The Impact of AI-Powered Technologies on Customer Satisfaction and OperationalEfficiency in the Hospitality Industry

MSc Research Project
Data Analytics

Joyal Johnson
Student ID: x22170936

School of Computing
National College of Ireland

Supervisor: Harshani Nagahamulla

National College of Ireland

MSc Project Submission Sheet

School of Computing

Student Name: Joyal Johnson

Student ID: X22170926

Programme: MSc. Data Analytics **Year:** 2023-2024

Module: MSc. Research Project

Supervisor: Harshani Nagahamulla

Submission Due Date: 14/12/2023

Project Title: The Impact of AI-Powered Technologies on Customer Satisfaction and OperationalEfficiency in the Hospitality Industry

Word Count: **Page Count:23**

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.
ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature: Joyal Johnson

Date: 14/12/2023

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST

Attach a completed copy of this sheet to each project (including multiple copies)	□
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	□
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	□

Assignments that are submitted to the Programme Coordinator Office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Table of Contents

The Impact of AI-Powered Technologies on Customer Satisfaction and OperationalEfficiency in the Hospitality Industry

Joyal Johnson
X22170936

Abstract

The aim of this study is to investigate the utilisation of AI-powered technologies in the hotel sector and assess their ability to enhance customer satisfaction and operational efficiency. The purpose of this study is to deepen the knowledge of the consequences of integrating AI technology in the worldwide hospitality sector, with a specific focus on the cultural and geographical complexities of the hotel industry. The primary objective of this study is to fill the existing gaps in the literature and identify the specific obstacles and potential advantages of using artificial intelligence in hospitality facilities. The study will rely on secondary qualitative data collection. This research study aims to offer valuable insights and practical solutions for the deployment of AI technology in the hotel business. The results are expected to provide insight into the correlation between consumer satisfaction and the use of AI-generated suggestions and services.

Keywords: Hospitality, Artificial Intelligence, Consumer Satisfaction, Personalized Recommendations, Operational Efficiency, Predictive Analysis.

1 Introduction

1.1 Background

The current assessment of robots and artificial intelligence (AI) in corporate services is having a substantial impact on business dynamics and operational norms. The impact of artificially intelligent robotic virtual agents (AIRVA) on consumer behaviour in the tourist and hospitality industries has not been well studied (Ukpabi et al., 2018). The introduction of digitization has enabled the construction of a robust infrastructure that can achieve sustainability in all scenarios (Ivanov et al., 2019). The hotel sector heavily depends on a reliable digital technology infrastructure to effectively gather and analyse top-notch consumer feedback. The utilisation of digital technology has proven to enhance hotel services by enabling efficient and well-informed decision-making (Sharma, 2021). Although there have been many studies exploring the effects of digital technologies on the hotel industry, there is a notable gap in the literature when it comes to investigating feedback systems based on digital technology in the hospitality sector (Ameen et al., 2021).

1.2 Research Aim

How effectively can AI-powered technologies, such as personalized recommendations, contribute to improving operational efficiency and consumer satisfaction in the Irish hospitality industry ?

1.3 Research Questions

RQ1: How efficient and accurate AI-powered technologies impact customer satisfaction and operational efficiency within the hospitality industry?

This research question explores the multifaceted impact of AI-powered technologies on customer satisfaction and operational efficiency in the hospitality industry. It considers the utilization of user-generated text reviews, text tag keywords indicating the purpose of the stay, and geo-locations for hotel selection, along with review ratings. The investigation delves into the cultural and geographical nuances, aiming to understand how these factors influence the effectiveness and adoption of AI technologies. The study seeks to identify both challenges and benefits associated with the implementation of AI in hotel facilities, providing a holistic perspective on the implications of artificial intelligence for enhancing the guest experience and operational processes in the hospitality sector.

2 Related Work

The literature survey explores the intersection of AI-powered technologies, customer satisfaction, and operational efficiency within the hospitality sector. Examining existing research on the adoption of AI in hospitality, customer preferences, and recommendation systems sets the stage for understanding the landscape. Noteworthy studies on Natural Language Processing (NLP) and sentiment analysis in hotel reviews contribute to the foundation, guiding the investigation into the potential impact of the developed system on enhancing user experience and operational processes in the hospitality industry.

2.1 Impact of Reviews on the Hotel Industry

The pervasive influence of the Internet has revolutionized daily life, impacting our holidays, communication patterns, and purchasing behaviors. This shift is evident in the proliferation of user-generated content on platforms like social media, review websites, blogs, and product fan pages. Notably, this transformation has redefined how individuals plan vacations and select accommodations, with hotel review websites emerging as a popular medium for sharing experiences and recommendations. In Malaysia, a vital player in the global hotel industry, understanding customer satisfaction is paramount. (Nicholas, et al., (2017)) studied this utilizes data from platforms like Trip Advisor, employing text analytics to discern factors influencing customer satisfaction and expectations within the dynamic landscape of hotel reviews.

To enhance the summarization of online hotel reviews by addressing the challenge of manually processing large volumes of reviews. The methodology involves constructing classifiers to identify helpful reviews, categorizing sentences by hotel features, and analysing sentiment to generate concise summaries. Unlike previous studies focusing solely on text summarization, this approach prioritizes selecting helpful reviews before categorizing sentences based on hotel features. (Chih-Fong Tsai et al., 2020) Experimental results on a TripAdvisor dataset show that the Random Forest (RF) classifier is effective, achieving over 70% accuracy and 80% Area Under the Curve (AUC). The research emphasizes the importance of high-quality summaries in assisting travel decisions and demonstrates the practicality of the constructed classifier. In the era of rapid internet growth, tourists heavily rely on online reviews for hotel information. (Arif Abdurrahman Farisi1 et al., 2019) This research tackles information overload through sentiment analysis, utilizing the Multinomial Naïve Bayes Classifier to categorize positive and negative opinions. By optimizing preprocessing, feature extraction, and selection, the approach attains an average F1-Score exceeding 91% with 10-fold cross-validation. Given the prevalence of social media, sentiment analysis becomes crucial for extracting opinions from user reviews. (Anis, Sarah et al., 2020) studied is introduces a sentiment detection approach using the Fuzzy C-means clustering algorithm on hotel reviews. Various classification techniques, including Naïve Bayes, K-Nearest Neighbour, Support Vector Machine, Logistic Regression, and Random Forest, are compared. The study explores sentiment analysis in the context of social media-driven tourism, providing insights into effective machine-learning models. As social media becomes integral globally, sentiment analysis offers a means to analyse vast data. Focusing on user comments and hotel reviews, (K. Shifullah et al., 2022) this research employs text pre-processing techniques and evaluates supervised machine learning algorithms such as Support Vector Machine, Naïve Bayes, Random Forest, and Logistic Regression. Logistic Regression proves the most accurate, achieving 86-89% accuracy, demonstrating the effectiveness of machine learning in predicting sentiment from hotel reviews. Hotel review analysis involves extracting insights from customer feedback using techniques like sentiment analysis and natural language processing. (Anis, S., et al., 2021) The abstract summarizes key findings, including common positive/negative aspects, frequently mentioned amenities, and sentiment trends. This concise overview aids stakeholders in making data-driven decisions to enhance hotel offerings. Considering the common practice of browsing online reviews before selecting a hotel, multiple criteria decision-making models provide valuable guidance. (Nie, Ru-xin, et al.,2020) studied is addresses gaps in existing models by considering textual reviews alongside customer ratings. A novel hotel selection model is proposed, leveraging a semantic mapping function and an evidence theory-based fusion method. The model is tested, demonstrating reliability and improved capability in managing conflicting attitudes and providing comprehensive hotel descriptions.

2.2 Trends in Hospitality Industry and the Types of Hotels

Friendliness is a crucial and fundamental aspect when it comes to attracting travellers. The frequent turnover of vacationers leads to the enhancement and advancement of all the available resources in a tourist destination, stimulating growth and development in the hospitality industry and elevating it to a higher standard of quality and offerings. Recently, the increasing popularity of a certain destination among travellers worldwide has become a common topic for research by numerous authors in the field of tourism and hospitality. In order to align with the rapid growth of hotel industry and overall social progress, it is necessary to address several questions regarding the benefits and drawbacks of this development.

2.3 AI-Powered Technologies in Hospitality

The digital revolution has brought about substantial transformations in hotel administration and operations. Digital technology is widely acknowledged as the principal driver for enhancing efficiency and creating profits in the business. The World Tourism Organisation (WTO) predicts that the world economy would maintain a steady growth rate of 2.2 percent year until 2030, as stated by Narayan et al. (2022). Hence, it is expected that the quantity of global tourists would experience a substantial rise, going from 1235 million in 2016 to 1800 million in 2026 (Ameen et al., 2021). To enhance an organization's capacities, digital transformation necessitates the integration of several technologies linked to information, computing, communication, and networking. This merger has a profound impact on the fundamental attributes of the organisation. The digital technologies that have the potential to significantly alter hotel management and value chains include Internet of Things (IoT), artificial intelligence (AI), robotics, blockchain, big data analytics, digital twins, augmented reality (AR), and virtual reality (VR). These solutions have the capacity to enhance various aspects of hotels, such as service, customer relationships, order process, competitiveness, service quality, flexibility, resource utilisation, innovation, capacity, and resource management (He and Zhang, 2023).

2.4 The Implementation of AI-Powered Technologies in the Hospitality Industry and its Impact on Consumer Satisfaction

The utilisation of artificial intelligence in the hotel business has sparked significant attention. A team of academics has commenced an inquiry into the prospective influence of artificial intelligence on the hotel sector, specifically concentrating on the extent of consumer satisfaction that AI may engender. Tan (2021) investigated the frequency of customised AI-powered suggestions in the hotel industry in their research. The findings suggest that visitors experienced greater overall satisfaction when they were provided with tailored recommendations for activities and dining choices. Research indicates that artificial intelligence (AI) has the potential to enhance organisations' ability to deliver superior service to their clients by accurately detecting and fulfilling their needs (Sharma, 2021). The study revealed that utilising an AI-powered chatbot to address consumer inquiries resulted in a substantial reduction in response durations and a notable increase in satisfaction ratings. This study proposes that to provide authentic and personalised service, the hospitality industry must achieve a suitable equilibrium between AI automation and human interaction (Tan and Wright, 2022)

2.5 AI-Powered Technologies for Enhancing Hospitality and Operational Efficiency

There is a lack of academic research examining the effects of artificial intelligence-driven technology on operational efficiency in the hotel industry in Ireland. Tan and

Wright (2022) conducted a study to examine the application of AI-driven predictive analysis in forecasting hotel demand. As per Campione (2021), the research findings demonstrated that accurately predicting future demand led to improved resource allocation, better inventory management, and optimised pricing methods. Hotels achieved cost reduction and improved operational efficiency by strategically optimising resource allocation based on guest demand. The findings indicate that employing data-driven decision-making has a positive impact on both supply chain management and menu enhancement. Tan (2021) suggests that restaurants might utilise artificial intelligence technology to assess client feedback and preferences, so enabling data-driven decision-making processes that improve both customer satisfaction and operational effectiveness.

2.6 Exploring Reviews and Market Positioning of AI-Powered Technologies

The scrutiny lies on the influence of user-generated reviews on social media sites, namely TripAdvisor, within the hospitality business. (Neirotti, Paolo et al., 2016) The study encompasses the years 2004 to 2012 and examines data from a total of 240 small and medium-sized hotels. The inquiry centres on the correlation between online ratings derived from user-generated reviews and the financial success of hotels. Remarkably, the findings demonstrate a subtle influence, where there are favourable consequences on the increase in revenue, but at the same time, unfavourable consequences on the decline in gross profit margins. The competition dynamics in the hospitality industry are undergoing a transformation, with a shift from focusing on individual profit margins to prioritising higher volumes and room occupancy rates. This change is driven by the growing importance of user-generated ratings. Online retailers are recognised as the main recipients of the value generated by social media features, which restricts the favourable effect on hotels' overall profitability. Nevertheless, strategic factors such as superior star ratings, reduced local competition, and expansion into less popular areas are found to amplify the advantages gained from online presence, resulting in increased gross and net profitability. The study continues by emphasising the management implications, emphasising the significance of strategically utilising social media characteristics to generate economic value and protect profit margins in response to intermediation by online platforms and distributors. The results emphasise the complex relationships between online presence and financial success in the hospitality sector, offering significant knowledge for hotel management tactics.

2.7 Research Niche

This research focuses on doing a thorough analysis of the implementation of AI-powered solutions in the worldwide hotel sector. The main objectives are to improve guest experiences and achieve financial savings. This research stands out from previous studies by recognising the importance of cultural, regional, and market aspects in influencing the adoption and use of AI technologies in hotels around the world. Our objective is to offer a comprehensive comprehension of the sector by considering these elements, analysing the viewpoints of consumers, management, and employees. By employing a multi-faceted approach, we may pinpoint deficiencies in existing literature and rectify them by providing valuable perspectives on the distinct challenges and advantages encountered by hotels in various geographical areas. Furthermore, our research has practical significance as it provides valuable insights for making strategic decisions in the hospitality industry and enables the efficient integration of AI solutions to enhance guest pleasure and operational effectiveness. This research aims to enhance knowledge in the field of hospitality management and the integration of AI technology. It provides useful insights that could encourage innovation and enhance the industry's performance.

3 Research Methodology

Both secondary qualitative data gathering methods will be used in this investigation. A mixed technique, which combines primary and secondary qualitative approaches, offers a thorough analysis of the subject under investigation. The literature research will offer the theoretical framework, while the gathering of primary data will analyze stakeholders' real-world experiences and viewpoints. The best way to comprehend the contextual relevance of AI technology in the hotel sector is to use qualitative research approaches. Researchers will possess the chance to discover cultural, regional, and sector-specific variations that might impact the acceptance of AI-driven solutions.

3.1 Dataset

The dataset, named "Hotel_Reviews.csv," is stored in the "input" directory and is loaded into a Pandas Data Frame. This dataset likely contains information related to hotel reviews, and the structure of the Data Frame implies a tabular organization with labeled rows and columns. Each row likely represents a specific entry, possibly a hotel review, and each column likely represents a different attribute or feature associated with the reviews, such as hotel details, reviewer information, review content, and ratings. The use of Pandas indicates that the dataset is being prepared for analysis and manipulation in Python, suggesting that it may be employed for tasks like exploratory data analysis or machine learning.

3.2 Preliminary Analysis

The preliminary analysis of a dataset stored in a Pandas Data Frame named 'df' with 515,738 rows and 17 columns. The dataset includes information about hotel reviews with columns such as Hotel_Address, Additional_Number_of_Scoring, Review_Date, Average_Score, Hotel_Name,Reviewer_Nationality,Negative_Review,review_Total_Negative_Word_Counts ,Total_Number_of_Reviews,Positive_Review,Review_Total_Positive_Word_Counts, Total_Number_of_Reviews_Reviewer_Has_Given,Reviewer_Score,Tags,days_since_review , lat and lng. The dataset has 515,738 rows and 17 columns. This preliminary analysis is useful for getting an overview of the dataset's structure and content before performing more in-depth exploration or analysis.

3.3 Data Pre-processing

The analysis of the dataset reveals that columns 'lat' and 'lng' contain 3,268 null values each. A glimpse into the dataset's cleanliness is gained through the examination of unique hotel names and reviewer nationalities. There are 1,492 unique hotel names, and a seemingly duplicated line, likely intended for reviewer nationalities, mistakenly repeats the count of unique hotel names. Despite this duplication, the information underscores the dataset's diversity in terms of both hotel names and reviewer nationalities. The presence of null values in the geographical columns suggests a need for handling missing data, while the enumeration of unique entities provides valuable insights into the dataset's structure and content.

3.4 Data processing

The process of converting raw data into information that may be used. I've attached a flowchart below that explains the different stages I'll be doing while processing the data.

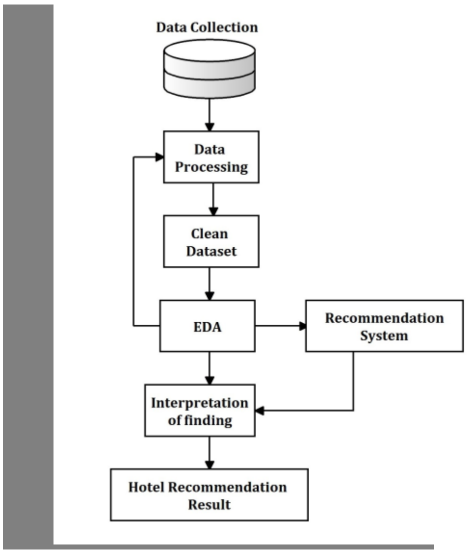


Fig.1 Data processing flowchart

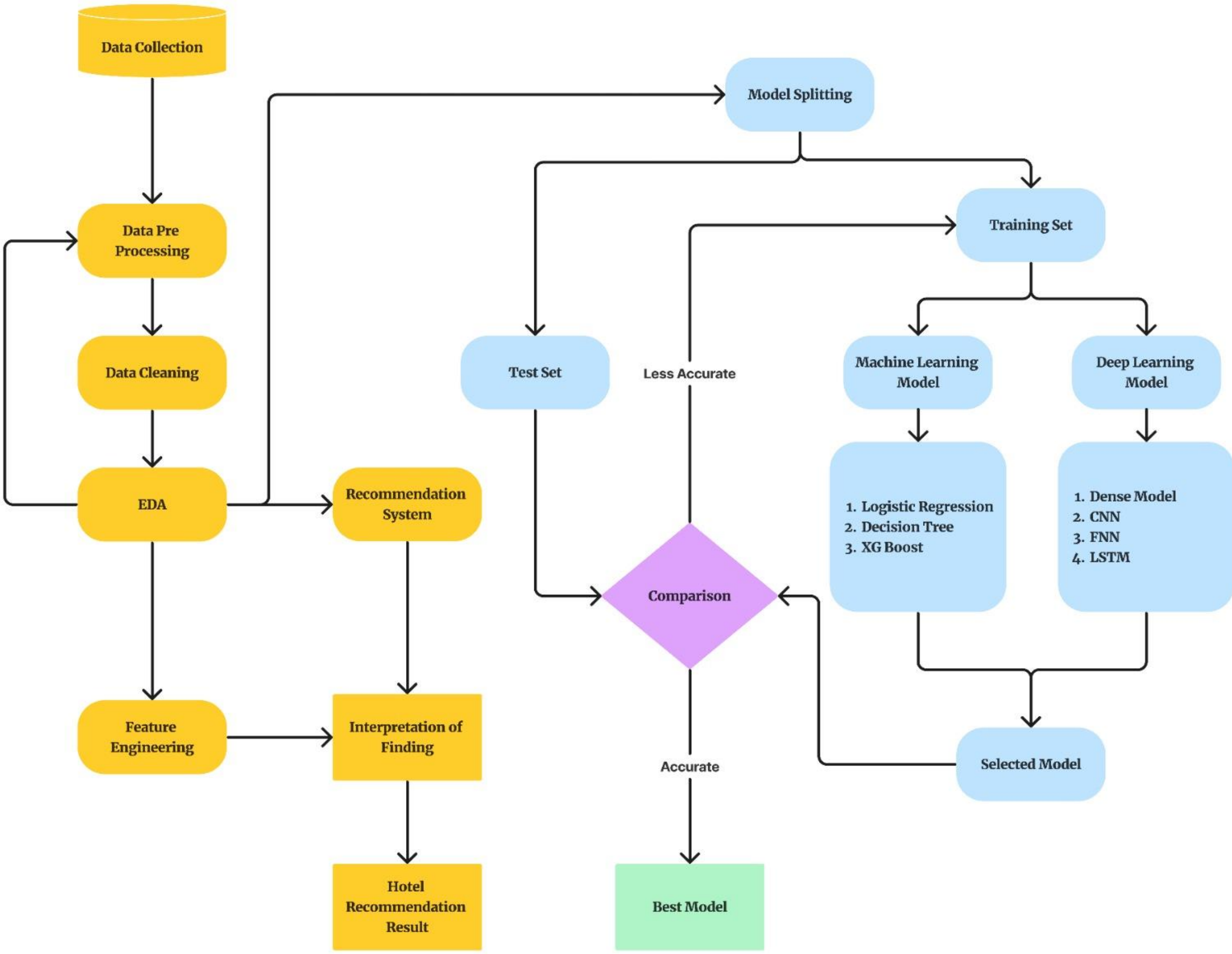
The *Fig* .1 illustrates the steps involved in building a hotel recommendation system. It starts with data collection, where various sources of information about hotels and user preferences are gathered. This data is then processed to clean and prepare it for analysis. Exploratory data analysis (EDA) involves examining the data to identify patterns, trends, and relationships between different variables. Based on the findings from EDA, a recommendation system is developed using appropriate machine learning algorithms. The system is trained on the prepared data and is evaluated to assess its performance.

3.5 Research Resources

Semi-structured interviews will be conducted with prominent figures in the hospitality sector, including managers of hotels, proprietors of restaurants, staff members, and patrons. The aim of these conversations is to acquire a more profound comprehension of people's viewpoints, feelings, and experiences about the application of AI-driven solutions, tailored suggestions, and predictive analysis in the hospitality sector. Semi-structured interviews will be conducted with prominent figures in the hospitality sector, including managers of hotels, proprietors of restaurants, staff members, and patrons.

To obtain insight into the current state of research and advancements in artificial intelligence technologies, personalized recommendations, predictive analysis, customer satisfaction, and operational efficiency in the hospitality industry, a thorough review of academic papers, research articles, industry reports, and other pertinent sources will be carried out. The author plans to accomplish this by reading credible scholarly publications on a variety of pertinent topics, including hotel management, artificial intelligence (AI), consumer behaviour, and operational efficiency. The author will use scholarly resources including Google Scholar, ProQuest, and JSTOR to research the application of AI in hotels.

4 Design Specification



- Because logistic regression is an easy-to-use yet highly successful technique for binary outcome prediction, it is the preferred choice for binary classification jobs. Understanding the effects of various characteristics on the target variable is made easier with the help of Logistic Regression, which yields findings that are simple to read.

- Combining the predictions of several weak learners to improve model performance and resilience is how XGBoost excels at ensemble learning.

- LSTMs are a unique class of RNNs that can learn long-term dependencies, which helps RNNs recall prior events and identify patterns over time to help explain their subsequent estimates. In terms of enhanced machine translation, language modelling, and multilingual language processing, LSTMs set records.

- The most popular use of Convolutional Neural Networks (CNNs, or ConvNets), a kind of deep neural networks, is the analysis of visual vision. Their additional uses include speech recognition, video comprehension, and natural language processing comprehension. Convolutional Neural Networks (CNNs) and Long Short-Term Memory (LSTM) were also used to enhance automatic picture captioning. As you can see, CNN assists with visual analysis, whereas RNN is more of a data processing tool that helps us anticipate our next move.

5 Implementation

5.1 Environmental setup: Install necessary Python libraries such as folium, word cloud, TensorFlow, XGboost using pip.

5.2 Data Preprocessing: Gather and preprocess textual data using NLTK for tasks like tokenization, stemming, and tagging.

- Prepare geospatial data for visualization using Folium.

5.3 Exploratory Data Analysis (EDA): Create word clouds to visualize frequent terms using the Word Cloud library.

- Use TF-IDF to analyse and understand word importance.

Latent Semantic Analysis (LSA): Implement LSA to uncover latent semantic features and relationships between words.

5.4 Machine Learning Models:

5.4.1 Logistic Regression: Implement logistic regression for binary classification tasks. Tune hyperparameters and evaluate model performance.

5.4.2 Decision Tree: Implement Decision Tree for classification tasks Tune hyperparameters such as tree depth and criterion. Evaluate the Decision Tree's performance using appropriate metrics.

5.4.3 XG Boost: Implement XGBoost for tasks like anomaly detection, regression, and classification. Fine-tune parameters and assess model performance.

When it comes to binary classification problems, like categorizing reviews into positive or negative feelings, logistic regression is a great fit. The model has a benefit in that it is easy to understand and straightforward, and the feature coefficients provide precise information on how each feature affects the classification choice. Because of its high processing efficiency, logistic regression works well with big datasets and in scenarios where resource constraints must be taken into consideration. Artificial neural networks (ANNs) are more successful than conventional statistical techniques like regression analysis in predicting binary outcomes, according to experimental data and scholarly study (Hornik et al., 1989). This claim is supported by George Tzougas and Konstantin Kutzkov (2023), who suggest that the use of neural network boosting in logistic regression makes it easier to examine non-linear missing interactions that are not captured by generalized linear models such as logistic regression. Deep Neural Networks are excellent at identifying detailed patterns and complex non-linear correlations in the data. They provide a large-capacity model that can learn hierarchical feature representations on its own. To obtain more accurate and lower test loss findings, the study used LSTM and Feedforward Neural Network models. Thus, we are using some deep learning models such as

5.5 Deep Learning Models

5.5.1 Dense Model**:** Design and implement a dense neural network for tasks like image categorization or natural language processing.

5.5.2 CNN**:** Employ CNNs for tasks like recommendation systems and interpreting structured matrices of user-item interactions.

5.5.3 FNN: Develop and deploy a Feedforward Neural Network (FNN) tailored for sentiment classification. Define the network architecture by specifying the number of layers and nodes in each layer. Strategically choose activation functions for each layer to enhance learning capabilities. Employ backpropagation alongside optimization algorithms, such as stochastic gradient descent, to train the FNN effectively. Evaluate the FNN's performance using relevant metrics to gauge its proficiency in sentiment.

5.5.4 LSTM**:** Implement LSTM networks for tasks involving sequential data such as natural language processing.

5.6 **Integration:** Combine different models and techniques based on project requirements.

- Ensure seamless communication and data flow between components.

5.7 **Model Evaluation and Validation:** Evaluate each model's performance using appropriate metrics.

- Validate models with test datasets to ensure generalizability.

5.8 **Testing and Deployment**: Conduct thorough testing to ensure model robustness.

- Deploy models in a suitable environment which can be any suitable cloud such as GCP, AWS or Azure, considering scalability and performance. The application has to be accessible for the end users.

6 VISUALISATION

6.1 Bar Chart with Nationality review

The visualization bar chart shows the number of hotel reviews by reviewer nationality. The x-axis shows the reviewer nationality, and the y-axis shows the number of reviews.

The bar chart shows that the top 5 nationalities of hotel reviewers are: United Kingdom, United States of America, Australia, Ireland, United Arab Emirates.

These nationalities account for over 50% of all hotel reviews. The remaining 50% of hotel reviews are from reviewers of over 100 different nationalities. Interestingly, the top 5 nationalities of hotel reviewers are also the top 5 countries in terms of the number of international tourist arrivals. This suggests that people from these countries are more likely to travel and stay in hotels than people from other countries.

There are a few possible explanations for this trend. First, these countries are all relatively wealthy and have high disposable incomes. This means that people from these countries have more money to spend on travel and leisure. Second, these countries all have well-developed tourism infrastructures, which makes it easier and more convenient for people to travel to these countries. Third, these countries are all popular tourist destinations with a lot to offer visitors, such as historical sites, natural attractions, and cultural experiences. Overall, the bar chart shows that people from the United Kingdom, United States of America, Australia, Ireland, and United Arab Emirates are the most likely to review hotels.

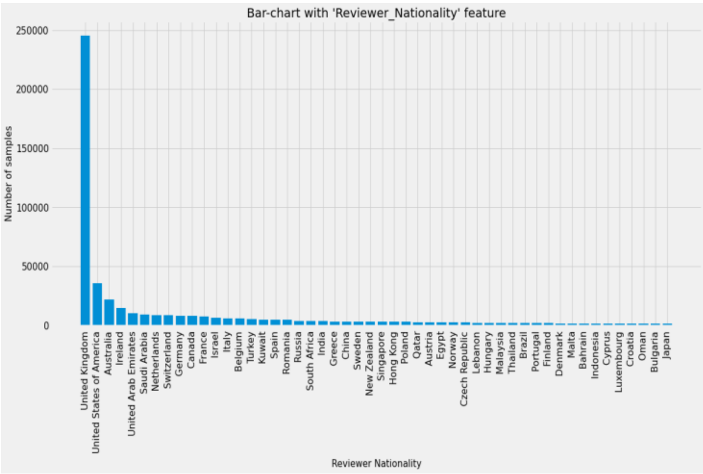


Fig.2 Bar Chart with Nationality review

6.2 Bar Chart with Year Feature

The number of hotel reviews collected from 2015 to 2017 has increased significantly, with 150,000 reviews collected in 2017. This trend is attributed to the internet making it easier for people to find and write reviews, increasing awareness of the importance of reviews when choosing a hotel, and hotels encouraging guests to write reviews through incentives like discounts or free upgrades. The year 2017 had the highest number of hotel reviews collected, likely due to the growing popularity of online travel booking websites and apps. These websites and apps make it easy for people to compare prices and book hotels online, as well as to write reviews of their experiences. This information can be used by hotel managers and potential guests to make informed decisions about which hotels to choose and when to travel. Hotel managers can track trends in the hotel review industry and identify areas for improvement, while potential guests can choose hotels with a good reputation and positive reviews.

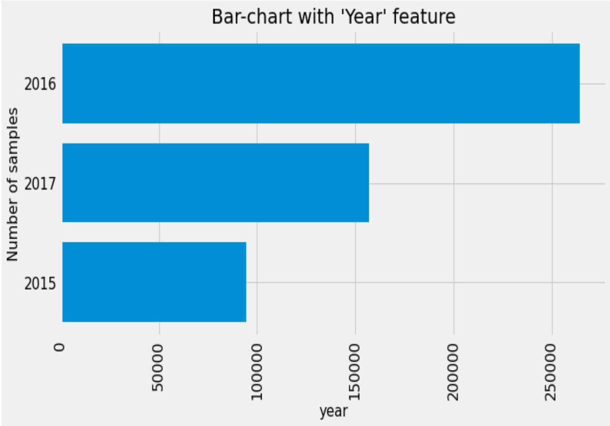


Fig.3 Bar Chart with Year Feature

6.3 Donut-Chart with Year Feature

It represents the year-wise getting of reviews in percentage. It is a donut chart with the year feature, and it shows the following: This means that the highest percentage of reviews were received in 2016, at 51.27%. The lowest percentage of reviews were received in 2015, at 18.33%.

There are a few possible reasons for this trend. First, 2016 was a popular year for travel, and more people may have been leaving reviews of their experiences. Second, online review platforms may have become more popular in 2017, and more people may have been using them to share their feedback. Third, hotels may have become more proactive in encouraging guests to leave reviews in 2015. Overall, the visualization shows that the percentage of reviews received by hotels has increased over time.

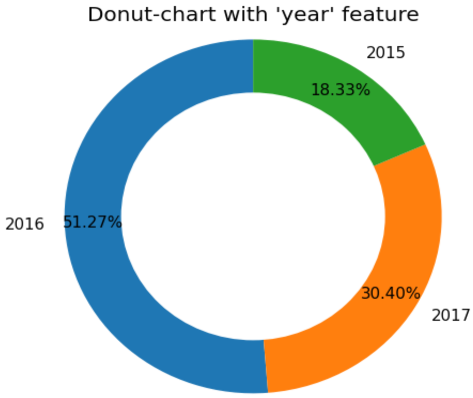


Fig. 4 Donut-Chart with Year Feature

6.4 Hotel Locations Using Folium Map

In our dataset, the features 'lat' and 'lng' represent the latitude and longitude coordinates of hotel locations, respectively. Utilizing the Folium Python library, I've created a geo chart that visually depicts the geographical distribution of hotels worldwide. By plotting these latitude and longitude coordinates on a map, the Folium library enables a dynamic and interactive representation of hotel locations across the globe. This visualization provides a spatial understanding of the distribution of hotels, offering insights into their geographic clustering and dispersion on a world map. The resulting geo chart enhances the exploration and analysis of hotel locations, contributing to a more comprehensive understanding of their global presence.

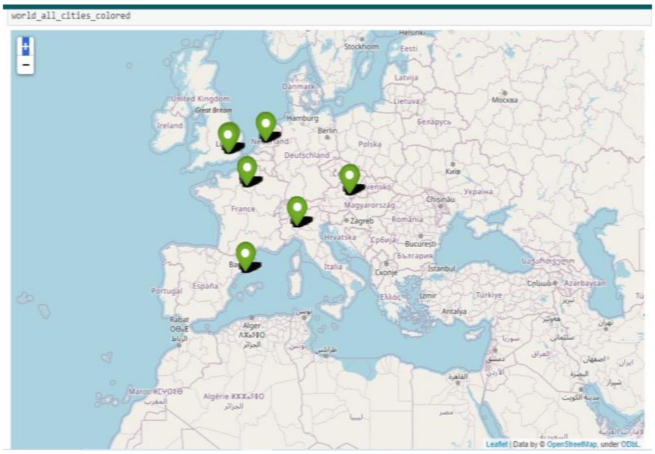


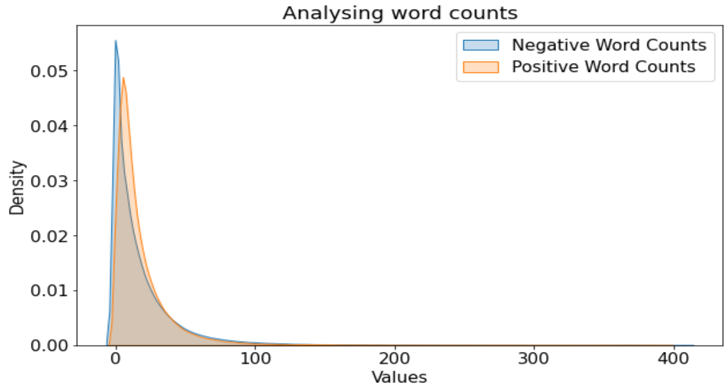
Fig. 5 Hotel Locations Using Folium Map

6.5 Analysing Word Count with Positive Word and Negative Word

The graph depicts the distribution of positive and negative words in hotel reviews, with the density of words on the y-axis and the words themselves on the x-axis. The most common positive words are "good", "clean", "comfortable", "friendly", and "great", while the most common negative words are "bad", "dirty", "noisy", "expensive", and "small". These words indicate that guests are most satisfied with the cleanliness, comfort, and friendliness of hotels, while they are most dissatisfied with the price and size of hotels.

The graph also shows a greater number of negative words than positive words, suggesting that guests are generally satisfied with their hotel experiences. The most common positive words are general words that describe a variety of aspects of a hotel experience, suggesting that guests are looking for a good overall experience. The most common negative words are more specific than the most common positive words, suggesting that guests are more likely to mention specific things they didn't like about their hotel experience.

The density of words on the graph is not evenly distributed, suggesting that some words are more common than others in hotel reviews. Understanding these words can help hotel managers identify areas where they can focus their efforts to improve the guest experience. The analysis of word counts based on positive and negative words separate using NLTK allows for a nuanced understanding of the word's distribution within the dataset.



6.6 Fig. 6 Analyzing positive and negative word counts

6.7 Bar Chart with Country Feature

The bar chart represents the distribution of hotel reviews by country. The x-axis shows the country, and the y-axis shows the number of reviews.

The bar chart shows that the top 6 countries with the most hotel reviews are:

United Kingdom, Spain, France, Netherlands, Austria, and Italy

These countries account for over 50% of all hotel reviews. The remaining 50% of hotel reviews are from reviewers of over 100 different countries.

This information can be used by hotel managers and potential guests to make informed decisions about which hotels to choose and when to travel. For example, hotel managers can use the information to identify the most popular countries for hotel reviews and to target their marketing campaigns to those countries. Potential guests can use the information to choose hotels that are likely to be popular with other travelers.

Overall, the bar chart shows that the United Kingdom, Spain, France, the Netherlands, and Austria are the most popular countries for hotel reviews. This is likely due to a few factors, including the popularity of these countries as tourist destinations, the large populations of these countries, and the growing economies of these countries.

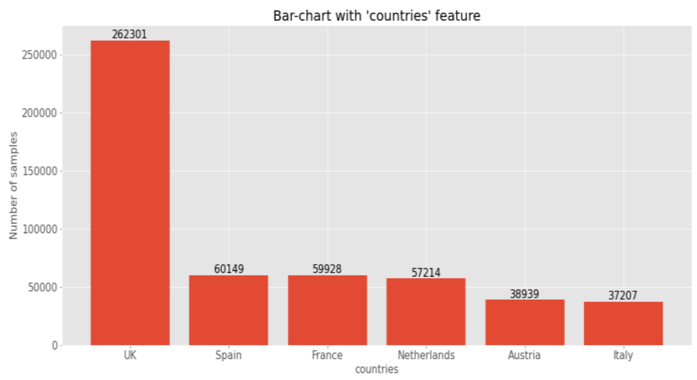


Fig. 7 Bar Chart with Country Feature

6.8 Word Cloud for the "Negative Review “and “Positive Review” feature

The image depicts a word cloud for the "Negative Review" and "Positive Review" features of a hotel review system. Word clouds are visual representations of words used to depict

keyword metadata on a dataset or visualize free form text. The most prominent words in the word cloud are "negative," "poor," "bathroom," "Wi-Fi," "coffee," and "bed," which are likely the most common words used in negative reviews of hotels. Words like "room," "spacious," "comfortable," and "clean" are prominently displayed for positive reviews. Word clouds can be created using various software tools and online services, where users provide the text, they want to analyze. Word clouds can be used for various purposes, such as highlighting room comfort, overall stay, location convenience, staff service, breakfast delight, value for money, and a friendly atmosphere.
Room comfort is highlighted by words like "room," "spacious," "comfortable," and "clean." Overall stay is rated highly, with words like "excellent," "wonderful," "enjoyable," and "pleasant" reflecting positive sentiment. Location convenience is a major positive point, with words like "convenient," "central," and "near attractions" indicating guests found the hotel's location and staff helpful and attentive. The breakfast spread is well-received, with words like "delicious," "superb," and "fantastic" indicating guests enjoyed the breakfast offerings Word clouds can be used for a variety of purposes, including:

- Identifying the most important topics in a dataset or collection of documents. For example, a word cloud of a news article might show the most common words in the article, which can give readers a quick overview of the main topics covered.
- Visualizing the relationships between different words and concepts. For example, a word cloud of a website's blog posts might show how often different keywords are used together, which can help users identify the website's main areas of expertise.
- Creating engaging and informative data visualizations. Word clouds can be used to create visually appealing and informative data visualizations that can be shared with others. For example, a word cloud of customer reviews of a hotel could be used to highlight the most common positive and negative feedback.

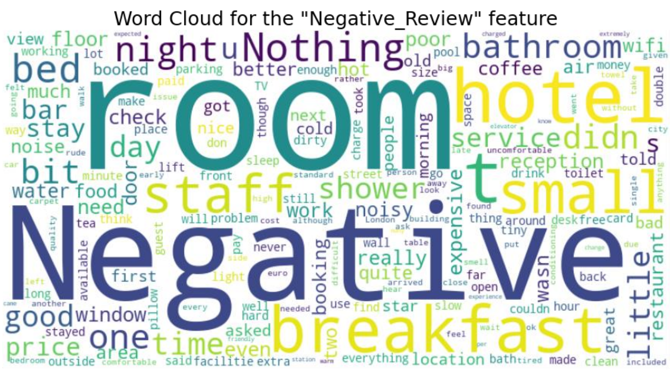


Fig. 8 Negative Review using Word Cloud

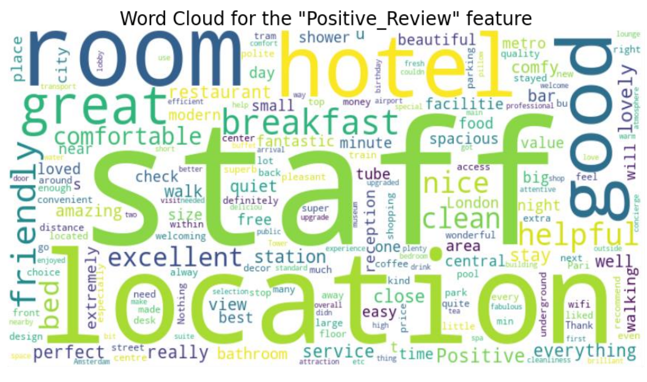


Fig. 9 Positive Review using Word Cloud

6.9 **Word Cloud for the "Tags" feature**

The image illustrates a word cloud representation of the "Tag" feature, where the size of each word corresponds to its frequency in describing the feature. Key highlighted words in the word cloud suggest popular tags for the "Tags" feature, including "Stayed," "Family," "Couple," "Single," "Trip," "Deluxe," "Solo," "Standard," "Classic," "Club," "Business," and "Comfort." These tags indicate the versatility of the "Tags" feature, serving various purposes such as visualizing and organizing data. Users can employ tags to label data points with relevant keywords, facilitating the creation of visualizations and reports. Additionally, tags are instrumental in uncovering insights and trends in data, as seen in their application by a news website for categorizing hotels under topics like "high class," "delux," or "classic." Furthermore, the word cloud emphasizes how tags contribute to improving search and discovery experiences. By tagging hotel pages with relevant keywords, companies enable customers to effortlessly find hotels through intuitive searches. In summary, the word cloud for the "Tags" feature underscores its versatility, illustrating its potential to enhance data visualization, reveal insights, and optimize search and discovery functionalities through the strategic use of highlighting words.

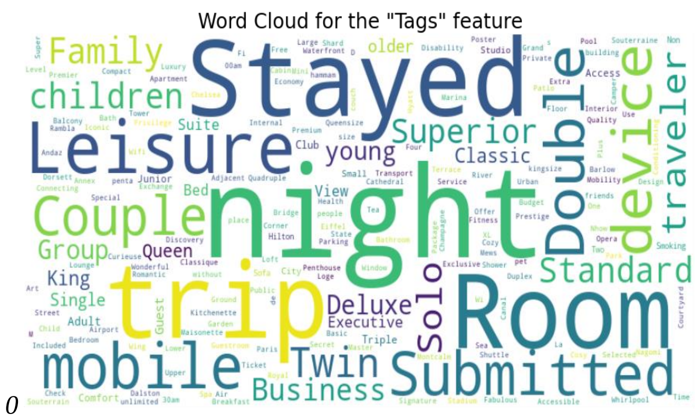


Fig. 10 Analyzing word counts Word Cloud for the "Tags" feature

7 Evaluation

The data that was acquired from the Kaggle dataset is analysed thematically. Thematic analysis, according to Snyder (2019), is a technique for classifying themes within a particular dataset. A thorough comprehension of the outcomes will be achieved via the examination of the focus group recordings. To ensure that the themes are in line with the data and capable of serving as a solid basis for further research projects, they will go through a thorough assessment and revision process.

7.1 **Experiment / Case Study 1**

Summary: Vacation in Spain to Unwind

Emily, a very occupied executive, has recently completed an extended duration of demanding business obligations. Seeking relaxation and renewal, she opts to organize a trip to Spain.
The objective is for Emily to locate a hotel that provides a serene and pleasurable atmosphere, as she is on vacation to relax and unwind. The hotel recommendation feature is incorporated into a travel recommendation system. The function requires two parameters.
1. Country: Spain.

2. Purpose of the trip: "I am taking a vacation to relax after a long period of work-related obligations."

Workflow of the system:

1. The function determines the user's purpose for a recreational journey to relax and alleviate stress caused by work.
2. It considers elements such as picturesque settings, spa amenities, leisure pursuits, and favorable feedback pertaining to tranquilly.
3. The system selects hotels in Spain that meet these criteria, with a focus on creating a peaceful and relaxing ambiance.

7.2 Experiment / Case Study 2

Summary: Business Trip to UK

John, a corporate executive, is organizing a seven-day business excursion to UK, where he will be participating in meetings in Milan and Florence. The aim of this is to effectively locate and reserve a hotel that meets specified criteria connected to business.

The function, which is a component of a travel recommendation system, examines the purpose and applies criteria to UK hotels that cater to the requirements of business travelers, such as accessibility to venues and availability of Wi-Fi. Terms such as "amicable," "hospitable," and "flawless" suggest that visitors had a sense of ease and comfort.

8 Results

8.1 Output for case study 1

```
recommend_hotel('Spain','I am on vacation to unwind following an extended period of business commitments.')
```

	Hotel_Name	Average_Score	Hotel_Address	Positive_Review
0	Hotel Casa Camper	9.6	Elisabets 11 Ciutat Vella 08001 Barcelona Spain	super friendly and helpful staff truly warm a...
1	Mercer Hotel Barcelona	9.5	Dels Lledo 7 Ciutat Vella 08003 Barcelona Spain	the hotel was very pretty and the staff were ...
2	The One Barcelona GL	9.4	277 Carrer de Proven a Eixample 08037 Barcelon...	staff couldn t do enough to help they were br...
3	Hotel Palace GL	9.4	Gran Via de les Corts Catalanes 668 Eixample 0...	very excellent class at hotel
4	Olivia Balmes Hotel	9.3	Balmes 117 Eixample 08008 Barcelona Spain	very friendly staff clean spacious and bright...
5	Aparthotel Arai 4 Superior	9.3	Avinyo 30 Ciutat Vella 08002 Barcelona Spain	everything is perfect here
6	Catalonia Passeig de Gr cia 4 Sup	9.2	Gran Via de les Corts Catalanes 644 Eixample 0...	good location
7	Hotel 1898	9.2	La Rambla 109 Ciutat Vella 08002 Barcelona Spain	i travelled with work colleagues you need to ...
8	Primero Primera	9.2	Doctor Carulla 25 Sarri St Gervasi 08017 Barce...	quiet and spacious
9	Catalonia Catedral	9.2	Arcs 10 Ciutat Vella 08002 Barcelona Spain	no positive

8.2 Output for case study 2

	Hotel_Name	Average_Score	Hotel_Address	Positive_Review
0	Haymarket Hotel	9.6	1 Suffolk Place Westminster Borough London SW1...	absolutely one of the best rooms ive stayed i...
1	41	9.6	41 Buckingham Palace Road Westminster Borough ...	everything staff fabulously friendly attentiv...
2	Charlotte Street Hotel	9.5	15 17 Charlotte Street Hotel Westminster Borou...	loved everything about this hotel charming an...
3	Milestone Hotel Kensington	9.5	1 Kensington Court Kensington and Chelsea Lond...	very old classic hotel and really feels luxur...
4	The Soho Hotel	9.5	4 Richmond Mews Westminster Borough London W1D...	great hotel with cool interior and excellent ...
5	Taj 51 Buckingham Gate Suites and Residences	9.5	Buckingham Gate Westminster Borough London SW1...	outstanding level of service great room size ...
6	Batty Langley s	9.4	12 Folgate Street City of London London E1 6BX UK	probably the best hotel in the city of london...
7	Rosewood London	9.4	252 High Holborn Holborn Camden London WC1V 7E...	the bed was super comfy the decor of the hote...
8	45 Park Lane Dorchester Collection	9.4	45 Park Lane Westminster Borough London W1K 1P...	amazing personalized service
9	Covent Garden Hotel	9.4	10 Monmouth Street Camden London WC2H 9HB UK	location charming decorations size of the roo...

8.3 Outputs of machine learning model

Model	Training Loss	Training AUC	Training Accuracy	Testing Loss	Testing AUC	Testing Accuracy	Training Time	Testing Time
Logistic Regression	0.094	0.999	0.985	0.438	0.898	0.814	5min 49s	50.4 s
Decision Tree	0.015	1	0.992	9.201	0.728	0.721	18.8 s	2.49 s
XGBoost	0.542	0.847	0.763	0.552	0.833	0.748	27min 20s	4min 44s

8.4 Outputs of deep learning models

Model	Evaluation Method	Loss	Accuracy	Evaluation Time
Dense Neural Network	Training	0.766	0.79	1s
Feedforward Neural Network	Testing	0.394	0.84	-
Convolutional Neural Network	Testing	0.434	0.797	0s
Long Short-Term Memory (LSTM)	Testing	0.3823	0.845	0s

9 Discussion

The work involves leveraging a hotel review dataset and employing natural language toolkit processing techniques using the NLTK package. The dataset is processed using NLTK's word tokenize function to extract and analyse the reviews. The goal is to categorize reviews into positive and negative. The use of a word cloud visually represents the processed data, aiding in the separation of positive and negative reviews.

The positive reviews are then utilized to recommend hotels to users. The user has the flexibility to specify their country of destination, allowing for more personalized recommendations. The recommendation system incorporates hotel names, average scores, and addresses. This approach enhances user decision-making by providing comprehensive information, allowing them to easily choose accommodations tailored to their preferences. By integrating NLTK's capabilities, enabling a more refined and accurate hotel recommendation system. This approach not only harnesses the power of natural language processing but also incorporates user customization through the selection of destination countries, fostering a more user-centric and satisfactory experience.

Machine Learning Models

The logistic regression model demonstrates an impressive performance on the training set, achieving near-perfect accuracy and an AUC of 0.999, indicative of an excellent fit to the training data. But, the model's effectiveness diminishes when applied to the testing set, as reflected by a lower accuracy of 0.898 and a decreased AUC of 0.814. This discrepancy suggests potential overfitting, where the model is overly tailored to the complications of the training data and struggles to generalize to unseen instances. The utilization of Tf-IDF features in a variant of logistic regression offers a marginal improvement in generalization, with a testing accuracy of 0.793 and a testing AUC of 0.823. While this model still exhibits signs of overfitting, the incorporation of Tf-IDF features aids in improving its adaptability to previously unseen data compared to the standard logistic regression.

The decision tree model, while achieving an impressive AUC of 1.000 on the training set, faces significant challenges in generalization, as evidenced by a testing AUC of 0.728 and a testing accuracy of 0.728. This obvious difference between training and testing performance highlights a clear case of overfitting, where the model has essentially memorized the training data without grasping the underlying patterns essential for effective predictions on new data. On the other hand, the XGBoost model strikes a better balance between training and testing performance, showcasing a solid AUC of 0.847 and an accuracy of 0.833 on the testing set. Although there is still room for improvement, XGBoost demonstrates better generalization capabilities compared to the decision tree, making it a more promising candidate for hotel review prediction.

Deep Learning Models

In the Dense neural network, during training, the model achieved a loss of 1.2751 and an accuracy of 0.7523. The loss represents the measure of the model's error, and the accuracy indicates the proportion of correctly classified instances. These metrics suggest that the Dense network has learned to predict with a moderate level of accuracy, with room for improvement. Further analysis, such as examining validation metrics and exploring potential overfitting, would be valuable for a detailed evaluation.

The Convolutional Neural Network (CNN) exhibited a lower loss of 0.394 during training, with an accuracy of 0.84. CNNs are particularly effective in capturing spatial hierarchies in data, and these metrics indicate a relatively successful learning process. However, it is essential to assess the model's performance on a separate test set to ensure its ability to generalize to unseen data.

In the Feedforward Neural Network (FNN), the training phase resulted in a loss of 0.434 and an accuracy of 0.797. Additionally, the test accuracy was reported as 0.855. These metrics suggest that the FNN has learned to classify instances with a high degree of accuracy during training and maintains this performance on an independent test set, indicating robust generalization.

For the Long Short-Term Memory (LSTM) model, the reported loss was 0.3823 with an accuracy of 0.843. LSTM networks are commonly used for sequence data, such as time series or natural language processing. The relatively higher loss and lower accuracy may indicate challenges in learning sequential dependencies in the provided data. Further investigation into the model architecture, hyperparameters, and potentially increasing the training duration could be explored to improve performance.

10 Conclusion and Future Work

Machine learning and deep learning algorithms have proven to be successful for sentiment classification, as demonstrated by the study of the hotel review dataset. The most successful classifier turned out to be logistic regression because of its ease of use and interpretability. Because of its ability to effectively capture pertinent patterns in the problem's linear structure, it makes it possible to clearly identify the variables driving both positive and negative feelings in hotel evaluations.

The feedforward neural network (FNN) demonstrated superior performance in handling unstructured data like natural language. Its capacity to automatically learn hierarchical features and relationships within the dataset played a pivotal role in discerning nuanced sentiments expressed in hotel reviews.

The hotel review recommendation system will capitalize on logistic regression's efficiency and interpretability as the primary driver. Logistic regression will act as the initial filter in the recommendation process, providing quick and accurate assessments of sentiments. This streamlined approach ensures a rapid and resource-efficient evaluation of reviews, allowing for prompt categorization of hotels based on overall sentiment.

Continuous improvement and monitoring are essential for the long-term success of the recommendation system. Regular updates to the models based on new data and feedback will maintain their relevance and effectiveness over time. Exploring advanced natural language processing techniques, sentiment analysis tools, and other state-of-the-art algorithms can contribute to ongoing enhancements, ensuring the recommendation system remains at the forefront of providing accurate and context-aware suggestions for an enhanced hotel selection experience.

11 References

- Ameen, N., Tarhini, A., Reppel, A. and Anand, A., 2021. Customer experiences in theage of artificial intelligence. *Computers in Human Behavior*, *114*, p.106548.
- Anis, S., Saad, S., Aref, M. (2021). Sentiment Analysis of Hotel Reviews Using Machine Learning Techniques. In: Hassanien, A.E., Slowik, A., Snášel, V., El-Deeb, H., Tolba, F.M. (eds) Proceedings of the International Conference on Advanced Intelligent Systems and Informatics 2020. AISI 2020. Advances in Intelligent Systems and Computing, vol 1261. Springer, Cham. https://doi.org/10.1007/978-3-030-58669-0_21
- Anis, Sarah & Ismail, Sally & Aref, Mostafa. (2020). Sentiment Analysis of Hotel Reviews Using Machine Learning Techniques. 10.1007/978-3-030-58669-0_21.
- Arif Abdurrahman Farisi1, Yuliant Sibaroni1 and Said Al Faraby1 Published under licence by IOP Publishing Ltd Journal of Physics: Conference Series, Volume 1192, The 2nd International Conference on Data and Information Science 15–16 November 2018, Bandung, Indonesia **Citation** Arif Abdurrahman Farisi *et al* 2019 *J. Phys.: Conf. Ser.* 1192 012024 **DOI** 10.1088/1742-6596/1192/1/012024 .
- Campione, R., 2021. How digitalization is revolutionizing relationships between customers and service providers in the hospitality industry: Personalization and gamification. *International Journal of Information, Business and Management*, *13*(1), pp.35-52.
- Chih-Fong Tsai, Kuanchin Chen, Ya-Han Hu, Wei-Kai Chen,Improving text summarization of online hotel reviews with review helpfulness and sentiment,Tourism Management,Volume 80,2020, 104122,ISSN 0261-5177,https://doi.org/10.1016/j.tourman.2020.10412.
- Chowdhury, Reataj & Deshpande, Abhijit. (2020). An Analysis of the Impact of Reviews on the Hotel Industry. Annals of Tropical Medicine and Public Health. 23. 10.36295/ASRO.2020.231742.
- George Tzougas & Konstantin Kutzkov. (2023). Enhancing Logistic Regression Using Neural Networks for Classification in Actuarial Learning. Algorithms, 16(2), 99–99. https://doi.org/10.3390/a16020099
- He, A.Z. and Zhang, Y., 2023. AI-powered touch points in the customer journey: a systematic literature review and research agenda. *Journal of Research in Interactive Marketing*, *17*(4), pp.620-639.
- Hornik, K., Stinchcombe, M., & White, H. (1989). Multilayer feedforward networks are universal approximators. Neural Networks, 2(5), 359–366. https://doi.org/10.1016/0893-6080(89)90020-8
- Ivanov, S. and Webster, C. eds., 2019. *Robots, artificial intelligence, and service automation in travel, tourism, and hospitality*. Emerald Publishing Limited.
- K. Shifullah et al., "Classification of Hotel Reviews Using Sentiment Analysis and Machine Learning," 2022 25th International Conference on Computer and

Information Technology (ICCIT), Cox's Bazar, Bangladesh, 2022, pp. 710-715, doi: 10.1109/ICCIT57492.2022.10054884.

- Narayan, R., Gehlot, A., Singh, R., Akram, S.V., Priyadarshi, N. and Twala, B., 2022. Hospitality Feedback System 4.0: Digitalization of Feedback System with Integration of Industry 4.0 Enabling Technologies. *Sustainability*, *14*(19), p.12158.
- Nicholas, Chan & Lee, Angela. (2017). Voice of Customers: Text Analysis of Hotel Customer Reviews (Cleanliness, Overall Environment & Value for Money). 104-111. 10.1145/3152723.3152739.
- Nie, Ru-xin, et al. "Hotel selection driven by online textual reviews: Applying a semantic partitioned sentiment dictionary and evidence theory." International Journal of Hospitality Management 88 (2020): 102495.
- Sharma, M.K., 2021. Artificial Intelligence's Prospects and Uses in Hospitality Industry: Literature Review. *Journal of Multidisciplinary Cases (JMC) ISSN 2799-0990*, *1*(01), pp.34-42.
- Singh, Inder. "A Review Paper on Hotel Industry-A Study of Trend in Hospitality Industry and the Types of Hotels." *International Research Journal of Engineering and Technology (IRJET)* 6.3 (2019): 7189-7193.
- Snyder, H., 2019. Literature review as a research methodology: An overview and guidelines. *Journal of business research*, *104*, pp.333-339.
- Tan, Y.S. and Wright, A.S., 2022. Exploring "Smart and Green" Concepts: A New Synergy for Hospitality. *Tourism and Hospitality*, *3*(1), pp.276-296.
- Tan, Y.S., 2021. Harvesting a New Synergy: Exploring 'Smart and Green'Hotels withinthe Context of the Hospitality Industry.
- uggero Sainaghi, Paul Phillips, Rodolfo Baggio, Aurelio Mauri, Cross-citation and authorship analysis of hotel performance studies, International Journal of Hospitality Management, Volume73,2018, Pages7584, ISSN0278319, https://doi.org/10.1016/j.ijhm.2018.02.004 .
- Ukpabi, D., Karjaluoto, H., Olaleye, S.A. and Mogaji, E., 2018. Dual perspectives on the role of artificially intelligent robotic virtual agents in the tourism, travel, and hospitality industries. In *EuroMed Academy of Business Conference Book of Proceedings*. EuroMed Press.