كليـــة العلـــوم
والتقنيـات – مراكش
FACULTÉ DES SCIENCES ET TECHNIQUE – MARRAKECH

DATA Embassy

A thesis submitted in fulfillment of the requirements for the degree of Master of Science and Techniques

in the

Data Science and Business Intelligence
Informatics Department

CADI AYYAD UNIVERSITY
FACULTY OF SCIENCES AND TECHNIQUES

MASTERS THESIS

Aspect-Based Sentiment Analysis of Marrakesh Hotel Reviews

Author

BOUASKAOUN Mohammed

Supervised by

Mr. KADDA Yasine, DATA Embassy, Marrakesh
Pr. BANOUAR Oumayma, FSTG, Marrakesh

Jury Members

Pr. BOURQUIA *Nawal*
Pr. BENHADDI *Maryam*
Pr. BANOUAR *Oumayma*

Academic year : 2022/2023

Acknowledgements

I would like to mention that the current thesis was written as part of the two-year full-time program in Data Science and Business Intelligence at the University of Cadi Ayyad, under the guidance of Mr. Kadda Yasine and Professor Banouar Oumayma. I am sincerely grateful for their invaluable assistance and support provided throughout this project. Their extensive knowledge and experience in the field greatly benefited our task. Equally important was their understanding and the emotional support they offered me, which proved essential in successfully completing this project.

I would also like to express my appreciation to my colleagues and friends, especially my friend Tareq, for their support and assistance at various stages of this program. Their contributions have been crucial to my progress. Additionally, I want to extend my deep gratitude to my family, particularly my parents, for their unwavering support and steadfast belief in my abilities.

Abstract

Understanding and extracting information from natural language text is a difficult task for machines. In the field of Artificial Intelligence, Natural Language Processing (NLP) plays a significant role in helping computers understand such texts.

This Master's thesis concentrates on applying different methods and techniques to classify the sentiment of hotel reviews in Marrakesh. It goes a step further by incorporating aspect-based sentiment analysis.

The thesis uses a dataset consisting of English reviews obtained from a popular review platform for Marrakesh hotels. To train and evaluate the proposed methods, the dataset is used, and three distinct approaches are employed: Term Frequency Inverse Document Frequency (TF-IDF), Recurrent Neural Networks with Bidirectional Long Short-Term Memory (LSTM), and Bidirectional Encoder Representations from Transformers (BERT).

A comprehensive series of experiments is conducted. It is observed that BERT, which solely relies on textual information, performs better than the other methods. Moreover, the models using Bidirectional LSTM show superior performance compared to those using TF-IDF.

While sentiment classification helps understand the overall satisfaction levels of customers, it is crucial for hotels to uncover the reasons behind positive or negative reviews. To address this, aspect-based sentiment analysis (ABSA) is employed, focusing on five key aspects: room, service, location, price, and food.

This analysis enables hotels to identify specific areas of improvement in their services, leading to effective solutions for addressing customer concerns.

Contents

List of Figures

List of Tables

List of Abbreviations

ABSA	**A**spect-**B**ased **S**entiment **A**nalysis
AI	**A**rtificial **I**ntelligence
BERT	**B**idirectional **E**ncoder **R**epresentations from **T**ransformers
BernoulliNB	**B**ernoulli **N**aive **B**ayes
DT	**D**eterminer
EMS	**E**uro **M**aghreb **S**ervice
FN	**F**alse **N**egatives
FP	**F**alse **P**ositives
JJ	**A**djective
LR	**L**ogistic **R**egression
LSTM	**L**ong **S**hort-**T**erm **M**emory
NLTK	**N**atural **L**anguage **T**ool**K**it
NN	**N**oun
NER	**N**amed **E**ntity **R**ecognition
NLP	**N**atural **L**anguage **P**rocessing
POS	**P**art-**O**f-**S**peech
PRP	**P**ronoun
RandomForestClassifier	**R**andom **F**orest **C**lassifier
RNNs	**R**ecurrent **N**eural **N**etworks
SVC	**S**upport **V**ector **C**lassifier
TF-IDF	**T**erm **F**requency **I**nverse **D**ocument **F**requency
TN	**T**rue **N**egatives
TP	**T**rue **P**ositives
URL	**U**niform **R**esource **L**ocator
VBZ	**V**erb (3rd person singular present tense)

Introduction

Machines have always struggled with understanding and extracting meaningful information from natural language text. However, with the rise of online platforms and user-generated content like reviews and feedback, there is an increasing need for automated analysis of textual data. In the field of Artificial Intelligence, Natural Language Processing (NLP) has emerged as a powerful tool for computers to effectively comprehend and process human language. This Master's thesis focuses on sentiment classification and aspect-based sentiment analysis of Marrakesh hotel reviews, aiming to provide valuable insights for the hospitality industry.

The hospitality industry heavily relies on customer satisfaction and feedback to maintain and improve the quality of its services. In today's digital age, online reviews have a significant impact on the reputation and success of hotels. Therefore, it is crucial for hotels to gain a deeper understanding of customer sentiments expressed in these reviews. However, manually analyzing a large volume of textual data is time-consuming and subjective. This calls for the development of automated techniques that can accurately classify sentiment and extract specific aspects related to customer experiences from hotel reviews.

This thesis proposes a comprehensive solution to address the challenges associated with sentiment classification and aspect-based sentiment analysis of Marrakesh hotel reviews. The study employs three different approaches: Term Frequency Inverse Document Frequency (TF-IDF), Recurrent Neural Networks with Bidirectional Long Short-Term Memory (LSTM), and Bidirectional Encoder Representations from Transformers (BERT). These approaches are evaluated and compared to determine their effectiveness in sentiment classification. Additionally, aspect-based sentiment analysis is conducted to uncover the underlying reasons behind positive or negative sentiments, focusing on five key aspects: room, service, location, price, and food.

To conduct the experiments, a dataset of English reviews is extracted from a popular review platform, specifically focusing on Marrakesh hotel reviews. This dataset is used to train and evaluate the sentiment classification models. Three different approaches, namely TF-IDF, LSTM, and BERT, are implemented and fine-tuned using the dataset. Evaluation metrics such as accuracy, precision, recall, and F1 score are used to measure the effectiveness of the sentiment classification models. Additionally, aspect-based sentiment analysis is performed by identifying specific aspects related to customer experiences in the reviews. These aspects are then analyzed for sentiment polarity, providing insights into the strengths and weaknesses of different hotels in Marrakesh.

The implementation of the proposed solution involves preprocessing the textual data, extracting features, training the models, and conducting evaluations. The implementation process is carefully designed to ensure the accuracy and reliability of the obtained results.

By utilizing sentiment classification and aspect-based sentiment analysis, this thesis aims to provide valuable insights for the hospitality industry, enabling hotels to identify specific areas of improvement in their services and effectively address customer concerns.

The thesis starts by providing the project context and background, followed by an overview of sentiment analysis in Chapter 2. In Chapter 3, we introduce the proposed framework, which includes the application framework, data collection, preprocessing, and the aspect-based sentiment analysis approach. Chapter 4 presents the results and analysis, comparing the performance of various approaches, and includes dashboard visualization. Finally, the conclusion summarizes the contributions, and future directions of the study.

Chapter 1

Project Context

1. Host organization
2. Background and motivation
3. Problem statement
4. Challenges

In the current era of digital technology, the hospitality industry encounters difficulties in efficiently comprehending and extracting valuable insights from the extensive textual data produced by online reviews. With the abundance of user-generated content, particularly hotel reviews, it has become crucial for hotels to analyze customer sentiments and identify specific areas where their services can be improved. However, manually analyzing these reviews is both time-consuming and subject to personal biases. Consequently, there is a need for automated techniques that can aid in sentiment classification and aspect-based sentiment analysis. This chapter aims to provide the project's background and context for conducting a comprehensive investigation into sentiment analysis and aspect-based sentiment analysis of Marrakesh hotel reviews.

1.1 Host Organization

1.1.1 Presentation of Data Embassy

Data Embassy is a production center owned by the EMS group, specializing in data capture, entry, processing, and document dematerialization. This center is capable of handling all types of raw information, whether physical or intangible, and offers cost-effective document transportation solutions between Europe and Morocco. With a workforce of 350 employees and an integrated training center, EMS is capable of quickly adapting to the needs of its clients by providing training for operators, project managers, technical managers, supervisors, and planners. Functional flexibility is one of the strengths of this group, allowing its employees to change positions without additional costs and to increase its workforce by accessing the labor market.

DATA Embassy

FIGURE 1.1: Data Embassy Logo

1.1.2 Presentation of EMS

The @rrom group has been present in the French market since 1997, initially focused on providing specialized press services through Minitel. In 2002, it established itself in France under the name EMS to diversify its services and expand its business activities. Since then, the company has experienced sustained growth, thanks to its position as a leader in their respective markets. EMS offers services related to information processing, such as consulting and/or production of document dematerialization, data entry and processing, electronic document management, and the management of marketing operations such as couponing, ODR, questionnaires, through its ISO 9001 certified production center "DATA EMBASSY". In 2012, the EMS group launched the DATA SHORE contact center in Marrakesh, equipped with 100 positions, aiming to establish a new center of excellence in multichannel customer relations.

FIGURE 1.2: Ems Logo

1.1.3 The services offered by Data Embassy

Data Embassy, the production center of the EMS group in Marrakech, has obtained ISO 9001-2008 certification from TUV Rheinland. This recognition confirms Data Embassy's commitment to providing a quality service in accordance with customer requirements and regulatory standards. By implementing an effective quality management system, Data Embassy ensures customer satisfaction through continuous process improvement and the application of the PDCA (Plan-Do-Check-Act) principle, also known as the Deming cycle.

1.1.4 Security and Data Confidentiality at Data Embassy

EMS and its production center, Data Embassy, have implemented a privacy policy that complies with French regulations and the recommendations of the CNIL (National Commission for Data Protection) to protect the collected and processed information and raw data. To ensure the security of this data, it is backed up twice daily on the production site's servers and on a separate and independent off-site location, using RAID 5 technology to ensure maximum reliability. Rigorous procedures are enforced to ensure the security of employees, the production site, and the assets entrusted by clients. In addition to preventive measures, the production center is also covered by liability insurance to safeguard against potential losses.

1.2 Background and Motivation

Artificial Intelligence (AI) has made significant advancements in recent years, especially in the field of Natural Language Processing (NLP). NLP focuses on teaching computers to understand and process human language, which is challenging due to the complexity and ambiguity of natural language. One important application of NLP is sentiment analysis, which involves determining the sentiment or emotional tone expressed in a given text. Sentiment analysis has gained attention because it can provide valuable insights by analyzing user-generated content like reviews, social media posts, and customer feedback.

In the hospitality industry, customer satisfaction and feedback are crucial for the reputation and success of hotels. With the rise of online review platforms, customers now have an easy way to express their opinions and share their experiences. Analyzing these reviews can provide valuable information for hotel management to understand customer sentiments, identify areas for improvement, and make informed business decisions. However, manually processing a large number of reviews is time-consuming, subjective, and prone to human errors. Therefore, there is a growing need for automated techniques that can accurately classify sentiment and extract specific aspects related to customer experiences from hotel reviews.

1.3 Problem statement

The main challenge at hand is the effective analysis and understanding of the sentiment expressed in Marrakesh hotel reviews, as well as capturing the specific aspects that drive these sentiments. Traditional sentiment analysis methods often fall short in providing detailed insights into the reasons behind positive or negative reviews, limiting the ability of hotel management to address specific concerns and enhance their services accordingly. Hence, there is a need for an advanced approach that incorporates aspect-based sentiment analysis to gain a deeper understanding of customer experiences in Marrakesh hotels.

Moreover, the large volume of textual data generated by hotel reviews presents scalability and efficiency challenges. Manual processing and interpretation of this vast amount of reviews become impractical. Therefore, it is necessary to develop automated techniques that can handle the large-scale analysis of Marrakesh hotel reviews, enabling efficient sentiment classification and aspect-based sentiment analysis.

Thus, the problem statement focuses on developing an advanced methodology for sentiment classification and aspect-based sentiment analysis of Marrakesh hotel reviews. This methodology should address the limitations of traditional sentiment analysis methods, handle the large-scale analysis of textual data, and evaluate the effectiveness of different NLP techniques in this specific context. By tackling these challenges, the study aims to provide valuable insights for hotel management, enabling them to improve the quality of services based on a comprehensive understanding of customer sentiments and the specific aspects that contribute to their experiences in Marrakesh hotels.

1.4 Challenges

During this study, I have encountered several challenges that have impacted the aspect-based sentiment analysis of Marrakesh hotel reviews. These challenges include the absence of specific labels for aspect-based sentiment analysis in the collected dataset, the difficulty in evaluating the accuracy of aspect detection and aspect classification methods, the intricacies of interpreting sentences with ironic or sarcastic tones, and the complexities associated with resolving pronouns in sentiment analysis.

The first challenge is that the dataset I collected for analysis doesn't have labels specifically for aspect-based sentiment analysis. This means that the reviews are not categorized according to the aspects they discuss or the sentiment expressed for each aspect. Without these labels, it's harder to train and test models accurately.

Secondly, there is a challenge related to the way I approach the task. I use scripts to find and classify aspects in the reviews, but I don't have a way to check if the aspects are identified and classified correctly. This means I can't really evaluate how well the model is performing in terms of aspect detection and aspect classification.

Another challenge lies in comprehending the opposing interpretation of sentences, particularly when dealing with instances of irony or sarcasm. Recognizing these nuanced sentiments can be tricky since there is no universally recognized symbol or punctuation mark to indicate such phrases. As a result, accurately capturing the true sentiment expressed in these cases poses a significant challenge.

Lastly, Resolving pronouns is another challenging task. Even though methods and algorithms can assist, determining the specific feature associated with emotion words when a sentence contains opinion terms and pronouns can be difficult.

In this chapter, we have explored the project context, including the host organization, background, motivation, problem statement, and challenges. These elements set the foundation for the subsequent chapters, where we delve deeper into addressing the identified problem and overcoming the challenges. By understanding the context surrounding the project, we can better appreciate the significance and relevance of the research to be conducted.

Next Chapter will focus on Sentiment Analysis and cover topics such as an overview of sentiment analysis, a literature review, sentiment classification approaches, and aspect-based sentiment analysis.

Chapter 2

Sentiment Analysis

1. Sentiment analysis overview
2. Literature review
3. Sentiment classification approaches
4. Aspect based sentiment analysis

Nowadays, people express their thoughts and feelings on social media, online shopping websites, and various other platforms. This generates a large amount of information on these platforms.

Many companies are interested in knowing what people are saying about them. They put in a lot of effort to promptly understand their customers' desires and offer them suitable services. This helps them learn what brings joy or dissatisfaction to their customers, enabling them to customize their products and services accordingly. Moreover, brands also want to gauge how their ads are affecting users.

Because of these reasons, sentiment analysis is growing increasingly significant with each passing day.

2.1 Sentiment Analysis Overview

Sentiment analysis, also known as opinion mining, is a valuable natural language processing (NLP) technique used to analyze textual data and determine the sentiment expressed within it. This technique enables businesses to classify data as positive, negative, or neutral, providing valuable insights into customer feedback and needs.

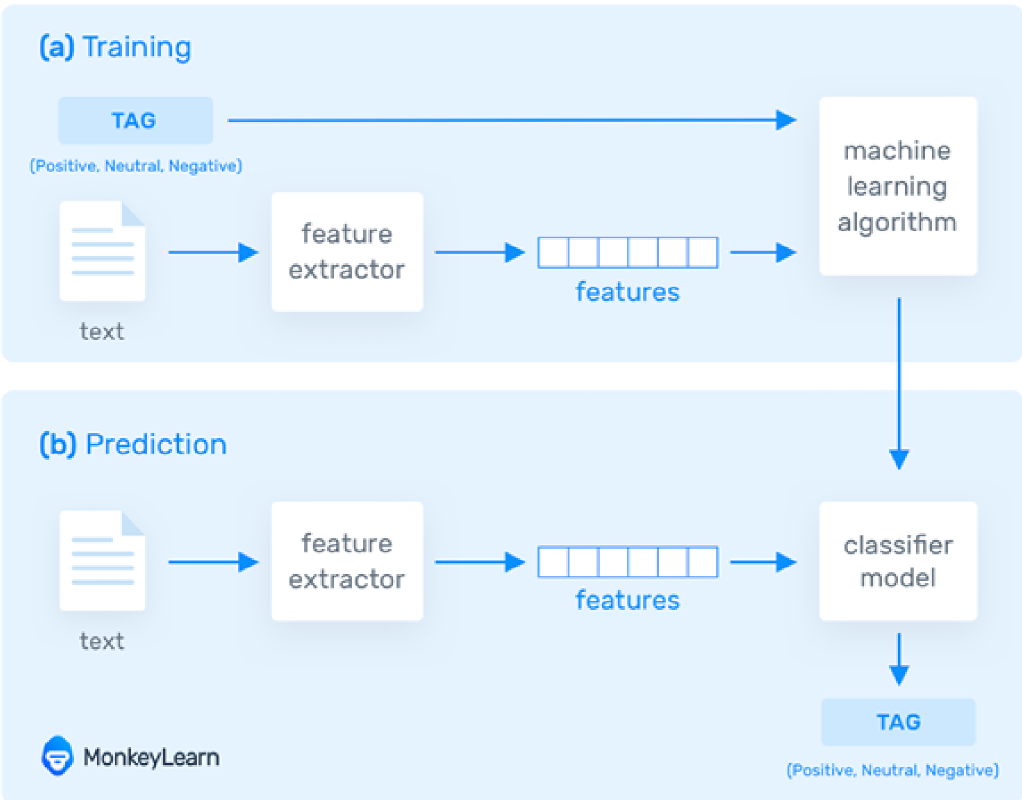

FIGURE 2.1: Sentiment Analysis Work flow, MonkeyLearn, Source

How Does Sentiment Analysis Work?

Firstly, data collection entails gathering the relevant textual information from diverse sources, such as social media posts, customer reviews, news articles, and survey responses. Once collected, the text data undergoes preprocessing, wherein it is cleaned and standardized to remove noise and extraneous details. This involves eliminating punctuation, converting text to lowercase, and filtering out commonly used words, known as stop words, amongst other normalization techniques.

Next, feature extraction is performed to identify meaningful attributes from the preprocessed text. These features can range from individual words (unigrams) to pairs of words (bigrams), part-of-speech tags, syntactic patterns, or other linguistic representations. The purpose of feature extraction is to capture relevant information that aids in determining sentiment.

The subsequent step involves sentiment classification, wherein a machine learning model is employed to categorize the sentiment of the text. Various techniques can be employed for this task, such as supervised learning, or deep learning. Supervised learning utilizes a labeled dataset, where each text sample is associated with a sentiment label, to train a machine learning model. This model learns patterns and relationships in the training data to predict the sentiment of new, unseen text samples. Deep learning models, such as recurrent neural networks RNNs(see Grossberg, 2013 and Salehinejad et al., 2017 for detailed studies and recent advances), leverage large amounts of labeled data to learn complex representations of text and accurately capture sentiment.

The output of sentiment analysis typically comprises a sentiment polarity score or a categorical sentiment label. The sentiment polarity score ranges from -1 to +1, where negative values indicate negative sentiment, positive values indicate positive sentiment, and values close to zero represent neutral sentiment. Alternatively, categorical classification labels the text as positive, negative, or neutral based on the sentiment analysis outcome.

2.2 Literature review

Sentiment analysis of hotel reviews has gained significant attention in recent years due to the exponential growth of user-generated content on online platforms. Researchers have employed various machine learning and deep learning techniques to accurately classify the sentiment expressed in these reviews. In their study, (Nohh et al., 2019) applied the Bernoulli Naive Bayes (BernoulliNB) classifier and support vector machines (SVMs) along with the term frequency-inverse document frequency (TF-IDF) representation to analyze hotel reviews. They achieved promising results, highlighting the effectiveness of this methods for sentiment classification tasks.

Another popular approach is the utilization of ensemble methods such as Random Forest.(Bompotas et al., 2020) investigated the effectiveness of Random Forest in sentiment analysis of hotel reviews by combining it with TF-IDF representation. Their findings revealed that the Random Forest algorithm achieved superior performance in terms of accuracy, precision, and recall, showcasing its robustness for sentiment classification.

Moreover, logistic regression has been extensively utilized for sentiment analysis tasks. (Reddy et al., 2022) employed logistic regression along with the TF-IDF representation to classify hotel reviews. Their results indicated that logistic regression achieved competitive performance in sentiment classification tasks, emphasizing its efficiency and interpretability.

In recent years, deep learning methods have also emerged as powerful tools for sentiment analysis. Long Short-Term Memory (LSTM), a recurrent neural network architecture, has shown promising results in this domain. (Ishaq et al., 2021) investigated the application of LSTM for sentiment analysis of hotel reviews. Their findings demonstrated that LSTM achieved superior performance compared to traditional machine learning methods, highlighting its ability to capture long-term dependencies in textual data.

Furthermore, the transformer-based model, BERT (Bidirectional Encoder Representations from Transformers), has gained significant popularity in sentiment analysis tasks. (González-Carvajal and Garrido-Merchán, 2020) introduced BERT as a state-of-the-art model for various natural language processing tasks, including sentiment analysis. Researchers have since applied BERT to hotel review sentiment analysis with remarkable success, as it effectively captures contextual information and semantic relationships in the text.

In conclusion, sentiment analysis of hotel reviews has witnessed advancements through the application of various machine learning and deep learning methods. The BernoulliNB (Farisi, Sibaroni, and Al Faraby, 2019), Random Forest (Anis, Saad, and Aref, 2021), SVC (Zainuddin and Selamat, 2014), and logistic regression algorithms (Hamdan, Bellot, and Bechet, 2015), along with the TF-IDF representation, have proven effective in sentiment classification. Additionally, LSTM (Priyantina and Sarno, 2019) and BERT (Wen, Liang, and Zhu, 2023) have emerged as powerful deep learning models, achieving state-of-the-art results in sentiment analysis tasks. These studies lay a solid foundation for the implementation and evaluation of sentiment analysis techniques in the context of hotel reviews.

2.3 Sentiment Classification Approaches

2.3.1 TF-IDF Approach

The TF-IDF (Term Frequency-Inverse Document Frequency)(Paltoglou and Thelwall, 2010) approach is a widely used method for feature extraction in natural language processing tasks, including sentiment classification.

2.3.1.1 Feature Extraction using TF-IDF

TF-IDF (Term Frequency-Inverse Document Frequency) is a widely used technique in Natural Language Processing (NLP) for feature extraction from textual data. It aims to capture the importance of individual words or terms within a document corpus. The process involves calculating two factors: the term frequency (TF) and the inverse document frequency (IDF).

The first step in TF-IDF feature extraction is to tokenize the text into individual words or terms(Luhn, 1957). Tokenization breaks down the text into meaningful units that can be further processed. Once tokenized, the term frequency (TF) is calculated by counting the number of occurrences of each term in a document. This step quantifies the importance of a term within a document.

$$\mathrm{TF}(t,d) = \frac{\mathrm{count}(t,d)}{\mathrm{count}(d)}$$

where:

- $\mathrm{count}(t,d)$ represents the number of times term t appears in document d.
- $\mathrm{count}(d)$ represents the total number of terms in document d.

The inverse document frequency (IDF) accounts for the significance of a term across the entire document corpus(Jones, 2014). IDF is calculated by taking the logarithm of the total number of documents divided by the number of documents containing the term.

$$\text{IDF}(t, D) = \log\left(\frac{\text{total number of documents}}{\text{number of documents containing term } t + 1}\right)$$

This calculation penalizes terms that appear frequently in the corpus and emphasizes rare or distinctive terms.

Finally, the TF-IDF score is computed by multiplying the term frequency (TF) with the inverse document frequency (IDF) for each term. This score represents the importance of a term within a specific document relative to the entire corpus. The resulting TF-IDF scores serve as features that capture the uniqueness and discriminative power of each term in the classification task.

$$\text{TF-IDF}(t, d, D) = \text{TF}(t, d) \times \text{IDF}(t, D)$$

where:

- t represents the term.
- d represents the document.
- D represents the corpus (collection of documents).

Once the TF-IDF features are computed for each term in the textual data, they are typically normalized to ensure that they are on a comparable scale. Common normalization techniques include L2 normalization, which divides each TF-IDF score by the Euclidean norm of the vector, or max normalization, which scales the scores between 0 and 1 based on the maximum value.

The prepared TF-IDF features can then be used as input for classification algorithms, such as logistic regression or support vector machines, to train models that can predict the sentiment of Marrakesh hotel reviews based on the extracted textual features.

2.3.1.2 Classification Models

In this study, several classification models or algorithms are utilized in conjunction with TF-IDF features to analyze and classify Marrakesh hotel reviews. These models include BernoulliNB (Bernoulli Naive Bayes), Random Forest Classifier, SVC (Support Vector Classifier) and Logistic Regression.

- **Bernoulli Naive Bayes :**

  The Bernoulli Naive Bayes algorithm calculates the probability that a document belongs to a particular class (positive or negative) based on the occurrence of words in the document.

  Let's denote the sentiment class as C (where C can be positive or negative), and let D be a document with words represented as features, denoted by $d_1, d_2, ..., d_n$. The algorithm estimates the probability $P(C|d_1, d_2, ..., d_n)$, which is the probability that the sentiment class is C given the presence or absence of each word in the document.

The formula for calculating this probability is:

$$P(C|d_1, d_2, ..., d_n) = \frac{P(C) \cdot \prod_{i=1}^{n} P(d_i|C)}{P(d_1, d_2, ..., d_n)}$$

Here, $P(C)$ represents the prior probability of class C, which is estimated by the frequency of C in the training data. $P(d_i|C)$ denotes the probability of word d_i given the sentiment class C, which is estimated using the training data as well.

The denominator $P(d_1, d_2, ..., d_n)$ acts as a normalizing factor and can be ignored during classification since it does not affect the decision of whether a document belongs to a positive or negative sentiment class.

The algorithm selects the class with the highest probability as the predicted sentiment class for the given document. It calculates $P(C|d_1, d_2, ..., d_n)$ for both positive and negative classes and assigns the class with the higher probability to the document. This process is repeated for each document in the dataset, resulting in sentiment predictions for all documents.

By using this approach, Bernoulli Naive Bayes effectively captures the probabilistic relationships between words and sentiment classes, allowing it to classify new documents based on their word features.

- **Random Forest algorithm :**

  The Random Forest algorithm(Breiman, 1996) operates by constructing an ensemble of decision trees, where each tree independently classifies the sentiment of a text input.

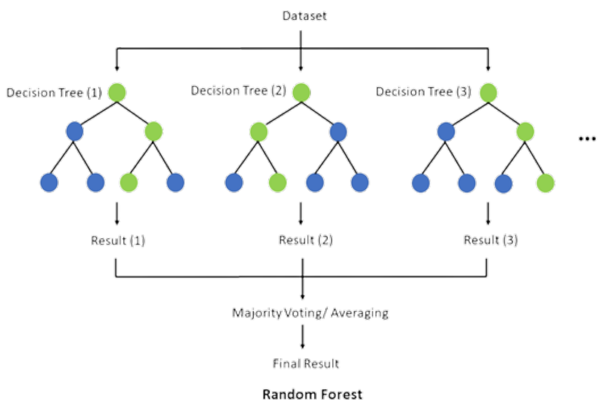

FIGURE 2.2: Random forest algorithm, TIBCO, Source

The algorithm starts by randomly selecting a subset of features from the input data, which are typically the words or n-grams present in the text. These features form the root node of the first decision tree. The decision tree recursively splits the data based on the selected features and their corresponding values, aiming to create homogeneous subsets of data in terms of sentiment. Each split is chosen by evaluating various metrics, such as Gini impurity or information gain, to maximize the separation of sentiment classes. This process continues until a specified depth or a termination condition is met for each decision tree.

Once the ensemble of decision trees is constructed, sentiment analysis is performed by aggregating the predictions of all the individual trees. Each tree assigns a sentiment label to the input text based on the features it selected and the learned decision rules. The final sentiment prediction is determined by either a majority voting scheme or by averaging the probabilities assigned by each tree. The Random Forest algorithm's strength lies in its ability to handle high-dimensional feature spaces, capture complex interactions between features, and mitigate overfitting through the ensemble approach.

- **SVC algorithm :** The SVC algorithm takes a set of training data consisting of labeled examples, where each example is associated with a sentiment label (positive or negative). It maps these examples into a high-dimensional feature space, where each feature represents a specific aspect of the input text. The algorithm then identifies the hyperplane that maximally separates the positive and negative examples, while also maintaining a margin of separation.

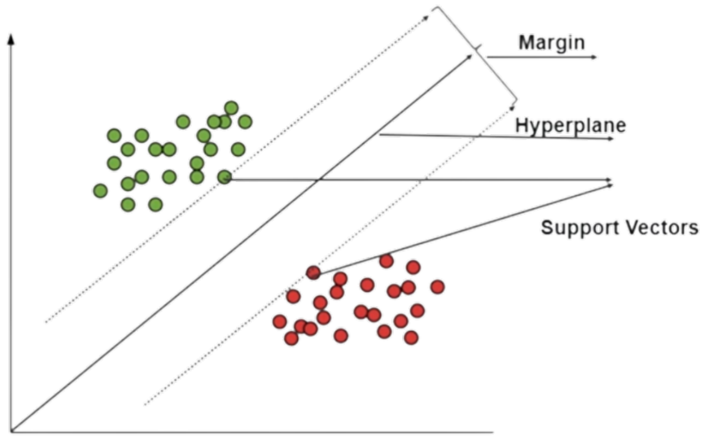

FIGURE 2.3: Support Vector Classifier, Edureka, Source

During the classification phase, the SVC algorithm uses the identified hyperplane to predict the sentiment of new, unlabeled data points. It assigns the sentiment label based on which side of the hyperplane the data point falls on. The SVC algorithm strives to find the hyperplane that not only separates the classes but also generalizes well to unseen data by minimizing the classification error and maximizing the margin between the classes.

In summary, the SVC algorithm for sentiment analysis constructs a hyperplane in a high-dimensional space to separate positive and negative sentiment examples. It learns the optimal hyperplane from labeled training data and uses it to classify new, unlabeled data points. By maximizing the margin and minimizing the classification error, the algorithm aims to make accurate predictions for sentiment analysis tasks.

- **Logistic Regression :**

  Logistic Regression is a popular linear classification algorithm that models the probability of belonging to a certain class. It uses a logistic function to map the input features to the probability values, enabling the classification of instances into different classes based on a threshold. Logistic Regression is computationally efficient and performs well in scenarios where the classes are linearly separable.

By employing these classification models alongside TF-IDF features, this study aims to evaluate their effectiveness in sentiment analysis of Marrakesh hotel reviews. The performance of each algorithm will be assessed based on metrics such as accuracy, precision, recall, and F1 score, providing valuable insights into the optimal approach for sentiment classification in the context of Marrakesh hotel reviews.

2.3.2 LSTM Approach

LSTM(Su and Jung, 2018) is a type of RNN that addresses the vanishing and exploding gradient problems by introducing a memory cell and various gating mechanisms. These mechanisms enable the LSTM to selectively remember and forget information over long sequences, making it well-suited for sentiment analysis tasks that involve analyzing the sentiment expressed across sentences or documents.

The LSTM unit consists of three main components: the input gate, the forget gate, and the output gate. These gates regulate the flow of information through the memory cell. Additionally, the LSTM introduces a cell state, which acts as a conveyor belt, enabling information to flow along the sequence with minimal disruption. The equations governing the LSTM cell are as follows:

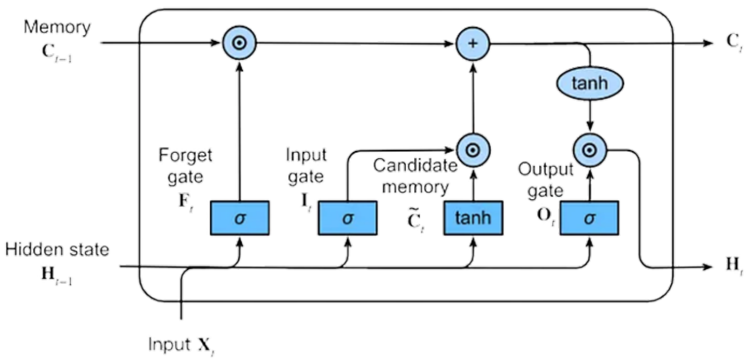

FIGURE 2.4: The Architecture of the LSTM Model, Medium, Source

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \tag{2.1}$$

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \tag{2.2}$$

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \tag{2.3}$$

$$\tilde{C}_t = \tanh(W_c \cdot [h_{t-1}, x_t] + b_c) \tag{2.4}$$

$$C_t = f_t \odot C_{t-1} + i_t \odot \tilde{C}_t \tag{2.5}$$

$$h_t = o_t \odot \tanh(C_t) \tag{2.6}$$

where x_t represents the input at time step t, h_t is the hidden state at time step t, f_t, i_t, and o_t denote the forget, input, and output gates respectively. The $\tilde{C}_t$ term represents the candidate cell state, C_t is the cell state at time step t, and σ and $\odot$ denote the sigmoid and element-wise multiplication operations respectively. W and b represent the weights and biases of the corresponding gates and cells.

2.3.3 BERT Approach

The Transformer design utilizes an encoder-decoder framework that doesn't depend on recurrence or convolutions for producing an output. The encoder takes an input sequence and converts it into a sequence of continuous representations. On the other hand, the decoder takes the encoder's output along with its own output from a previous time step and generates a sequence of outputs.

Here is an image that visualizes the architecture:

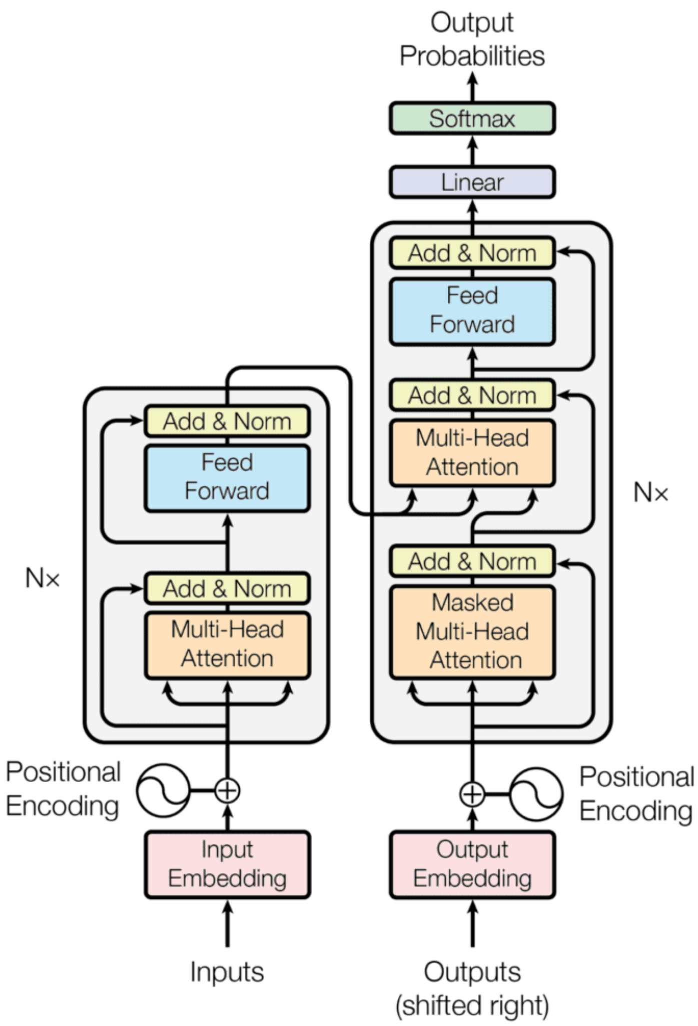

FIGURE 2.5: The Transformer Model Architecture (Vaswani et al., 2017)

As can be seen in Figure 2.5, the data traverses the network as follows: the input and output tokens are converted to vectors using learned embeddings. Because all tokens are processed simultaneously, positional information is lost. This issue is solved by adding vectors that carry information regarding the relative or absolute position of the tokens. These vectors are called positional encodings. The model consists of an encoder-decoder structure. The encoder and the decoder consist of N stacked Feed-Forward and Multi-Head Attention layers. The inputs to the encoder first proceed through the Multi-Head Attention layer, which consists of self-attention layers that examine the other words in the input sentence while encoding one specific word. The encoded input then moves to the Feed-Forward layer. The decoder consists of the same elements. However, between the Feed-Forward and the Multi-Head Attention layer, an additional attention layer has been inserted to focus on the most relevant parts of the input sentence. The Transformer architecture is designed to perform machine translation experiments. However, the authors have shown that it also generalizes to other natural language processing tasks. The activations in the output can be used for tasks such as speech recognition, question answering summarization, language understanding, and many more.

BERT(Munikar, Shakya, and Shrestha, 2019), short for Bidirectional Encoder Representations from Transformers, is an advanced deep learning model created by researchers at Google. It has gained widespread popularity in the field of natural language processing. BERT has become an integral part of the Google search engine since 2020, playing a crucial role in handling all English language search queries. It is based on the Transformer model and was first introduced in a paper titled "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding" (Devlin et al., 2018).

One notable feature of BERT is its bidirectional nature. Unlike traditional "unidirectional" models that consider only the context to the left of a word, BERT takes into account both the left and right context of a word during text processing. This bidirectional approach greatly enhances BERT's performance on tasks that require a comprehensive understanding of the word's surrounding context, such as natural language understanding and machine translation.

2.4 Aspect Based sentiment Analysis

Aspect-Based Sentiment Analysis (ABSA) (Afzaal et al., 2019) is an advanced approach that addresses the limitations of traditional sentiment analysis by focusing on the identification and analysis of specific aspects or attributes within a given text and their associated sentiments. Unlike traditional sentiment analysis, which provides an overall sentiment score, ABSA aims to uncover the sentiments expressed towards different aspects of a product, service, or experience. By breaking down the text into fine-grained aspects, ABSA enables a more comprehensive understanding of the nuanced opinions and sentiments expressed by users.

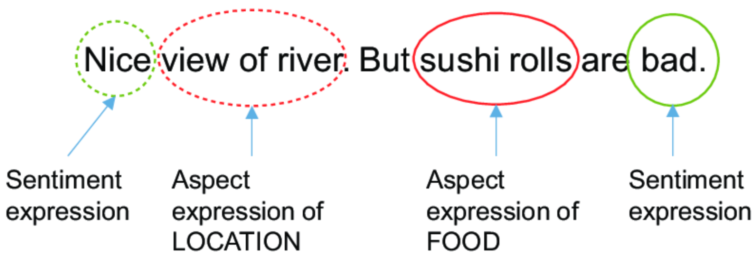

FIGURE 2.6: An example of ABSA, (Noh, Park, and Park, 2019)

ABSA involves two primary steps: aspect extraction and sentiment classification. In the aspect extraction phase, the goal is to identify the aspects or features mentioned in the text that are relevant to the sentiment analysis. This step often relies on techniques such as named entity recognition or syntactic parsing. Once the aspects are identified, the sentiment classification phase assigns sentiment polarities (positive, negative, or neutral) to each aspect, indicating the sentiment associated with that particular feature.

The application of ABSA is particularly beneficial in domains where analyzing multiple aspects of a product or service is crucial. For example, in the context of hotel reviews, ABSA can help identify sentiments towards various aspects such as staff behavior, room cleanliness, amenities, and food quality. This granular analysis provides valuable insights for businesses to prioritize and improve specific areas based on customer feedback.

In this chapter, we talked about the basics of sentiment analysis and why it's important in research. We looked at a lot of previous studies to learn more about what people already know. We also studied different ways to figure out the sentiment in text. We talked about a special technique called aspect-based sentiment analysis, which helps us understand the sentiment towards specific things. The information we learned in this chapter will help us in the next chapters when we use and test these sentiment analysis techniques.

Chapter 3

Proposed Framework

1. Application framework
2. Data collection and preprocessing
3. Aspect based sentiment analysis approach
4. Sentiment classification approach

In order to develop the aspect-based sentiment analysis system, two essential components need to be constructed: the aspect detector and the sentiment classifier. The following sections will provide an overview of the application's framework and elaborate on these two main components.

3.1 Application Framework

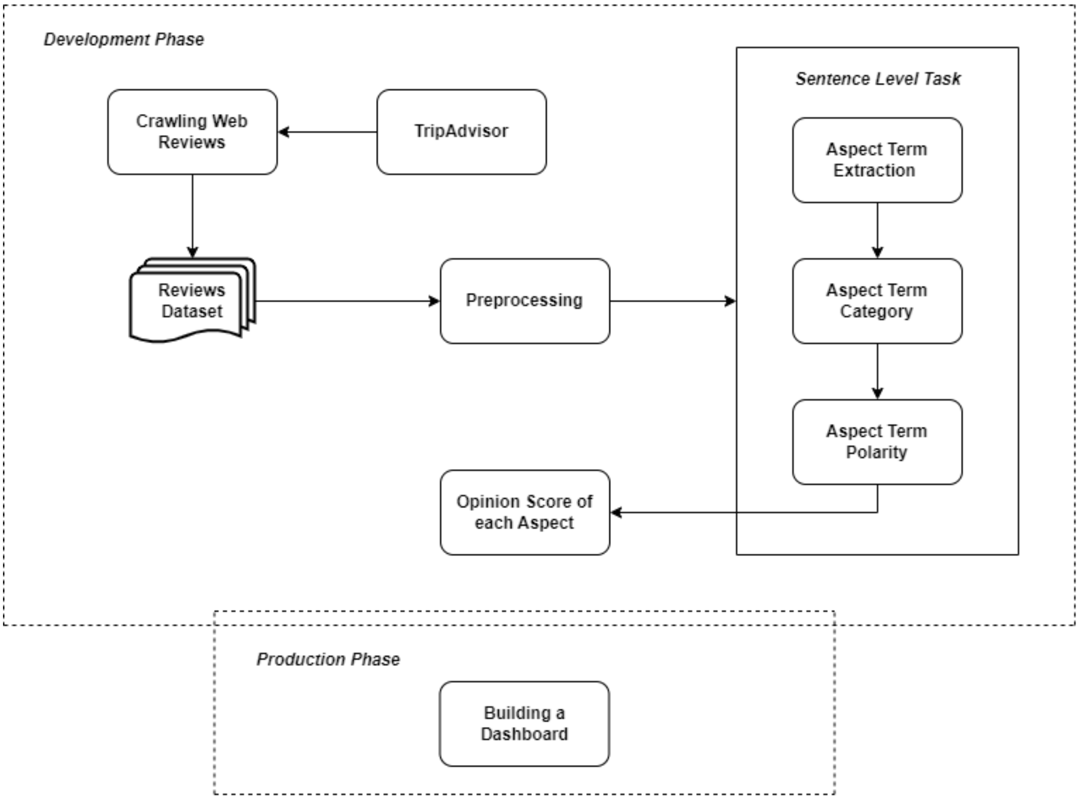

FIGURE 3.1: Architecture of our Review Processing system

The process of conducting aspect-based sentiment analysis of hotel reviews involves several key steps. Firstly, This study and analysis focused on the TripAdvisor website, which was selected due to its widespread usage and extensive collection of hotel reviews. To accomplish this, a specialized scraper was created utilizing the Selenium Python library. This results in a reviews dataset, which serves as the foundation for subsequent analysis.

To ensure data quality and usability, preprocessing is performed on the reviews dataset. This step involves cleaning the data, removing irrelevant information, and standardizing the text format. Once the dataset is preprocessed, aspect term extraction is employed to identify and extract specific aspects or features mentioned in the reviews. These aspects could be related to various aspects of a hotel experience, such as service, food, or price.

After aspect term extraction, the next step is to categorize the extracted aspects into different categories. This categorization helps in organizing and understanding the different aspects that are being discussed in the reviews. Furthermore, aspect term polarity is determined, which involves identifying whether the sentiment expressed towards each aspect is positive, negative, or neutral.

Finally, an opinion score is assigned to each aspect, indicating the strength or intensity of sentiment associated with that particular aspect. This score provides a quantitative measure of the sentiment expressed in the reviews. By following this comprehensive process of data collection, preprocessing, aspect term extraction, aspect term category classification, aspect term polarity determination, and opinion score calculation, aspect-based sentiment analysis of hotel reviews can be effectively conducted, leading to valuable insights and understanding of customer sentiments and preferences.

3.2 Data Collection and Preprocessing

In order to conduct the analysis of textual content, it is essential to have a dataset that contains relevant information. This dataset should ideally consist of text along with corresponding ratings that can be annotated. For this study, TripAdvisor was chosen as the primary data source. TripAdvisor is an American travel and restaurant website company that was established in 2000. It provides a platform for users to share reviews of hotels and restaurants, make accommodation bookings, and access other travel-related content.

Although there are alternative options such as Google Places reviews, TripAdvisor was preferred due to its popularity and the extensive amount of content it possesses. With 19 years of operation, TripAdvisor has accumulated a larger volume of data compared to Google Places. Since TripAdvisor does not offer direct access to its data, a web scraper was developed specifically to collect the required information from Marrakech Hotel reviews on the TripAdvisor platform.

3.2.1 Data Collection

To ensure a comprehensive dataset for the Sentiment Analysis of Marrakesh hotel reviews, data was gathered by scraping information from TripAdvisor, a popular online travel platform. The web scraping process was facilitated using the Selenium framework, which is commonly used for automating web browsers.

The data scraping procedure followed a systematic approach. Initially, hotel links were extracted from a provided URL, which served as the starting point for collecting data. These links were organized and returned as a list of dictionaries, making it easy to access and scrape individual hotel reviews.

The next step involved scraping the reviews for each hotel from TripAdvisor. The scraping algorithm accessed the page of each hotel and extracted relevant information in a meticulous manner. This process resulted in a dataset consisting of a list of dictionaries, where each dictionary contained key information related to a specific review.

3.2.2 Exploratory Data Analysis

I gathered a large dataset of around 55,000 English reviews from TripAdvisor about Marrakesh hotels. These reviews were collected carefully to include a wide range of opinions and perspectives. This extensive collection of reviews provides a solid foundation for conducting a detailed analysis of people's sentiments towards Marrakesh hotels.

The screenshot displays a review from TripAdvisor for the hotel "Sol Oasis Marrakesh".

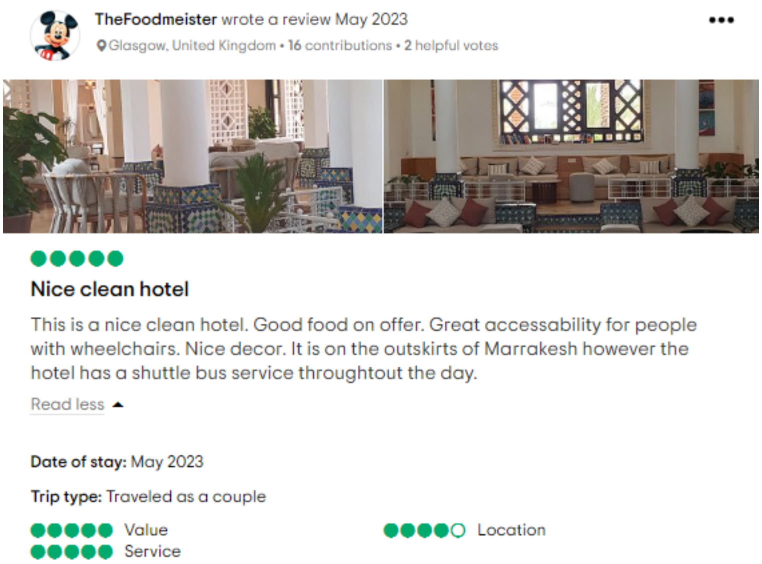

FIGURE 3.2: A review from TripAdvisor

Variable name	Value
Hotel Name	Sol Oasis Marrakech
Review Date	May 2023
Review Rating	5
Review Title	Nice clean hotel
Review Text	This is a nice clean hotel. Good food ...
Date Of Stay	May 2023
Trip Type	Traveled as a couple
Reviewer Location	Glasgow, United Kingdom

TABLE 3.1: Details of a review for a hotel on TripAdvisor

The table provides a comprehensive snapshot of a hotel review from TripAdvisor. It includes essential details such as the hotel's name, review date, rating, title, and text. The additional information about the reviewer's stay, trip type, and location adds context to the review. This dataset will be invaluable for conducting sentiment analysis on hotel reviews and gaining a deeper understanding of customer experiences.

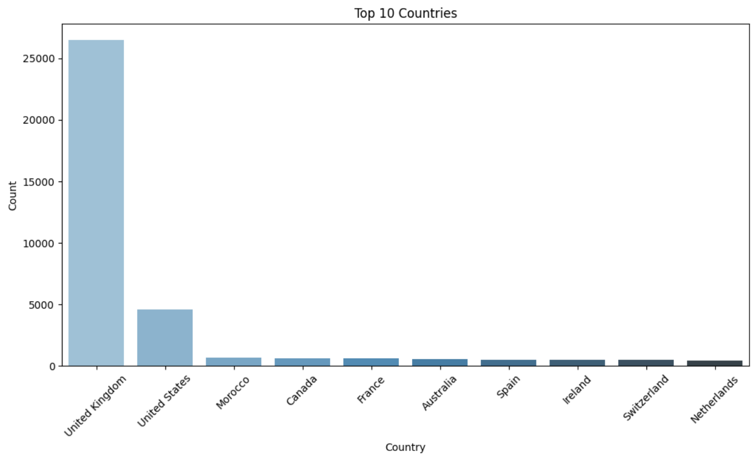

FIGURE 3.3: Country Distribution Across Dataset

The bar plot showcases the top 10 countries based on their occurrences in the dataset. The dataset reveals that the United Kingdom holds the highest count, surpassing 25,000 occurrences, making it the most prevalent country. Following closely behind is the United States, with an approximate count of 5,000.

The reason for the dominance of the United Kingdom and the United States is because we specifically focused on collecting English reviews while scraping information from TripAdvisor.

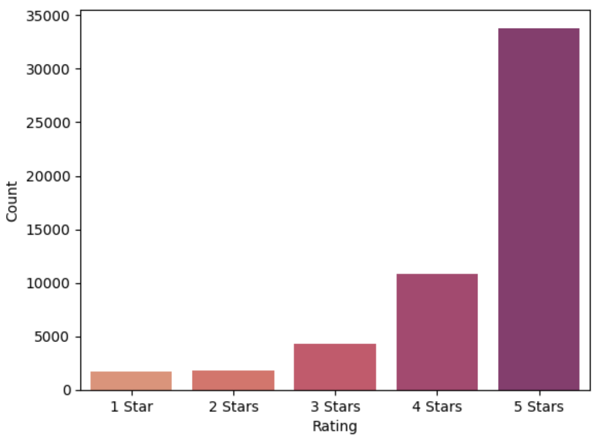

FIGURE 3.4: Overall Rating Distribution

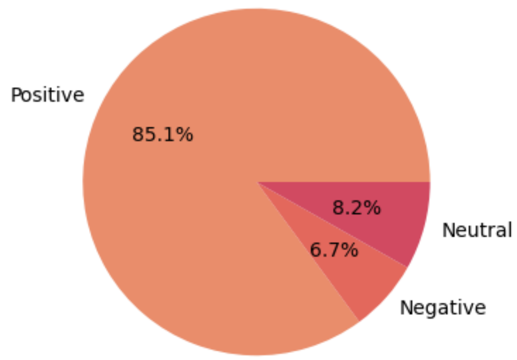

FIGURE 3.5: Overall Sentiment Distribution

The distribution of ratings in the hotel reviews dataset reveals an interesting pattern. The majority of customers have expressed highly positive sentiments, as indicated by the substantial count of ratings at the highest end of the scale, specifically 4 and 5. This suggests that a significant proportion of guests have had exceptional experiences and have been highly satisfied with the hotels. However, it is essential to note the presence of lower ratings, particularly 1 and 2, which indicate instances of negative sentiment and dissatisfaction. While these lower ratings form a smaller portion of the overall distribution, they still signify areas where improvements are necessary to address customer concerns and enhance their experiences. This distribution showcases the range of sentiments expressed by customers in hotel reviews, emphasizing the need to understand both the positive aspects and the areas for improvement in order to gain a comprehensive understanding of customer satisfaction levels.

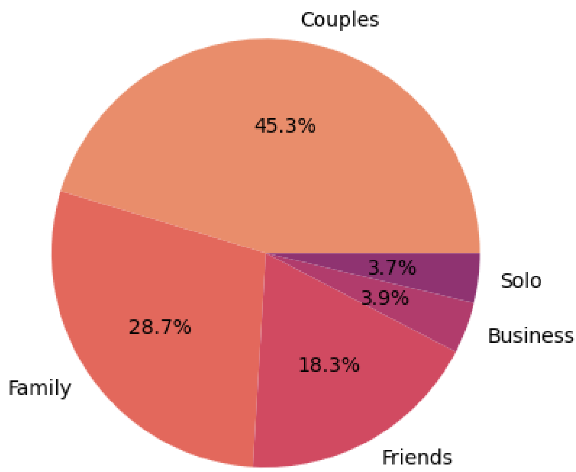

FIGURE 3.6: Trip type Distribution

The trip type distribution across the dataset provides valuable insights into the demographics and preferences of the reviewers. Among the various trip types, the most common category observed is 'couple', indicating that a significant number of guests have visited the hotels for romantic getaways or vacations. Following 'couple', the next prevalent trip type is 'family', suggesting that a considerable portion of the reviewers consisted of families seeking accommodation options that catered to their specific needs. Additionally, the 'friends' category demonstrates that a substantial number of reviewers embarked on trips with their friends, possibly indicating a desire for shared experiences and group activities. Interestingly, the 'business' and 'solo' trip types have notably lower counts compared to the other categories, implying that the dataset predominantly comprises leisure-oriented reviews. Understanding the distribution of trip types is crucial for hoteliers to tailor their services and amenities to cater to the diverse needs and preferences of their target audience.

3.2.3 Preprocessing techniques

Data preparation is a critical step that ensures the quality, reliability, and suitability of the data for analysis. It helps to uncover meaningful insights, improve model performance, and facilitate informed decision-making based on the results of sentiment analysis on hotel reviews.

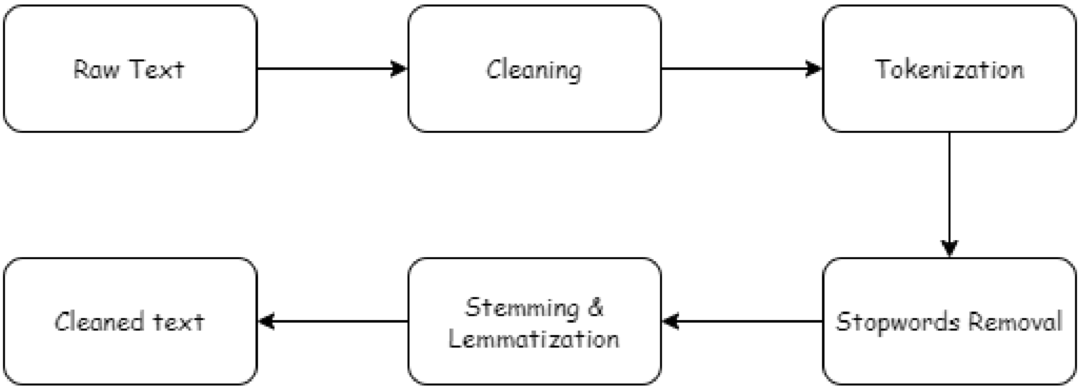

FIGURE 3.7: Text Preprocessing

3.2.3.1 NLTK and SpaCy

NLTK and SpaCy are popular Python libraries for NLP tasks. NLTK offers a comprehensive suite of tools, allowing fine-grained control over the NLP pipeline. It provides functionalities like tokenization, stemming, POS tagging, and syntactic parsing. NLTK's strength lies in its flexibility and extensive resources. SpaCy, on the other hand, focuses on efficiency and ease of use. It excels in large-scale data processing and real-time applications. SpaCy offers functionalities such as tokenization, POS tagging, dependency parsing, named entity recognition, and entity linking. It prioritizes speed and performance with its streamlined pipeline. Both libraries have pre-trained models for various tasks. The choice between them depends on specific project requirements, with NLTK being ideal for research and experimentation and SpaCy for production-level applications.

3.2.3.2 Tokenization

Tokenization is a fundamental task in natural language processing (NLP) that involves breaking down a text into smaller units called tokens. These tokens typically correspond to words, but they can also represent other linguistic units like sentences or subwords. Tokenization plays a crucial role in various NLP tasks, including sentiment analysis on hotel reviews. By segmenting a text into tokens, it enables further analysis and processing at a granular level. Tokenization helps to handle complex linguistic structures, punctuation marks, and special characters, thereby transforming unstructured text into a structured format that can be easily understood by machine learning algorithms.

"This hotel is my absolute favorite."

Tokenization

↓

This	hotel	is	my	absolute	favorite

3.2.3.3 Stopwords removal

Stopwords removal is a common preprocessing technique in Natural Language Processing (NLP) that involves eliminating commonly occurring words that do not carry significant meaning for the task at hand. Stopwords are words such as "the," "is," "and," and "in" that appear frequently in a language but often contribute little to the overall understanding of the text.

By removing stopwords, several benefits can be achieved. Firstly, it can reduce the dimensionality of the text data, which is especially useful when working with large datasets. Removing stopwords helps in focusing on the more meaningful and informative words, allowing for more efficient and effective analysis. Secondly, eliminating stopwords can enhance the accuracy and performance of various NLP tasks, such as text classification, sentiment analysis, and information retrieval. By removing noise words, the model can concentrate on the important words that carry more semantic value, leading to better results.

"This hotel is my absolute favorite."

Stopwords removal

↓

~~This~~	hotel	~~is~~	~~my~~	absolute	favorite

3.2.3.4 Stemming

Stemming is a technique used in natural language processing (NLP) to reduce words to their base or root form, known as a stem. Stemming aims to normalize words by removing prefixes or suffixes. This process helps to consolidate variations of the same word, enabling more effective analysis and comparison of text data.

changing → chang	waited → wait	information → inform

3.2.3.5 Lemmatization

Lemmatization is a linguistic process commonly used in natural language processing (NLP) to reduce words to their base or dictionary form, known as a lemma. Unlike stemming, which focuses on removing prefixes or suffixes, lemmatization takes into account the morphological analysis of words and considers the context and part of speech (POS) of each word. By applying lemmatization, words are transformed into their canonical forms, which helps in reducing inflected or derived words to their common root.

$\boxed{\text{chang} \rightarrow \text{change}}$ $\boxed{\text{wait} \rightarrow \text{wait}}$ $\boxed{\text{was} \rightarrow \text{be}}$

3.2.3.6 Part of Speech

Part of speech (POS) in natural language processing (NLP) refers to the grammatical category or syntactic role that a word plays within a sentence. It is a fundamental concept in NLP used to analyze and understand the structure of language. POS tagging involves labeling each word in a text with its appropriate part of speech, such as noun, verb, adjective, adverb, pronoun, preposition, conjunction, or interjection. This process is crucial for various NLP tasks, including syntactic parsing, semantic analysis, information extraction, and machine translation. POS tagging helps in disambiguating words with multiple meanings, capturing syntactic relationships, and facilitating subsequent analysis and interpretation of text data. By identifying the part of speech of each word, NLP systems can gain insights into the grammatical structure and meaning of sentences, enabling more accurate language understanding and processing.

"This hotel is my absolute favorite."

$\downarrow$ *Part-of-speech*

$\boxed{\text{This} \rightarrow \text{DT}}$ $\boxed{\text{hotel} \rightarrow \text{NN}}$ $\boxed{\text{is} \rightarrow \text{VBZ}}$ $\boxed{\text{my} \rightarrow \text{PRP}}$ $\boxed{\text{absolute} \rightarrow \text{JJ}}$ $\boxed{\text{favorite} \rightarrow \text{NN}}$

3.3 Aspect Based Sentiment Analysis Approach

Aspect-based sentiment analysis is a technique used to analyze the sentiment expressed towards different aspects or features of a product or service in text data, such as hotel reviews. The goal is to extract and quantify the sentiment associated with specific aspects mentioned in the reviews. The process involves several steps: preprocessing, noun extraction, aspect association and sentiment analysis.

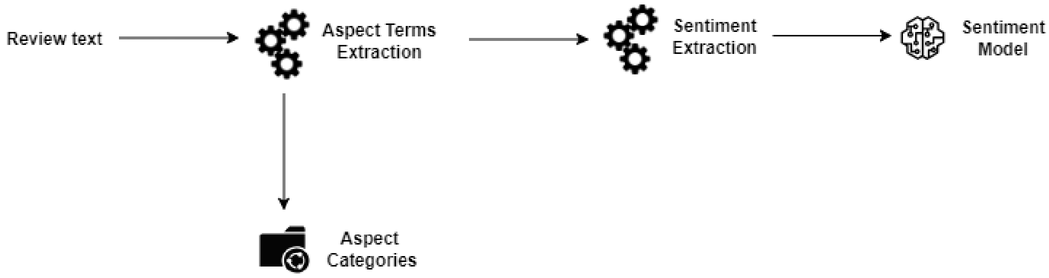

FIGURE 3.8: Aspect based sentiment analysis flow

The process begins with preprocessing the text to ensure data cleanliness. This preprocessing step help improve the accuracy of subsequent analysis.

Next, the text is tokenized to extract individual words or tokens. The function **get_noun()** is applied to the tokenized text, leveraging part-of-speech tagging to identify nouns that represent the aspects of interest. This step allows for the identification of key aspects mentioned in the reviews.

Algorithm 1 get_noun(lines)

procedure GET_NOUN(*lines*)
  $tokenized \leftarrow \text{word\_tokenize}(lines)$
  $tagged\_words \leftarrow \text{pos\_tag}(tokenized)$
  $nouns \leftarrow \{\}$
  for $(word, pos)$ **in** $tagged\_words$ **do**
    if $\text{pos}[:2] == \text{'NN'}$ **then**
      $nouns \leftarrow \text{add}(nouns, word)$
  return *nouns*

To associate alternative names or synonyms with each aspect, the function **get_similar_words()** is utilized. This function computes similarity scores between the extracted nouns and the predefined aspects. If the similarity score surpasses a predefined threshold (0.50 in this case), the noun is considered associated with the corresponding aspect. This step enables a more comprehensive analysis by capturing different expressions referring to the same aspect.

Algorithm 2 get_similar_words(nouns, aspects)

```
procedure GET_SIMILAR_WORDS(nouns, aspects)
    aspect_classes ← {aspect : [aspect] for aspect in aspects}
    for noun in nouns do
        scores ← []
        for aspect in aspects do
            aspect_token ← nlp(aspect)
            noun_token ← nlp(noun)
            similarity_score ← aspect_token.similarity(noun_token)
            scores ← append(scores, similarity_score)
        index ← scores.index(max(scores))
        aspect_name ← aspects[index]
        if max(scores) > 0.50 then
            value ← aspect_classes[aspect_name]
            value ← append(value, noun)
            aspect_classes[aspect_name] ← list(set(value))
        else
            pass
    return aspect_classes
```

After the aspects and their alternative names have been identified, the sentiment analysis phase takes place. The function **get_sentiment()** is employed to determine the sentiment expressed towards each aspect in the original text. It uses a combination of question-answering and sentiment analysis models to obtain answers to predefined questions about each aspect and determine the sentiment label (positive, negative, or neutral) associated with the answers.

Algorithm 3 get_sentiment(aspect_classes, text)

```
procedure GET_SENTIMENT(aspect_classes, text)
    sentiment_dict ← {aspect : 0 for aspect in aspect_classes}
    for aspect in aspect_classes do
        alt_names ← aspect_classes[aspect]
        for name in alt_names do
            question ← f"What do you think about the {name}?"
            print(question)
            QA_input ← {'question' : question,' context' : text}
            qa_result ← qa_model(QA_input)
            print(qa_result)
            answer ← qa_result['answer']
            sent_result ← sent_model(answer)
            print(sent_result)
            sentiment ← sent_result[0]['label']
            if sentiment == 'NEGATIVE' then
                sentiment, score ← 'Negative', −1
            else if sentiment == 'POSITIVE' then
                sentiment, score ← 'Positive', 1
            else
                sentiment, score ← 'Neutral', 0
            value ← sentiment_dict[aspect] + score
            sentiment_dict[aspect] ← value
    return sentiment_dict
```

.

Algorithm 4 compute(text, aspects)

```
procedure COMPUTE(text, aspects)
    preprocess_text ← preprocess(text)
    noun_list ← get_noun(preprocess_text)
    aspect_classes ← get_similar_words(noun_list, aspects)
    sentiment_result ← get_sentiment(aspect_classes, text)
    return sentiment_result
```

The sentiment results are aggregated and scored for each aspect using the **aspect_sentiment()** function. The reviews content is processed, and for each review, the sentiment score for each aspect is computed. Positive sentiment scores are incremented for aspects receiving positive feedback, while negative sentiment scores are incremented for aspects receiving negative feedback. The scores are accumulated for all reviews.

Algorithm 5 aspect_sentiment(aspects)

```
procedure ASPECT_SENTIMENT(aspects)
    reviews_content ← df
    aspect_score ← {asp : {'positive' : 0,' negative' : 0} for asp in aspects}
    if reviews_content.to_dict() then
        for text in reviews_content['Review'] do
            sentiment_result ← compute(text, aspects)
            for result in sentiment_result do
                score ← sentiment_result[result]
                if score > 0 then
                    aspect_score[result]['positive'] ← aspect_score[result]['positive'] + score
                else if score < 0 then
                    aspect_score[result]['negative'] ← aspect_score[result]['negative'] − score
                else
                    pass
    else
        print('No data')
    result_list ← [[k,' positive', v['positive']] for k, v in aspect_score.items()]
    result_list.extend([[k,' negative', v['negative']] for k, v in aspect_score.items()])
    aspects_df ← pd.DataFrame(result_list, columns = ['aspect',' sentiment',' score'])
```

Finally, the sentiment scores are visualized using a bar plot. The **aspect_sentiment()** function generates a pandas DataFrame, **aspects_df**, containing information about each aspect, its sentiment (positive or negative), and the corresponding score. The seaborn library is then used to create a bar plot, where each aspect is represented by a bar, categorized by sentiment and displaying the corresponding score.

By following this aspect-based sentiment analysis process, insights can be gained into how hotel guests perceive different aspects of their experience. The visual representation of sentiment scores allows for easy comparison and identification of the aspects that contribute most positively or negatively to overall guest sentiment.

In conclusion, aspect-based sentiment analysis of hotel reviews involves preprocessing the text, extracting nouns, associating them with relevant aspects, performing sentiment analysis on user feedback, and visualizing the sentiment scores. The process enables a deeper understanding of customer sentiment towards specific aspects of a hotel and facilitates valuable insights for businesses to improve their services.

3.4 Sentiment Classification Approach

The final step in the process of aspect-based sentiment analysis is sentiment classification. This step involves determining the sentiment expressed towards each aspect identified in the hotel reviews. In Chapter 2 of the thesis, detailed explanations of the methods employed for sentiment classification are provided. These methods include Term Frequency Inverse Document Frequency (TF-IDF), Recurrent Neural

Networks with Bidirectional Long Short-Term Memory (LSTM), and Bidirectional Encoder Representations from Transformers (BERT).

In this chapter, we have presented the proposed framework for sentiment analysis and aspect-based sentiment analysis in the context of this study. The application framework provides a structured approach for the implementation of the study, ensuring consistency and clarity in the subsequent steps.

Next Chapter will focus on presenting the Results and Analysis, including evaluation metrics, results from classic machine algorithms, LSTM, and BERT models, results comparison, and dashboard visualization.

Chapter 4

Results and Analysis

1. Evaluation metrics
2. Classic machine learning algorithms results
3. LSTM results
4. BERT results
5. Results comparison
6. Dashboard visualization

This chapter focuses on presenting the results and analysis of the sentiment analysis and aspect-based sentiment analysis techniques applied in this study. The chapter begins by introducing the evaluation metrics used to assess the performance of the models. It then proceeds to present the results obtained from classic machine algorithms, LSTM, and BERT models. A thorough comparison of the results is conducted to identify the best model for our case. Additionally, the chapter shows a visualization of the results through a dashboard, providing a comprehensive overview of the findings.

4.1 Evaluation Metrics

In the field of sentiment classification, evaluating the performance of models is crucial to assess their effectiveness in predicting sentiments accurately. Several evaluation metrics are commonly used to measure the performance of sentiment classification models, including accuracy, precision, recall, and F1 score.

	Actual Positive	**Actual Negative**
Predicted Positive	True Positives (TP)	False Positives (FP)
Predicted Negative	False Negatives (FN)	True Negatives (TN)

TABLE 4.1: Confusion Matrix

Accuracy is a widely used metric that calculates the proportion of correctly classified instances over the total number of instances. It provides a general overview of the model's overall correctness in predicting sentiment labels. However, accuracy alone may not be sufficient when dealing with imbalanced datasets, where the distribution of positive and negative sentiment labels is uneven.

$$\text{Accuracy} = \frac{\text{True Positives} + \text{True Negatives}}{\text{True Positives} + \text{True Negatives} + \text{False Positives} + \text{False Negatives}}$$

Precision measures the proportion of correctly predicted positive sentiment instances out of all instances predicted as positive. It quantifies the model's ability to correctly identify positive sentiment, minimizing false positives. Precision is particularly valuable when the consequences of false positive predictions are significant, such as in applications where misclassifying negative sentiment as positive can have detrimental effects.

$$\text{Precision} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Positives}}$$

Recall, also known as sensitivity or true positive rate, measures the proportion of correctly predicted positive sentiment instances out of all actual positive instances. It captures the model's ability to identify and retrieve positive sentiment correctly, minimizing false negatives. Recall becomes crucial when missing positive sentiment instances can lead to missed opportunities or vital insights.

$$\text{Recall} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}}$$

F1 score is a combined metric that balances both precision and recall. It computes the harmonic mean of precision and recall, providing a single value that represents the overall performance of the model. F1 score is commonly used when there is an equal importance placed on precision and recall, seeking a balance between correctly predicting positive sentiment and minimizing false positives and false negatives.

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

In the context of sentiment classification of Marrakesh hotel reviews, these evaluation metrics will be employed to assess the performance of different sentiment classification models. The metrics will help measure the accuracy, precision, recall, and overall effectiveness of the models in capturing the sentiments expressed by customers. By carefully evaluating and comparing these metrics, we can identify the most suitable model for accurately predicting sentiments and providing valuable insights for the improvement of hotel services in Marrakesh.

4.2 Classic Machine Learning Algorithms Results

Model	Accuracy
Random Forest	0.779039
SVC	0.854990
Logistic Regression	0.859444
BernoulliNB	0.754026

TABLE 4.2: Accuracy scores of different models

Among these models, both SVC and Logistic Regression demonstrated higher accuracy scores compared to the other two models. This indicates that they performed better in correctly predicting the sentiment expressed in the hotel reviews.

SVC, a machine learning algorithm that finds an optimal hyperplane to separate data points into different classes, achieved an accuracy score of 85.50%. This suggests that SVC is effective in capturing the underlying patterns and relationships in the data, leading to more accurate sentiment predictions. The algorithm's ability to handle non-linear relationships between the features and the target variable might have contributed to its superior performance.

Similarly, Logistic Regression, a widely-used classification algorithm, achieved a slightly higher accuracy score of 85.94%. This model applies the logistic function to calculate the probability of a sample belonging to a particular class. In the context of sentiment analysis, Logistic Regression likely leveraged the inherent linearity in the features to make accurate predictions.

Comparatively, Random Forest Classifier and BernoulliNB achieved lower accuracy scores of 77.90% and 75.40%, respectively. Random Forest Classifier utilizes an ensemble of decision trees, whereas BernoulliNB is a probabilistic algorithm based on the assumption of independence between features. These models might have struggled to capture the complex relationships and patterns within the hotel review data, resulting in lower accuracy rates.

Based on the obtained results, both SVC and Logistic Regression demonstrated superior performance in sentiment analysis of hotel reviews. However, considering the slightly higher accuracy achieved by Logistic Regression.So we will use it to train our model.

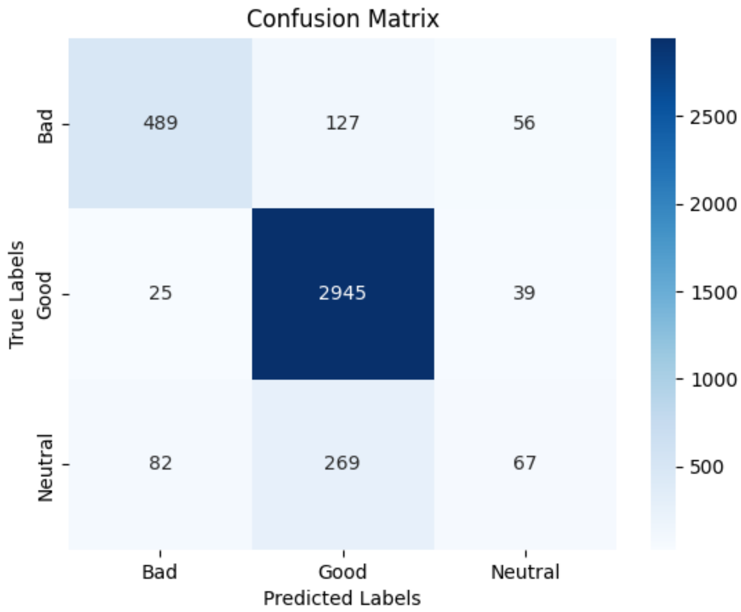

FIGURE 4.1: Confusion Matrix for Logistic Regression

Class	precision	recall	f1-score	support
Bad	0.82	0.73	0.77	672
Good	0.88	0.98	0.93	3009
Neutral	0.41	0.16	0.23	418
accuracy			0.85	4099
macro avg	0.71	0.62	0.64	4099
weighted avg	0.82	0.85	0.83	4099

TABLE 4.3: Classification Report of Logistic Regression

The confusion matrix provides a comprehensive overview of the performance of the sentiment analysis model on hotel reviews. The matrix reveals the model's predictions against the actual sentiments of the reviews, categorized into three classes: "Bad," "Good," and "Neutral."

The diagonal elements of the matrix represent the correct predictions, also known as true positives (TP) and true negatives (TN). In this case, the model accurately classified 489 instances as "Bad," 2945 instances as "Good," and 67 instances as "Neutral." These results demonstrate the model's ability to correctly identify the sentiment of the majority of the hotel reviews.

However, the off-diagonal elements represent the model's incorrect predictions, known as false positives (FP) and false negatives (FN). The model mistakenly classified 127 negative reviews as positive ("Bad" misclassified as "Good"), 82 positive reviews as negative or neutral ("Good" misclassified as "Bad" or "Neutral"), and 269 neutral reviews as either positive or negative ("Neutral" misclassified as "Bad" or "Good").

When examining the precision metric, which measures the accuracy of positive predictions, we find that the model achieved a precision of 0.82 for the "Bad" class, 0.88 for the "Good" class, and 0.41 for the "Neutral" class. The recall metric, indicating the ability to correctly identify positive instances, yielded values of 0.73 for "Bad," 0.98 for "Good," and 0.16 for "Neutral." These results suggest that the model performed well in accurately identifying positive reviews but struggled with the "Neutral" class.

The F1-score, which considers both precision and recall, provides a balanced measure of the model's performance. The F1-score for the "Bad" class was 0.77, indicating a reasonable balance between precision and recall. The "Good" class achieved an impressive F1-score of 0.93, reflecting a strong ability to identify positive reviews accurately. However, the "Neutral" class had a lower F1-score of 0.23, highlighting the model's difficulty in correctly classifying neutral sentiments.

Overall, the model demonstrated an accuracy of 0.85, correctly predicting the sentiment of 85% of the hotel reviews. However, it is important to note that the model's performance varied across different sentiment classes. The macro average precision, recall, and F1-score were 0.71, 0.62, and 0.64, respectively, indicating that the model's performance was relatively weaker in classifying the minority classes.

In conclusion, while the sentiment analysis model exhibited strong performance in identifying positive sentiments, it faced challenges in correctly classifying negative and neutral sentiments. These findings highlight potential areas for improvement in the model, such as enhancing the classification of neutral reviews to enhance the overall performance and accuracy of sentiment analysis on hotel reviews.

4.3 LSTM Results

Before presenting the results the model implemented follows a sequential architecture based on LSTM (Long Short-Term Memory) networks. It involves several steps, starting with the tokenization of the textual data using the Keras Tokenizer. The tokenized sequences are then padded to ensure uniform length, which is necessary for LSTM's input requirements. The sentiment labels are encoded using one-hot encoding.

The model itself consists of an Embedding layer, which learns meaningful representations for words based on their contextual information. A Bidirectional LSTM layer is employed to capture both forward and backward dependencies in the text, enhancing the model's ability to understand the sentiment. Dropout layers are utilized to mitigate overfitting by randomly dropping out units during training.

A fully connected Dense layer follows the LSTM layer, incorporating L2 regularization to control the magnitude of weights and biases. Another Dropout layer is included to further prevent overfitting. The model concludes with a Dense layer using softmax activation, which outputs probabilities for each sentiment class.

During training, the model is optimized using the Adam optimizer with a specified learning rate. The categorical cross-entropy loss function is used to measure the discrepancy between predicted and actual sentiment labels. The model's performance is evaluated using accuracy as the metric.

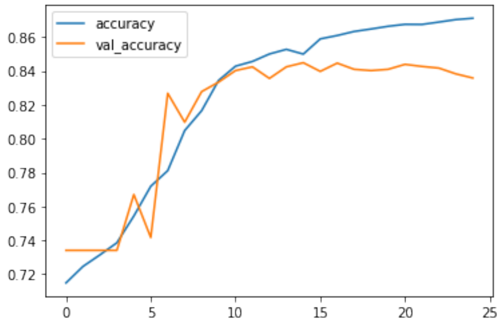

FIGURE 4.2: Accuracy Plot

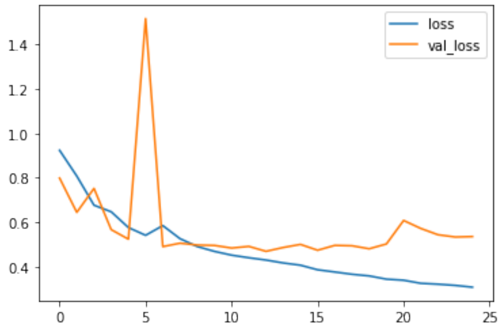

FIGURE 4.3: Loss Plot

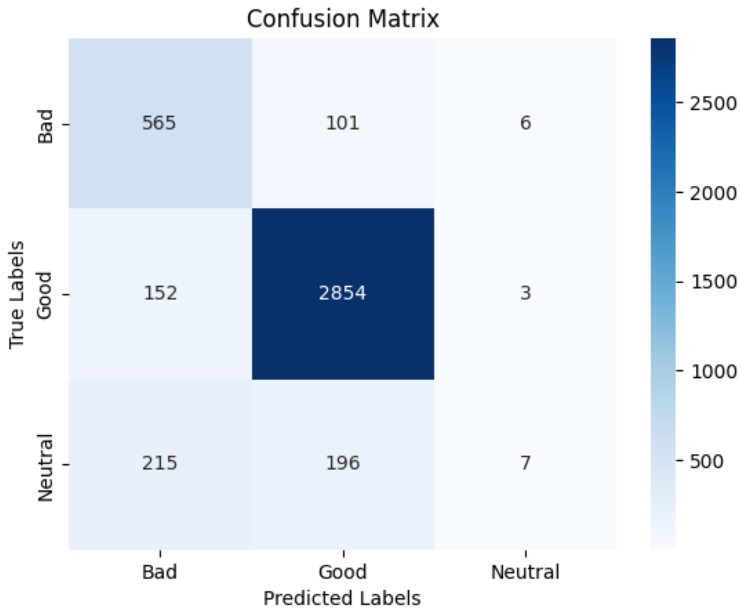

FIGURE 4.4: Confusion Matrix for LSTM

Class	precision	recall	f1-score	support
Bad	0.61	0.84	0.70	672
Good	0.91	0.95	0.93	3009
Neutral	0.44	0.02	0.03	418
accuracy			0.84	4099
macro avg	0.65	0.60	0.55	4099
weighted avg	0.81	0.84	0.80	4099

TABLE 4.4: Classification Report of LSTM

Looking at the matrix, we can observe that the model achieved the highest number of correct predictions for class 1 (positive sentiment) with 2,854 samples correctly classified. The second-highest number of correct predictions was for class 0 (negative sentiment) with 565 samples correctly classified. However, the model struggled to predict class 2 (neutral sentiment), as only 7 samples were correctly classified.

Moving on to the classification report, we can see the precision, recall, and F1-score for each sentiment class. The precision is the proportion of correctly predicted samples out of the total predicted samples for a specific class. The recall is the proportion of correctly predicted samples out of the total actual samples for a specific class. The F1-score is the harmonic mean of precision and recall, providing an overall assessment of the model's performance.

For class 0 (negative sentiment), the precision is 0.61, indicating that 61% of the predicted negative reviews were actually negative. The recall is 0.84, meaning that the model captured 84% of the actual negative reviews. The F1-score is 0.70, which represents the balance between precision and recall for this class.

For class 1 (positive sentiment), the precision is 0.91, indicating that 91% of the predicted positive reviews were indeed positive. The recall is 0.95, implying that the model captured 95% of the actual positive reviews. The F1-score is 0.93, which indicates a high level of accuracy and effectiveness in predicting positive sentiment.

For class 2 (neutral sentiment), the precision is 0.44, suggesting that only 44% of the predicted neutral reviews were actually neutral. The recall is extremely low at 0.02, indicating that the model struggled to capture actual neutral reviews. Consequently, the F1-score for this class is only 0.03, reflecting the poor performance in predicting neutral sentiment.

Overall, the model achieved an accuracy of 0.84, correctly classifying 84% of the reviews in the dataset. The weighted average of the precision, recall, and F1-score across all classes is 0.81, indicating a reasonably good performance for sentiment analysis. However, it's worth noting that the model's performance varies across different sentiment classes, with neutral sentiment being the most challenging to predict accurately.

4.4 BERT Results

BERT is a state-of-the-art pre-trained language model that has achieved impressive performance in various natural language processing (NLP) tasks, including sentiment analysis.

The model's architecture followed a standard approach, beginning with data preprocessing. The dataset was split into training, validation, and test sets. The text data was then tokenized, converting it into a suitable format for the subsequent steps.

Next, the tokenized data was encoded to create input features for the model. These features included input token IDs, attention masks, and other tensors necessary for training and evaluation. The encoded data was converted into tensors and organized into datasets, enabling efficient handling during the training and validation processes.

For sentiment analysis, a specific model designed for sequence classification was employed. This model was initialized with pre-trained weights and configured to accommodate the sentiment analysis task, which involved classifying reviews into positive, negative, or neutral sentiments. The model was trained through multiple epochs, with each epoch consisting of a training loop. During training, the model's parameters were adjusted based on the calculated gradients, and the training loss was tracked.

To optimize the model's performance, an optimizer was employed along with a learning rate scheduler. The optimizer adjusted the model's parameters to minimize the loss, while the scheduler dynamically adjusted the learning rate to aid convergence. After each training epoch, the model's performance was evaluated using a separate validation set. This evaluation process involved calculating the validation loss, generating predictions, and comparing them with the true labels.

Various metrics, such as F1 score and accuracy per class, were computed to assess the model's performance. These metrics provided insights into the model's ability to classify reviews accurately. Throughout the training process, the progress, training loss, validation loss, and F1 score were logged and monitored.

Class	precision	recall	f1-score	support
Bad	0.82	0.85	0.83	310
Good	0.95	0.94	0.95	1513
Neutral	0.48	0.46	0.47	227
accuracy			0.88	2050
macro avg	0.75	0.75	0.75	2050
weighted avg	0.88	0.88	0.88	2050

TABLE 4.5: Classification Report of Bert

The classification report provides valuable insights into the performance of the sentiment analysis model based on the application of BERT. The report presents precision, recall, and F1-score metrics for each sentiment class, as well as the overall accuracy of the model.

Looking at the precision values, it is evident that the model performs well in classifying sentiment class 2 (positive sentiment) with a precision score of 0.95. This indicates that when the model predicts a review as positive, it is correct approximately 95% of the time. The precision for sentiment class 0 (negative sentiment) is also relatively high at 0.82, suggesting that the model has a good ability to identify negative sentiment.

However, the precision score for sentiment class 1 (neutral sentiment) is relatively lower at 0.48. This implies that the model tends to misclassify some neutral reviews as either positive or negative. Further analysis may be required to understand the reasons behind this misclassification and identify potential areas of improvement.

Regarding recall, sentiment class 2 (positive sentiment) demonstrates a recall score of 0.94, indicating that the model effectively identifies the majority of positive reviews in the dataset. Sentiment class 0 (negative sentiment) also shows a reasonable recall score of 0.85, suggesting that the model successfully captures a significant portion of negative sentiment.

On the other hand, sentiment class 1 (neutral sentiment) exhibits a lower recall score of 0.46. This implies that the model struggles to correctly classify some neutral reviews, potentially mislabeling them as positive or negative. Exploring additional strategies or incorporating domain-specific knowledge might be beneficial in improving the recall for neutral sentiment.

The F1-score, which is the harmonic mean of precision and recall, provides a comprehensive evaluation metric for each sentiment class. Sentiment class 2 (positive sentiment) achieves the highest F1-score of 0.95, indicating excellent overall performance. Sentiment class 0 (negative sentiment) has a reasonably balanced F1-score of 0.83, while sentiment class 1 (neutral sentiment) exhibits a relatively lower F1-score of 0.47.

The overall accuracy of the model is reported as 0.88, indicating that approximately 88% of the reviews in the evaluation dataset were classified correctly by the model. This suggests a reasonably high level of accuracy, but it's worth noting that accuracy alone may not provide a complete picture, especially when dealing with imbalanced datasets.

In conclusion, the classification report highlights the strengths and weaknesses of the sentiment analysis model based on the application of BERT. While the model demonstrates excellent performance in classifying positive sentiment, it faces challenges in accurately identifying neutral sentiment. These findings offer valuable insights for further analysis and potential enhancements to the sentiment analysis model.

4.5 Results Comparison

Metric	Logistic Regression	LSTM	BERT
Precision (Bad)	0.82	0.61	**0.82**
Precision (Good)	0.88	0.91	**0.95**
Precision (Neutral)	0.41	0.44	**0.48**
Recall (Bad)	0.73	0.84	**0.85**
Recall (Good)	**0.98**	0.95	0.94
Recall (Neutral)	0.16	0.02	**0.46**
F1-Score (Bad)	0.77	0.70	**0.83**
F1-Score (Good)	0.93	0.93	**0.95**
F1-Score (Neutral)	0.23	0.03	**0.47**
Accuracy	0.85	0.84	**0.88**

TABLE 4.6: Comparison of Models for Sentiment Analysis

After analyzing the classification reports of the three models (Logistic Regression, LSTM, and BERT), it can be concluded that the BERT model outperforms the other models and is the most suitable choice for this task.

The BERT model exhibits consistently high precision, recall, and F1-scores across all three classes (Bad, Good, and Neutral). It achieves an accuracy of 0.88, indicating its ability to correctly classify hotel reviews with a high degree of accuracy. This performance is particularly evident in the "Bad" and "Good" classes, where BERT achieves precision scores of 0.82 and 0.95, respectively, along with recall scores of 0.85 and 0.94. These metrics demonstrate the model's proficiency in accurately identifying negative and neutral sentiments in hotel reviews.

In comparison, both Logistic Regression and LSTM models show relatively lower performance in certain areas. While Logistic Regression performs well in predicting the "Good" class, it struggles with the "Neutral" class, where its recall and F1-scores are notably low. The LSTM model exhibits similar limitations, with lower recall and F1-scores for the "Neutral" class. This suggests that both models may have difficulty distinguishing and capturing the nuances of neutral sentiments in hotel reviews.

BERT, on the other hand, utilizes a state-of-the-art transformer-based architecture that incorporates contextual information effectively, allowing it to capture more complex patterns and dependencies within the text. This enables BERT to better understand the sentiment expressed in hotel reviews and make more accurate predictions. Furthermore, BERT's performance is reinforced by its ability to leverage pre-training on large amounts of textual data, which enhances its understanding of natural language and contextual relationships.

Considering the overall performance and the ability to handle different sentiment classes, it is evident that the BERT model surpasses the other models in this sentiment analysis task. Its superior performance in accurately identifying negative and positive sentiments, along with its high accuracy, make it the recommended choice for analyzing hotel reviews on TripAdvisor. By selecting the BERT model, you can enhance the accuracy and reliability of sentiment analysis results, leading to more valuable insights for hotel management and decision-making processes.

4.6 Dashboard Visualization

In addition to the sentiment classification and aspect-based sentiment analysis conducted in this thesis, a key aspect of the research involved the development of a comprehensive dashboard for visualizing and analyzing the results. The dashboard serves as a user-friendly interface to explore and interpret the findings, providing a holistic view of the sentiment distribution and aspect-level insights derived from Marrakesh hotel reviews.

The purpose of this section is to present and discuss the design, functionality, and significance of the dashboard in facilitating the interpretation and analysis of the sentiment classification and aspect-based sentiment analysis results. It goes beyond simply presenting screenshots and delves into the key features, interactive elements, and the added value it brings to the research.

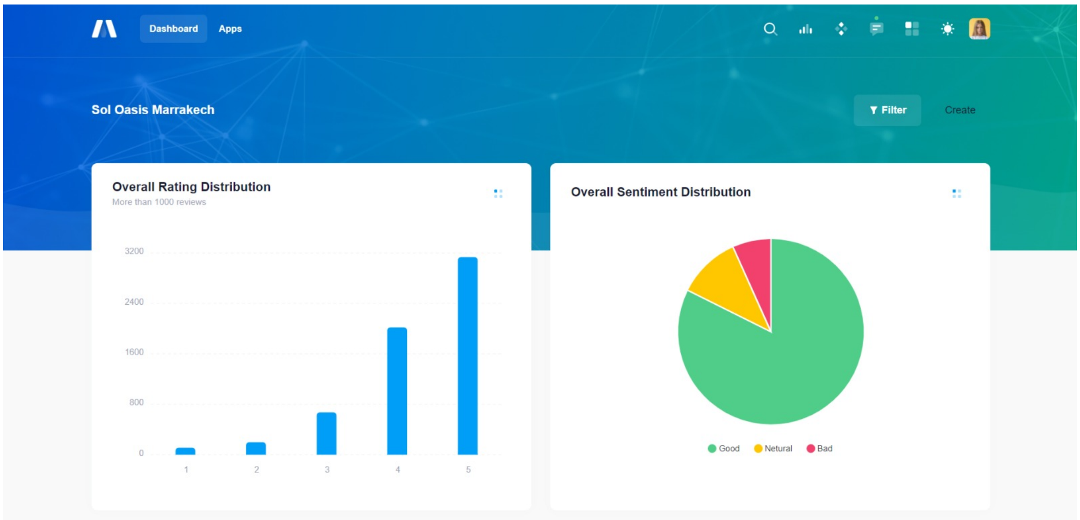

FIGURE 4.5: Dashboard screen 1

- Overall Ratings Distribution: This graph visually represents the distribution of overall ratings given by users in the Sol Oasis Marrakesh reviews.
- Overall Sentiment Distribution: This graph illustrates the distribution of overall sentiment expressed in the Sol Oasis Marrakesh reviews.

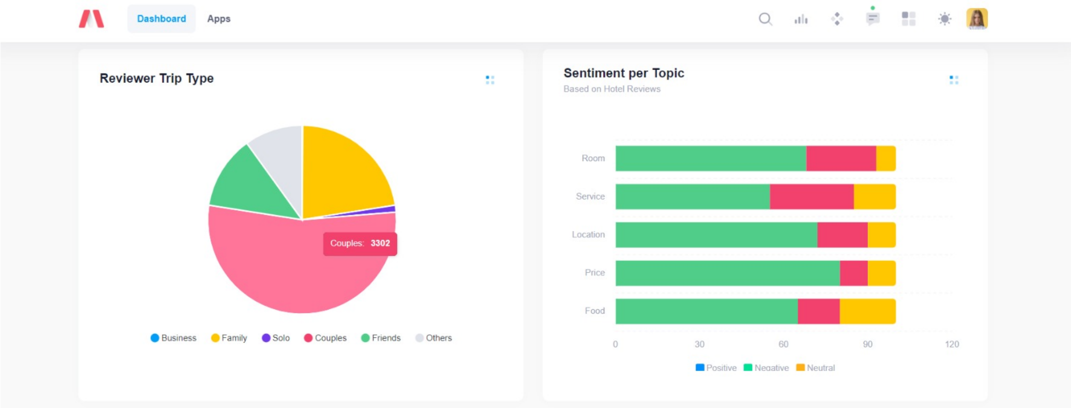

FIGURE 4.6: Dashboard screen 2

- Reviewer Trip Type Distribution: This graph displays the distribution of reviewer trip types in the Sol Oasis Marrakesh reviews.
- Sentiment per Topic: This graph showcases the sentiment distribution per topic in the Sol Oasis Marrakesh reviews.

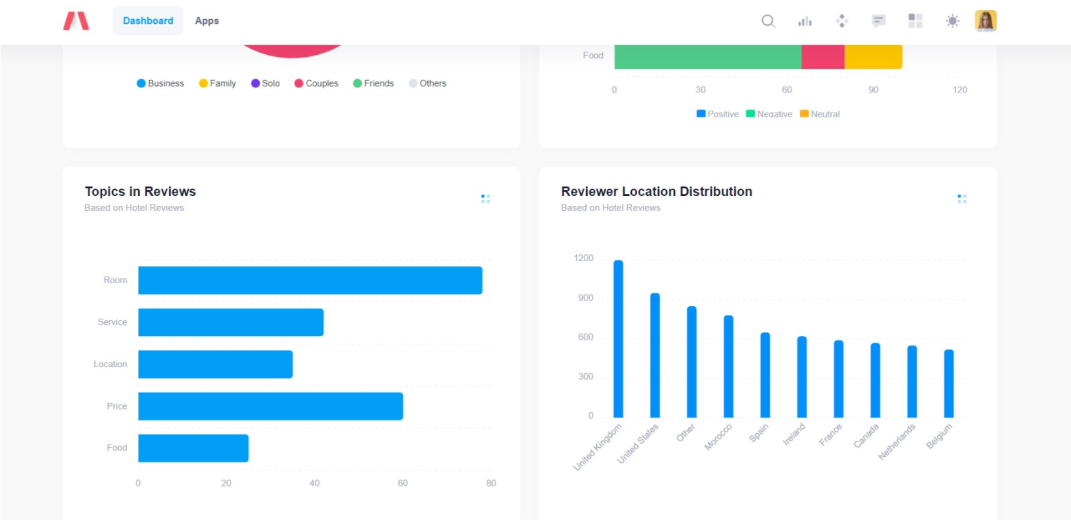

FIGURE 4.7: Dashboard screen 3

- Reviewer Location Distribution: This graph showcases the geographical distribution of the reviewers who have provided feedback on the Sol Oasis Marrakesh.
- Topics in Reviews: This section of the dashboard displays the identified topics within the Sol Oasis Marrakesh.

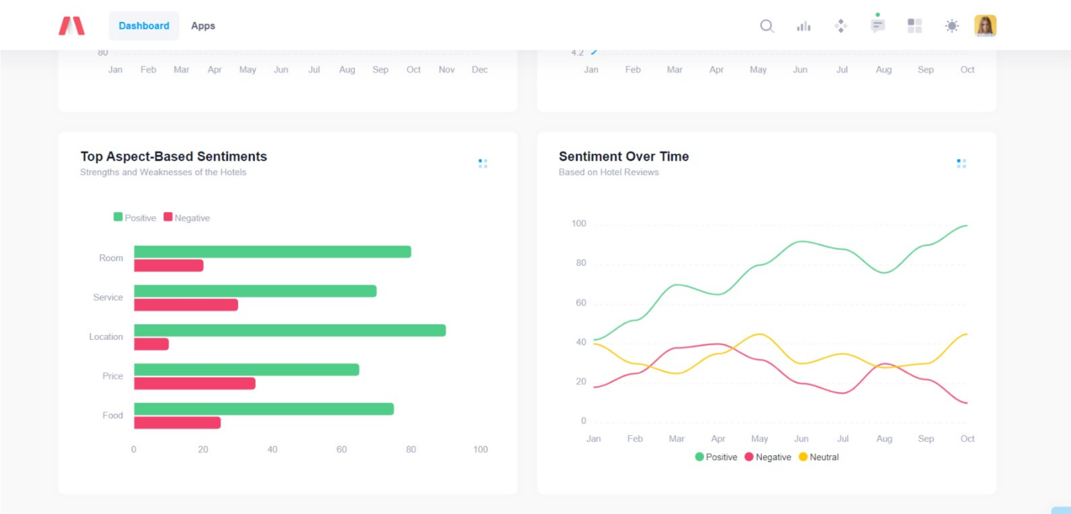

FIGURE 4.8: Dashboard screen 4

- Topic-based Sentiment Analysis: This graph showcases the sentiment analysis results based on different topics extracted from Sol Oasis Marrakesh reviews.
- Sentiments Over Time: This graph displays the sentiment trends over time for Sol Oasis Marrakesh reviews.

Key Functionalities:

- Sentiment Distribution: The dashboard displays an overview of the sentiment distribution across the hotel reviews, allowing users to gain insights into the overall sentiment polarity and distribution of positive, negative, and neutral sentiments.
- Aspect-Level Analysis: The dashboard enables users to delve deeper into the aspect-based sentiment analysis results. It provides visualizations that highlight the sentiment polarity associated with specific aspects such as room, service, location, price, and food. Users can easily identify the strengths and weaknesses of different aspects based on the sentiment distribution.
- Interactive Filters: The dashboard incorporates interactive filters to enable users to customize their analysis based on specific criteria. Users can filter the data based on time periods, star ratings, or specific hotels, providing a more focused and tailored analysis.
- Trend Analysis: The dashboard includes visualizations that showcase the sentiment trends over time, allowing users to identify any temporal patterns or fluctuations in customer sentiments.
- Comparative Analysis: Users can compare sentiment distributions and aspect-level sentiments between different hotels, aiding in benchmarking and competitive analysis.

In this chapter, we have presented the results and analysis of the sentiment analysis techniques applied in this study. The evaluation metrics have provided a quantitative assessment of the performance of the models. We have examined the results obtained from classic machine algorithms, LSTM, and BERT models, enabling us to gauge their effectiveness in sentiment analysis.

Chapter 5

Conclusion and Future Work

In this Master of Science thesis, we focused on sentiment classification of Marrakesh hotel reviews, employing aspect-based sentiment analysis to gain deeper insights. Through the utilization of a dataset comprising English reviews extracted from a popular review platform, we explored three distinct approaches: TF-IDF, LSTM, and BERT.

Our experimental results revealed that BERT, leveraging solely textual information, outperformed the other methods, showcasing superior performance in sentiment classification. Furthermore, models utilizing Bidirectional LSTM demonstrated better performance compared to those employing TF-IDF. This indicates the effectiveness of deep learning techniques in capturing the nuanced sentiment expressed in hotel reviews.

While sentiment classification provides an understanding of overall customer satisfaction levels, we recognize the importance of uncovering the underlying reasons behind positive or negative reviews. Therefore, aspect-based sentiment analysis was employed, focusing on key aspects such as room, service, location, price, and food. This analysis enables hotels to identify specific areas of improvement, thereby facilitating effective solutions for addressing customer concerns and enhancing their overall service quality.

Although this Master of Science thesis has made significant strides in sentiment classification and aspect-based sentiment analysis of Marrakesh hotel reviews, several avenues for future research and improvement remain open.

Integration of Multimodal Data: Currently, our analysis primarily relies on textual information. Integrating other modalities such as images, videos, or audio could provide a more comprehensive understanding of customer sentiments, allowing for a richer analysis and potentially improved classification accuracy.

Fine-Grained Aspect Analysis: Our study focuses on five key aspects, but there may be other important aspects specific to Marrakesh hotels that warrant consideration. Conducting a more granular analysis by incorporating additional aspects, or even allowing the model to learn the aspects from the data, could provide more detailed insights for hotel management.

Interpretability of Model Decisions: Deep learning models often lack interpretability, making it challenging to understand the factors that contribute to their predictions. Future work should explore techniques for interpreting and explaining the decisions made by the sentiment classification models. This would enable hotels to gain deeper insights into the reasons behind the sentiment expressed in reviews.

Real-Time Sentiment Analysis: Building a system that can perform sentiment analysis in real-time, continuously monitoring and analyzing new reviews as they arrive, would provide hotels with up-to-date information and allow for prompt response to customer feedback. Investigating streaming data processing and online learning techniques would be essential in developing such a system.

In conclusion, this Master of Science thesis has contributed to the field of sentiment analysis and aspect-based sentiment analysis of hotel reviews. By employing advanced techniques such as BERT and Bidirectional LSTM, we have achieved promising results. However, further research is necessary to address the future work outlined above and advance the state-of-the-art in sentiment analysis for the hospitality industry.

Bibliography

Afzaal, Muhammad et al. (2019). "Multiaspect-based opinion classification model for tourist reviews". In: *Expert Systems* 36.2, e12371.

Anis, Sarah, Sally Saad, and Mostafa Aref (2021). "Sentiment analysis of hotel reviews using machine learning techniques". In: *Proceedings of the International Conference on Advanced Intelligent Systems and Informatics 2020*. Springer, pp. 227–234.

Bompotas, Agorakis et al. (2020). "A sentiment-based hotel review summarization using machine learning techniques". In: *Artificial Intelligence Applications and Innovations. AIAI 2020 IFIP WG 12.5 International Workshops: MHDW 2020 and 5G-PINE 2020, Neos Marmaras, Greece, June 5–7, 2020, Proceedings 16*. Springer, pp. 155–164.

Breiman, Leo (1996). "Bagging predictors". In: *Machine learning* 24, pp. 123–140.

Devlin, Jacob et al. (2018). "Bert: Pre-training of deep bidirectional transformers for language understanding". In: *arXiv preprint arXiv:1810.04805*.

Edureka (Source). *Support vector machine*. Available at: https://www.edureka.co/blog/support-vector-machine-in-python/.

Farisi, Arif Abdurrahman, Yuliant Sibaroni, and Said Al Faraby (2019). "Sentiment analysis on hotel reviews using Multinomial Naïve Bayes classifier". In: *Journal of Physics: Conference Series*. Vol. 1192. 1. IOP Publishing, p. 012024.

González-Carvajal, Santiago and Eduardo C Garrido-Merchán (2020). "Comparing BERT against traditional machine learning text classification". In: *arXiv preprint arXiv:2005.13012*.

Grossberg, Stephen (2013). "Recurrent neural networks". In: *Scholarpedia* 8.2, p. 1888.

Hamdan, Hussam, Patrice Bellot, and Frederic Bechet (2015). "Lsislif: Crf and logistic regression for opinion target extraction and sentiment polarity analysis". In: *Proceedings of the 9th international workshop on semantic evaluation (SemEval 2015)*, pp. 753–758.

Ishaq, Abid et al. (2021). "Extensive hotel reviews classification using long short term memory". In: *Journal of Ambient Intelligence and Humanized Computing* 12, pp. 9375–9385.

Jones, George (2014). "Karen". In: *The Use of Personal Narratives in the Helping Professions*. Routledge, pp. 247–251.

Luhn, Hans Peter (1957). "A statistical approach to mechanized encoding and searching of literary information". In: *IBM Journal of research and development* 1.4, pp. 309–317.

Medium (Source). *An Intuitive Explanation of LSTM*. Available at: https://medium.com/@ottaviocalzone/an-intuitive-explanation-of-lstm-a035eb6ab42c/.

MonkeyLearn (Source). *How Does Sentiment Analysis Work?* Available at: https://monkeylearn.com/sentiment-analysis/.

Munikar, Manish, Sushil Shakya, and Aakash Shrestha (2019). "Fine-grained sentiment classification using BERT". In: *2019 Artificial Intelligence for Transforming Business and Society (AITB)*. Vol. 1. IEEE, pp. 1–5.

Noh, Yunseok, Seyoung Park, and Seong-Bae Park (2019). "Aspect-based sentiment analysis using aspect map". In: *Applied Sciences* 9.16, p. 3239.

Nohh, Nur Hidayu Md et al. (2019). "Sentiment Analysis towards Hotel Reviews". In: *Open International Journal of Informatics* 7.Special Issue 2, pp. 1–19.

Paltoglou, Georgios and Mike Thelwall (2010). "A study of information retrieval weighting schemes for sentiment analysis". In: *Proceedings of the 48th annual meeting of the association for computational linguistics*, pp. 1386–1395.

Priyantina, Reza Amalia and Riyanarto Sarno (2019). "Sentiment analysis of hotel reviews using Latent Dirichlet Allocation, semantic similarity and LSTM". In: *Int. J. Intell. Eng. Syst* 12.4, pp. 142–155.

Reddy, YCA Padmanabha et al. (2022). "Classification of Hotel Reviews using Machine Learning Techniques". In: *2022 8th International Conference on Smart Structures and Systems (ICSSS)*. IEEE, pp. 1–5.

Salehinejad, Hojjat et al. (2017). "Recent advances in recurrent neural networks". In: *arXiv preprint arXiv:1801.01078*.

Su, Haonan and Cheolkon Jung (2018). "Perceptual enhancement of low light images based on two-step noise suppression". In: *IEEE Access* 6, pp. 7005–7018.

TIBCO (Source). *Random Forest*. Available at: `https://www.tibco.com/reference-center/what-is-a-random-forest`.

Vaswani, Ashish et al. (2017). "Attention is all you need". In: *Advances in neural information processing systems* 30.

Wen, Yu, Yezhang Liang, and Xinhua Zhu (2023). "Sentiment analysis of hotel online reviews using the BERT model and ERNIE model—Data from China". In: *Plos one* 18.3, e0275382.

Zainuddin, Nurulhuda and Ali Selamat (2014). "Sentiment analysis using support vector machine". In: *2014 international conference on computer, communications, and control technology (I4CT)*. IEEE, pp. 333–337.