King Saud University
College of Computer and Information Sciences
Department of Information Technology

IT362: Principles of Data Science

Semester-2, 1447H

Phase 1: Data Collection Research and Assessment

Topic: Sentiment analysis of reviews of beauty centers

Prepared by

Section No	Name	ID
73059	Leen Almuqhim	445201272
	Remas Alayiedh	445202140
	Lubna Alqifari	445202104
	Aljoury Alreemi	445200074
	Khadijah Alshehri	445202138

Supervised by: Lama Alsudias

1. Introduction:

With the widespread use of digital platforms, online reviews have become an essential source for understanding customer experiences and opinions across various services, including beauty centres such as salons and cosmetic clinics. Many customers rely on these reviews before making booking decisions or selecting a specific beauty center.

These reviews often express different sentiments such as satisfaction, dissatisfaction, or neutrality and are usually written as unstructured textual data. Analyzing large volumes of such data manually is time-consuming and inefficient.

This project aims to analyse the sentiments expressed by customers in online reviews of beauty centres using sentiment analysis techniques in order to understand overall customer satisfaction trends and extract insights that support service improvement and decision-making.

Research Question:

What sentiments do customers express in online reviews of beauty centres?

2. Literature review:

Study	Problem Focus	Dataset	Methods / Models	Key Findings
Study 1 [1]	Difficulty of manually analyzing millions of unstructured reviews	4M Amazon reviews (Kaggle); balanced subset of 500K	ML: LR, SVM, DT, RF, MNB, BNB;DL: LSTM;Features: BoW, TF-IDF, N-grams	TF-IDF with N-grams improved results; LR & SVM reached 91.3%; LSTM achieved highest accuracy (93.3%)
Study 2 [2]	Limited research using advanced NLP for aspect-level insights	2,782 annotated Trustpilot reviews from 5 platforms; 14 aspects	ABSA with POS tagging & bi-terms;Models: LSTM, BERT, DistilBERT, RoBERTa, XLNet	Transformer models outperformed others; shipping & trust were key drivers of sentiment; shipping most influential
Study 3 [3]	Traditional sentiment analysis fails to capture aspect-level opinions	Sephora skincare reviews (top 10 products)	Keyword-based aspect detection;VADER + TextBlob;TF-IDF + SMOTE;XGBoost	Strong performance for positive/neutral sentiment; weaker on negative sentiment; TF-IDF lacks contextual depth
Study 4 [4]	Need to compare ML and DL effectiveness across datasets	Amazon product reviews with star ratings	ML: LR, NB, RF;DL: RNN, CNN	ML models outperformed DL; RF & LR achieved up to 99%; lexicon-based methods performed worst
Study 5 [5]	Lack of aspect-based analysis for a single product across reviews	Wireless earbud reviews from e-commerce platforms	DL & ensembles: LSTM, BERT;Voting+BERT, Bagging+BERT	Ensemble models improved accuracy; Bagging+BERT performed best; BERT strongest base model

Common Datasets or Features Used:

Across the reviewed studies, customer reviews serve as the primary data source for sentiment analysis, highlighting their importance in understanding user opinions in e-commerce and service contexts. Several studies relied on Amazon product reviews, while others used reviews from platforms such as Trustpilot and Sephora or focused on specific product categories like wireless earbuds.

In terms of feature representation, traditional machine learning approaches commonly employed Bag-of-Words, TF-IDF, and N-grams, whereas deep learning and transformer-based models relied on contextual embeddings generated by models such as BERT. Additionally, some studies applied aspect-based sentiment analysis to capture sentiment related to specific attributes (e.g., price, quality, shipping), while others focused on overall document-level sentiment classification.

Differences in Modelling Approaches:

The reviewed studies adopt diverse modelling strategies for sentiment analysis. Some focus on traditional machine learning classifiers, such as Logistic Regression, SVM, Random Forest, and Naive Bayes, which emphasize simplicity and interpretability but depend heavily on manual feature engineering.

Other studies employ deep learning models, including LSTM, RNN, and CNN, to better capture sequential and contextual information in text. More recent work shifts toward transformer-based models, such as BERT, DistilBERT, RoBERTa, and XLNet, which provide richer contextual representations and improved performance. In addition, one study applies ensemble learning techniques, combining models such as Voting and Bagging with BERT to enhance robustness and accuracy.

Furthermore, some studies extend beyond document-level analysis by using aspect-based sentiment analysis, enabling more detailed insights but often relying on manual annotation or keyword-based extraction, which limits scalability and generalizability.

Strengths and Limitations of Existing Work:

The reviewed studies demonstrate several strengths. Deep learning and transformer-based models, particularly BERT and its variants, consistently outperform traditional approaches by capturing contextual and nuanced sentiment. Aspect-based sentiment analysis provides fine-grained and actionable insights by linking sentiment to specific features, while ensemble learning methods improve model robustness and classification accuracy.

However, notable limitations remain. Traditional machine learning models struggle with contextual understanding, while deep learning and transformer-based methods require large, annotated datasets and significant computational resources, limiting their practical deployment. Many aspect-based approaches depend on manual annotation or keyword-based extraction, reducing automation and generalizability. Additionally, model performance often varies across datasets and domains, making cross-domain transfer challenging.

Missing, Limited, and Unclear Aspects in Previous Work:

Despite progress in sentiment analysis research, several gaps persist. Many studies focus on general sentiment trends rather than detailed aspect-level analysis across multiple reviews. There is also limited research that systematically compares integrated or ensemble deep learning models under consistent experimental settings. Scalability remains a concern due to reliance on manual annotation and the high computational cost of transformer models. Furthermore, the impact of data preprocessing and model interpretability is often unclear. Additionally, most existing studies concentrate on product-based reviews, with limited attention given to service-oriented domains, where customer experiences are more subjective and context-dependent.

How Our Project Will Build Upon or Differ from Existing Studies:

Our project builds on existing sentiment analysis research by focusing on a service-based domain, specifically beauty centers, rather than the product-focused reviews commonly analyzed in prior work. While many studies centered on e-commerce products, this project emphasizes customer experiences in salons and cosmetic clinics, where feedback is more experience-driven.

Unlike several prior studies that emphasize either large-scale generic sentiment classification or highly complex deep learning and ensemble models, our project adopts a practical and interpretable sentiment analysis approach tailored to service reviews and combines textual sentiment analysis with numerical ratings to examine the relationship between star ratings and expressed sentiment. It also explicitly addresses real-world data challenges, such as short or unclear reviews and potential review bias. By focusing on service quality factors (e.g., cleanliness, pricing, and staff behavior), the project aims to deliver actionable insights that support service improvement and customer satisfaction in the beauty industry.

Overall, this project extends prior sentiment analysis research by applying established techniques to a new service-oriented application domain, with an emphasis on practical usability and interpretability rather than solely optimizing model performance.

3. Data sources:

Data sources: The dataset consists of unstructured textual reviews (Unstructured Text Data), where each individual review represents one observation (reviews record). Each review is linked to its corresponding beauty centre using a place_id collected earlier from the Google Places API. Number of observations: The dataset contains 2,889 reviews collected from 700 beauty salons and spas, where reviews were gathered separately for each place.

Features and Data Types:

• **place_id:** A string identifier used to link each review to its corresponding beauty centre.

• **review_text:** Text data representing the user's written review; this is the primary input for sentiment analysis.

• **rating:** A numeric value ranging from 1 to 5 representing the user's overall rating of the service.

• **time:** A numeric timestamp indicating when the review was published.

• **language:** A categorical variable indicating the language of the review.

• **relative_time_description:** A categorical variable describing the relative posting time (e.g., "one week ago," "one month ago"). The data was integrated at the review level without applying any preprocessing or cleaning steps at this stage, in order to preserve the dataset in its raw form.

The dataset includes several potential sources of bias related to its origin and collection method:

1. **Representation** Bias The data represents only users who chose to write reviews on Google, which may result in unequal representation of certain groups, such as individuals who are less digitally active or who do not typically leave online reviews. In addition, users with extremely positive or negative experiences are more likely to submit reviews, potentially underrepresenting moderate or neutral experiences.

2. **Measurement** Bias The dataset relies on subjective user opinions and ratings, which may be influenced by temporary emotions, isolated incidents, or personal expectations. As a result, the reviews may not consistently reflect the actual or long-term quality of the services provided.

3. **Historical** Bias Some reviews may be outdated and may not reflect the current condition of the beauty centers, especially if there have been changes in management, staff, or service quality. Using such reviews may reinforce perceptions based on past conditions that are no longer valid.

4. Objectives:

Insights:

• Understand overall sentiment trends in online beauty center reviews

• Identify key factors associated with positive and negative sentiments, such as service quality, pricing, cleanliness, and staff behavior

Questions:

Main research question:

What sentiments do customers express in online reviews of beauty centers?

• What are the most common sentiments expressed in beauty center reviews?

• Do positive or negative sentiments dominate customer opinions?

• Is there a relationship between numerical ratings and textual sentiments?

Tasks:

• Classify textual reviews into positive, negative, and neutral sentiments

• Analyze sentiment distribution across a large set of beauty center reviews

• Extract general patterns related to customer satisfaction and dissatisfaction

5. Method:

To achieve the objectives of this project, the following methodology will be applied:

1. Data Collection

Online reviews of beauty centers will be collected from Google Mapsusing application programming interfaces (APIs).

2. Text Preprocessing

The collected textual data will be prepared by removing irrelevant characters, normalizing text, and handling duplicated or very short reviews to ensure better text quality for analysis.

3. Labeling (Sentiment Annotation)

Each review will be assigned a sentiment label (positive, negative, or neutral) based on semantic meaning.

The labeling process will be supported by numerical ratings, where higher ratings generally indicate positive sentiment and lower ratings indicate negative sentiment, in addition to semantic analysis of the review text. This step enables supervised or semi-supervised sentiment classification.

4. Sentiment Analysis

Sentiment analysis techniques will be applied to classify reviews according to their labeled sentiment categories.

5. Insights Extraction

The classification results will be analyzed to identify overall sentiment trends and to understand key factors influencing customer satisfaction and dissatisfaction.

6. Challenges & Recommendations:

Challenges:

1. Difficulty collecting data from review platforms that rely on dynamic content loading.

2. Presence of reviews written in different languages, which may affect the accuracy of sentiment analysis.

3. Very short or unclear reviews that limit accurate sentiment classification.

4. Review bias, as customers with strong positive or negative experiences are more likely to leave reviews.

Recommendations:

- Use multiple data sources to increase review diversity.
- Apply language detection techniques and process each language appropriately.
- Increase dataset size to reduce the impact of short reviews.
- Consider potential biases when interpreting final results.

7. References:

[1] A. Alqurafi and T. Alsanoosy, “Measuring customers’ satisfaction using sentiment analysis: Model and tool,” Journal of Computer Science, vol. 20, no. 4, pp. 419–430, 2024.

[2] L. Davoodi, J. Mezei, and M. Heikkilä, “Aspect-based sentiment classification of user reviews to understand customer satisfaction of e-commerce platforms,” Electronic Commerce Research, 2025, doi: 10.1007/s10660-025-09948-4.

[3] S. A. Gul and F. Ahmad, “A comparative study on sentiment analysis approaches using machine learning classifiers,” International Journal of Innovative Research in Technology (IJIRT), 2024.

[4] L. Ashbaugh and Y. Zhang, “A comparative study of sentiment analysis on customer reviews using machine learning and deep learning,” Computers, vol. 13, no. 12, Art. no. 340, 2024.

[5] “Sentiment analysis of consumer reviews on online shopping platforms using integrated deep learning models,” Sustainable Computing: Informatics and Systems, 2025.