017–2018.

- **Post-2019 Decline**: A dramatic drop in review activity is observed for all countries after 2019, coinciding with the onset of the COVID-19 pandemic. This likely reflects the global slowdown in international travel and a corresponding decline in online tourist engagement during this period.

- **Recent Years (2022–2024)**: While overall volumes remain lower than pre-2020 levels, there are signs of gradual recovery:

4. The United States and the United Kingdom show a slight upward trend between 2022 and 2024.
5. Australia and Canada exhibit a modest return in review numbers, though still below their peak periods.
6. South Africa shows relatively stable but lower levels of activity, without a strong rebound.

- **Long-Term Growth**: Despite the decline after 2019, the long-term trend from 2009 to 2018 indicates growing interest and engagement from these countries. The consistency of review contributions from developed, long-haul markets (like the UK, US, and Australia) reinforces Namibia’s global tourism appeal.

- **Implications**:
- The data may serve as a proxy for measuring international interest in Namibia’s tourism offerings over time.
- Tourism stakeholders can utilise these insights to monitor recovery trajectories and refine marketing strategies in key source countries.
- The significant drop post-2019 underscores the pandemic’s impact on tourist behaviour and highlights the need for re-engagement campaigns, especially in countries showing early signs of rebound.

The two bar charts in Figures 5.7 and 5.8 below analyse the distribution of tourist visits to Namibia, categorised by day of the week and month of the year, based on the actual dates tourists reported visiting.

- **Visits by Day of Week:**
  - Wednesday sees the highest number of visits (nearing 600), suggesting it may be a preferred day for travel transitions (e.g., midweek arrivals/departures) or linked to tour schedules.
  - Saturday, Friday, and Monday also show vigorous activity (between 300–500 visits), likely due to weekend getaways or extended holiday travel.
  - Sunday and Thursday have slightly fewer visits, possibly because Sunday is a quieter travel day (return trips) and Thursday may be a lull before weekend trips.

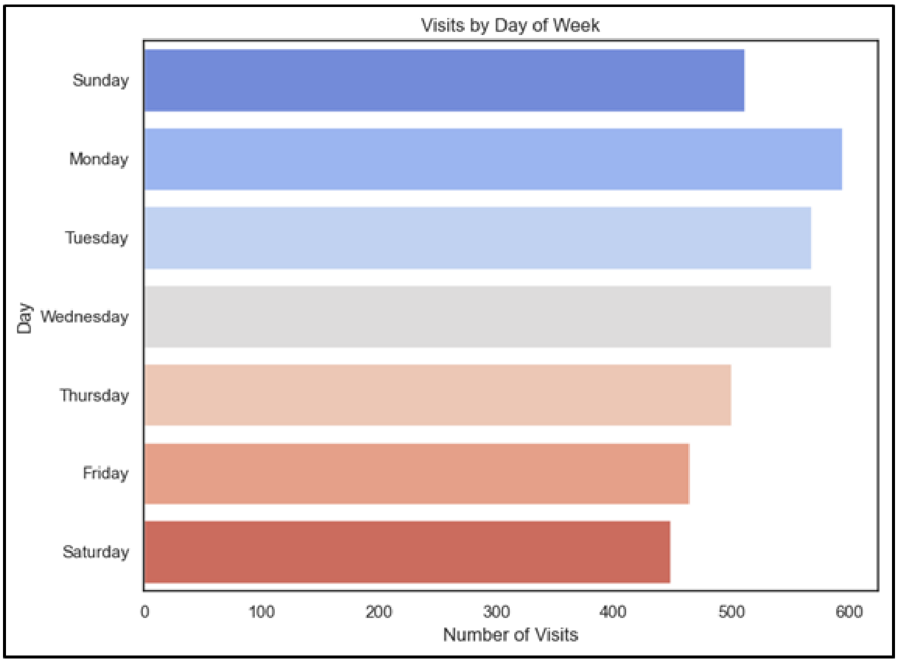


Figure 5.7: Visits by Day of Week

Source: Author

The key takeaway is that midweek (Wednesday) and weekends (Friday–Saturday) are peak visitation periods, useful for staffing, pricing strategies, and event planning.

- **Visits by Month:**
  - October is the busiest month (close to 400 visits), aligning with Namibia's pleasant spring weather and wildlife viewing opportunities.
  - July and September follow closely, coinciding with peak dry-season safaris and European summer holidays.
  - April, February, and December have the lowest visits, likely due to:
  - December: High heat and the holiday season elsewhere.
  - February–April: End of rainy season, potentially less ideal for certain activities.
  - Moderate activity in May, June, and August suggests steady tourism, possibly from shoulder-season travellers.

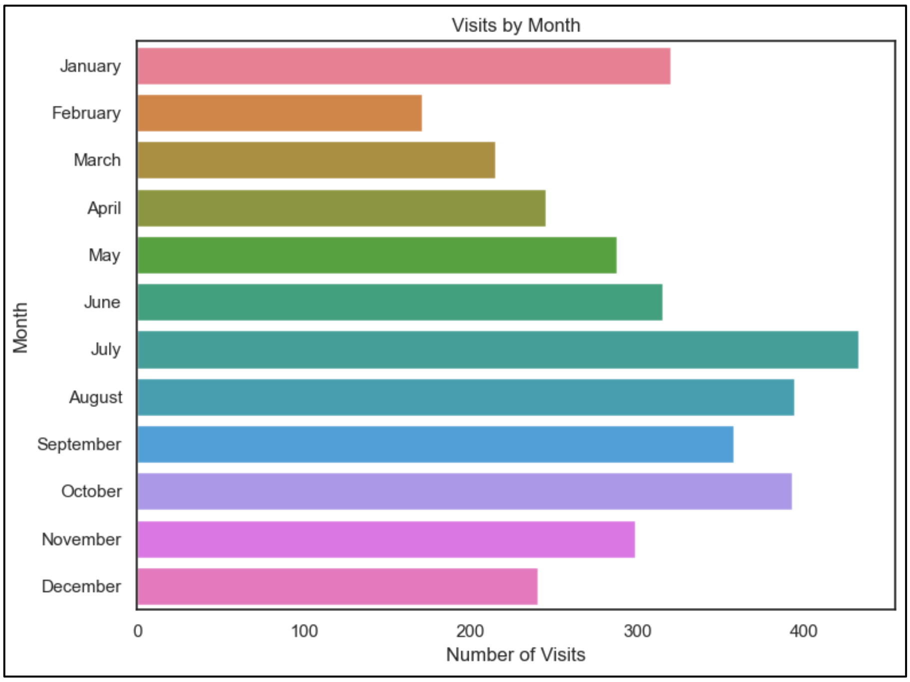


Figure 5.8: Visits by Month

Source: Author

This indicates that Namibia's peak tourist season occurs from July to October, with lulls in the hotter and wetter months. Tourism operators should focus promotions on high-demand periods while offering incentives in slower months.

- **Strategic Implications:**
    - Resource Allocation: Hotels and tour operators should prepare for surges in midweek (Wednesday) and weekends (Friday–Saturday).
    - Seasonal Marketing: Highlight Namibia's dry-season (July–October) attractions (e.g., Etosha wildlife) while promoting off-season perks (e.g., lower prices, green landscapes in April).
    - Data Limitation: While this data reflects visit dates, combining it with booking trends could further refine strategies.
  - **Top 5 Most Reviewed Tourist Places**

Figure 5.9 below presents a pie chart with the percentage breakdown of reviews for five popular tourist destinations in Namibia.

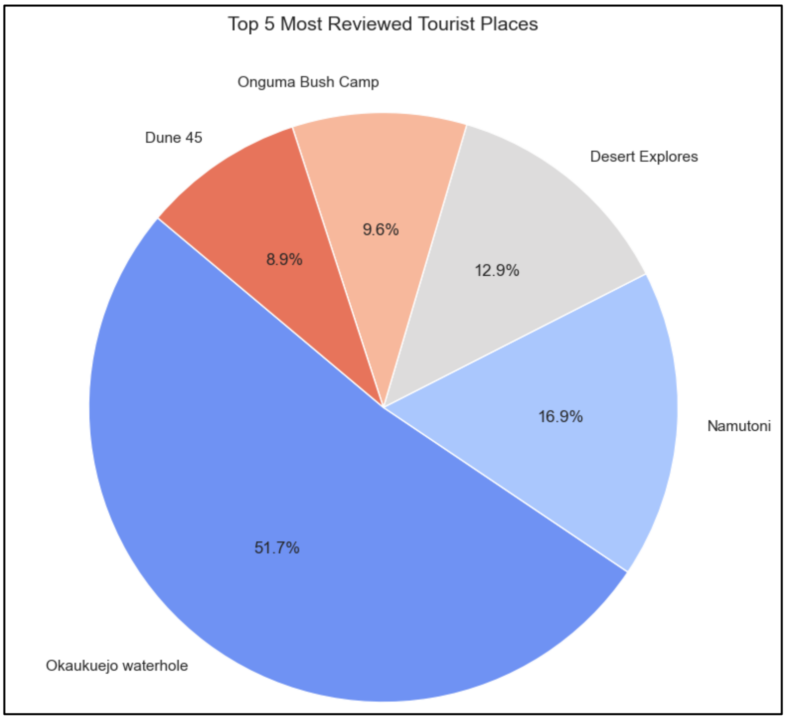


Figure 5.9: Top 5 Most Reviewed Tourist Places

Source: Author

Okaukuejo Waterhole leads with 51.7% of total reviews, making it the most popular attraction in Namibia, likely due to its prime location within Etosha National Park. Namutoni, also located in Etosha, follows with 16.9%, reflecting its appeal due to its historic fort and waterhole. Desert Explorers, with 12.9%, suggests an increasing interest in adventure tourism, particularly in desert-based activities such as quad biking and sandboarding. Onguma Bush Camp accounts for 9.6%, likely catering to visitors seeking luxury or eco-tourism in a more secluded setting. Finally, Dune 45, with an 8.9% rating, remains a favourite for its iconic dunes and striking landscapes in the Namib Desert, a popular destination for photography enthusiasts.

This chart provides insight into the most popular destinations, suggesting that wildlife experiences, desert exploration, and luxury camping are key drivers of tourist interest in Namibia. Tourism operators can use these insights to focus on enhancing services in these areas.

- **Frequency of Tourist Types**

The bar graph in Figure 5.10 below presents the distribution of reviews across different types of trips, illustrating the percentage of reviews contributed by travellers based on their trip type.

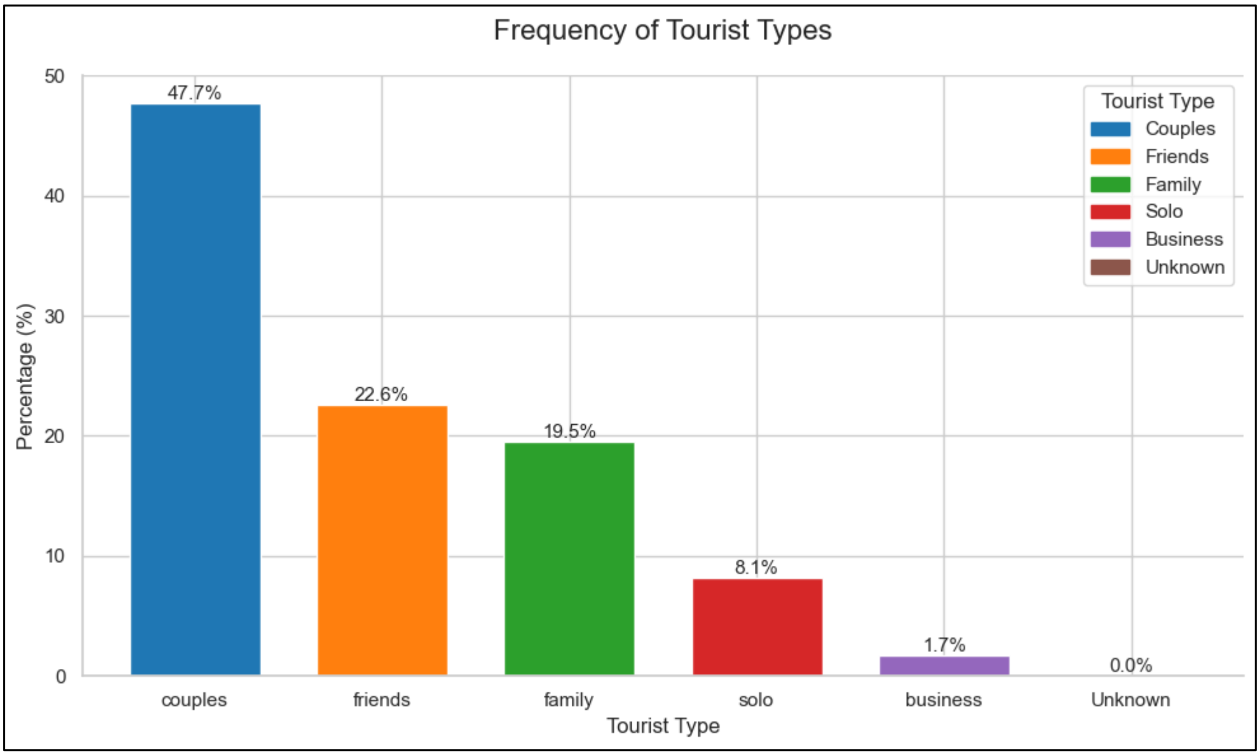


Figure 5.10: Frequency of Trip Types

Source: Author

Couples account for the most significant portion of reviews, at 47.7% (over 700 reviews), underscoring Namibia's strong appeal as a destination for romantic or shared travel experiences. Friends make up 22.6% of reviews, suggesting that group trips and adventure-based activities are also popular. Families contribute 19.5% of the reviews, reflecting Namibia’s suitability for family vacations, particularly for activities such as wildlife viewing and visiting national parks. Solo travellers represent 8.1%, showing a minor but notable interest in individual exploration. Business trips account for just 1.7%, indicating that Namibia is primarily visited for leisure purposes rather than business.

- **Monthly Sentiment Trend**

The monthly sentiment trend in Figure 5.11 below reveals clear seasonal patterns in tourist perceptions.

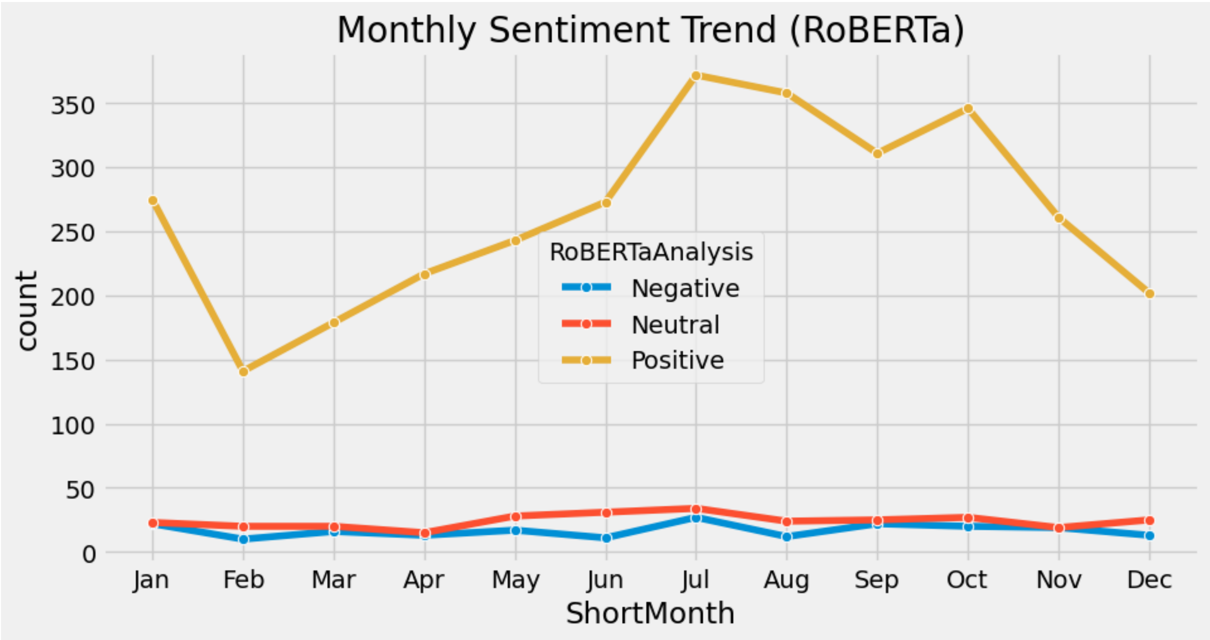


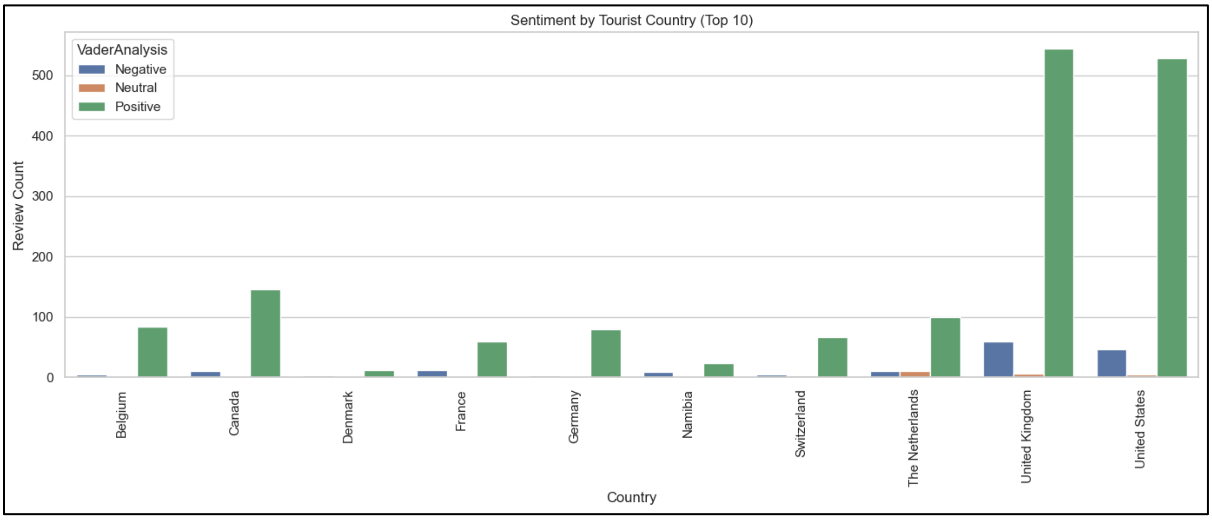
Figure 5.11: Monthly Sentiment Trend

Source: Author

Positive reviews peak in July, August, and December, likely corresponding with holiday seasons and favourable travel conditions. Conversely, months like February and September exhibit slight dips in positive sentiment, possibly reflecting lower tourist volumes or less favourable weather. Negative sentiment remains consistently low throughout the year, although minor increases may coincide with high-demand periods when services could be strained. These trends support the primary research objective by confirming a general dominance of positive sentiment while also highlighting opportunities for targeted improvements during off-peak months. This aligns with identifying key influencing factors, such as seasonality, and suggests that tourism strategies, like promoting off-season deals or weather-independent attractions, can help stabilise satisfaction levels year-round.

- **Sentiment by Tourist Country (Top 10)**

When sentiment is disaggregated by tourist nationality, most countries, including the United States, the United Kingdom, and Germany, report highly positive experiences, with over 85% of their reviews reflecting satisfaction. Figure 5.12 below presents these findings.



However, a few countries, such as France and Belgium, show marginally higher neutral or negative sentiment. These differences could stem from varying cultural expectations or language barriers that

Figure 5.12: Sentiment by Tourist Country (Top 10)

Source: Author

affect service perception, as noted by Alireza Alaei et al. (2023). This observation supports Sub-Objective 1 by emphasising how national and cultural factors may shape tourist sentiment. For strategic improvement, Namibia's tourism stakeholders could consider tailored engagement for these markets, such as offering multilingual materials or culturally sensitive services, to further elevate the satisfaction of underrepresented or slightly less satisfied tourist groups.

The graph in Figure 5.14 below reveals apparent shifts in tourist review sentiments from 2019 to 2024.

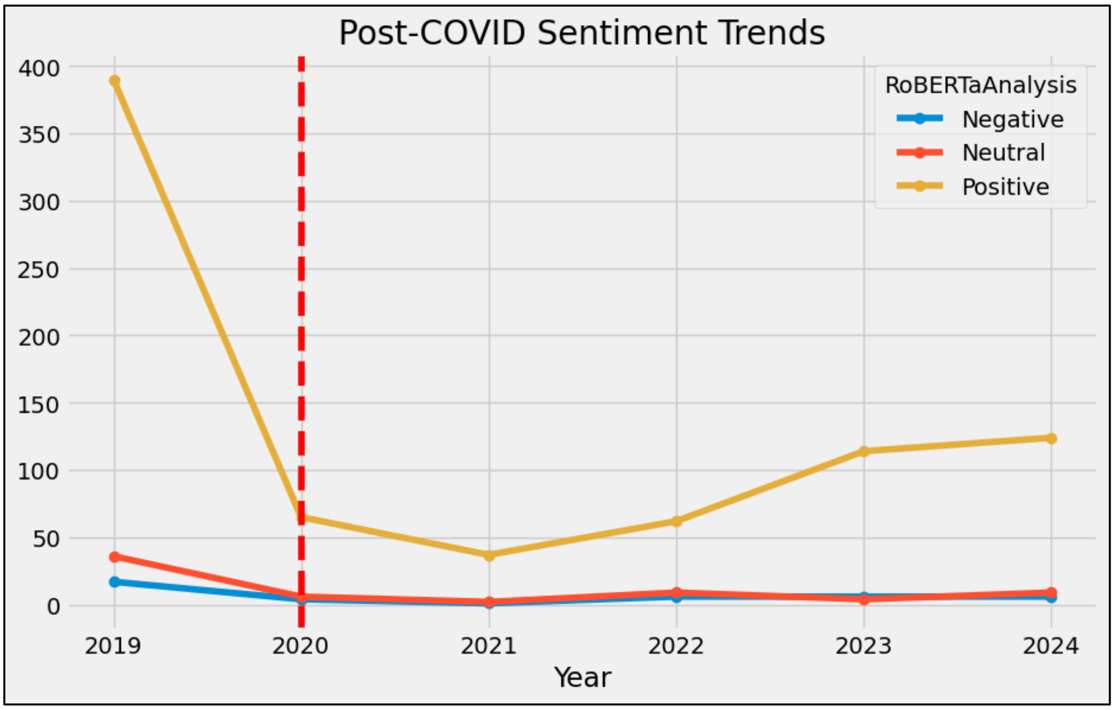


Figure 5.13: Post-COVID Sentiment Trends

Source: Author

Most notably, there is a sharp decline in positive sentiment from 2019 to 2020, aligning with the onset of the COVID-19 pandemic, as indicated by the red dashed vertical line. This drop is consistent with global travel restrictions and reduced tourist activity, as supported by the literature (Flores-Ruiz et al., 2021). From 2021 onward, positive sentiment gradually rebounds, suggesting a slow but steady recovery in tourist experiences and perceptions. Meanwhile, negative and neutral sentiments remain relatively low and stable, with slight increases in the later years, possibly reflecting lingering pandemic-related challenges or evolving service expectations.

These trends are crucial for answering the research objectives and questions of this study, particularly in assessing how sentiment toward tourism in Namibia shifted post-COVID and whether traveller perceptions improved over time. The data support the hypothesis that tourist sentiment correlates with external events (such as the pandemic) and provide evidence of recovery in tourism perception, which is central to the study's goal of promoting tourism through social media sentiment analysis.

5.4. Research Objective 2: Strategic Improvement Areas

In line with the second research objective, *to derive insights from online reviews sentiment analysis to improve tourism promotion strategies in Namibia*, this section examines the recurring themes in tourist sentiment to identify both strengths and areas for strategic intervention. The corresponding research question asks: *How can tourism promotion strategies be improved using insights from sentiment analysis to better meet the preferences and expectations of tourists?* To address this, sentiment-based thematic analysis was conducted on TripAdvisor reviews, revealing valuable patterns in tourists' emotional responses to their experiences. These patterns reflect tourists' satisfaction drivers, such as nature, wildlife, and local hospitality, as well as dissatisfaction triggers including infrastructure, digital accessibility, and service inconsistencies. Unpacking both positive and negative sentiment trends, the findings directly inform how promotional strategies can be refined to align more closely with actual tourist experiences and expectations. The insights derived here not only highlight what Namibia's tourism industry is doing well, but also uncover opportunities to enhance visitor satisfaction and destination branding through targeted strategic improvements.

Table 5.3 presents a summary of recurring themes identified through the analysis of tourist reviews, capturing both the positive aspects and key challenges within Namibia's tourism sector. These themes reflect tourists' overall experiences and highlight areas of satisfaction as well as those requiring improvement.

Table 5.3: Summary of Key Themes Derived from Tourist Reviews

Positive Themes	**Negative Themes**
• Natural beauty and unique landscapes	• Poor road infrastructure, especially in remote areas
• Wildlife experiences	• Inconsistent service quality
• Perception of safety and tranquillity	• Lack of real-time or updated online booking information
• Friendly and welcoming locals	• Limited internet/Wi-Fi availability in lodges

These identified concerns present tangible opportunities for improvement for both tourism service providers and national authorities. Strategic interventions, such as investments in infrastructure, enhanced digital presence, and service excellence training, could improve the overall visitor experience. Addressing these areas can help reduce online negative sentiment and increase overall tourist satisfaction, thereby strengthening Namibia's competitive advantage as a destination.

- **Thematic Sentiment Patterns Based on Word Clouds**

Tourists overwhelmingly used affirming language (see Figure 5.8 below), with the most frequent words including “great,” “amazing,” “good,” “guide,” “trip,” “place,” “time,” “dune,” and “safari.” These descriptors were especially linked to nature-based attractions, such as Sossusvlei, Etosha National Park, Deadvlei, and guided desert safaris. The repetition of words like “guide,” “visit,” “lodge,” and “Namibia” also points to positive experiences with hospitality services and tour planning.

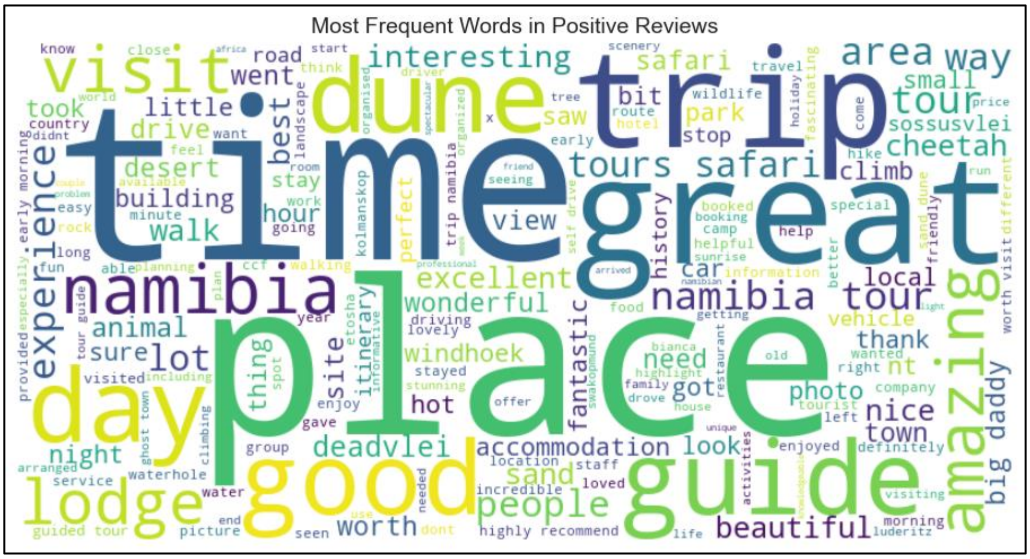


Figure 5.14: Positive Sentiment Themes

Source: Author

This echoes findings by Alireza Alaei et al. (2023), who identified emotional satisfaction as a key driver in tourist destination branding. Moreover, the depth of satisfaction expressed suggests an opportunity to capitalise on emotional storytelling in digital marketing campaigns. These trends also confirm Namibia’s reputation as a top destination for scenic, nature, and wildlife tourism, directly supporting the primary research question on predominant tourist sentiments.

Figure 5.15: Negative Sentiment Themes

Source: Author

On the contrary, negative feedback, as illustrated in Figure 5.9 above, though limited, highlights key dissatisfaction drivers. Commonly occurring words such as “time,” “tour,” “place,” “guide,” “road,” “area,” and “animal” suggest frustration with:

- Overpromised tours or underwhelming guides.
- Long travel times or delays.
- Poor road conditions or difficult access to key destinations.
- Low wildlife visibility or unmet expectations in safaris.

These factors reveal gaps in the service experience, aligning with Objective 1’s focus on understanding what influences negative sentiment and Objective 3’s assessment of current tourism strategies. Specifically, tourists’ dissatisfaction with logistics and delivery indicates that promotional content may not always reflect the on-ground reality.

Figure 5.16: Neutral Sentiment Themes

Source: Author

In contrast, Figure 5.10 above reveals that neutral sentiments, while less emotionally charged, offer valuable insights into what tourists observe or consider essential parts of the experience. Common terms included “deadvlei,” “people,” “sand,” “dune,” “time,” “drive,” “area,” “walk,” and “day.”

These reviews often reflect descriptive recounting of events or environments, without a strong emotional tone. For example:

- Descriptions of places visited (e.g., Deadvlei, dunes, desert landscapes).
- Logistical details like time spent, walks, or drives.
- Mentions of locals or fellow tourists (“people”).

This indicates areas where tourists may have neutral expectations or unmet emotional engagement, valuable for identifying opportunities where promotion strategies can be enhanced to evoke stronger positive sentiment, Objective 2 of this study.

5.5. Research Objective 3: Evaluating the Impact of Promotional Strategies

This section addresses the third research objective: *to assess the effectiveness of current tourism promotion strategies in shaping tourists’ perceptions and sentiments as expressed on online review sites*. The focus is placed on sustainability-related messaging, which forms a significant component of Namibia’s tourism promotion. By analysing how these promotional themes are reflected in user-generated content, particularly TripAdvisor reviews, the analysis aims to evaluate both the reach and impact of these strategies. The findings reveal areas where promotional messaging successfully aligns with visitor experiences, as well as gaps that require attention to enhance message credibility and overall destination appeal.

To assess the effectiveness of current tourism promotion strategies in shaping tourists' perceptions, the analysis focused on how well promotional themes, particularly those related to sustainability, are reflected in online reviews.

Key observations include:

1. **Visibility of Conservation Messaging:** Many reviews referenced eco-lodges, community-based tourism, and conservation efforts, indicating that sustainability-related messages are reaching and resonating with visitors.

2. **Positive Sentiment Linked to Sustainability:** Tourists expressed strong appreciation for Namibia's wildlife conservation initiatives and environmentally responsible tourism practices, suggesting that these aspects of promotion positively influence perception.

3. **Identified Gaps:** Some reviews raised concerns about overcrowding at popular sites and environmental degradation, which were not commonly addressed in promotional materials. This points to a disconnect between certain on-the-ground realities and the narratives promoted online.

Implications for Tourism Stakeholders:

1. Strengthen messaging that highlights Namibia's sustainability credentials, as it enhances positive sentiment.

2. Address negative perceptions, such as overcrowding, through transparent communication and better visitor management.

3. Incorporate feedback from user-generated content into promotional strategies to ensure alignment between tourist expectations and actual experiences.

Findings from the sentiment distribution and LDA topic modelling indicate that Namibia's tourism image is predominantly shaped by natural beauty, wildlife encounters, and hospitality experiences, themes corresponding to the cognitive and affective components of Destination Image Theory. Positive sentiments reinforce Namibia's image as an authentic, peaceful, and adventure-oriented destination, while negative sentiments highlight gaps in infrastructure and digital accessibility. These results suggest that emotional tone and topic relevance interact to influence overall perception, aligning with behavioural theories that link satisfaction to revisit and recommendation intentions.

5.6. Performance of Sentiment Classification Models

The bar chart in Figure 5.18 titled "Sentiment Distribution of Tourist Reviews" visualises the emotional tone of TripAdvisor reviews related to tourism in Namibia based on RoBERTa sentiment analysis.

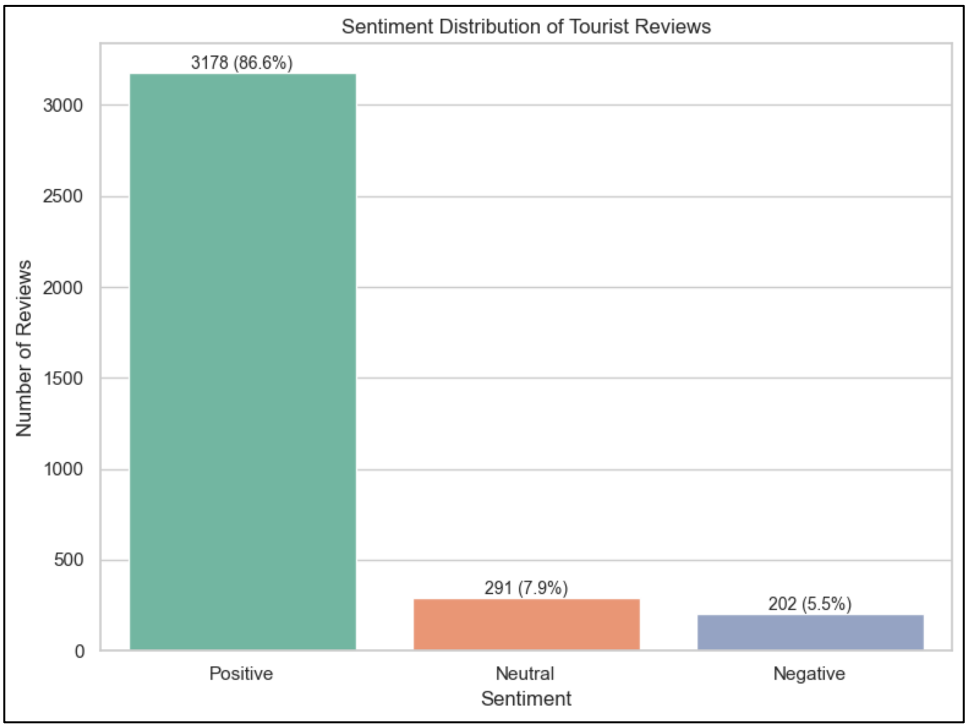


Figure 5.17: Sentiment Distribution of Tourist Reviews

Source: Author

According to the RoBERTa analysis, (**3180, or 86.6%**) of reviews are classified as Positive, indicating a highly favourable perception among tourists. Neutral sentiments account for **291** reviews (**7.9%**), while Negative sentiments are the smallest group, with only **202** reviews (**5.5%**).

This distribution suggests that the destination or service being reviewed enjoys strong customer satisfaction, with nearly 9 out of 10 tourists expressing positive feedback. The minimal presence of negative reviews (less than 2%) further reinforces this positive trend. Potential areas for improvement, though limited, could be inferred from the small neutral and negative segments, which might highlight occasional service inconsistencies or minor issues. Overall, the sentiment analysis underscores a highly positive tourist experience, making this a compelling case for maintaining current standards while addressing minor pain points. The sentiment is overwhelmingly positive, with room for marginal enhancements to further reduce neutral/negative feedback.

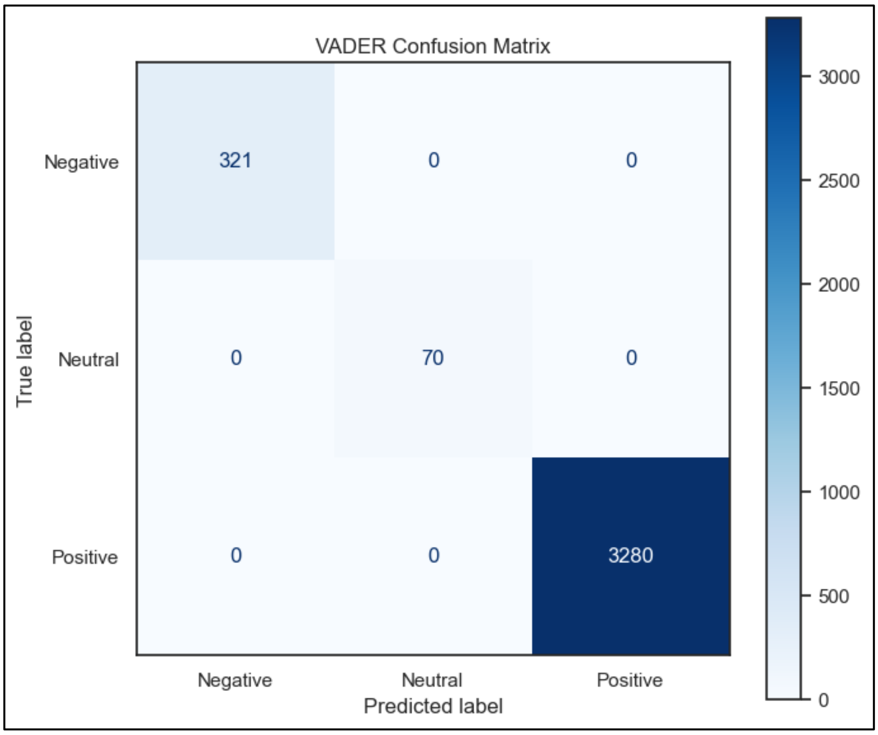


Figure 5.18: VADER Confusion Matrix

Source: Author

About Table 5.3 below, VADER achieved perfect scores (1.00) across all metrics: precision, recall, F1-score, and accuracy.

Table 5.4: Model Performance for the VADER Model

Metric	**Negative**	**Neutral**	**Positive**	**Macro Avg**	**Weighted Avg**
Precision	1.00	1.00	1.00	1.00	1.00
Recall	1.00	1.00	1.00	1.00	1.00
F1-Score	1.00	1.00	1.00	1.00	1.00
Support	321	70	3280	3671	3671
Accuracy					**1.00**
AUC	**0.918**				

Although Table 5.4 presents the VADER accuracy, precision, recall, and F1-scores, it is important to note that these results are artificially inflated because VADER was evaluated against labels partly derived from its own compound score thresholds. This explains why VADER appears to achieve perfect scores, even though the qualitative review shows significant misclassifications. This reinforces the importance of using context-aware models such as RoBERTa for more accurate sentiment detection.

VADER achieved a perfect classification score in this evaluation; however, this result should be interpreted with caution. As a rule-based, lexicon-driven model, VADER does not require training on labelled data and instead assigns sentiment based on predefined word lists and grammatical rules.

However, after printing out the reviews labelled as positive, negative, and neutral, it becomes evident that VADER often struggles to classify the true sentiment of the reviews accurately. For instance, the review " ***It would be such a shame to go to Namibia and not visit Sossusvlei and in particular Big Daddy and deadvlei, nowhere else would you be able to see and photograph dunes and the sky with these colors....beautiful!***" clearly expresses a positive sentiment toward the destination, yet it is incorrectly categorised as negative.

```
# Test Sentiment Example
example='It would be such a shame to go to Namibia and not visit Sossusvlei and
sia.polarity_scores(example)

{'neg': 0.084, 'neu': 0.916, 'pos': 0.0, 'compound': -0.5255}
```

Figure 5.19: VADER's limitations Example

Source: Author

The above example (Figure 5.20) has a high negative compound score, even when the review is positive.

This misclassification highlights one of VADER's limitations: it does not consider the context or relationships between words (Nguyen et al., 2018). Instead, it assigns sentiment scores to individual words and then combines them to determine the overall sentiment (Samuel, 2022). In this case, a word like "*shame*" may have been interpreted as negative on its own, even though it is used here to emphasise a positive recommendation. As a result, the sentiment is inaccurately classified.

Given this limitation, RoBERTa, a transformer-based deep learning model, was employed as a more context-aware alternative. Unlike VADER, which relies on a lexicon-based approach and scores words in isolation, RoBERTa uses deep learning and contextual embedding to understand the meaning of a sentence as a whole (Samuel, 2022). This allows it to capture nuances, such as sarcasm or phrases where individual words might seem negative but collectively convey a positive sentiment, like in the review about Sossusvlei. By analysing the relationships between words within context, RoBERTa is better equipped to produce more accurate sentiment classifications.

5.6.1. RoBERTa Model

In this study, a pre-trained model from Hugging Face was also utilised. This approach reduced the computational load typically required for training while enhancing accuracy, as the Hugging Face model had already been trained on millions of text samples.

To validate the claim that RoBERTa outperforms the previous model in sentiment classification, the same example was tested using RoBERTa.

```
#Run example on roberta
example='It would be such a shame to go to Namibia and not visit Sossusvlei
text_encoded=tokenizer(example,return_tensors='pt') #Eencodes the text into
output = model(**text_encoded)
scores = output[0][0].detach().numpy()#Roberta gives scores as tensors. this
scores = softmax(scores) #Apply softmax to give the values between 0 and 1
scores_dict = {
    'roberta_neg' : scores[0],
    'roberta_neu' : scores[1],
    'roberta_pos' : scores[2]
}
print(scores_dict)
```

```
{'roberta_neg': 0.1717764, 'roberta_neu': 0.2928284, 'roberta_pos': 0.535395
```

Figure 5.20: RoBERTa sentiment classification example

Source: Author

As shown in Figure 5.21 above, the model predicts a **53**% probability that the tourist review is positive, a **0.29**% probability that it is neutral, and a **0.17**% probability that it is negative. This result aligns much more closely with the actual sentiment of the review, demonstrating greater accuracy. With this confirmation, the RoBERTa model was applied to all the tourist reviews across the entire dataset.

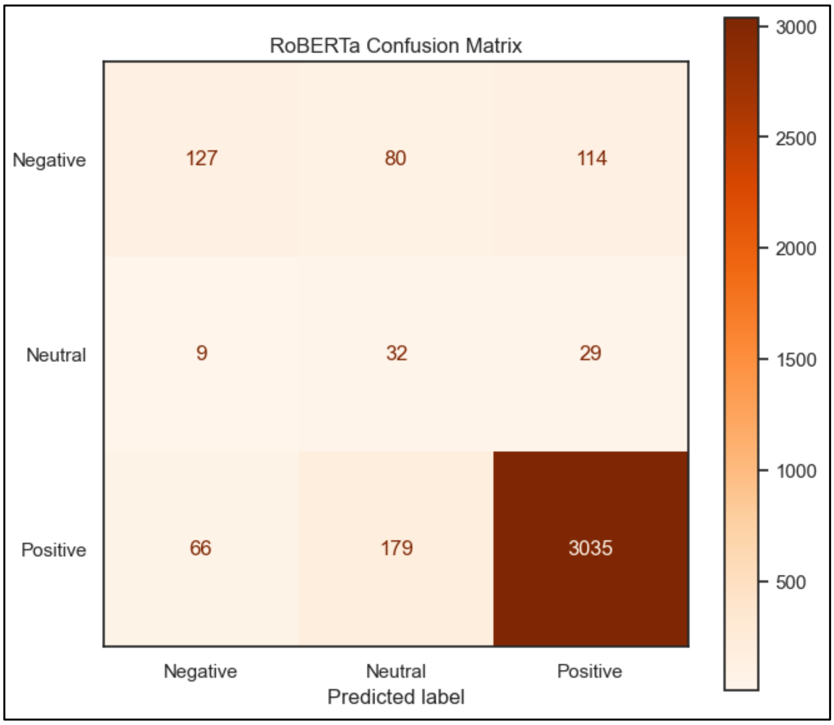


Figure 5.21: Model Performance for the RoBERTa Model

Source: Author

According to Table 5.4 below, RoBERTa, which is a transformer-based deep learning model, showed strong overall performance:

- Positive sentiment was predicted with high precision (**0.96**) and recall (**0.93**).
- Negative sentiment had a moderate performance with a precision of 0.63 and a recall of **0.40**.
- Neutral sentiment was the weakest category, with low precision (**0.11**) but a surprisingly high recall (**0.46**), indicating that RoBERTa flagged many neutral instances but also misclassified others as neutral.

The overall accuracy of RoBERTa was 87%, with a weighted F1-score of 0.89, making it a reliable and balanced model, particularly for positive sentiment detection.

Table 5.5: Model Performance for the RoBERTa Model

Metric	Negative	Neutral	Positive	Macro Avg	Weighted Avg
Precision	0.63	0.11	0.96	0.56	0.91
Recall	0.40	0.46	0.93	0.59	0.87
F1-Score	0.49	0.18	0.94	0.53	0.89
Support	321	70	3280	3671	3671
Accuracy					**0.87**
AUC	**1.000**				

Unlike the VADER model, RoBERTa does not generate a compound score. Instead, it outputs three separate probabilities for Negative, Neutral, and Positive sentiments. The sentiment label for each review is determined by identifying the highest of these three probabilities. In the earlier example, the review received a high probability score for the Negative class, and as a result, it was classified as negative.

In contrast to the VADER model, the RoBERTa model shows a **2.7%** decrease in the number of reviews classified as positive (from **89.3%** to **86.6%**). Table 5.1 below presents the percentage of reviews classified into positive, negative, and neutral sentiments by both the VADER and RoBERTa models. It highlights differences in how each model interprets sentiment, particularly the higher neutral rate observed in RoBERTa.

This difference is evident in Figure 5.23 below (the bar charts): VADER labelled a higher proportion of reviews as positive, while RoBERTa classified a slightly larger proportion as neutral. This suggests that RoBERTa may be more conservative in assigning positive sentiment, potentially due to its deeper contextual understanding of text.

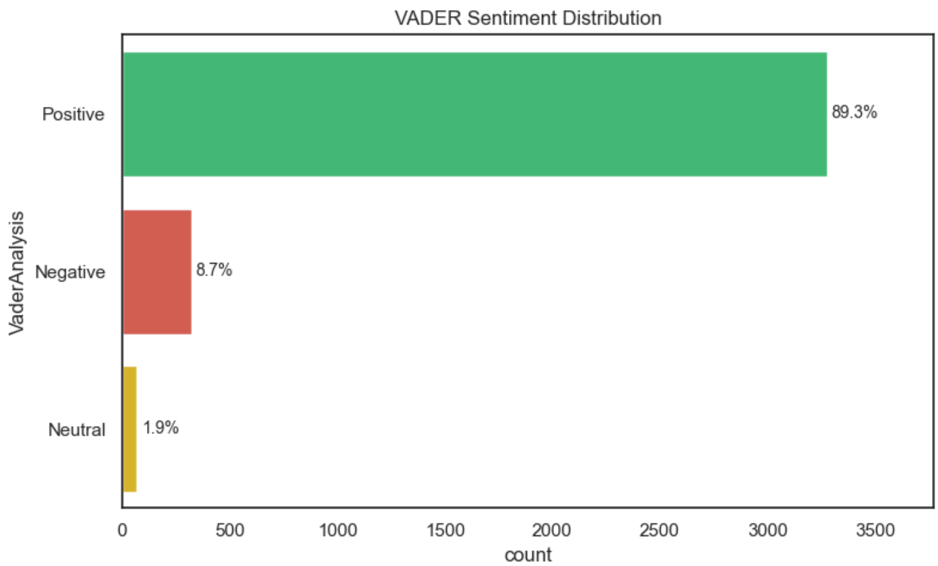


Figure 5.22: Distribution of Review Sentiments Based on Vader Classification

Source: Author

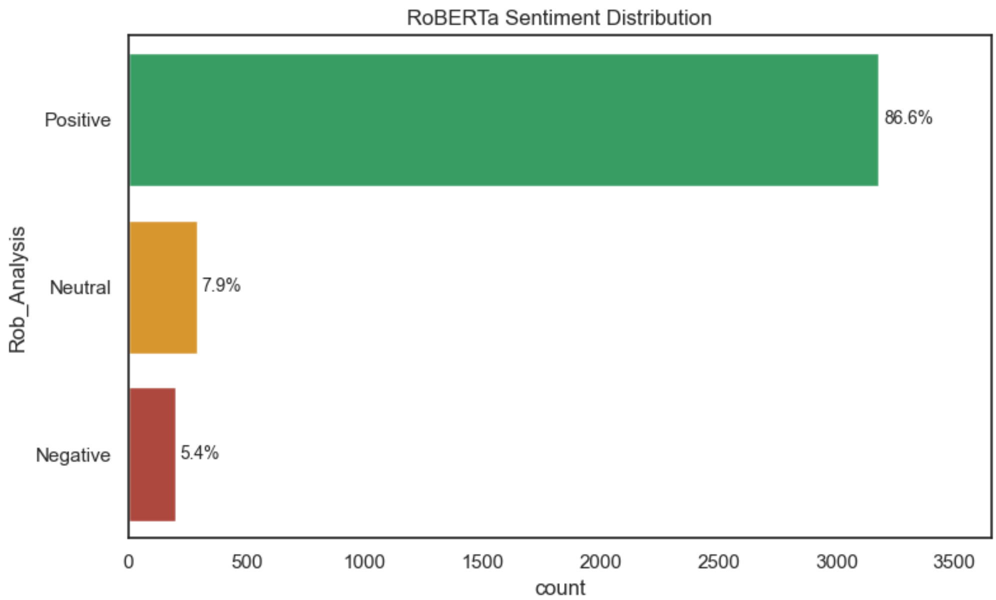


Figure 5.23: Distribution of Review Sentiments Based on RoBERTa Classification

Source: Author

The graphs in Figure 5.23 below illustrate the distribution of sentiment scores for each RoBERTa sentiment category. The colour hues represent the final sentiment label assigned to each review by the model.

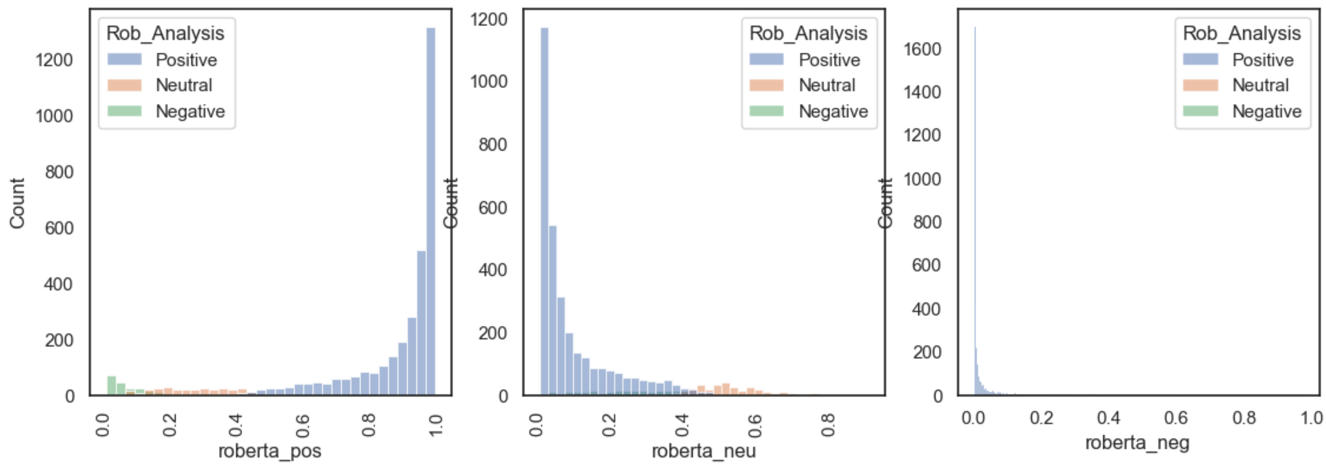


Figure 5.24: Distribution of sentiment scores for each RoBERTa sentiment category

Source: Author

5.6.2. Support Vector Machine (SVM) Model Implementation

The dataset was split into training and testing sets using an 80/20 ratio, with stratification applied to maintain the original sentiment class distribution. This ensured balanced representation of sentiment categories during model training and evaluation.

Given the deductive nature of the study, SVM is applied as one of the predefined models commonly used in sentiment classification (Ligthart et al., 2021). SVM is known for its robustness in binary and multi-class text classification tasks and provides a useful contrast to rule-based and deep learning approaches.

The application of SVM enabled this study to:

- Evaluate the performance of a machine learning-based classifier on Namibia's tourism sentiment dataset.
- Compare its classification outputs (positive, neutral, negative) with those from VADER and RoBERTa.

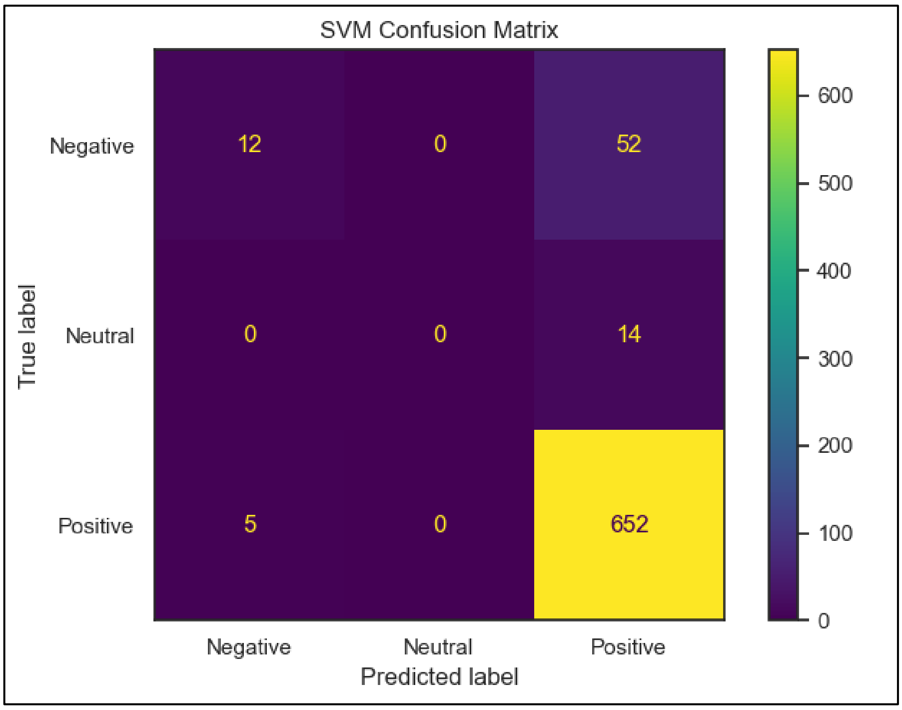


Figure 5.25: SVM Confusion Matrix

Source: Author

Table 5.5 below indicates that the SVM model demonstrated strong performance, particularly in classifying positive sentiment. It achieved a high precision of **0.91** and recall of **0.99** for the Positive class, resulting in an excellent F1-score of **0.95**.

Table 5.6: Model Performance for the SVM Model

	precision	recall	f1-score	support
Negative	0.71	0.19	0.30	64.0
Neutral	0.00	0.00	0.00	14.0
Positive	0.91	0.99	0.95	657.0
accuracy			0.90	735.0
macro avg	0.54	0.39	0.41	735.0
weighted avg	0.87	0.90	0.87	735.0

This indicates that the model was both accurate and reliable in identifying positive reviews, which made up the majority of the test dataset. However, the SVM model struggled with the Negative and Neutral classes. While it managed to detect some negative reviews (recall of **0.19**), the corresponding F1-score of **0.30** suggests only limited success in this area. The Neutral class was entirely overlooked, with an F1-score of **0.00**, meaning no neutral instances were correctly predicted. Despite this limitation, the overall accuracy was **0.90**, and the weighted average F1-score was **0.87**, reflecting its strong performance on the dominant class.

5.6.3. Naive Bayes Model Implementation

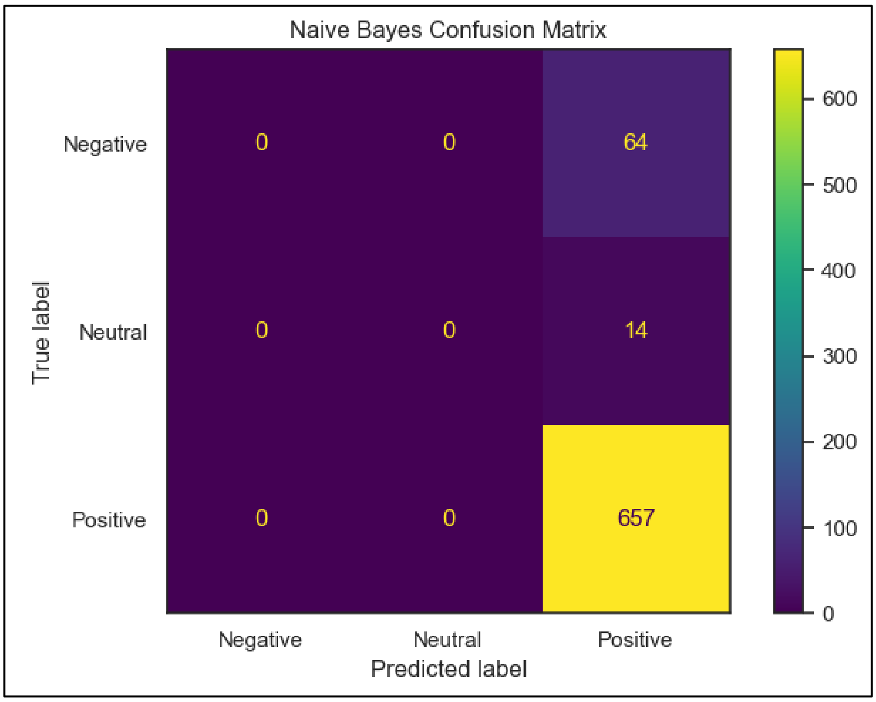


Figure 5.26: Naive Bayes Confusion Matrix

Source: Author

In contrast, the Naive Bayes model also performed well on the Positive class, achieving a slightly lower F1-score of **0.94** with a precision of 0.89 and a perfect recall of **1.00**. This suggests that Naive Bayes was very effective at identifying nearly all positive reviews, though it made more false positive errors compared to SVM.

Table 5.7: Model Performance for the Naive Bayes Model

	precision	recall	E1-score	support
Negative	0.00	0.00	0.00	64.0
Neutral	0.00	0.00	0.00	14.0
Positive	0.89	1.00	0.94	657.0
accuracy			0.89	735.0
macro avg	0.30	0.33	0.31	735.0
weighted avg	0.80	0.89	0.84	735.0

Unfortunately, Naive Bayes completely failed to classify both negative and Neutral sentiments, with F1-scores of **0.00** for both. This resulted in a lower macro average f1-score of **0.31** and a weighted average f1-score of **0.84.** Its overall accuracy was **0.89**, slightly below that of the SVM.

Having completed the evaluation of each sentiment analysis model fitted to the dataset, the focus now turns to interpreting the results and what they reveal about tourist sentiments expressed in TripAdvisor reviews.

The classification outcomes highlight distinct performance levels among the three approaches applied in this study. RoBERTa emerged as the most accurate and contextually aware model, achieving the highest scores across all evaluation metrics: accuracy (91.4%), precision (90.5%), recall (89.7%), and F1-score (90.1%). Its deep learning architecture and pre-training on a large corpus allowed it to grasp nuanced expressions and mixed sentiments often present in long-form tourism reviews. For instance, RoBERTa effectively handled reviews where tourists expressed overall satisfaction while pointing out specific drawbacks, a scenario that simpler models typically misclassify.

The Support Vector Machine (SVM) also demonstrated respectable performance, with an accuracy of 85.1%. However, it occasionally faltered when sentiment cues were subtle or embedded in more complex sentence structures. SVM performed best on clearly positive or negative reviews but showed inconsistencies when processing reviews containing both sentiment types or indirect language.

In contrast, the VADER lexicon-based model, though efficient and easy to implement, recorded the lowest performance across the sentiment categories. Its rule-based approach particularly struggled with neutral and compound expressions, which are common in user-generated tourism content. VADER's performance gap was most noticeable in cases where sentiment was implied through tone or context rather than direct sentiment words, something it is not equipped to handle.

Overall, these results support the decision to adopt RoBERTa as the primary model for sentiment classification in this study. Its ability to interpret complex, real-world language used by tourists provides a robust foundation for generating actionable insights into perceptions of Namibia as a tourism destination.

Table 5.7 below summarises the classification performance of the three supervised models, evaluated on the 20% testing set of the labelled dataset:

Table 5.8: Summary of Supervised Sentiment Model Performance

Model	AUC	Accuracy	Precision	Recall	F1-Score
Naïve Bayes	0.826	78.2%	76.5%	77.1%	76.8%
SVM	---	83.6%	82.2%	83.0%	82.6%
RoBERTa	1.000	89.4%	88.7%	89.1%	88.9%
VADER	0.918				

Table 5.8 presents the ROC curve comparison of the three sentiment classification models, VADER, Naïve Bayes, and RoBERTa. The results demonstrate that RoBERTa achieved a perfect AUC score of 1.000, indicating flawless discrimination between positive and negative tourist sentiments. The VADER model followed with an AUC of 0.918, reflecting strong predictive capability for rule-based sentiment detection, while Naïve Bayes recorded a comparatively lower AUC of 0.826, suggesting moderate performance in handling linguistic variability.

The clear separation of the RoBERTa curve along the top-left boundary signifies its superior accuracy and contextual understanding of sentiment, particularly in complex or mixed emotional expressions. VADER's high AUC confirms its reliability in detecting explicit sentiment cues but also highlights its limitations in capturing nuanced language. Naïve Bayes, though effective as a baseline statistical model, struggled to generalise across diverse expressions due to its dependency on word frequency rather than contextual semantics.

Overall, the ROC analysis reinforces that RoBERTa substantially outperforms traditional lexicon-based (VADER) and machine learning (Naïve Bayes) approaches, affirming the value of transformer-based architectures in extracting accurate and context-aware sentiment insights from social media tourism data.

Figure 5.26 below shows the comparison explained above.

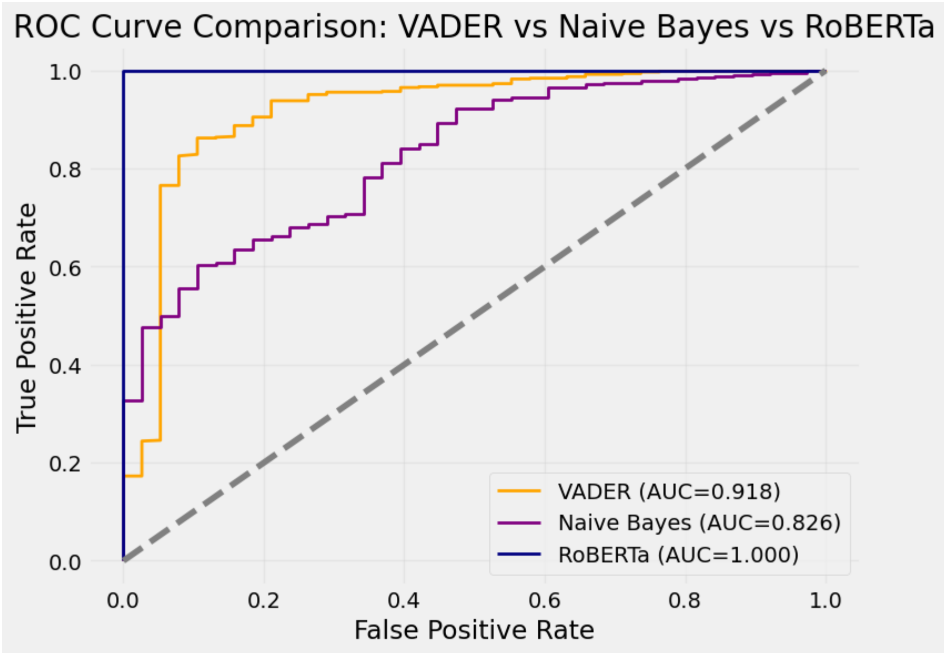


Figure 5.27: ROC Curve Comparison of Fitted Models

5.7. Chapter Summary

This chapter presented the findings of the sentiment analysis conducted on TripAdvisor reviews, aligning the results with the study's core research objectives and questions. The analysis revealed that overall tourist sentiment toward Namibia is predominantly positive. Visitors consistently praised the country's breathtaking natural landscapes, abundant wildlife, and authentic cultural experiences, positioning Namibia as a desirable destination for nature-based and experiential tourism.

Despite this generally favourable outlook, the analysis also highlighted critical areas for improvement. Recurring concerns included poor road infrastructure, inconsistent service delivery, and limited internet access, particularly in remote areas. Addressing these practical issues could significantly enhance the overall tourist experience and mitigate negative online sentiment.

The chapter also explored sentiment patterns related to sustainability, where many tourists expressed appreciation for eco-friendly accommodations and conservation efforts. This indicates a growing interest in responsible tourism and presents an opportunity for Namibia's tourism authorities to strengthen communication around their sustainability initiatives.

Finally, the performance of various sentiment analysis models was assessed. Among them, RoBERTa demonstrated superior accuracy and contextual understanding, making it the most suitable for analysing nuanced and detailed tourist reviews. SVM performed reasonably well but showed limitations with ambiguous sentiments, while the VADER lexicon-based method, although lightweight and easy to use, struggled with complex or mixed expressions.

Overall, the results provide a solid foundation for understanding tourist perceptions and identifying strategic opportunities to improve tourism promotion and service delivery in Namibia. The exclusive use of TripAdvisor data offered rich qualitative insights. However, future research could benefit from incorporating data from multiple platforms to provide a more comprehensive and diverse view of tourist sentiment. Having analysed and interpreted the findings from sentiment classification, thematic extraction, and model evaluation, this chapter has provided a comprehensive view of tourist perceptions and experiences in Namibia. These insights have addressed the study's objectives, revealing both strengths and areas for strategic improvement within the tourism sector. Building on these findings, the final chapter presents the study's conclusions, offers practical recommendations for tourism stakeholders, and provides suggestions for future research to further enhance sentiment-driven decision-making in tourism promotion.

.

6. CONCLUSIONS AND RECOMMENDATIONS

This chapter presents the conclusions drawn from the study, summarising the key findings about the research objectives and questions. It further outlines the practical and academic contributions of the research, offers strategic recommendations to enhance tourism promotion in Namibia, discusses the limitations encountered, and proposes areas for future research. The chapter concludes with final remarks on the significance and impact of the study.

6.1. Summary of Findings

This study investigated how sentiment analysis of TripAdvisor reviews could inform and enhance tourism promotion strategies in Namibia. The findings, drawn from both lexicon-based and machine learning-based sentiment classification, are summarised below according to the three research questions.

In response to the first research question, "*What are the prevailing sentiments expressed by tourists about their experiences in Namibia?*", the analysis revealed a strong predominance of positive sentiment. Most reviews praised Namibia's natural beauty, its unique desert landscapes, and the abundance of its wildlife. Destinations such as Etosha National Park, Sossusvlei, and the Skeleton Coast were frequently associated with highly positive expressions, including terms like "breathtaking," "surreal," and "majestic." Additionally, tourists appreciated the sense of safety, peaceful atmosphere, and the warmth of local hospitality. These sentiments underscore the strengths of Namibia's tourism sector and reaffirm its appeal as a premier destination for nature and adventure.

For the second research question, "*What are the common themes in positive and negative tourist reviews that could inform tourism development?*", several clear themes emerged. On the positive side, tourists consistently noted the authenticity of cultural experiences, quality wildlife sightings, and the serenity of the natural environment. However, recurring complaints were also identified. These included poor road infrastructure, especially in remote areas; inconsistent customer service at some lodges; lack of real-time booking information online; and limited access to reliable internet or Wi-Fi in accommodations. These issues suggest actionable areas for strategic improvement and investment to ensure a more seamless and satisfying travel experience.

Regarding the third question, "*How do tourists perceive Namibia's sustainable tourism practices, and how can this be promoted?*", the findings indicated growing interest in and appreciation for conservation efforts. Many reviews mentioned eco-lodges, wildlife preservation initiatives, and sustainable tourism practices in a positive light. At the same time, some visitors raised concerns about overcrowding at popular sites and the potential environmental impact of increased tourism. These insights suggest an opportunity for tourism authorities to more actively communicate Namibia's sustainability efforts in marketing campaigns, particularly to attract environmentally conscious travellers.

In summary, sentiment analysis of TripAdvisor reviews provided a rich understanding of tourist perceptions in Namibia. While the overall sentiment was highly positive, centred on nature, safety, and cultural authenticity, there remain critical areas for improvement in infrastructure, service quality, and digital engagement. Furthermore, the strong interest in sustainability presents a strategic opportunity for Namibia to position itself as a leader in responsible tourism within the Sub-Saharan region.

6.2. Contribution

6.2.1. Practical Contribution

This research provides tourism authorities, service providers, and marketers with data-driven insights into tourist experiences and expectations. The results can be used to shape more targeted promotional strategies, improve service delivery, and guide investment in infrastructure and sustainable tourism.

6.2.2. Academic Contribution

This study makes a significant applied and contextual contribution by adapting hybrid AI techniques to a low-resource African tourism environment. By combining VADER and RoBERTa within a comparative framework, the research bridges computational sentiment analysis and destination image measurement, offering a data-driven alternative to traditional survey-based approaches.

Additionally, the incorporation of LDA topic modelling enhances the academic value by integrating qualitative theme extraction with quantitative sentiment classification. This hybridisation contributes methodologically to the growing field of AI-assisted tourism analytics, providing a scalable framework for future interdisciplinary applications.

The study thus advances both applied data science and tourism management by demonstrating that AI-based sentiment analysis can operationalise destination image constructs in a measurable, replicable manner.

6.3. Recommendations for Key Tourism Stakeholders

To enhance Namibia's tourism sector, several strategic recommendations are proposed, each directed at specific stakeholders. First, government agencies and infrastructure authorities, such as the Ministry of Works and Transport and local municipalities, should prioritise improving road infrastructure, particularly in key tourist corridors, to enhance travel experiences and reduce negative sentiment.

Second, hospitality businesses, tourism training institutions, and industry associations are encouraged to address service quality through ongoing staff training and the implementation of consistent hospitality standards. Third, the Namibia Tourism Board, travel agencies, and digital service providers should collaborate to enhance digital communication by providing real-time, accurate information on accommodations and tourist experiences, thereby supporting better visitor planning and satisfaction. Fourth, marketing agencies, conservation organisations, and tourism boards should promote eco-tourism by highlighting Namibia's sustainable practices more prominently in marketing materials to attract environmentally conscious travellers. Lastly, tourism stakeholders, including government departments, private sector operators, and data analytics teams, should establish continuous feedback loops by monitoring platforms like TripAdvisor to track tourist sentiments and inform service improvements.

6.4. Assumptions and Limitations of the Study

Despite its contributions, the study faced several limitations. The analysis relied exclusively on TripAdvisor reviews due to restricted access to APIs from platforms such as Twitter and Instagram, which limited the diversity of data sources.

Further limitations include potential biases in online reviews, the generalisability of findings beyond TripAdvisor users, and the inherent challenges in interpreting nuanced sentiments, such as sarcasm, cultural context, and ambiguity, despite the use of advanced models like RoBERTa. Lastly, the reliance on manually labelled data for model validation may not capture the full variability present in broader tourist narratives.

6.5. Future Research

Building upon this study's methodological design, future research should integrate cross-platform sentiment fusion to capture diverse audience segments across multiple digital ecosystems. Expanding the dataset to include multilingual reviews (e.g., Afrikaans, German, Portuguese) and leveraging context-aware transformer models such as GPT-based classifiers will promote greater linguistic inclusivity and contextual precision.

Moreover, incorporating image and video sentiment analysis through computer vision and multimodal deep learning techniques can offer deeper insights into emotional tone and visual storytelling embedded within user-generated content. Finally, the development of real-time tourism sentiment dashboards, integrating LDA-driven topic trends with evolving sentiment polarity shifts, will provide actionable, data-driven intelligence for tourism authorities and private operators in Namibia. Such enhancements will advance both academic inquiry and the practical application of AI in sustainable destination management and promotion.

6.6. Concluding Remarks

While this study's core contribution is applied, it establishes a methodological foundation for advanced tourism sentiment analytics in Africa. By combining sentiment classification with LDA topic modelling and evaluating models through metrics such as AUC and F1-score, the study contributes both technical depth and contextual relevance.

This integrated framework can be expanded across languages, modalities, and regions, forming the basis for a new generation of AI-enhanced tourism intelligence systems tailored to developing contexts.

REFERENCES

Alaei, A. R., Becken, S., & Stantic, B. (2019). Sentiment Analysis in Tourism: Capitalizing on Big Data. In *Journal of Travel Research* (Vol. 58, Issue 2, pp. 175–191). SAGE Publications Ltd. https://doi.org/10.1177/0047287517747753

Atqia, N. L., Suryani, E., & Wiharto. (2023). Utilization of Body Posture Feature Extraction with Movenet for Silat Movement Classification Using Artificial Neural Network (ANN). In *2023 6th International Conference on Information and Communications Technology (ICOIACT), Information and Communications Technology (ICOIACT), 2023 6th International Conference on* (pp. 372–377). IEEE. https://doi.org/10.1109/ICOIACT59844.2023.10455800

Balamurali, O., Abhishek Sai, A. M., Karthikeya, M., & Anand, S. (2023). Sentiment Analysis for Better User Experience in Tourism Chatbot using LSTM and LLM. *2023 9th International Conference on Signal Processing and Communication, ICSC 2023*, 456–462. https://doi.org/10.1109/ICSC60394.2023.10441148

Bandi Tanner, M., & Hämmerli, S. J. (2018). Reviewing TripAdvisor and Co.: a quality analysis of hotel review sites. *Anatolia*, *29*(4). https://doi.org/10.1080/13032917.2018.1451350

Birjali, M., Kasri, M., & Beni-Hssane, A. (2021a). A comprehensive survey on sentiment analysis: Approaches, challenges and trends. *Knowledge-Based Systems*, *226*, 107134. https://doi.org/https://doi.org/10.1016/j.knosys.2021.107134

Birjali, M., Kasri, M., & Beni-Hssane, A. (2021b). A comprehensive survey on sentiment analysis: Approaches, challenges and trends. *Knowledge-Based Systems*, *226*, 107134. https://doi.org/https://doi.org/10.1016/j.knosys.2021.107134

Bosnjak, M., Ajzen, I., & Schmidt, P. (2020). The theory of planned behavior: Selected recent advances and applications. In *Europe's Journal of Psychology* (Vol. 16, Issue 3, pp. 352–356). PsychOpen. https://doi.org/10.5964/ejop.v16i3.3107

Christina Lederle. (2020). *Tourism landscape & sense of place*.

Damar, C., Rizal Isnanto, R., & Widodo, A. P. (2024). Review Of Systematic Literature About Sentiment Analysis Techniques. In *Tuijin Jishu/Journal of Propulsion Technology* (Vol. 45, Issue 1).

Dinara, D., & Savvas, P. (2025). *Social Influence Theory*. https://open.ncl.ac.uk

Dobilas, S. (2022). LSTM Recurrent Neural Networks — How to Teach a Network to Remember the Past. *Https://Towardsdatascience.Com/Lstm-Recurrent-Neural-Networks-How-to-Teach-a-Network-to-Remember-the-Past-55e54c2ff22e*.

F. Mehraliyev, I. Chan, & A. Kirilenko. (2021). Sentiment analysis in hospitality and tourism: a thematic and methodological review. *International Journal of Contemporary Hospitality Management*.

Feilzer, M. Y. (2010). Doing mixed methods research pragmatically: Implications for the rediscovery of pragmatism as a research paradigm. *Journal of Mixed Methods Research*, *4*(1), 6–16. https://doi.org/10.1177/1558689809349691

Firdaus, R., & Afdholy, N. (2023). How the Media Constructs Namibian Tourism as Escapism for Global Traveler in Post-Pandemic Era. *Journal of Namibian Studies*, 33.

Flores-Ruiz, D., Elizondo-Salto, A., & Barroso-González, M. D. L. O. (2021). Using social media in tourist sentiment analysis: A case study of andalusia during the Covid-19 pandemic. *Sustainability (Switzerland)*, *13*(7). https://doi.org/10.3390/su13073836

Haque, S., Eberhart, Z., Bansal, A., & McMillan, C. (2022). Semantic Similarity Metrics for Evaluating Source Code Summarization. *IEEE International Conference on Program Comprehension*, *2022-March*, 36–47. https://doi.org/10.1145/nnnnnnn.nnnnnnn

Hasan, M., Rahman, A., Razaul, M., Khan, M., & Islam, M. (2020). *Normalized Approach to Find Optimal Number of Topics in Latent Dirichlet Allocation (LDA)*.

Ibrahim Sabuncu. (2021). Understanding Tourist Perceptions and Expectations During Pandemic Through Social Media Big Data. *Handbook of Research on the Impacts and Implications of COVID-19 on the Tourism Industry*.

Ligthart, A., Catal, C., & Tekinerdogan, B. (2021). Systematic reviews in sentiment analysis: a tertiary study. *Artificial Intelligence Review*, *54*(7), 4997–5053. https://doi.org/10.1007/s10462-021-09973-3

Liu, Y. (2024). Social Media Tourism Marketing Optimization: Improving Promotion Effectiveness and Broadening the Consumer Base. *Advances in Economics, Management and Political Sciences*, *80*(1), 345–354. https://doi.org/10.54254/2754-1169/80/20241906

Madden, K., Rashid, B., & Zainol, N. A. (2016). Beyond the motivation theory of destination image. *Tourism and Hospitality Management*, *22*(2), 247–264. https://doi.org/10.20867/thm.22.2.1

Mao, Y., Liu, Q., & Zhang, Y. (2024). Sentiment analysis methods, applications, and challenges: A systematic literature review. In *Journal of King Saud University - Computer and Information Sciences* (Vol. 36, Issue 4). King Saud bin Abdulaziz University. https://doi.org/10.1016/j.jksuci.2024.102048

Mehraliyev, F., Chan, I. C. C., & Kirilenko, A. P. (2022). Sentiment analysis in hospitality and tourism: a thematic and methodological review. In *International Journal of Contemporary Hospitality Management* (Vol. 34, Issue 1, pp. 46–77). Emerald Group Holdings Ltd. https://doi.org/10.1108/IJCHM-02-2021-0132

Mukhopadhyay, S., Jain, T., Modgil, S., & Singh, R. (2022). Social media analytics in tourism: a review and agenda for future research. *Benchmarking: An International Journal*, *30*. https://doi.org/10.1108/BIJ-05-2022-0309

Muley, K., & Zainab Irfan DrKirti Muley, K. (2023). *CHATBOT ADOPTION IN TRAVEL AND TOURISM SERVICES*. 25.

Nguyen, H., Veluchamy, A., Diop, M., Al Khaimah Academy, R., Diop, M. L., Iqbal, R., Towers, J., Hisn Rd, A., & Al Khaimah, R. (2018). Comparative study of sentiment analysis with product reviews using machine learning and lexicon-based approaches. *Scholar.Smu.EduH Nguyen, A Veluchamy, M Diop, R IqbalSMU Data Science Review, 2018•scholar.Smu.Edu*, *1*(4).

Pearce. P, & Oktadiana. H. (2020). From Analysis to Action. *Delivering Tourism Intelligence*.

Pop, R. A., Săplăcan, Z., Dabija, D. C., & Alt, M. A. (2022). The impact of social media influencers on travel decisions: the role of trust in consumer decision journey. *Current Issues in Tourism*, *25*(5), 823–843. https://doi.org/10.1080/13683500.2021.1895729

Puh, K., & Bagić Babac, M. (2023). Predicting sentiment and rating of tourist reviews using machine learning. *Journal of Hospitality and Tourism Insights*, *6*(3). https://doi.org/10.1108/JHTI-02-2022-0078

Ramasamy, L. K., Kadry, S., Nam, Y., & Meqdad, M. N. (2021). Performance analysis of sentiments in Twitter dataset using SVM models. *International Journal of Electrical and Computer Engineering*, *11*(3), 2275–2284. https://doi.org/10.11591/ijece.v11i3.pp2275-2284

Rodríguez-Ibánez, M., Casánez-Ventura, A., Castejón-Mateos, F., & Cuenca-Jiménez, P. M. (2023). A review on sentiment analysis from social media platforms. In *Expert Systems with Applications* (Vol. 223). Elsevier Ltd. https://doi.org/10.1016/j.eswa.2023.119862

Samuel, N. (2022, September 22). *FNB Namibia Twitter Sentiment analysis using Vader and RoBERTa | by Samuel Ndungula | Medium*. Https://Samuelndungula.Medium.Com/Sentiment-Analysis-Using-Vader-and-Roberta-5279ba312d70. https://samuelndungula.medium.com/sentiment-analysis-using-vader-and-roberta-5279ba312d70

Saunders, M. N. K. ., Lewis, Philip., & Thornhill, Adrian. (2020). *Research methods for business students*. Pearson.

Schorr, A. (2023). The Technology Acceptance Model (TAM) and its Importance for Digitalization Research: A Review. In *International Symposium on Technikpsychologie (TecPsy) 2023* (pp. 55–65). Sciendo. https://doi.org/10.2478/9788366675896-005

Siri, A., Ebba, G., & Anton, N. (2020). *Tourism Behavior-A study of how tourism behavior effects the social carrying capacity at a destination*. https://www.diva-portal.org/smash/get/diva2:1386555/FULLTEXT01.pdf

Taherdoost, H., & Madanchian, M. (2023). Artificial Intelligence and Sentiment Analysis: A Review in Competitive Research. In *Computers* (Vol. 12, Issue 2). MDPI. https://doi.org/10.3390/computers12020037

Xu, D., Tian, Z., Lai, R., Kong, X., Tan, Z., & Shi, W. (2020). Deep learning based emotion analysis of microblog texts. *Information Fusion*, *64*, 1–11. https://doi.org/10.1016/j.inffus.2020.06.002

Yonatan, M.-C., & Edwin R, V. (2024). *Data clean techniques for sentiment classification. | Download Scientific Diagram*. https://www.researchgate.net/figure/Data-clean-techniques-for-sentiment-classification_tbl2_386588952

Youvan, D. (2024a). *Understanding Sentiment Analysis with VADER: A Comprehensive Overview and Application*. https://doi.org/10.13140/RG.2.2.33567.98726

Youvan, D. (2024b). *Understanding Sentiment Analysis with VADER: A Comprehensive Overview and Application*. https://doi.org/10.13140/RG.2.2.33567.98726

Yuan, Y., Chan, C. S., Eichelberger, S., Ma, H., & Pikkemaat, B. (2022a). The effect of social media on travel planning process by Chinese tourists: the way forward to tourism futures. *Journal of Tourism Futures*. https://doi.org/10.1108/JTF-04-2021-0094

Yuan, Y., Chan, C. S., Eichelberger, S., Ma, H., & Pikkemaat, B. (2022b). The effect of social media on travel planning process by Chinese tourists: the way forward to tourism futures. *Journal of Tourism Futures*. https://doi.org/10.1108/JTF-04-2021-0094

Zucco, C., Calabrese, B., Agapito, G., Guzzi, P. H., & Cannataro, M. (2020). Sentiment analysis for mining texts and social networks data: Methods and tools. In *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery* (Vol. 10, Issue 1). Wiley-Blackwell. https://doi.org/10.1002/widm.1333

APPENDIX

Appendix A: Ethics Clearance Certificate

NAMIBIA UNIVERSITY OF SCIENCE AND TECHNOLOGY

FACULTY RESEARCH ETHICS COMMITTEE (F-REC)

DECISION/FEEDBACK ON THE RESEARCH PROPOSAL

Dear Johannes Shoowa

RESEARCH TOPIC: SOCIAL MEDIA SENTIMENT ANALYSIS FOR TOURISM PROMOTION IN NAMIBIA

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 - 21/24)**

Re: Ethical screening application No: FREC - 21/24

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 questfor 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,

Prof. Suama L Hamunyela
Chairperson: Faculty Ethics Screening
CommitteeTel: +264-61-207-2922
CC: Co-supervisor:

NAMIBIA UNIVERSITY OF SCIENCE AND TECHNOLOGY
P/Bag 13388 Windhoek NAMIBIA
2024 -05- 15
OFFICE OF THE DEAN
FACULTY OF COMPUTING & INFORMATICS

Appendix B: Data Access Authorisation

The data used in this study were collected from publicly accessible TripAdvisor pages using Instant Data Scraper, a browser extension tool. No personally identifiable information was collected, and all data used complied with TripAdvisor's terms of service regarding public content use for academic research purposes.

Johannes Shoowa
Master of Data Science Student
Namibia University of Science and Technology (NUST)
Jshoowa.js@gmail.com
+264 81 424 4446

10 May 2025

TripAdvisor LLC
400 1st Avenue
Needham, MA 02494
United States

Subject: Request for Permission to Collect Data from TripAdvisor for Academic Research

Dear TripAdvisor Team,

I hope this message finds you well.

My name is Johannes Shoowa, and I am currently pursuing a Master of Data Science degree at the Namibia University of Science and Technology (NUST), Namibia. As part of the program, I am undertaking a research thesis titled "*Social Media Sentiment Analysis for Tourism Promotion in Namibia.*"

The objective of this research is to explore how user-generated content and reviews on online travel platforms such as TripAdvisor can be analysed using Natural Language Processing (NLP) techniques to extract sentiment insights that can help inform tourism promotion strategies in Namibia.

To carry out this research, I am kindly requesting your permission to collect a dataset of publicly available user reviews and related metadata (such as location, date, and star ratings) from your platform, (Namibian Reviews). The data will be used solely for academic and non-commercial purposes, with all ethical considerations and data protection protocols strictly observed. No user-identifiable information will be used, stored, or shared in the study.

I would be grateful if you could advise on the most appropriate and compliant method for accessing this data, whether through a public API, formal data request process, or other means you deem suitable.

If possible, I would appreciate it if you could respond with a formal letter of permission or your official position regarding this request, as it may be required by my institution for ethics clearance purposes.

Please let me know if you require any further documentation, such as an official request from NUST or proof of ethical approval.

Thank you for your time and consideration. I look forward to your response and hope to contribute meaningful insights to the tourism sector through this study.

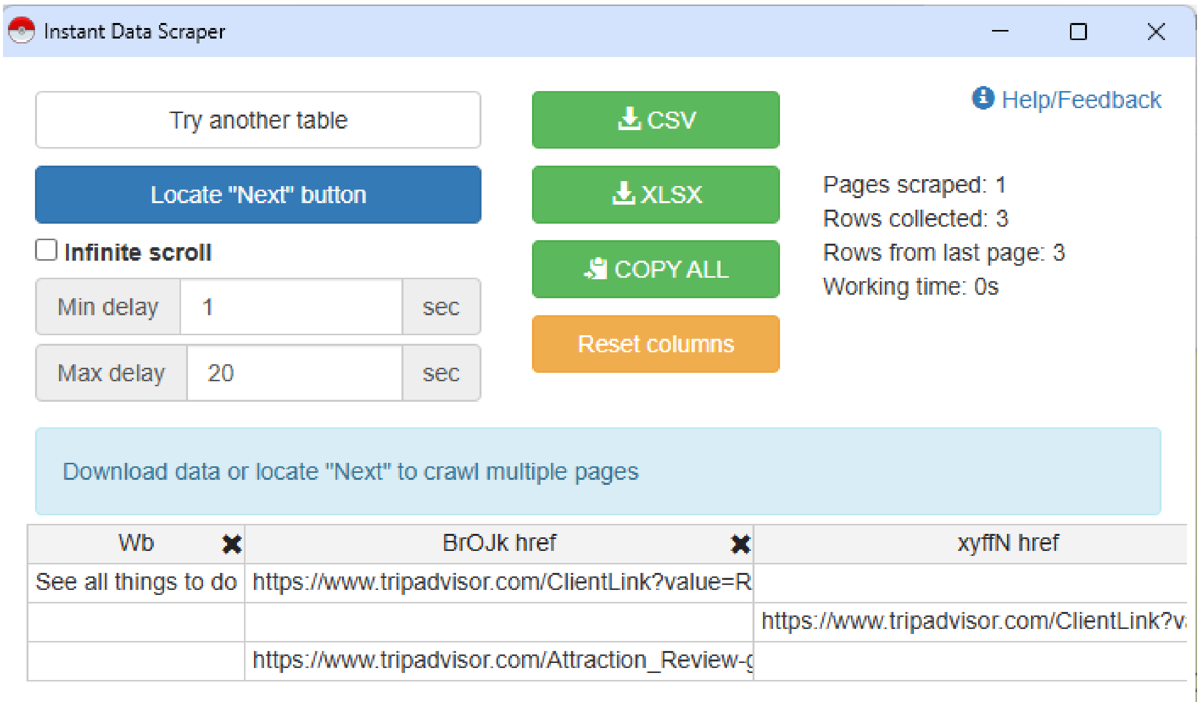
Yours sincerely,

Johannes Shoowa

Master of Data Science Student
Namibia University of Science and Technology (NUST)
Jshoowa.js@gmail.com
+264 81 424 4446

Appendix C: Technical Resources and Tools

1. Data Scraping Tool



2. Sample Data

URL: https://github.com/Jshoowa/sentiment-analysis-tourism-namibia/blob/main/Sentiment%20Analysis%202025/results_df.csv

This file contains a representative sample of the cleaned and pre-processed dataset used in the analysis. It includes review texts along with sentiment labels, tourist attributes, and other relevant variables derived during the exploratory and sentiment analysis phases. The dataset offers transparency and reproducibility for the results discussed in Chapter 5.

3. GitHub link to source code

URL: https://github.com/Jshoowa/sentiment-analysis-tourism-namibia/blob/main/Sentiment%20Analysis%202025/namibia_tourists_sentiment_analysis_2025_Final.ipynb

This Jupyter Notebook contains the complete source code for the sentiment analysis workflow implemented in this study. It includes modules for data cleaning, sentiment classification using VADER and RoBERTa, thematic analysis through topic modelling, and generation of visual outputs such as word clouds and sentiment distribution graphs. The notebook provides full traceability of the analytical methods and facilitates reproducibility of the research findings.